UNDERSTANDING PROTEIN DYNAMICS, BINDING AND ALLOSTERY FOR DRUG DESIGN

EDITED BY: Guang Hu, Pemra Doruker, Hongchun Li, and Ebru Demet Akten
PUBLISHED IN: Frontiers in Molecular Biosciences

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88966-848-9
DOI 10.3389/978-2-88966-848-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

UNDERSTANDING PROTEIN DYNAMICS, BINDING AND ALLOSTERY FOR DRUG DESIGN

Topic Editors:
Guang Hu, Soochow University, China
Pemra Doruker, University of Pittsburgh, United States
Hongchun Li, Shenzhen Institutes of Advanced Technology (CAS), China
Ebru Demet Akten, Kadir Has University, Turkey

Citation: Hu, G., Doruker, P., Li, H., Akten, E. D., eds. (2021). Understanding Protein Dynamics, Binding and Allostery for Drug Design. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-848-9

Table of Contents

frontiers in Molecular Biosciences

EDITORIAL
published: 21 April 2021
doi: 10.3389/fmolb.2021.681364

Editorial: Understanding Protein Dynamics, Binding and Allostery for Drug Design

Guang Hu[1]*, *Pemra Doruker*[2], *Hongchun Li*[3] and *Ebru Demet Akten*[4]

[1] Center for Systems Biology, Department of Bioinformatics, School of Biology and Basic Medical Sciences, Soochow University, Suzhou, China, [2] Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States, [3] Research Center for Computer-Aided Drug Discovery, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [4] Department of Bioinformatics and Genetics, Faculty of Engineering and Natural Sciences, Kadir Has University, Istanbul, Turkey

Keywords: allostery, drug discovery, molecular dynamics simulation, elastic network model, protein structure network

Editorial on Research Topic

Understanding Protein Dynamics, Binding and Allostery for Drug Design

Proteins as molecular machines have dynamic structures sampling various conformational states, which determine their functionality, ligand binding, and allosteric properties. Allosteric communication as an intrinsic property of proteins (Gunesakaran et al., 2004) can be triggered by physical (protein-protein interaction, ligand-binding) and chemical (mutations, post-translational modification) events happening distant from the orthosteric site (Zhang et al., 2020). Protein dynamics and conformational transitions govern allosteric communication between distinct sites, which is central for the regulation of protein function, signal transduction, and approaches in drug discovery.

Computational modeling and simulations at hierarchical levels of complexity have become requisite for unraveling the link between protein structure, dynamics and function, as well as toward designing agents that regulate protein function. Being an active research field for over a century, allostery has become pivotal today in the pursuit of designing allosteric modulators with specificity. Simultaneously, a wealth of computational methods and tools have emerged for uncovering how protein dynamics affects ligand binding events and allosteric communication at the molecular scale.

In this Research Topic, a total of 20 works has been compiled that will be introduced here based on the diverse computational approaches employed including molecular dynamics (MD) simulations, elastic network models (ENM), hybrid and integrated methods, and protein structure networks (PSN).

MD simulation is the most popular computational method used in complementing experimental techniques as it captures the behavior of proteins in full atomistic detail for understanding binding and allosteric events (Hollingsworth and Dror, 2018), as well as molecular aspects of diseases and their treatments. Khan et al., performed an integrated computational study on the estrogen receptor alfa (ERα), which have been observed to be recurrent in metastatic breast cancer patients. The impact of experimentally-reported ERα polymorphisms was studied using techniques such as mCSM stability and binding affinity analysis (Pires et al., 2013) and MD simulations for revealing the dynamical effects on receptor structure. Another cancer type, Glioblastoma (GBM), is the most common and aggressive intracranial malignant brain tumor. Histone deacetylase 1 (HDAC1) is a promising target for therapy of GBM, and macrocyclic peptides

OPEN ACCESS

Edited and reviewed by:
Francesco Luigi Gervasio, University College London, United Kingdom

***Correspondence:**
Guang Hu
huguang@suda.edu.cn

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *16 March 2021*
Accepted: *25 March 2021*
Published: *21 April 2021*

Citation:
Hu G, Doruker P, Li H and Demet Akten E (2021) Editorial: Understanding Protein Dynamics, Binding and Allostery for Drug Design. Front. Mol. Biosci. 8:681364. doi: 10.3389/fmolb.2021.681364

have gained great attention due to their remarkable inhibitory selectivity on HDAC1. Zhang et al., employed molecular modeling approaches including molecular docking and MD simulation, along with protein-ligand interaction fingerprints and per-residue binding free energy analysis to explore the binding of a typical macrocyclic peptide FK228 to both HDAC1 and HDAC6. The blockade of programmed death receptor 1 (PD-1) has become a promising therapeutic approach in cancer immunotherapy. In Liu et al.'s work, the binding features of PD-1 with Nivolumab, a humanized IgG4 antibody approved by the US FDA, were investigated using MD simulations. The computational analysis suggested that the N-terminal loop of PD-1 serves as an important gatekeeper for the anti-PD-1 antibody binding, which might be a potential target for anti-PD-1 antibody design.

In the brief research report by Du et al., the structure of Oncostatin M and Receptor (OSM–OSMR) complex was generated as a potential therapeutic target for inflammatory bowel disease. Eight "hot spots" residues and six potential binding sites at the OSM-OSMR interface were predicted using computational alanine scanning and FTMap (Kozakov et al., 2015) analysis, which might be useful to guide further experimental studies and drug design. The article from Shi et al., compared the binding behaviors of four triazole-based inhibitors to sterol 14α demethylase enzyme (CYP51), and identified potential key binding sites. Besides, some possible tunnel pathways of the inhibitors in these CYP51-inhibitor complexes were proposed. Using MD simulations, Du et al. gave a detailed analysis of thermostability factors of barley limit dextrinase, including eight salt-bridges.

Dudas et al. used the hybrid MDeNM (Costa et al., 2015) method that combines MD simulations with classical normal mode analysis for efficient exploration of the conformational space of full-length K-Ras4B. Unbiased conformational sampling was carried out for the GDP- and GTP-bound states to determine the interaction details between a flexible tail-like hypervariable region (HVR) and the catalytic domain of K-Ras4B. The results elucidate the molecular details of a population shift mechanism between the autoinhibited state of the catalytic domain (GDP-bound) and its active state (GTP-bound). Salawu introduced DESP method by combining MD simulations and deep neural networks for enhanced sampling of conformation spaces. Guzel et al. studied allosteric signaling and communication pathways in the exit tunnel of the bacterial ribosome. Due to the huge size of this supramolecular machine, they used coarse-grained MD simulations (RedMD) (Górecki et al., 2009) and an ENM-based hybrid method (ClustENM) (Kurkcuoglu et al., 2016) for conformer generation. Suboptimal pathways based on the contact topology of conformers elucidated the allosteric signaling in the ribosomal tunnel.

Normal mode analysis based on ENM facilitates the study of protein dynamics and allosteric effects in a high-throughput manner (Krieger et al., 2020). Ayyildiz et al., integrated ENM-based methods with computational solvent mapping and sequence/structural alignments for identifying potential allosteric sites. By applying the method to three glycolytic enzymes, the specific interface regions connecting the subunits were predicted as promising target sites for allosteric regulation and species-specific drug design. Two widely used ENMs-Gaussian (GNM) (Bahar et al., 1997) and anisotropic network model (ANM) (Atilgan et al., 2001) - were used by Khairallah et al., to investigate the collective motions of the 6-pyruvol tetrahydropterin synthase (PTPS) enzyme and its allosteric properties. In particular, large fluctuations of the N-terminal domain and its allosteric role were discussed in the context of infectious disease treatment.

Abdizadeh et al. used an ENM-based perturbation response scanning (PRS) (Atilgan and Atilgan, 2009) to identify the residues that trigger dissociation between interacting protein pairs. Based on a set of 25 protein complexes, such residues were identified to be located either at regions with large conformational changes or at parts of the protein that are structurally unaffected. Interestingly, the interfacial residues were responsible for the dissociation in only a few of the complexes. Furthermore, they categorized four modes of dissociation based on PRS and electrostatic effects. Jernigan et al. proposed two ENM-based approaches for designing peptide-based inhibitors using the influenza protein hemagglutinin (HA) as the case study. Based on available experimental structures conformationally most variable region of HA was identified as a potential target for diverse ligands. Furthermore, the empirical contact potentials including an ENM-based entropy term were found successful in ranking the free energies of peptide/proteins designed against HA.

Topological description of protein structures has become a popular tool to quantify protein structures and dynamics (Liang et al., 2020). By means of joint dynamical/topological description of SARS-COV2 spike protein, Halder et al., gave a thorough presentation on the network-dynamical systemic approach to protein function. In addition, their work highlighted the advantages of the side-chain network analysis in studying subtle conformational changes with an emphasis related to allostery. Gosu et al., used PSN to study the MD ensemble of Myeloid differentiating factor 88 (Myd88) and AlloSigMA (Guarnera et al., 2017) to describe allosteric effects of S34Y and R98C variants. These two mutations were shown not only to induce a large conformational change of Myd88, but also influence the interaction with other death domains in TLR downstream signaling. Wang et al., combined PSN with some structural dynamics approaches such as molecular docking, MD simulations, and free energy calculation to study specific interactions between crocin and tubulin. Their results pointed to a common residue motif (val175-Xxx176-Asp177) that could serve as a potential binding site. Yan et al. designed the ANCA webserver for constructing and analyzing PSN for interpretation of functional residues and allosteric regulation.

Three review papers complete our Research Topic in experimental and computational prospects. Isothermal titration calorimetry (ITC) is a technique that measures the thermodynamics of binding reactions and reaction kinetics. The review by Wang et al., give a broad overview of the use of ITC to measure the strength, mode, and

association/dissociation kinetics of enzyme inhibitors, as well as its potential applications on allostery and drug design. Jaffe describes the morpheeins model for allosteric regulation by focusing on porphobilinogen synthase. This model characterizes the dynamic nature of quaternary structures/assemblies formed by homo-multimeric proteins with its implications/applications for drug discovery. Lastly, Verkhivker et al., provide a comprehensive overview on the state-of-the-art approaches that advance quantitative characterization of allosteric mechanisms in proteins, including experiment-guided Markovian models, simulation-based multiscale approaches and machine learning. These frontier technologies not only provide tools for the studying of molecular basis of allosteric mechanisms, but also help the discovery of allosteric modulators for therapeutically important protein targets.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This study was supported by the National Natural Science Foundation of China (31872723, 22007097) and a project funded by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions.

ACKNOWLEDGMENTS

We would like to thank all authors, editors, and reviewers of this Research Topic.

REFERENCES

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80, 505–515. doi: 10.1016/S0006-3495(01)76033-X

Atilgan, C., and Atilgan, A. R. (2009). Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput. Biol.* 5:e000544. doi: 10.1371/journal.pcbi.1000544

Bahar, I., Atilgan, A. R., and Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des.* 2, 173–181. doi: 10.1016/S1359-0278(97)00024-2

Costa, M. G. S., Batista, P. R., Bisch, P. M., and Perahia, D. (2015). Exploring free energy landscapes of large conformational changes: molecular dynamics with excited normal modes. *J. Chem. Theory Comput.* 11, 2755–2767. doi: 10.1021/acs.jctc.5b00003

Górecki, A., Szypowski, M., Długosz, M., and Trylska, J. (2009). RedMD-redcuded molecular dynamics. *J. Comput. Chem.* 30, 2364–2374. doi: 10.1002/jcc.21223

Guarnera, E., Tan, Z. W., Zheng, Z., and Berezovsky, I. N. (2017). AlloSigMA: allosteric signaling and mutation analysis server. *Bioinformatics* 33, 3996–3998. doi: 10.1093/bioinformatics/btx430

Gunesakaran, K., Ma, B., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? *Proteins Struct. Funct. Bioinform.* 57, 433–443. doi: 10.1002/prot.20232

Hollingswort, S. A., and Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron* 99, 1129–1143. doi: 10.1016/j.neuron.2018.08.011

Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., et al. (2015). The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* 10, 733–755. doi: 10.1038/nprot.2015.043

Krieger, J. M., Doruker, P., Scott, A. L., Perahia, D., and Bahar, I. (2020). Towards gaining sight of multiscale events: utilizing network models and normal modes in hybrid methods. *Curr. Opin. Struct. Biol.* 64, 34–41. doi: 10.1016/j.sbi.2020.05.013

Kurkcuoglu, Z., Bahar, I., and Doruker, P. (2016). ClustENM: ENM-based sampling of essential conformational space at full atomic resolution. *J. Chem. Theory Comput.* 12, 4549–4562. doi: 10.1021/acs.jctc.6b00319

Liang, Z., Verkhivker, G. M., and Hu, G. (2020). Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *Brief Bioinformat.* 21 815–835. doi: 10.1093/bib/bbz029

Pires, D. E., Ascher, D. B., and Blundell, T. L. (2013). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 30, 335–342. doi: 10.1093/bioinformatics/btt691

Zhang, Y., Doruker, P., Kaynak, B., Zhang, S., Krieger, J., Li, H., et al. (2020). Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Curr. Opin. Struct. Biol.* 62, 14–21. doi: 10.1016/j.sbi.2019.11.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 24 January 2020
doi: 10.3389/fmolb.2019.00159

Dynamics Insights Into the Gain of Flexibility by Helix-12 in ESR1 as a Mechanism of Resistance to Drugs in Breast Cancer Cell Lines

*Abbas Khan[1], Ashfaq-Ur-Rehman[1], Muhammad Junaid[1], Cheng-Dong Li[1], Shoaib Saleem[2], Fahad Humayun[1], Shazia Shamas[3], Syed Shujait Ali[4], Zainib Babar[5] and Dong-Qing Wei[1,6,7]**

[1] State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, [2] National Center for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan, [3] Department of Zoology, University of Gujrat, Gujrat, Pakistan, [4] Centre for Biotechnology and Microbiology, University of Swat, Mingora, Pakistan, [5] School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China, [6] Peng Cheng Laboratory, Shenzhen, China, [7] Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education, Shanghai, China

OPEN ACCESS

Edited by:
Guang Hu,
Soochow University, China

Reviewed by:
Weiwei Xue,
School of Pharmaceutical Sciences, Chongqing University, China
Alessandra Magistrato,
Italian National Research Council (CNR), Italy

***Correspondence:**
Dong-Qing Wei
dqwei@sjtu.edu.cn

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: 31 October 2019
Accepted: 18 December 2019
Published: 24 January 2020

Citation:
Khan A, Ashfaq-Ur-Rehman, Junaid M, Li C-D, Saleem S, Humayun F, Shamas S, Ali SS, Babar Z and Wei D-Q (2020) Dynamics Insights Into the Gain of Flexibility by Helix-12 in ESR1 as a Mechanism of Resistance to Drugs in Breast Cancer Cell Lines. Front. Mol. Biosci. 6:159. doi: 10.3389/fmolb.2019.00159

Incidents of breast cancer (BC) are on the rise on a daily basis and have proven to be the most prevelant cause of death for women in both developed and developing countries. Among total BC cases diagnosed after menopause, 70% of cases are Estrogen Receptor (ER) positive (ER-positive or ER+). Mutations in the LBD (ligand-binding domain) of the ER have recently been reported to be the major cause of resistance to potent antagonists. In this study, the experimentally reported mutations K303R, E380Q, V392I, S463P, V524E, P535H, P536H, Y537C, Y537N, Y537S, and D538G were analyzed, and the most significant mutations were shortlisted based on multiple analyses. Initial analyses, such as mCSM stability, occluded depth analysis, mCSM-binding affinity, and FoldX energy changes shortlisted only six mutations as being highly resistant. Finally, simulations of force field-based molecular dynamics (MD on wild type (WT) ERα) on six mERα variants (E380Q, S463P, Y537S, Y537C, Y537N, and D538G) were carried out to justify mechanism of the resistance. It was observed that these mutations increased the flexibility of the H12. A bonding analysis suggested that previously reported important residue His524 lost bonding upon mutation. Other parameters, such as PCA (principal component analysis), DCCM (dynamics cross-correlation), and FEL (free energy landscape), verified that the shortlisted mutations affect the H12 helix, which opens up the co-activator binding conformation. These results provide deep insight into the mechanism of relative resistance posed to fulvestrant due to mutations in breast cancer. This study will facilitate further understanding of the important aspects of designing specific and more effective drugs.

Keywords: ESR1 estrogen receptor, mutation, resistance, simulation, molecular docking, molecular dynamics (MD) simulation

INTRODUCTION

Breast cancer (BC) is the primary contributor to a rise in global female mortality rates. It has been reported that 70% of BC cases diagnosed after menopause are Estrogen Receptor (ER) positive (ER+). The human aromatase (HA) enzyme produces estrogens (17-β-estradiol or estrone) (Magistrato et al., 2017) mainly after menopause, and its inactivity increases the level of estrogens in malignant tissues (Liang and Shang, 2013; Sgrignani et al., 2014, 2015, 2016; Magistrato et al., 2017). These hormones have a pro-oncogenic effect by either stimulating cell proliferation or decreasing apoptosis when binding to ERα as an agonist (Liang and Shang, 2013). The endocrine behavior of ER + BC is mainly determined by the deficiency of estrogen, which is caused by the inhibition of downregulators (SERDs), selective modulators of ERα (SERMs), or Human aromatase. Furthermore, as SERMs do leads to ERα ubiquitination and degradation, it also covers the substrate-binding site and alters receptor by changing its conformation (Osborne et al., 2004).

The ligand, which facilitates estrogen activity in several essential physiological processes, regulates ERα, which is a transcription factor and nuclear hormone receptor (Nilsson and Gustafsson, 2011; Lai et al., 2015). The ligand-binding domain (LBD), as well as the DNA-binding domain of ERα (among five separate functional parts), has been determined crystallographically. ERα becomes stable through the binding of either agonists or antagonists under physiological conditions and acts as a dimer. The essential structural element of each LBD monomer is Helix 12 (H12) and can be seen by observing the crystal structures. H12 behaves as a molecular switch between the active and inactive conformation of the receptor. H12 occludes the ligand-binding site on the binding of estrogen to helix packing H3, H5/6, and H11 (Brzozowski et al., 1997; Jordan et al., 2015). It remains stable with ERα's agonist (active) conformation. However, when an antagonist binds, it inhibits H12 from assuming the active conformation, and H12 travels to a groove made by H3 and H5. This relates to antagonistic (inactive) conformation (Joseph et al., 2016). In recent clinical scenarios, there are several effective antagonistic uses of ERα: (i) tamoxifen is a SERM that is active in its metabolites but inactive in peripheral tissues; and (ii) fulvestrant, which is also a SERD, without inactivity regulates ERα, which experiences reduced pharmacokinetic properties (e.g., low water solubility) (Nilsson and Gustafsson, 2011; van Kruchten et al., 2015). In recent decades, the use of tamoxifen by BC patients has reduced the death rate by 25–30%. In ER+ BC patients, 40% get resistance through disease progression and prolonged therapy (Jensen and Jordan, 2003). Recently, Robinson et al. (2013), Toy et al. (2013), Merenbakh-Lamin et al. (2013), and Jeselsohn et al. (2014) reported ERα polymorphisms (mERαs) in the LBD between H9 and H10 (S463P), close to the estrogen binding site (i.e., E380Q) and in the loop that connects H11 and H12 (i.e., L536Q, L536R, Y537C, Y537N, Y537S, and D538G). These polymorphisms occur at a significant rate in relapsed metastatic patients. However, it is rare or absent in untreated patients that have a primary tumor (Robinson et al., 2013; Toy et al., 2013; Jeselsohn et al., 2014). Experimental studies have suggested the possible role of these polymorphisms in inherited resistance to treatments of an endocrine nature (Liang and Shang, 2013) by developing novel features during BC to avoid the use of therapeutics. The reported occurrences of a particular ERα polymorphism vary from case to case, 21–36% of cases in D538G, 5–33% in Y537N, and 13–22% in Y537S, while there are less occurrence of other polymorphisms (Robinson et al., 2013; Toy et al., 2013; Jeselsohn et al., 2014). A double mutant D538G and Y537S was also seen in a few cases (Chandarlapaty et al., 2016). The overall survival time of patients is not dependent on the abundance of these mutations (26 months D538G, 20 months Y537S, and 15 months for double mutant Y537S and D538G) (Chandarlapaty et al., 2016), with the most aggressive isoform being Y537S (De Savi et al., 2015; Lai et al., 2015). A previous study by Pavlin et al. performed a simulation-based study of these mutations, but their analysis utilized only a single simulation tool (Pavlin et al., 2018).

In the multifactorial nature of diseases like cancer, various factors determine the positive, and negative reaction to drugs. The ultimate objective is therefore to investigate the factors directly involved in the development of BC drug resistance and to overcome this problem (Magistrato et al., 2017; Spinello and Magistrato, 2017). So far, computational studies on mERαs and their mechanism have been inadequate to explain K303R, E380Q, V392I, S463P, V524E, P535H, P536H, Y537C, Y537N, Y537S, and D538G polymorphisms that are most repeated isoforms (Delfosse et al., 2012; Robinson et al., 2013; Toy et al., 2013; Fanning et al., 2016; Joseph et al., 2016). It was indicated that these are constitutively active mutants but with a different molecular mechanism (Delfosse et al., 2012; Toy et al., 2013). Other mutants can also have different activation pathways that lead to therapy responses that are mutant dependent and are yet to be studied (Delfosse et al., 2012; Spoerke et al., 2016).

In silico, methods to predict structural implications of mutations will be beneficial in understanding mechanisms of drug resistance for quantitative estimation of the phenotypic resistance outcomes (Khan et al., 2019). To systematically understand the effects (protein stability changes, flexibitliy drift, and protein ligand interaction) of these mutations, we performed *in silico* saturation mutagenesis. Additionally, we also assessed the impacts of mutations on the relative sidechain solvent accessibility, depth, and the residue-occluded packing density. Extremely detrimental mutations were selected and analyzed for changes in their interatomic interactions that might explain the destabilizing effects. To explore further the vibrational entropy and enthalpy changes of flexible conformations, we employed an empirical force field-based method, FoldX, a coarse-grained normal mode analysis (NMA)-based elastic network contact model, ENCoM, and a consensus predictor that integrates normal mode approach with graph-based distance matrix in the mutating residue environment. Finally, simulations of force field-based molecular dynamics on wild type (WT) ERα and six variants (E380Q, S463P, Y537S, Y537N, and D538G) were carried out to justify mechanism of the resistance. To cure all types of metastatic BC types, this detailed investigation advocates advancement in precision medicine.

MATERIALS AND METHODS

Comparative Modeling, Quality Assessment, and Model Refinement

A crystallized structure of Estrogen receptor alpha (ESR1) was downloaded from RCSB (**PDB ID 1GWQ**) (Petrella et al., 2011). Structural topology was analyzed for the coordinate's defects. MolProbity (Chen et al., 2010) was used to evaluate the quality of the constructed structure, and atomic conflicts were resolved by energy minimization using the Steepest descent and conjugate gradient algorithms. Using YASARA (Land and Humble, 2018), energy minimization was conducted. The water molecules were completely removed before any further analysis. Pymol was used for visualization (Scientific and San Carlos, 2002). The Fulvestrant structure with accession ID CID802 was obtained from the PubChem database (Kim et al., 2015). The mutant models and sidechains of the mutants were optimized using FoldX (Schymkowitz et al., 2005). Molecular docking of the ESR1 with fulvestrant was performed by using a Schrödinger suite. A glide tool (Friesner et al., 2006) was used for docking of the fulvestrant. Since the structures of fulvestrant and ZB716 are similar, the same protocol using a flexible docking simulation was performed with the Induced Fit protocol (IFD) method as previously reported (Guo et al., 2018). The docking complexes were refined with the protein–ligand interaction refinement tool in the Schrödinger suite. Heirarchial optimization of the complexes that consider the systematic sampling of ligand position, conformation, and orientations along with the proteins residues was performed. All these calculations were performed on the apo structures (wild and mutant) obtained from MD simulation.

Effects of Mutations on Protein Stability and Interactions

mCSM (http://biosig.unimelb.edu.au/mcsm/) (Pires et al., 2013), SDM (Pandurangan et al., 2017), and FoldX were used to understand the impact of mutations on the thermodynamic stability of the protein. For SDM, mutant-protein models were produced using FoldX, which considers preserved conservation angle laws while identifying the most likely mutant residue sidechain rotamers. To determine the energy fold change upon inducing mutations, FoldX utilizes a linear combination of empirical terms to calculate the effect of mutations on the protein structure in kcal/mol. FoldX uses the following equation to calculate each energy term.

$$\begin{aligned}\Delta G = {} & a\,.\,\Delta G_{vdW} + b\,.\,\Delta G_{solH} + c\,.\,\Delta G_{solP} + d\,.\,\Delta G_{wb} \\ & + \Delta e\,.\,\Delta G_{Hbond} + f\,.\,\Delta G_{el} + g\,.\,\Delta G_{kon} + h\,.\,T\Delta S_{mc} \\ & + \Delta i\,.\,T\Delta S_{sc} + l\,.\,\Delta G_{clash}\end{aligned}$$

In this expression (a ... l) are relative weights of the different energy terms used for the free energy calculation. Each term in the above equation is defined in the original manuscript (Schymkowitz et al., 2005).

The effect of mutations on the protein–ligand affinity, ESR1-Fulvestant, was determined by using mCSM-lig (Pires et al., 2016). The mCSM-lig server analyzed only residues within 10Å of the interatomic distance to fulvestrant. The stability changes were further compared with predictions from other computational tools in order to estimate the reliability of the predictions.

Changes in Vibrational Entropy and Normal Mode Analysis

To evaluate the implications of mutations in flexible conformations on protein stability, we used FoldX, an empirical force field method that computes free energy changes between the protein's native and mutant forms, and an elastic network contact model (ENCoM) (Frappier et al., 2015), a coarse grain NMA strategy that takes into account the nature of the amino acids and aids in calculating vibrational entropy changes upon mutations. We have also used DynaMut (Rodrigues et al., 2018), a protein stability consensus predictor based on ENCoM's predicted vibrational entropy changes and the stabilization changes predicted by mCSM's graph-based signature method.

Conformational Changes

Conformational changes and their impacts on biophysical properties of the proteins were estimated using SDM (Pandurangan et al., 2017). The interatomic distances between each residue and fulvestrant in the protein-ligand complex were measured and included in the analysis. For all mutations, secondary structure switches in mutants, changes in relative solvent accessibility, residue depth in Å, and residue-occluded packing densities were determined.

Molecular Dynamics Simulation

In order to estimate the dynamic behavior of fulvestrant at the active site of native and mutant receptors, an all-atoms simulation using an Amber14 package was carried out (Salomon-Ferrer et al., 2013) with the ff14SB force field. For MD simulation, seven systems have been prepared, including a wild type and six complex systems with fulvestrant. Each system was solvated with a rectangular TIP3P water box and neutralized by adding counter ions. The steepest descent minimization method was used for energy minimization followed by conjugate gradient minimization of 3,000 steps. Each system was then gradually heated for 200 ps−300 K. Weak restraints for 2 ns were used to balance each system's density, followed by the constant pressure of 2 ns for system's balance. A constant pressure of using Langevin approach was used (1 atm, 300 K) (Zwanzig, 1973). To evaluate long-range electrostatic interactions the Particle Mesh Ewald (PME) algorithm with default settings in AMBER14 was used (Darden et al., 1993; Essmann et al., 1995). The threshold distances for long-range electrostatic interactions and Van der Waals were set to 10.0 Å, and for hydrogen covalent bonds the SHAKE algorithm was used. (Ryckaert et al., 1977). A total of 700 ns simulations were carried out using pmemd.cuda (Gotz et al., 2012). CPPTRAJ and PYTRAJ packages in Amber 14 were used to evaluate the MD trajectories. We also performed 200 ns simulation for each apo system.

Unsupervised Clustering of MD Trajectories and Gibbs Free Energy Calculation

Motion in the trajectories from both wild and mutant systems was calculated by using an unsupervised machine-learning technique known as Principle component analysis (PCA) (Pearson, 1901). For this purpose, a CPPTRAJ package in Amber was used. The reference structure was subjected to the translational as well as rotational motions. The positional covariance matrix for atomic coordinates, as well as its eigenvectors, were calculated. The diagonal matrix of eigenvalues was obtained by diagonalizing the matrix with the help of orthogonal coordinate transformation. The eigenvector and its eigenvalue suggested the principal component of the trajectory and highlighted the principal dominant global motion of the structures.

The free energy landscape (FEL) was calculated by using the first two PCs (PC1 and PC2). Deep valleys plot was used to draw and understand the native and metastable states of each system (Hoang et al., 2004). In this study, FEL was calculated using the following equation based on the first two principal components:

$$\Delta G\,(X) = -K_B T ln P(X)$$

where X suggests the response organizes taken by the primary the two principal components, KB implies the Boltzmann steady, and P(X) is the dispersion of the framework's likelihood on the first two principal components.

Dynamic Cross-Correlation Map (DCCM) Analysis

Using dynamics cross-correlation maps, the time-subordinate corresponded movements of C-α atom could be plotted. We, therefore, implemented DCCM to comprehend the highly connected movement of the C-alpha atoms when the ligand is bounded.

$$C_{ij} = \frac{<\Delta r_i \Delta r_j>}{<<\Delta r_i^2><\Delta r_j^2>>}$$

The equation above contains various elements of DCCM plots. D_{ri} and D_{rj} exemplify the vector of displacement of atoms *i* and *j* while **<...>** symbolizes the average trajectory.

Binding Affinity Calculations

The wild and mutant ESR1 systems free binding energy was figure out by using the script MMPBSA.PY (Hou et al., 2012; Miller et al., 2012; Xu et al., 2013; Sun et al., 2014, 2018; Chen et al., 2016), and these energies were calculated by considering 500s napshots from the MD trajectory. The binding free energy was determined as:

$$\Delta G_{bind} = \Delta G_{complex} - \left[\Delta G_{receptor} + \Delta G_{ligand}\right]$$

where ΔG_{bind} is the absolute free binding energy, and the rest of the parts are the free energy of the complex, the protein, and the ligand. Every segment's free energy was determined by utilizing the given equation:

TABLE 1 | List of selected mutations and their respective regions (Helix).

Index	Mutation	Region
1.	K303R	Helix-1
2.	E380Q	Helix-12
3.	V392I	Helix-3
4.	S463P	Helix-12
5.	V524E	Helix-11
6.	P535H	Helix-11
7.	P536H	Helix-11
8.	Y537C	Helix-12
9.	Y537N	Helix-12
10.	Y537S	Helix-12
11.	D538G	Helix-12

$$G = G_{bond} + G_{ele} + G_{vdW} + G_{pol} + G_{npol} - TS$$

G_{bond}, G_{ele} and G_{vdW} indicate interactions among bonded, electrostatic, and van der Waals states, whereas G_{pol} and G_{npol} demonstrate the polar and non-polar binders to the free energy figured by the certain solvent method of the generalized Born (GB) with SASA perceptible to solvents. Ordinary mode investigation determined the entropic commitment of TS.

RESULTS

Structure Preparation

Structural coordinates of the ligand-binding domain (LBD) of the ESR1 monomer were downloaded from RCSB using PDB ID (1GWQ). An initial visual analysis of the structure revealed some residues were missing while others had defects. The missing residues were identified by comparing them with the primary amino acid sequence of ESR1. Missing residues were added, and structural refinement with YASARA and Fold-X tools was performed. The final structure was subjected to 200 ns simulation to obtain the most stable conformation. Eleven mutations reported by different experimental studies were included as indicated in **Table 1**. The mutations were spotted in different helices of ESR1, and six mutations were found in Helix 12. Only six mutations E380Q, S463P, Y537S, Y537C, Y537N, and D538G were selected for MD simulation and post-simulation analyses. The selection of these mutations is based on pre-MD simulation analyses, which revealed substantial information about their specificity. In order to get better insight into the mechanism, an initial structure was prepared. The 3D structure of ESR1, its different domains, Helices pattern, and the 2D structure of fulvestrant are shown in **Figure 1**.

Analysis of Free Energy Changes Predicted by Different Computational Tools

Multiple well-implemented algorithms were implied to predict the stability changes associated with each of the mutations in the ESR1 structure. Structure-based stability changes predictors FoldX, mCSM, and SDM were used to calculate effect the

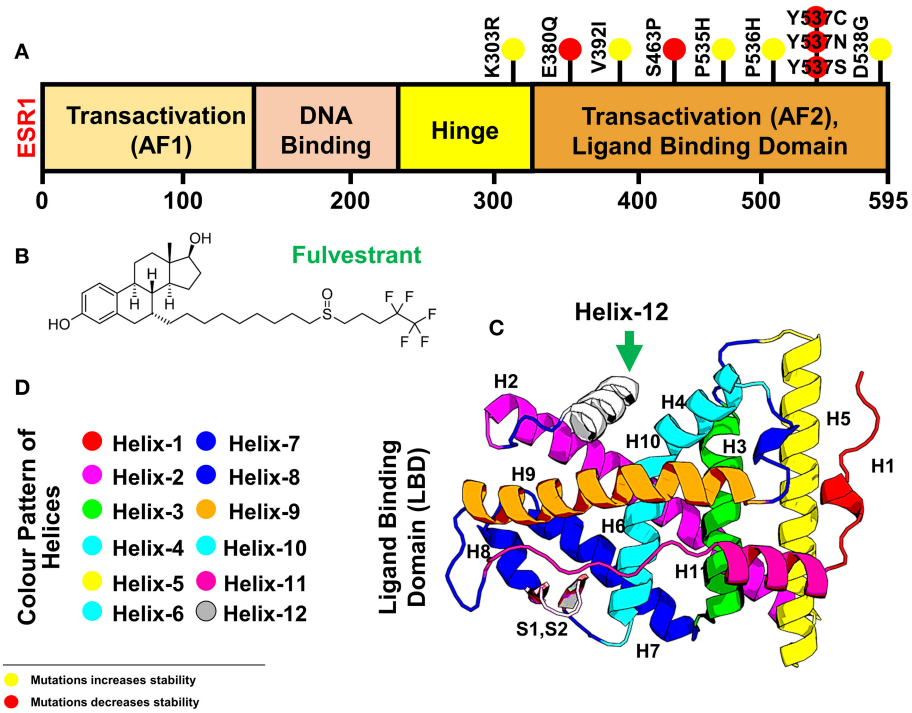


FIGURE 1 | Figure is showing ESR1 and the resistant drug Fulvestrant **(A)** showing different domains [Transactivation domain (AF1), DNA Binding Domain, Hinge Domain, and Transactivation (AF2) or Ligand Binding domain] of ESR1, **(B)** showing the 2D topology of resistant drug Fulvestrant, and **(C)** showing the 3D modeled structure of ESR1 and its different Helices. Helix 12 is highlighted in a green color. **(D)** The color pattern of different Helices shown in **(B)**. The yellow spot represents the resistant mutations with negligible effect on the drug, while the red color spot shows mutations with a significant level of resistance effect.

possible mutations at each residue position in the LBD of ESR1. The rationale for performing these analyses is to understand how mCSM and SDM, being structure-based predictors of stability change upon mutations and relate to sequence-based methods as well as how vibrational entropy changes in normal mode perturbations.

Mutations-Stability Correlation Analysis

mCSM uses graph-based signatures, where the primary feature is distances between different atoms. It also uses the common pharmacophoric features and converts it into digits where its mutant pharmacophore count is compared with the wild one. Wild type and mutant residues are represented as pharmacophore frequency vectors. These feature vectors are appended to experimentally important thermodynamics features, such as pH, solvent accessibility, and temperature. Herein, using mCSM, stability changes upon mutations were calculated, and average changes ranging from 0.823 to −3.033 kcal/mol was reported. Mutations, such as E380Q with stability fold change −1.192 kcal/mol and S463P with stability fold change of −0.689 kcal/mol, Y537S, Y537N, and Y537C with stability fold changes of −1.899, −1.66, and −0.566 kcal/mol, respectively, were found to be primarily affected in the highest fold with a destabilizing effect. In addition, mutation D538G with stability change of −0.545 kcal/mol was also clustered as destabilizing mutants. Mutations E380Q and S463P lie near or in the active site (Helix 3 and Helix 11), thus posing high level of resistance to the drug.

On the other hand, mutations Y537S, Y537N, Y537C, and D538G are spotted in Helix-12, whose flexibility is affected by these mutations and thus pose significant flexibility drift. It has been previously reported that the flexibility and replacement of Helix-12 can cause major destruction on the binding of ligands in the active site. The active site residues lie on Helix-3 and Helix-11, which is affected by the motion of Helix-12. According to the results, the remaining mutations, such K303R, V392I, V524E, P535H, and P536H, produce the opposite effect (i.e., stability) and do not induce major changes in affinity or reported to be least influenced. DynaMut, DUET, CUPSAT, and I-mutant also rationalized the destabilizing effects of the six mutations E380Q, S463P, Y537S, Y537C, Y537N, and D538G. All the stability results predicted by different servers are given in **Table 2**.

Impact of Mutations on Flexible Conformations and Changes in Vibrational Entropy (ΔS)

FoldX calculated the stability changes between the wild type and each mutant in the lowest energy conformation. It optimizes the sidechain rotamers of the mutant residues to attain a low energy state and calculates the change in free energy between the states. It can be seen from **Table 2** that the mCSM suggested mutations E380Q, S463P, Y537S, Y537C, Y537N, and D538G are more likely

TABLE 2 | The table shows the stability results predicted by different servers and softwares (DUET, ENCoM, DynaMut, mCSM, SDM, and FoldX).

Index	Mutation	DUET	ΔΔG ENCoM	ΔΔG DynaMut	mCSM ΔΔG	SDM ΔΔG	FoldX	Outcome
1.	K303R	0.116	−0.025	−0.217	−0.719	−0.17	−0.03	Stabilizing
2.	E380Q	−1.192	0.03	0.416	−1.482	−0.4	−0.29	Destabilizing
3.	D538G	−0.445	0.028	0.265	0.008	−0.54	0.92	Destabilizing
4.	Y537S	−1.899	−0.027	−0.095	−0.215	−0.39	3.22	Destabilizing
5.	Y537N	−1.369	−0.809	−0.077	−0.315	−1.66	2.41	Destabilizing
6.	H524E	−0.243	−0.087	−0.198	−0.73	−0.11	2.02	Destabilizing
7.	V392I	−0.431	−0.6	−1.726	−0.264	0.16	−0.86	Stabilizing
8.	S463P	−0.495	−0.606	−1.188	−0.689	−0.09	3.69	Destabilizing
9.	P535H	−0.036	−0.08	−0.056	−0.312	−1.1	1.65	Destabilizing
10.	L536H	−0.044	−0.413	−0.174	−0.788	−0.77	0.16	Stabilizing
11.	Y537C	−0.566	−0.559	0.076	−0.164	0.27	2.22	Destabilizing

to affect the energy changes in the higher fold than the others. It can be confirmed that mutations in helix-12 are primarily the major contributor in the energy changes profile.

On the other hand, fully flexible conformers of the mutants were sampled to compute the difference in vibrational entropy(ΔS) between the mutants and wild type. The average vibrational entropy change was observed from −0.038 to 1.011. All the calculations were carried out in kcal/mol.K^{-1}. The maximum vibrational entropy (ΔS) changes were induced by mutation Y537N, followed by S463P, Y537C, L536H, D538G, Y537S, and E380Q. It can be seen in **Figure 2** that the six mutations Y537N, S463P, Y537C, D538G, Y537S, and E380Q induced higher flexibility than those of the others. These results, along with the other, i.e., stability changes, clearly pointing out the importance of these six mutations. It has been previously reported that these mutations specifically in the Helix-12 pose major resistance to treatment in breast cancer. It can be seen that mutations D538G, Y537N, Y537S, and Y537C in helix 12 destabilizes the protein conformation by inducing significant flexibility drift.

Furthermore, mutations, such as E380Q and S463P, increase the residual rigidity in some helices while alternative inducing flexibility in some residues. These changes in flexibility (red) and rigidity (blue) are mapped onto the corresponding protein structure and presented in **Figure 2**. In order to get further insight into the phenomena, we conducted a simulation analysis of each apo system for a total of 200 ns. The results suggested that the mutations induced stability and conformational changes in the structure of protein. The root mean square fluctuation (RMSF) was calculated to confirm the flexibles regions and residual flexibility changes upon mutations. It was noticed that mutations, such as D538G, Y537N, Y537S, and Y537C, in helix 12 expand its motion. The results of RMSD, RMSF, b-factor, radius of gyration, and cross-correlation analysis are given in **Figures S1–S5**. These results suggest that the major fluctuation and destabilization effect was caused by E380Q, S463P, Y537S, Y537C, Y537N, and D538G. Thus, it can be inferred that the antagonist and co-activator binding conformation is stabilized by these mutations. When compared to the wild type apo helix-12, this region of the six mutations E380Q, S463P, Y537S, Y537C, Y537N, and D538G was found to have greater flexibility (RMSF **Figure S2**).

To give further insight into conformational changes induced by these mutations, secondary structure switches in mutants, changes in relative solvent accessibility, residue depth in Å, and residue-occluded packing densities were determined. From the maximum destabilizing mutations, increases in RSA, residue-occluded packing densities, and decrease in depth were observed and are tabulated in **Table S1**.

Herein, using mCSM-lig, affinity changes upon mutations were calculated, and average changes ranged from 0.823 to −3.033 kcal/mol was reported. Mutations, such as E380Q with affinity fold change −1.399 kcal/mol, S463P with affinity fold change of −1.305 kcal/mol, Y537S, Y537N, and Y537C with affinity fold changes of −1.098, −0.878, and −0.931 kcal/mol, were found to be primarily affected in highest fold with destabilizing effect. In addition, mutation D538G with an affinity fold change of −0.909 kcal/mol was also clustered in destabilizing mutants. Mutations E380Q, S463P, Y537S, Y537N, Y537C, and D538G were spotted in Helix-12 whose flexibility is affected by these mutations and thus poses a significant flexibility drift. It has been previously reported that the flexibility and replacement of Helix-12 can cause major destruction on the binding of ligands in the active site (Kuang et al., 2018). The active site residues lie on Helix-3 and Helix-11, which is affected by the motion of Helix-12. According to mCSM results, the remaining mutations, such K303R, V392I, V524E, P535H, and P536H, do not induce significant changes in affinity or reported to be least influenced. The results obtained from these analyses are tabulated in **Table 3**.

Interaction Analysis

Important residues, such as Glu353, Arg394, and His524, are important for antagonist activity. Upon docking with a wild type and mutants, these residues were considered for the bonding analysis. In case of wild type as given in **Figure 3,** a strong hydrogen bond with His524 and Glu353 can be easily seen, while, in case of mutations, such bonds are especially absent with His524. Thus, we speculate that these mutations also disturb the bonding pattern of fulvestrant with ESR1. Furthermore, other hydrophobic and electrostatic interactions are also formed and

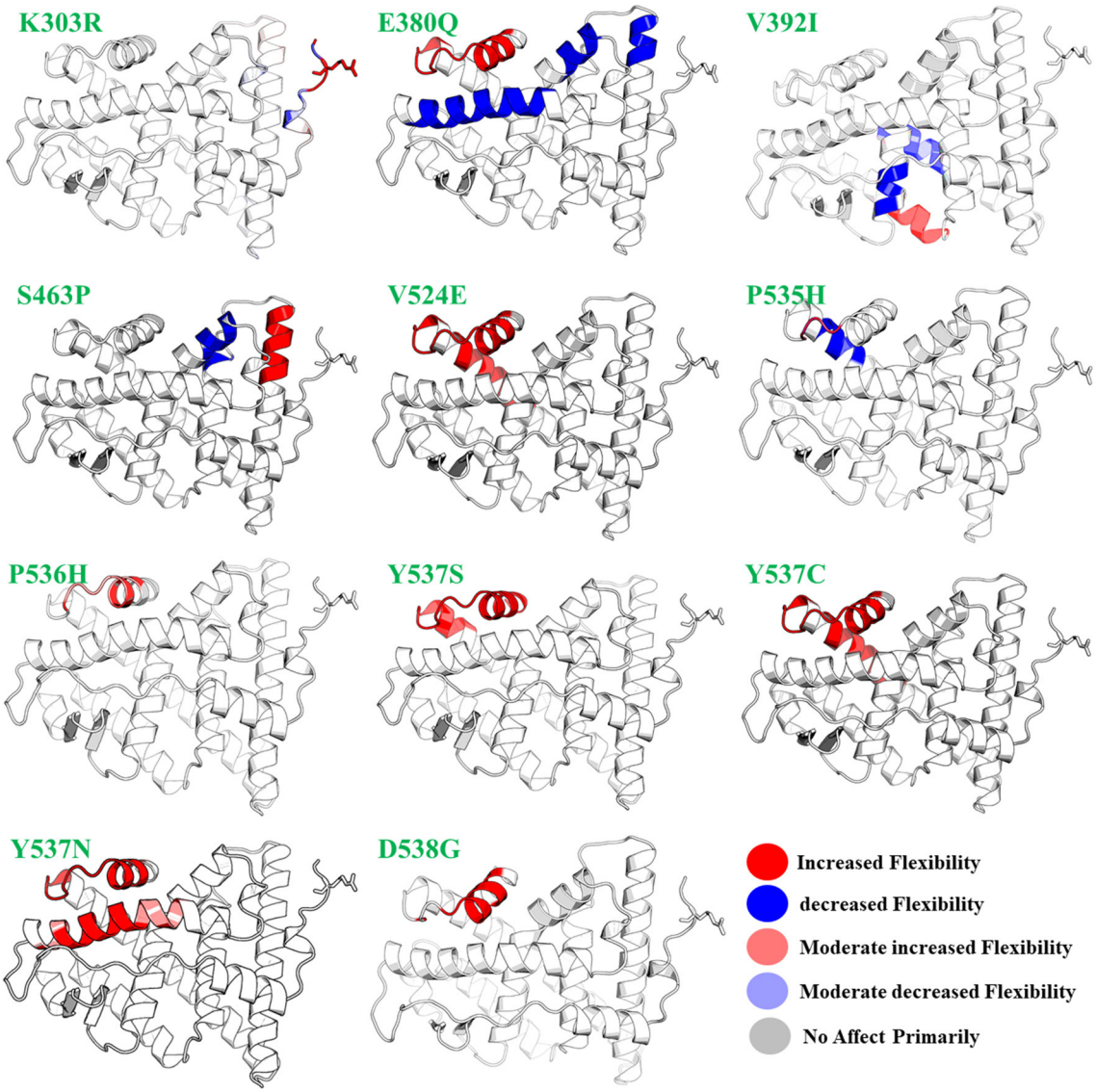


FIGURE 2 | The effect of mutations on the flexibility of different residues. Different colors represent different levels of flexibility.

lost. It is also important that targeting Cys530 residue could help in resolving the resistance posed by these mutations (Furman et al., 2019). The number of hydrogen bonds before and after the MD simulation and their lengths are given in **Table S2**.

Dynamics and Conformational Transition of Wild and Mutated ESR1

Dynamics features of all the systems, including RMSD, RMSF, per-residue RMSF analysis, radius of gyration, and distances of important atoms/residues of both wild and mutant systems were calculated after a total of 1,400 ns of simulation time. Different effects of these mutations differentially affected the dynamics of these systems. While calculating RMSD, it was found that the wild system attained the steady-state soon after reaching 10 ns. While compared to the wild type, these systems (D538G, E380Q, and S463P) and the RMSD values are relatively high, which is due to the induction of mutation in the H12 helix, and thus affect the stability of the protein by targeting the specific residues. It has been discussed in our results that these mutations, compared to the other Y537S/N/C, did not affect the protein significantly. On the other hand, mutations induced at Y537S, Y537C, and Y537N greatly affected the protein dynamics and stability. It can be seen that the mutations induced in the H12 at Y537 residue posses higher RMSD than the wild system, which, as has been discussed in the above results, means that these mutations significantly affect the protein dynamics, and thus stability is mostly influenced. These results are consistent with the results predicted by different methods that the mutation Y537S/N/C affects the protein in a higher fold as compared to the other three. Previous studies based on other antagonists also suggested that

TABLE 3 | FoldX energy changes and mCSM-lig ligand binding affinity fold change prediction upon mutation in ESR1.

Index	Mutations	ΔΔS ENCoM	Affinity fold change	Docking score
1.	Wild	00	0.00	−10.5
2.	K303R	0.032	−0.445	−9.54
3.	E380Q	0.338	−1.399	−6.27
4.	D538G	0.355	−0.909	−8.97
5.	Y537S	0.340	−1.098	−8.63
6.	Y537N	1.011	−0.878	−8.87
7.	H524E	0.109	−0.429	−8.7
8.	V392I	0.25	0.154	−10.5
9.	S463P	0.758	−1.305	−9.54
10.	P535H	0.1	0.481	−6.27
11.	L536H	0.517	0.444	−8.97
12.	Y537C	0.699	−0.931	−8.63

For the wild type, the experimental concentration already reported 0.138 nM was used.

Y537S/N/C is less stable than the others. These mutations are previously prioritized to be primarily treated for the successful treatment of breast cancer. Thus, the results we obtained are consistent with the results predicted in other algorithms and pervious studies (Jeselsohn et al., 2015; Kuang et al., 2018). All the RMSDs are given in **Figure 4.**

Furthermore, residual fluctuations and fluctuation of the H12 helix residues were calculated as RMSF and per-residue RMSF. Analyses of the root mean square fluctuations (RMSFs) by residue in the wild and mutant complexes were compared. It was found that fluctuation in the H1 helix was observed due to its continues loop structures and accommodated by few helix residues. During the simulation time, fluctuation in these residues was observed to be higher. We also observed high dynamic activity in a loop preceding H9 helix (residues 160–166). The most important feature of these wild and complex systems was understanding the flexibility of H12 helix in all the systems. It has been previously shown that hydrogen bonds formed by His524 residue with the antagonist could reduce the motion of H11 and H12 helices and thus maintain the antagonist conformation of the protein and avoid the binding of co-activator which could lead to the reduced activity of this co-complex in breast cancer. Here, a per-residue RMSF analysis was correlated with the mutations, and the RMSF of each residue lies in the H12 helix was calculated. As given in **Figure 5**, it can be easily observed that the mutation affected the residual flexibility of the H12 helix in a higher fold when compared with the wild system. It can be seen that residual flexibility of each of these residues in H12 helix is higher than that of the wild system. The flexibility of H12 increases with the simulation time. Thus, these mutations substantially affected this helix.

Experimental studies proposed the mechanism that the movement of H12 is mainly responsible for the conformational transition between agonist and antagonist and the binding of a co-activator. Thus, the loss of important hydrogen and other bonds with the enhanced flexibility reduces the efficacy of the antagonist and thus fails to restrain the movement of the H12 helix and halt the binding of co-activator to the ESR1. Therefore, the increased flexibility favors the binding of a co-activator by supporting the agonist conformation, which was already reported in the previous study. These hotspot mutations, specifically Y537S, Y537C, and Y537N, affect the flexibility of the H12 in the highest fold and are shown in **Figure 5** (right panel). Previous results also suggested that these mutations primarily destabilizes the binding of antagonists and reduces the interactions with the H12 by enhancing the conformational transition. Our results also support the early results reported by different studies: the role of H12 in terms of its flexibility is the primary cause of resistance (Merenbakh-Lamin et al., 2013; Robinson et al., 2013; Toy et al., 2013), and Y537S/C/N are significantly involved in the resistance to fulvestrant (O'Leary et al., 2018).

To further demonstrate the mechanism of resistance posed by the H12 helix, we calculated the distance between the ligand and residue His524, an essential parameter in determining the role of agonist and antagonist conformation stabilization. As given in **Figure 6**, the distance between the ligand and His524 is minimal in the wild system, while in the other systems with mutations this distance was found to be increased range from 0.15 to 0.18 nm in different mutations. The average distance in the mutated systems was ∼0.15 nm. Hence, the movement of H11 and H12 is highly correlated, and we speculate that the loss of important interactions due to the flexibility could allosterically affect the H11. It has also been previously shown that this is due to the rotation of His524 from a –guache to +guache conformation. However, the bond between Glu339 of H3 and Lys531 of H11 ties together and avoid the unwinding of the protein. Here, the fluctuation of bond distance between the ligand and His524 is also an important factor in determining the stability of each conformation (agonist and antagonist) because a shift in the orientation between agonist and antagonist has been previously reported to be associated with interaction with His524 too. The radius of gyration is given in **Figure 5**.

Trajectories Transitions and Dynamic Motions

A PCA (Principal component analysis) was applied to study and used to evaluate the distinct protein conformational states in a principal component (PC) phase space during the molecular dynamics simulations. Trajectories were projected onto a two-dimensional subspace using the first three eigenvectors, i.e., PC1, PC2, and PC3, to understand the conformational transitions of the complexes. **Figure 7** shows that all the complexes attained two conformational states on the subspace differently colored (**Figure 7**). The conjoined distributions of the principal components of the complexes discovered that the energetically unstable conformational state blue neared convergence and attaining a stable conformational state red color. Consequently, different periodic jumps are required for the transition of different conformations in mutants.

To further understand the transition mechanism of mutants and wild complexes from metastable to native states, eigenvectors

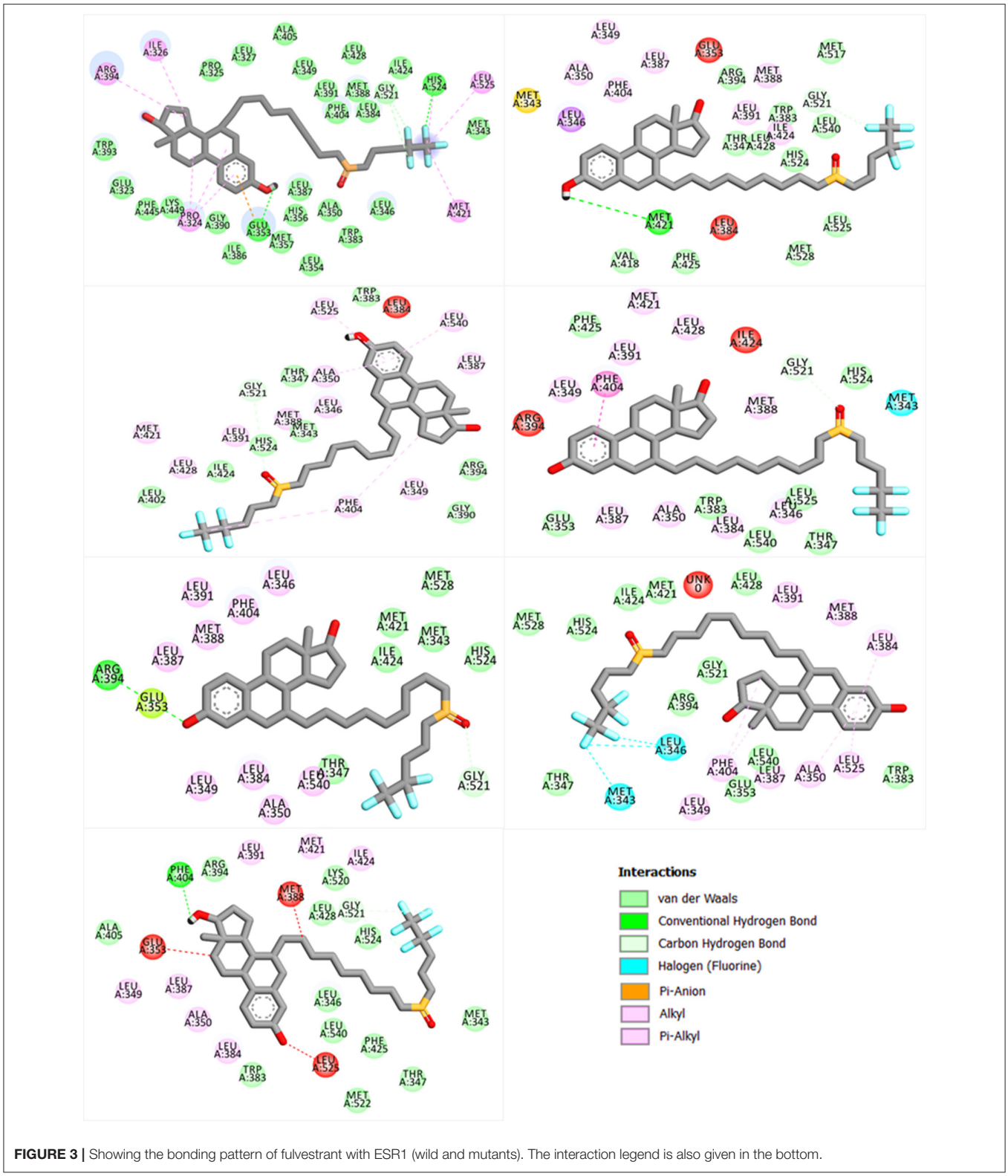


FIGURE 3 | Showing the bonding pattern of fulvestrant with ESR1 (wild and mutants). The interaction legend is also given in the bottom.

(the first two) were used to calculate and to plot the free energy landscape (FEL) of the 200 ns trajectory time. Low energy states were extracted to understand structural evolution. As shown in **Figure 8**, the lowest Gibbs energy states are highlighted using a dark purple color, while the numbers represent the positions of the structural coordinates sampled from that locus on the FEL plot. The FEL plot showed that the wild type attained four different energy states (three metastable and one native). On

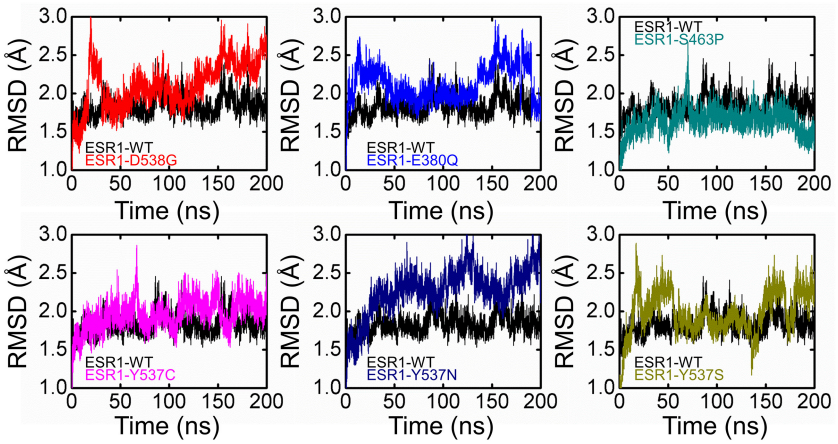


FIGURE 4 | Root Mean square deviation of all the systems compared with the wild type. The black color is showing the wild while the rest of the colors represent the mutant systems.

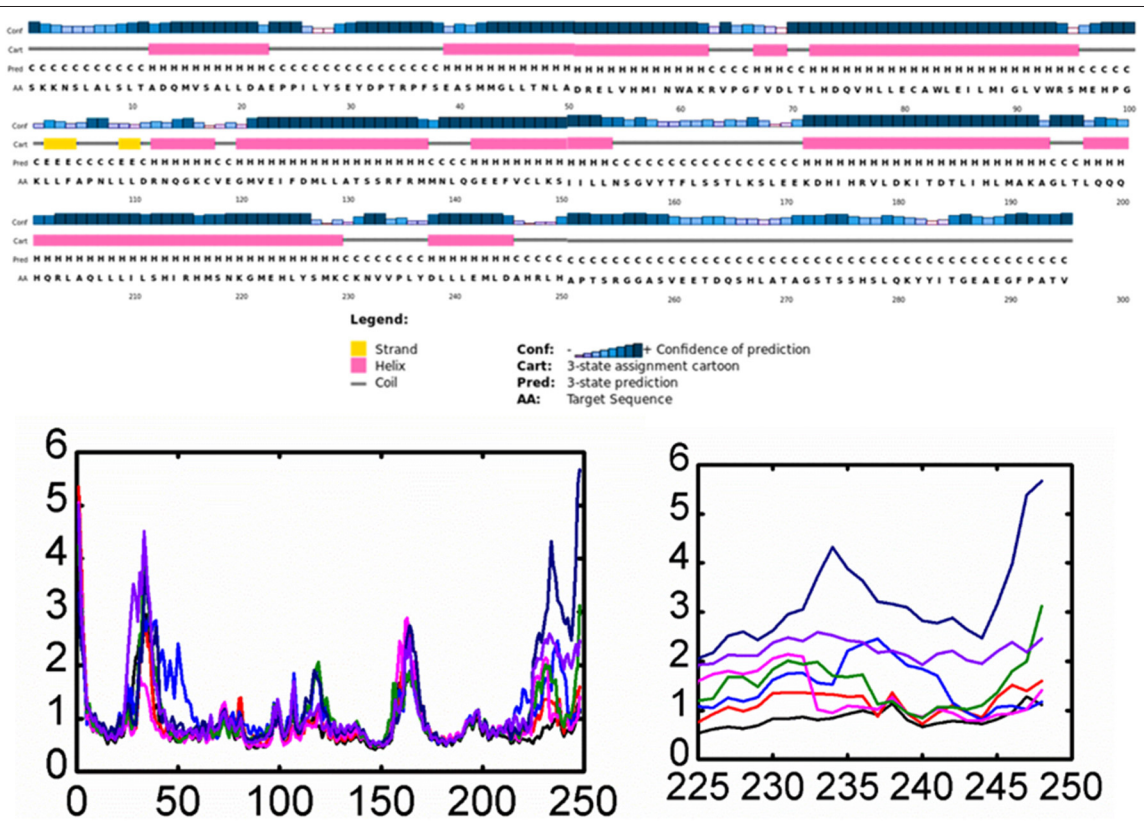


FIGURE 5 | Root Mean square fluctuation of all the systems compared with the wild type. The left panel is showing the RMSF and the complete structure of ESR1. The right panel is showing the fluctuation of Helix-12 which is confirming the flexibility drift caused by the mutations. Each system is shown in a different color.

the other hand, mutations D538G, S463P, E380Q, and Y537S were found to have three states, including one native and two metastable states. Mutations Y537C and Y537N probably formed two metastable states only. The result indicates the frequent transition of the conformations in the mutants compared to wild type. Mutations adopted multiple metastable states during their structural evolution in MD simulations and were separated by low- and high-energy barriers, respectively. Interestingly, mutations like Y537C, Y537N, and Y537S were observed in its profoundly transition state by observing the RMSD plot.

To further examine the residue's correlative motions, trajectories were subjected to a dynamics cross-correlation

analysis. A diverse pattern of correlations was observed in all the systems. The atomic displacement in mutants was observed to be high when compared to wild. It can be seen that highly atomic displacement of H12 in Y537C/N/S is experienced, while in the case of E380Q and S463P these motions are observed in multiple atoms of different residues. The results given in **Figure 9** clearly show that all mutants display different correlated motions than the wild complex.

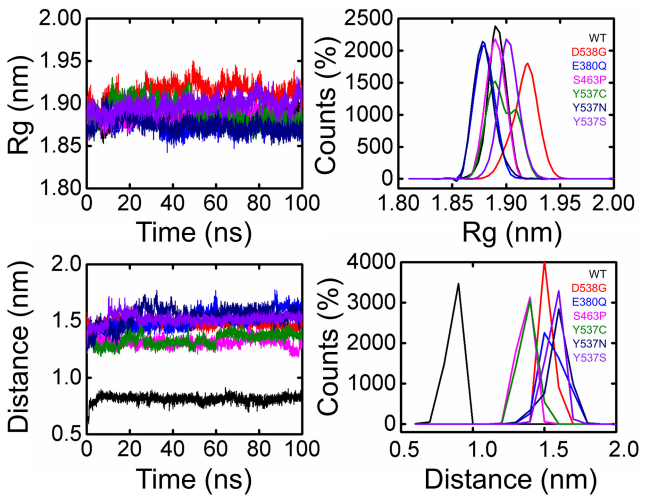


FIGURE 6 | Radius of gyration and Distance of 5,000 snapshots obtained from each system are shown in different color.

DISCUSSION

In practice, ER proved to be the prime target for BC therapy, but poor response or complete resistance developes during the course of treatment, making treatment a grim challenge and BC lethal. Comprehension of these mechanisms at a cellular and genetic level is of paramount importance to evade this muddle and come up with an effective treatment. Fulvestrant is a potent antagonist, and it has been characterized to reduce the burden of breast cancer. The resistance to fulvestrant, caused by genetic aberrations in ESR1, has been reported (Shi et al., 2014; Akhmetova et al., 2015; Khan et al., 2018), but the molecular mechanism subsequently leading to resistance has not yet been elucidated. Here, we used a logical approach to understand the mechanism's underlying resistance to targetting ER, using structure-guided approaches. The present examination gives insight into the mechanism behind the fulvestrant resistance, which could help in designing new anti-BC drugs. We adopted an extensive computational procedure to unveil the molecular mechanisms of resistance offered by ESR1 mutations

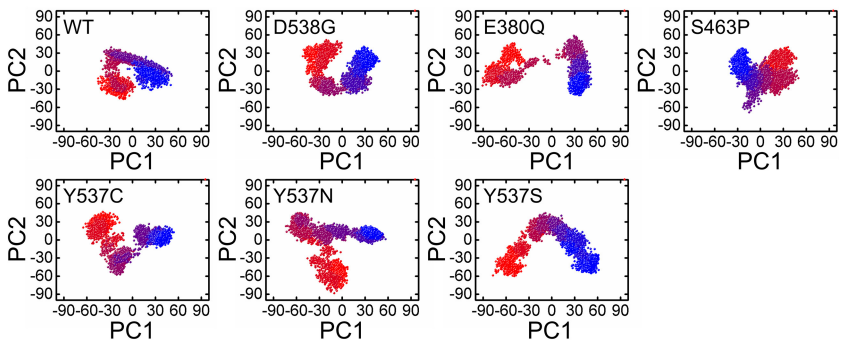


FIGURE 7 | Principle component analysis of 5000 snapshopts obtained from MD simulation of all the systems.

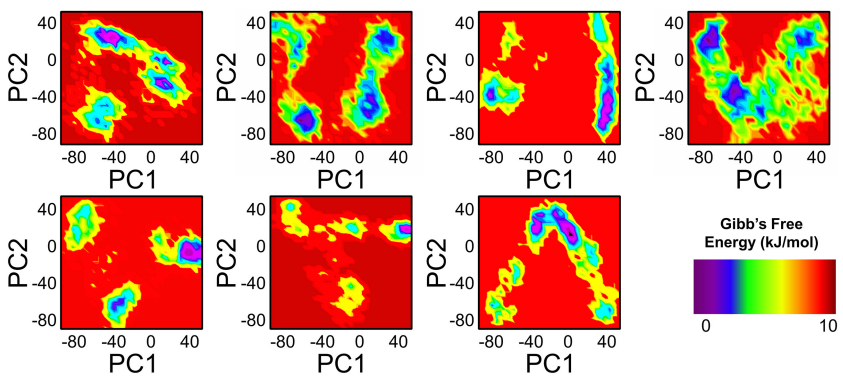


FIGURE 8 | FEL of PC1 and PC2 obtained from MD simulation of all the systems.

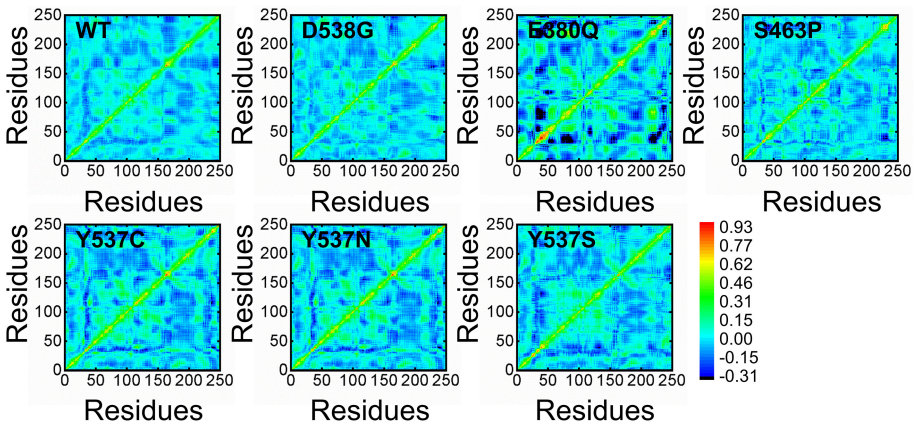


FIGURE 9 | Distance cross-correlation matrix of both wild and mutant systems.

to fulvestrant. The activity of a native protein is affected by an aberration that can occur anywhere, not only in the residues of the active site. Previous studies demonstrated that such changes have a remarkable impact on the structure and action of fulvestrant. In the present investigation, the Y537S, Y537C, and Y537N mutations significantly influence the activity of the fulvestrant drug. Using multiple servers, such as mCSM, SDM, DUET and many others, the servers reported that six mutations have significantly affected the activity of the fulvestrant. Also, a molecular dynamics simulation revealed that the structural stability and flexibility directly correlated to the mutation. All these analyses suggested that the flexibility of H12 could open up the co-activator confirmation due to enhanced flexibility. Differences in the docking and free energies also clarified the distortion caused by these mutations. Furthermore, the principal component analysis, free energy landscape, and dynamics cross-correlation analysis also clarified the dominant motions, native and metastable state, and correlated motions in ESR1 (wild and mutant systems). These methods have previously been used by different studies to understand protein dynamics, mechanim of resistance, and drug interaction (Du et al., 2016, 2019; Yang et al., 2018).

In conclusion, we quantified the impact of reported mutations K303R, E380Q, V392I, S463P, V524E, P535H, P536H, Y537C, Y537N, Y537S, and D538G in the activity of fulvestrant. This study clarified how these mutations alter structural properties, binding affinity, stability, and resistance in breast cancer treatment. Our results provide further understanding into the factors associated with drug resistance in breast cancer cell lines and thus provide a useful pathway for the development of new medications for treatment of breast cancer.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

AK conceptualized the methodology. AK, A-U-R, and MJ did the analysis. FH, SSa, SA, C-DL, SSh, and D-QW wrote the manuscript. AK and ZB revised the manuscript and performed all the revised analyses. D-QW supervised the study. All authors approved the manuscript.

FUNDING

D-QW was supported by the grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (Contract Nos. 61832019, 61503244), the Natural Science Foundation of Henan Province (162300410060), and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The computations were partially performed at the Center for High-Performance Computing, Shanghai Jiao Tong University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2019.00159/full#supplementary-material

REFERENCES

Akhmetova, A., Kozhamkulov, U., Bismilda, V., Chingissova, L., Abildaev, T., Dymova, M., et al. (2015). Mutations in the pncA and rpsA genes among 77 *Mycobacterium tuberculosis* isolates in Kazakhstan. *Int. J. Tubercul. Lung Dis.* 19, 179–184. doi: 10.5588/ijtld.14.0305

Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R., Bonn, T., Engström, O., et al. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389, 753–758. doi: 10.1038/39645

Chandarlapaty, S., Chen, D., He, W., Sung, P., Samoila, A., You, D., et al. (2016). Prevalence of ESR1 mutations in cell-free DNA and outcomes in metastatic breast cancer: a secondary analysis of the BOLERO-2 clinical trial. *JAMA Oncol.* 2, 1310–1315. doi: 10.1001/jamaoncol.2016.1279

Chen, F., Liu, H., Sun, H., Pan, P., Li, Y., Li, D., et al. (2016). Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking. *Phys. Chem. Chem. Phys.* 18, 22129–22139. doi: 10.1039/C6CP03670H

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Biolo. Crystallogr.* 66, 12–21. doi: 10.1107/S0907444909042073

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: an N·log(N) method for ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397

De Savi, C., Bradbury, R. H., Rabow, A. A., Norman, R., de Almeida, C., Andrews, D. M., et al. (2015). Optimization of a novel binding motif to (E)-3-(3, 5-difluoro-4-((1 R, 3 R)-2-(2-fluoro-2-methylpropyl)-3-methyl-2, 3, 4, 9-tetrahydro-1 H-pyrido [3, 4-b] indol-1-yl) phenyl) acrylic Acid (AZD9496), a potent and orally bioavailable selective estrogen receptor downregulator and antagonist. *J. Med. Chem.* 58, 8128–8140. doi: 10.1021/acs.jmedchem.5b00984

Delfosse, V., Grimaldi, M., Pons, J-L., Boulahtouf, A., Le Maire, A., Cavailles, V., et al. (2012). Structural and mechanistic insights into bisphenols action provide guidelines for risk assessment and discovery of bisphenol A substitutes. *Proce. Natl. Acad. Sci. U.S.A.* 109, 14930–14935. doi: 10.1073/pnas.1203574109

Du, J., Wang, X., Dong, C. H., Yang, J. M., and Yao, X. (2016). Computational study of the binding mechanism of actin-depolymerizing factor 1 with actin in *Arabidopsis thaliana*. *PLoS ONE* 11:e0159053. doi: 10.1371/journal.pone.0159053

Du, Q., Qian, Y., Yao, X., and Xue, W. (2019). Elucidating the tight-binding mechanism of two oral anticoagulants to factor Xa by using induced-fit docking and molecular dynamics simulation. *J. Biomol. Struct. Dyn.* 37: 1–9. doi: 10.1080/07391102.2019.1583605

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh ewald method. *J. Chem. Phys.* 103, 8577–8593. doi: 10.1063/1.470117

Fanning, S. W., Mayne, C. G., Dharmarajan, V., Carlson, K., Martin, T. A., Novick, S. J., et al. (2016). Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. *Elife* 5:e12792. doi: 10.7554/eLife.12792.038

Frappier, V., Chartier, M., and Najmanovich, R. J. (2015). ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability. *Nucleic Acids Res.* 43, W395–W400. doi: 10.1093/nar/gkv343

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein– ligand complexes. *J. Med. Chem.* 49, 6177–6196. doi: 10.1021/jm051256o

Furman, C., Hao, M. H., Prajapati, S., Reynolds, D., Rimkunas, V., Zheng, G., et al. (2019). Estrogen receptor covalent antagonists: the best is yet to come. *Cancer Res.* 79, 1740–1745. doi: 10.1158/0008-5472.CAN-18-3634

Gotz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *J. Chem. Theory Comput.* 8, 1542–1555. doi: 10.1021/ct200909j

Guo, S., Zhang, C., Bratton, M., Mottamal, M., Liu, J., Ma, P., et al. (2018). ZB716, a steroidal selective estrogen receptor degrader (SERD), is orally efficacious in blocking tumor growth in mouse xenograft models. *Oncotarget* 9:6924. doi: 10.18632/oncotarget.24023

Hoang, T. X., Trovato, A., Seno, F., Banavar, J. R., and Maritan, A. (2004). Geometry and symmetry presculpt the free-energy landscape of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7960–7964. doi: 10.1073/pnas.0402525101

Hou, T., Li, N., Li, Y., and Wang, W. (2012). Characterization of domain–peptide interaction interface: prediction of SH3 domain-mediated protein–protein interaction network in yeast by generic structure-based models. *J. Proteome Res.* 11, 2982–2995. doi: 10.1021/pr3000688

Jensen, E. V., and Jordan, V. C. (2003). The estrogen receptor: a model for molecular medicine. *Clin. Cancer Res.* 9, 1980–1989.

Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M., and Schiff, R. (2015). ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nat. Rev. Clin. Oncol.* 12, 573–583. doi: 10.1038/nrclinonc.2015.117

Jeselsohn, R., Yelensky, R., Buchwalter, G., Frampton, G., Meric-Bernstam, F., Gonzalez-Angulo, A., et al. (2014). Emergence of constitutively active estrogen receptor-α mutations in pretreated advanced estrogen receptor–positive breast cancer. *Clin. Cancer Res.* 20, 1757–1767. doi: 10.1158/1078-0432.CCR-13-2332

Jordan, V. C., Curpan, R., and Maximov, P. Y. (2015). Estrogen receptor mutations found in breast cancer metastases integrated with the molecular pharmacology of selective ER modulators. *JNCI: J. Natl. Cancer Inst.* 107:djv075. doi: 10.1093/jnci/djv075

Joseph, J. D., Darimont, B., Zhou, W., Arrazate, A., Young, A., Ingalla, E., et al. (2016). The selective estrogen receptor downregulator GDC-0810 is efficacious in diverse models of ER+ breast cancer. *Elife* 5:e15828. doi: 10.7554/eLife.15828.026

Khan, M. T., Khan, A., Rehman, A. U., Wang, Y., Akhtar, K., Malik, S. I., et al. (2019). Structural and free energy landscape of novel mutations in ribosomal protein S1 (rpsA) associated with pyrazinamide resistance. *Sci. Rep.* 9:7482. doi: 10.1038/s41598-019-44013-9

Khan, M. T., Malik, S. I., Ali, S., Sheed Khan, A., Nadeem, T., Zeb, M. T. et al. (2018). Prevalence of pyrazinamide resistance in Khyber Pakhtunkhwa, Pakistan. *Microb. Drug Resist.* 24, 1417–1421. doi: 10.1089/mdr.2017.0234

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2015). PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213. doi: 10.1093/nar/gkv951

Kuang, Y., Siddiqui, B., Hu, J., Pun, M., Cornwell, M., Buchwalter, G., et al. (2018). Unraveling the clinicopathological features driving the emergence of ESR1 mutations in metastatic breast cancer. *NPJ Br. Cancer* 4:22. doi: 10.1038/s41523-018-0075-5

Lai, A., Kahraman, M., Govek, S., Nagasawa, J., Bonnefous, C., Julien, J., et al. (2015). Identification of GDC-0810 (ARN-810), an orally bioavailable selective estrogen receptor degrader (SERD) that demonstrates robust activity in tamoxifen-resistant breast cancer xenografts. *J. Med. Chem.* 58, 4888–4904. doi: 10.1021/acs.jmedchem.5b00054

Land, H., and Humble, M. S. (2018). YASARA: a tool to obtain structural guidance in biocatalytic investigations. *Methods Mol. Biol.* 1685, 43–67. doi: 10.1007/978-1-4939-7366-8_4

Liang, J., and Shang, Y. (2013). Estrogen and cancer. *Ann. Rev. Physiol.* 75, 225–240. doi: 10.1146/annurev-physiol-030212-183708

Magistrato, A., Sgrignani, J., Krause, R., and Cavalli, A. (2017). Single or multiple access channels to the CYP450s active site? An answer from free energy simulations of the human aromatase enzyme. *J. Phy. Chem. Lett.* 8, 2036–2042. doi: 10.1021/acs.jpclett.7b00697

Merenbakh-Lamin, K., Ben-Baruch, N., Yeheskel, A., Dvir, A., Soussan-Gutman, L., Jeselsohn, R., et al. (2013). D538G mutation in estrogen receptor-α: a novel mechanism for acquired endocrine resistance in breast cancer. *Cancer Res.* 73, 6856–6864. doi: 10.1158/0008-5472.CAN-13-1197

Miller, B. R., III, McGee, T. D. Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* 8, 3314–3321. doi: 10.1021/ct300418h

Nilsson, S., and Gustafsson, J. Å. (2011). Estrogen receptors: therapies targeted to receptor subtypes. *Clin. Pharmacol. Therap.* 89, 44–55. doi: 10.1038/clpt.2010.226

O'Leary, B., Cutts, R. J., Liu, Y., Hrebien, S., Huang, X., Fenwick, K., et al. (2018). The genetic landscape and clonal evolution of breast cancer resistance to palbociclib plus fulvestrant in the PALOMA-3 trial. *Cancer Dis.* 8, 1390–1403. doi: 10.1158/2159-8290.CD-18-0264

Osborne, C., Wakeling, A., and Nicholson, R. (2004). Fulvestrant: an oestrogen receptor antagonist with a novel mechanism of action. *Br. J. Cancer* 90:S2. doi: 10.1038/sj.bjc.6601629

Pandurangan, A. P., Ochoa-Montaño, B., Ascher, D. B., and Blundell, T. (2017). SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* 45, W229–W235. doi: 10.1093/nar/gkx439

Pavlin, M., Spinello, A., Pennati, M., Zaffaroni, N., Gobbi, S., Bisi, A., et al. (2018). A computational assay of estrogen receptor α antagonists reveals the key

common structural traits of drugs effectively fighting refractory breast cancers. *Scienti. Rep.* 8:649. doi: 10.1038/s41598-017-17364-4

Pearson, K. (1901). LIII. on lines and planes of closest fit to systems of points in space. *Lond. Edinbur. Dublin Philosoph. Magaz. J. Sci.* 2, 559–572. doi: 10.1080/14786440109462720

Petrella, S., Gelus-Ziental, N., Maudry, A., Laurans, C., Boudjelloul, R., and Sougakoff, W. (2011). Crystal structure of the pyrazinamidase of *Mycobacterium tuberculosis*: insights into natural and acquired resistance to pyrazinamide. *PLoS ONE* 6:e15785. doi: 10.1371/journal.pone.0015785

Pires, D. E., Ascher, D. B., and Blundell, T. L. (2013). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 30, 335–342. doi: 10.1093/bioinformatics/btt691

Pires, D. E., Blundell, T. L., and Ascher, D. B. (2016). mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Scienti. Rep.* 6:29575. doi: 10.1038/srep29575

Robinson, D. R.,Wu, Y. M., Vats, P., Su, F., Lonigro, R., Cao, X., et al. (2013). Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Gene.* 45, 1446–1451. doi: 10.1038/ng.2823

Rodrigues, C. H., Pires, D. E., and Ascher, D. B. (2018). DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* 46, W350–W355. doi: 10.1093/nar/gky300

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5

Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013). An overview of the amber biomolecular simulation package. *Wiley Interdiscipl. Rev.* 3, 198–210. doi: 10.1002/wcms.1121

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The foldx web server: an online force field. *Nucleic Acids Res.* 33(Suppl. 2), W382–W388. doi: 10.1093/nar/gki387

Scientific, D., and San Carlos, C. (2002). *The PyMOL Molecular Graphics System*. DeLano, WL, 2.40–0.44.

Sgrignani, J., Bon, M., Colombo, G., and Magistrato, A. (2014). Computational approaches elucidate the allosteric mechanism of human aromatase inhibition: a novel possible route to small-molecule regulation of CYP450s activities? *J. Chem. Inform. Model.* 54, 2856–2868. doi: 10.1021/ci500425y

Sgrignani, J., Cavalli, A., Colombo, G., and Magistrato, A. (2016). Enzymatic and inhibition mechanism of human aromatase (CYP19A1) enzyme a computational perspective from QM/MM and classical molecular dynamics simulations. *Mini Rev. Med. Chem.* 16, 1112–1124. doi: 10.2174/1389557516666160623101129

Sgrignani, J., Iannuzzi, M., and Magistrato, A. (2015). Role of water in the puzzling mechanism of the final aromatization step promoted by the human aromatase enzyme. insights from QM/MM MD simulations. *J. Chem. Inform. Model.* 55, 2218–2226. doi: 10.1021/acs.jcim.5b00249

Shi, W., Chen, J., Feng, J., Cui, P., Zhang, S., Weng, X., et al. (2014). Aspartate decarboxylase (PanD) as a new target of pyrazinamide in *Mycobacterium tuberculosis*. *Emerg. Microbes Infect.* 3:e58. doi: 10.1038/emi.2014.61

Spinello, A., and Magistrato, A. (2017). An omics perspective to the molecular mechanisms of anticancer metallo-drugs in the computational microscope era. *Exp. Opin. Drug Dis.* 12, 813–825. doi: 10.1080/17460441.2017.1340272

Spoerke, J. M., Gendreau, S., Walter, K., Qiu, J., Wilson, T., Savage, H., et al. (2016). Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant. *Nat. Commun.* 7:11579. doi: 10.1038/ncomms11579

Sun, H., Duan, L., Chen, F., Liu, H., Wang, Z., Pan, P., et al. (2018). Assessing the performance of MM/PBSA and MM/GBSA methods. 7 entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* 20, 14450–14460. doi: 10.1039/C7CP07623A

Sun, H., Li, Y., Tian, S., Xu, L., and Hou, T. (2014). Assessing the performance of MM/PBSA and MM/GBSA methods. 4. accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* 16, 16719–29. doi: 10.1039/C4CP01388C

Toy, W., Shen, Y., Won, H., Green, B., Sakr, R. A., Will, M., et al. (2013). ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat. Gene.* 45, 1439–1445. doi: 10.1038/ng.2822

van Kruchten, M., de Vries, E. G., Glaudemans, A. W., van Lanschot, M., Kema, I. P., et al. (2015). Measuring residual estrogen receptor availability during fulvestrant therapy in patients with metastatic breast cancer. *Cancer Dis.* 5, 72–81. doi: 10.1158/2159-8290.CD-14-0697

Xu, L., Sun, H., Li, Y., Wang, J., and Hou, T. (2013). Assessing the performance of MM/PBSA and MM/GBSA methods. 3. the impact of force fields and ligand charge models. *J. Phys. Chem.* 117, 8408–8421. doi: 10.1021/jp404160y

Yang, F., Zheng, G., Fu, T., Li, X., Tu, G., Li, Y., et al. (2018). Prediction of the binding mode and resistance profile for a dual-target pyrrolyl diketo acid scaffold against HIV-1 integrase and reverse-transcriptase-associated ribonuclease H. *Phys. Chem. Chem. Phys.* 20, 23873–23884. doi: 10.1039/C8CP01843J

Zwanzig, R. (1973). Nonlinear generalized langevin equations. *J. Statist. Phys.* 9, 215–220. doi: 10.1007/BF01008729

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

BRIEF RESEARCH REPORT
published: 04 March 2020
doi: 10.3389/fmolb.2020.00029

Molecular Simulation of Oncostatin M and Receptor (OSM–OSMR) Interaction as a Potential Therapeutic Target for Inflammatory Bowel Disease

Qingqing Du[1], *Yan Qian*[1]* *and Weiwei Xue*[2]*

[1] Department of Pharmacy, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China, [2] School of Pharmaceutical Sciences, Chongqing Key Laboratory of Natural Product Synthesis and Drug Research, Chongqing University, Chongqing, China

Therapeutics targeting cytokines such as the oncostatin M (OSM)-mediated inflammation represent a potential strategy for the treatment of inflammatory bowel disease (IBD). Despite the investigation of the specific role of the interactions between OSM and the receptor (OSMR) in IBD pathogenesis, the 3D structure of the OSM–OSMR complex remains elusive. In this work, the interaction mode between OSM and OSMR at atomic level was predicted by computational simulation approach. The interaction domain of the OSMR was built with the homology modeling method. The near-native structure of the OSM–OSMR complex was obtained by docking, and long-time scale molecular dynamics (MD) simulation in an explicit solvent was further performed to sample the conformations when OSM binds to the OSMR. After getting the equilibrated states of the simulation system, per-residue energy contribution was calculated to characterize the important residues for the OSM–OSMR complex formation. Based on these important residues, eight residues (OSM: Arg100, Leu103, Phe160, and Gln161; OSMR: Tyr214, Ser223, Asp262, and Trp267) were identified as the "hot spots" through computational alanine mutagenesis analysis and verified by additional MD simulation of R100A (one of the identified "hotspots") mutant. Moreover, six cavities were detected at the OSM–OSMR interface through the FTMap analysis, and they were suggested as important binding sites. The predicted 3D structure of the OSM–OSMR complex and the identified "hot spots" constituting the core of the binding interface provide helpful information in understanding the OSM–OSMR interactions, and the detected sites serve as promising targets in designing small molecules to block the interactions.

Keywords: inflammatory bowel disease, oncostatin M and oncostatin M Receptor, protein-protein docking, molecular dynamics simulation, binding sites prediction

OPEN ACCESS

Edited by:
Guang Hu,
Soochow University, China

Reviewed by:
Fei Ye,
Zhejiang Sci-Tech University, China
Yong Wang,
University of Copenhagen, Denmark

***Correspondence:**
Yan Qian
cqqianyan@hospital.cqmu.edu.cn
Weiwei Xue
xueww@cqu.edu.cn

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *24 December 2019*
Accepted: *11 February 2020*
Published: *04 March 2020*

Citation:
Du Q, Qian Y and Xue W (2020) Molecular Simulation of Oncostatin M and Receptor (OSM–OSMR) Interaction as a Potential Therapeutic Target for Inflammatory Bowel Disease. Front. Mol. Biosci. 7:29. doi: 10.3389/fmolb.2020.00029

INTRODUCTION

Inflammatory bowel diseases (IBDs) are complex chronic inflammatory conditions of the gastrointestinal tract that are driven by perturbed signal pathways of cytokines such as tumor necrosis factor (TNF)-α and IL-6 (Neurath, 2014). Nowadays, anti-TNF antibodies (such as infliximab and golimumab) are mainstay therapies for IBD (Choi et al., 2017). However, there are

still more than 40% of patients who are non-responsive to anti-TNF agents, making the discovery of alternative therapeutic targets a priority (Kim et al., 2017). One of those potential targets, oncostatin M (OSM)-mediated inflammation, has gained a lot of interest (Verstockt et al., 2019). It is found that high pretreatment expression of OSM is strongly associated with failure of anti-TNF therapy of patients with IBD, which revealed the role of the receptor (OSMR) as part of a unique pathway that contributes to the chronicity of intestinal inflammation (West et al., 2017).

OSM belongs to the IL-6 family, and the activation of the OSM signal pathway is highly determined by the high affinity of OSM to the receptor (OSMR) (Adrian-Segarra et al., 2018a,b). The crystal structure of OSM reveals that the protein comprises four α helices ranging from 15 to 22 amino acids in length (termed A, B, C, and D) and linked by polypeptide loops (**Figure 1A**) (Deller et al., 2000). The OSMR is a member of the IL-6 receptor family that transduces signaling events of OSM (Yu et al., 2019). Currently, available antibodies, such as GSK315234 and GSK2330811, have already been proven to affect the OSM signal (Verstockt et al., 2019). Although neutralizing OSM antibodies are being developed and should be considered as a novel proof-of-concept trial in IBD patients (West et al., 2017), these developed biological medicines are large, complex, and relatively fragile molecules, which make them difficult and expensive to produce and administer on a large scale (Monaco et al., 2015).

In recent years, development of small molecule modulators targeting protein–protein interactions (PPIs) has emerged as a promising therapeutic intervention in complex diseases (Nero et al., 2014; Nim et al., 2016; Weng et al., 2019). In selecting biologically relevant protein–protein interfaces, the availability of computer-aided drug design (CADD) approach has led to the discovery of small molecules either stabilizing or disrupting the biological processes (Arkin et al., 2014; Laraia et al., 2015). The critical role for OSM in antipathogen immunity has not been described, and targeting OSM–OSMR may offer inhibition of the inflammatory pathology while preserving protective immunity (Verstockt et al., 2019). These hypotheses stimulate the idea of identification of small molecular inhibitors against the OSM–OSMR interface, which might provide safer and more broadly effective alternatives to conventional antibodies targeting monomeric macromolecules. To discover ligands specifically disrupting the OSM–OSMR interface, the information of the protein–protein interactions is needed. Unfortunately, the 3D structure of the OSM–OSMR complex remains elusive (Kim et al., 2017). It is of paramount importance to understand the details of the OSM and OSMR complex formation as well as the potential binding site between the protein–protein interface.

In this work, molecular simulation approaches aimed at filling the aforementioned gap were performed to accelerate the discovery of small molecules targeting OSM–OSMR. Starting from the crystal structure of OSM (Deller et al., 2000) and the model of the OSMR [a protein-binding region was built using the leukemia inhibitory factor receptor (LIFR) crystal structure (Huyton et al., 2007) as a template], the near-native conformation of the OSM–OSMR complex was obtained through protein–protein docking. The docking conformation was further sampled through long-time scale (1 μs) molecular dynamics (MD) simulation to get the equilibrated binding states. Based on the simulation trajectory, per-residue binding free energy decomposition (Tu et al., 2018; Wang et al., 2019) and computational alanine scanning (CAS) (Huo et al., 2002) analysis were carried out to identify the protein–protein interface "hotspots." Using one of the identified "hotspots" (Arg100) as an example, an additional 500 ns of MD simulation was performed to investigate the stability of the R100A mutant complex. Finally, the "hotspots" were mapped to the seven binding sites located at the OSM–OSMR interface detected using FTMap (Kozakov et al., 2011), and three of them were suggested as important target sites for future designs of small molecular modulators in the OSM–OSMR interaction.

MATERIALS AND METHODS

Structure Preparation

Construction of OSM Missing Loop

The crystal structure and sequence of OSM were obtained from the PDB database (PDB code: 1EVS) (Deller et al., 2000). Residues from 1 to 3 and 135 to 155 (highlighted in red color in **Table S1**) were missing in the resolved crystal structure. The coordinates of the missing fragments of the OSM structure were constructed using the optimization-based approach (Fiser et al., 2000) in *Modeler* (Webb and Sali, 2016).

Homology Modeling of OSMR

The full-length sequence of the OSMR was obtained from the NCBI database (GenBank: AAI25210.1) (Strausberg et al., 2002). Then the sequence of the OSMR was submitted to search a template structure with the *BLAST* algorithm (Schaffer et al., 2001). Searching result showed that the sequence identity between the OSMR and LIFR was higher than 30%, especially in the protein-binding domain (57%). Therefore, based on the crystal structure (PDB code: 2Q7N) (Huyton et al., 2007) of the LIFR (residues from 201 to 383), 10 homology models of the OSMR protein-binding domain was constructed using *Modeler* (Webb and Sali, 2016).

Protein–Protein Docking

OSM–OSMR docking was performed using the protein docking module of the latest version of Rosetta (Alford et al., 2017). Before docking, the PDB structures of OSM and OSMR were first formed through the script of *clean_pdb.py*. The formed structures of the two proteins were refined by running the Rosetta relax protocol, and the PDB files consisting of refined OSM and OSMR were generated. Then, according to the knowledge of the residues of OSM for OSMR binding detected by site mutagenesis studies (Adrian-Segarra et al., 2018b), the generated two complexes were loaded into *PyMOL* (Schrödinger, 2010) and with OSM reoriented to contact with the OSMR. To ensure low-energy starting side-chain conformations for docking, further prepacking of the OSM and OSMR complexes were conducted. Finally, 10,000 poses were calculated for the OSM–OSMR interactions using the Monte Carlo (MC) refinement method

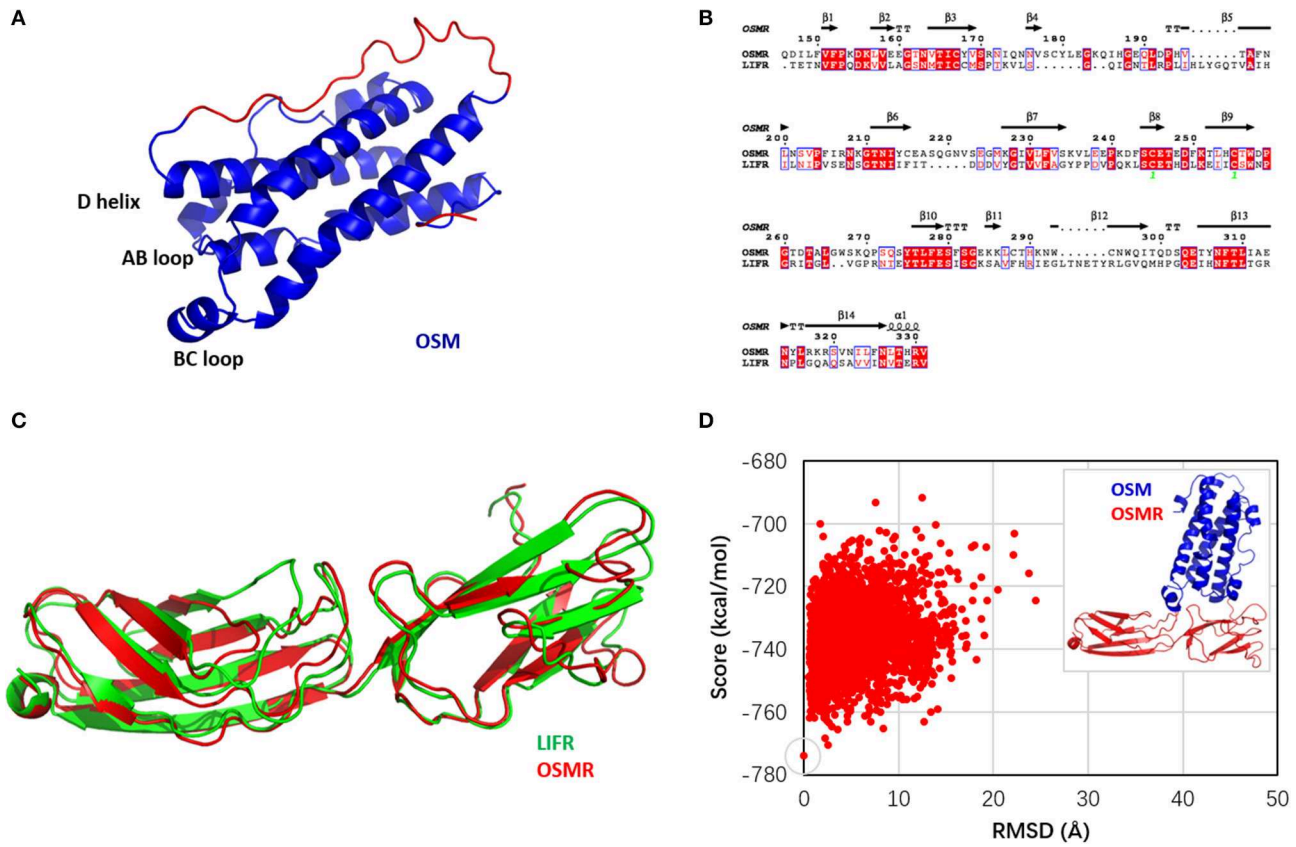


FIGURE 1 | (A) Structure of oncostatin M (OSM); the modeled fragments are colored in red. **(B)** Sequence alignment between oncostatin M receptor (OSMR) and leukemia inhibitory factor receptor (LIFR). **(C)** Structural alignment of OSMR homology model (red) and LIFR crystal structure (green). **(D)** Docking funnel of OSM and OSMR. Inset: the top scoring conformation as near-native OSM–OSMR structure.

(Gray et al., 2003), with the pre-packed conformation as a starting point.

Docking Funnel Analysis

With *InterfaceAnalyzer* mover in *RosettaScripts* (Fleishman et al., 2011), the RMSD was calculated from the heavy atoms of the interface residues (**Table S2**) using each pose of the top five scorers as a reference structure (Chaudhury et al., 2011). The docking funnel was then identified through plotting total_score against RMSD. Finally, the top scoring structure with the lowest RMSD was selected as the successful pose of the OSM–OSMR complex.

Molecular Dynamics Simulation

Molecular dynamics (MD) simulation was performed with GPU-accelerated *PMEMD* in *AMBER14* (Babin et al., 2014). The selected near-native structure of OSM–OSMR from Rosetta docking was used as the initial conformation for MD simulation. The LEaP (Wang et al., 2006) was applied to assign *AMBERff14SB* force field parameters (Maier et al., 2015) for the two proteins, and two disulfide bonds in OSM and one disulfide bonds in the OSMR were identified and added. The complex was immersed into a rectangular periodic box of TIP3P (Hornak et al., 2006) water molecules, and the system was neutralized with two chloride ions. The distance between any protein atom and the edge of the box was set to 10 Å, and the prepared system contains 86,446 atoms per periodic cell. Starting from the representative snapshot of wild type OSM–OSMR, additional MD simulation was performed on the R100A complex using the same setup.

MM/GBSA Binding Free Energy

The binding free energy (ΔG_{tol}) between OSM and OSMR was estimated by the end-point molecular mechanics generalized Born surface area (MM/GBSA) approach (Kollman et al., 2000) as below:

$$\Delta G_{\text{tol}} = \Delta E_{\text{vdW}} + \Delta E_{\text{ele}} + \Delta G_{\text{pol}} + \Delta G_{\text{nonpol}} \quad (1)$$

where ΔE_{vdW} and ΔE_{ele} are the van der Waals and electrostatic interaction energies, and ΔG_{pol} and ΔG_{nonpol} are the polar and non-polar solvent energies, respectively. ΔE_{vdW} and ΔE_{ele} were calculated using *AMBER ff14SB* (Maier et al., 2015) in the gas phase. ΔG_{pol} was calculated by solving the GB equation (Onufriev et al., 2004) with the dielectric constants of solute and solvent set to 1 and 80, respectively. ΔG_{nonpol} was calculated by $\Delta G_{\text{nonpol}} = \gamma \times SASA$, where $\gamma = 0.0072$, and *SASA* is referred

to the solvent-accessible area and determined using a water probe radius of 1.4 Å (Sitkoff et al., 1994).

To further analyze the energy contribution between OSM and OSMR at a per-residue basis ($\Delta G_{\text{calc}}^{\text{per-residue}}$), the total binding free energy was decomposed by:

$$\Delta G_{calc}^{\text{per-residue}} = \Delta E_{\text{vdW}}^{\text{per-residue}} + \Delta E_{\text{ele}}^{\text{per-residue}} + \Delta G_{\text{pol}}^{\text{per-residue}} + \Delta G_{\text{nonpol}}^{\text{per-residue}} \quad (2)$$

The definition of each term in Equation (2) is similar as in Equation (1), except that *SASA* was computed by recursively approximating a sphere around an atom, starting from an icosahedron (ICOSA) (Babin et al., 2014).

Computational Alanine Scanning Mutagenesis

Computational alanine scanning (CAS) mutagenesis was widely used to characterize the "hotspots" associated to protein–protein interactions (Huo et al., 2002). The whole process included the generation of mutated snapshots, and the binding free energy difference ($\Delta\Delta G_{\text{calc}}$) between the wild type (WT) and mutant (MUT) complex is calculated below

$$\Delta\Delta G_{\text{calc}} = \Delta G_{\text{MUT}} - \Delta G_{\text{WT}} \quad (3)$$

where G_{WT} andG_{MUT} refer to the MM/GBSA binding free energy of the WT and MUT complexes, respectively. Snapshot(s) of the WT of OSM-OSM and LIF-LIFR complex were collected from the last 500-ns trajectory and the crystal structure 2Q7N (Huyton et al., 2007), respectively. Alanine mutation was generated by truncating the selected mutation residue at Cγ and by replacing Cγ with a hydrogen atom at a 1.09-Å distance from Cß along the direction of the Cγ-Cß bond (Huo et al., 2002).

Detection of Druggable Binding Sites

Based on the representative snapshot of the OSM–OSMR structure derived from the long-time simulation and the crystal structure of LIF–LIFR (Huyton et al., 2007), FTMap (Kozakov et al., 2011) was employed to detect the druggable binding site in the protein–protein interaction complexes. FTMap uses a fragment-based mapping algorithm that implements an efficient fast Fourier transform (FFT) correlation approach to search a global protein surface for potential druggable binding sites. The fragments include 16 small organic probe molecules (benzene, cyclohexane, ethane, ethanol, isopropanol, isobutanol, acetone, acetaldehyde, dimethyl ether, acetonitrile, urea, methylamine, phenol, benzaldehyde, acetamide, and N, N-dimethylformamide) of varying sizes, shapes, and polarities (Kozakov et al., 2015).

RESULTS AND DISCUSSION

Modeled Structures of OSM and OSMR

The missing structures of OSM (**Table S1**), including the N-terminal fragment (1–3, AAI) and loop (135–155, SDTAEPTKAGRGASQPPTPTP), were built and refined using *Modeler* (Webb and Sali, 2016) because sequence identity between the loops of OSM and LIF (SKYHVGHVDVTYGPDTSGKDV) was only 10.3%. In addition, structural alignment indicated that the conformations of the two terminals that link the loops in the crystal structures of 1EVS and 2Q7N was significantly different (**Figure S1**). Therefore, the missing loop of OSM was predicted based on its own crystal structure 1EVS. Homology modeling approach in *Modeler* (Webb and Sali, 2016) was applied to provide the 3D structure of the OSMR binding domain (146–331) using the LIFR crystal structure (PDB code: 2Q7N) (Huyton et al., 2007) as a template. **Figure 1B** shows that the sequence identity between the OSMR and LIFR binding domain was 57%. As a result, 10 models were predicted for OSM and OSMR, respectively, and the model for each of them (**Figures 1A,C**) was selected by picking the structure with the best DOPE assessment score considering the Lennard–Jones potential and GBSA implicit solvent interaction (Shen and Sali, 2006).

Prediction of OSM–OSMR Interaction Profiles

The Near-Native Conformation of OSM–OSMR Complex

To predict the OSM–OSMR binding funnel, RosettaDock was used to sample 10,000 poses from the starting position. The starting position was estimated according to the knowledge of binding site residues identified by site mutagenesis studies (Adrian-Segarra et al., 2018b), as the presence of a docking funnel is considered to be the most robust measure of success in a docking simulation (Chaudhury et al., 2011). Here, the top five scorers of the OSM–OSMR complex were used as references to plot the docking score of all 10,000 poses as a function of RMSD (**Figure S2**). One of the top five structures presenting the most reasonable docking funnel, in which the near-native conformations consistently have better scores than the non-native conformations (Chaudhury et al., 2011), is shown in **Figure 1D**. Therefore, the top scoring structures with the lowest RMSD in **Figure 1D** was selected as the initial conformation of OSM–OSMR for further studies. In addition, given that the structure of OSM is very similar to that of LIF, and the OSMR is modeled using the LIFR as the template, the structure of the OSM–OSMR complex was modeled based on the crystal structure of the LIF–LIFR complex. The calculated RMSD between the template-based and docking structures of the OSM–OSMR was 3.37 Å, suggesting that the two modeled structures are very similar with each other (**Figure S3A**). However, several spatial clashes were found between the interface of OSM and OSMR in the template-based OSM–OSMR complex (**Figure S3B**). As a result, it is proposed that the docking pose of the OSM–OSMR complex is more suitable for further investigation.

The Simulated Equilibration States of OSM–OSMR Complex

Starting from the docking conformation, 1 μs of all-atom MD simulation was performed for OSM–OSMR in explicit water. The time evolution of the RMSD of the Cα atom of proteins

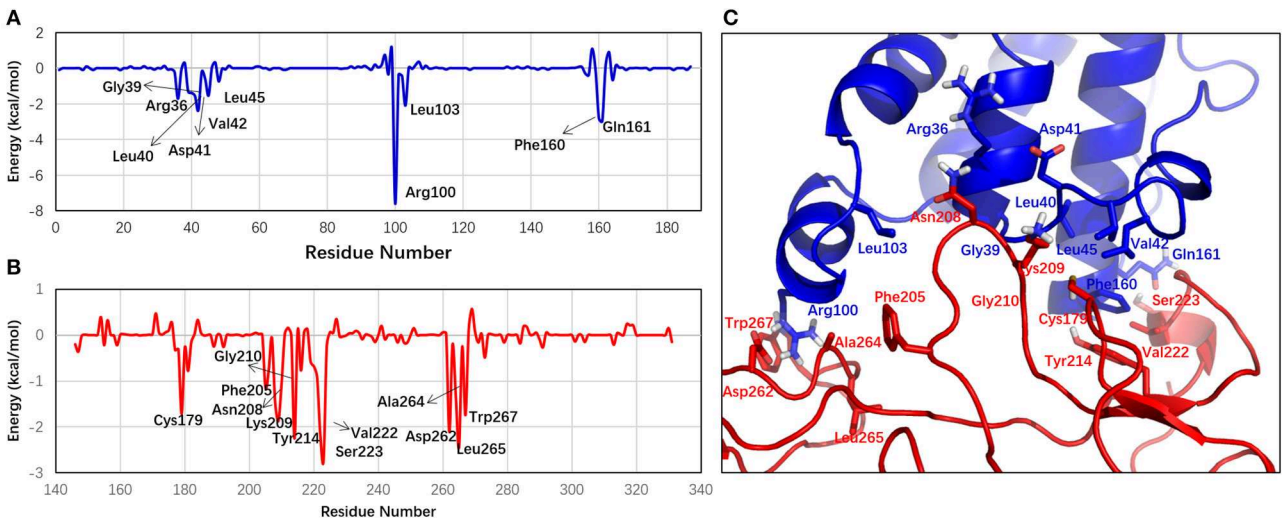


FIGURE 2 | Per-residue energy profiles in **(A)** OSM and **(B)** OSMR contribute to the formation the complex. **(C)** The cartoon representation of the interaction mode of OSM–OSMR interface. Only the important residues (the absolute energy contribution ≥1 kcal/mol) are labeled.

with respect to the initial coordinates of the docking pose is shown in **Figure S4A**. The RMSD values of OSM (∼4 Å) and OSMR (∼5 Å) showed that the two protein partners underwent conformation changes over the course of the simulation. In addition, compared with OSM and OSMR, the higher RMSD of the complex (∼6 Å) suggested that the rotation of the two-partner orientation occurred. The extended root mean square fluctuation (RMSF) analysis of the protein residues indicated that the loop residues (135–155) in OSM were more flexible during the simulation; however, the interface residues in both OSM and OSMR were stabilized due to the non-bond interactions (Huang et al., 2019) between the two proteins (**Figures S4B,C**). Compared with the RMSF analysis of the OSM residues (**Figure S4B**) with the plot of B-factor of the LIF residues (12–180) in the crystal structure 2Q7N (**Figure S5**) indicated that OSM shares a similar structural fluctuation with LIF, especially in the loop region (135–155).

The Thermodynamics Properties of OSM–OSMR Complex

To characterize the thermodynamics properties between OSM and OSMR interaction, the snapshots derived from the last 500-ns equilibrated trajectory were used to estimate the MM/GBSA (Kollman et al., 2000) binding free energy. The decomposed energy terms of the total binding free energy (G_{tol}) indicated that electrostatic interaction energy (E_{ele}, −338.29 ± 44.73 kcal/mol), van der Waals interaction energy (E_{vdW}, −82.02 ± 7.17 kcal/mol), and non-polar solvent energy (G_{nonpol}, −11.74 ± 1.08 kcal/mol) play important roles in the formation of the protein–protein complex, whereas polar solvent energies (G_{polar}, 383.11 ± 43.06 kcal/mol) were unfavored for the interaction.

In addition, per-residue energy decomposition analysis was performed to identify the important residues for the OSM–OSMR complex formation. The residues with an absolute energy contribution of more than 0.5 kcal/mol are listed in **Table S3**. The chart of the per-residue interaction energy and the interaction mode between OSM and OSMR are further shown in **Figure 2**. The per-residue energy decomposition analysis successfully predicted five residues in OSM (AB loop: Gly39, Leu40, Lys44, and Leu45; D helix: Phe160) reported by experiments, which played specific roles in activating OSMR signaling (Adrian-Segarra et al., 2018a,b). In addition, seven new residues (Arg36, Asp41, Val42, Arg100, Leu103, Gln161, and Leu164) in OSM were predicted as the important ones that contribute to the protein–protein interaction. Moreover, the 18 residues (Cys179, Leu181, Phe205, Ile206, Asn208, Lys209, Gly210, Tyr214, Glu216, Gln219, Gly220, Asn221, Val222, Ser223, Asp262, Ala264, Leu265, and Trp267) characterized in the OSMR were informative in experimentally verifying these residues, which may play an important role in OSM and OSMR interaction (Huang et al., 2019).

"Hot Spots" Located at OSM–OSMR Interface

In the context of protein–protein interaction, residues that made major contribution to the binding of free energy were termed as "hot spots," which can be determined by alanine scanning mutagenesis (Zerbe et al., 2012). These "hot-spots" are highly interesting since the protein–protein interaction could be disrupted by targeting them (Grosdidier and Fernandez-Recio, 2008). Herein, to find the "hot spots" located at the OSM–OSMR interface, the computational alanine scanning (CAS) mutagenesis calculation was conducted on the residues with an absolute energy contribution of more than 1 kcal/mol identified by the per-residue energy decomposition analysis (Yang et al., 2018; Du et al., 2020). There were eight "hot spots" (OSM: Arg100, Leu103, Phe160, and Gln161; OSMR: Tyr214, Ser223, Asp262, and Trp267) with a relative binding

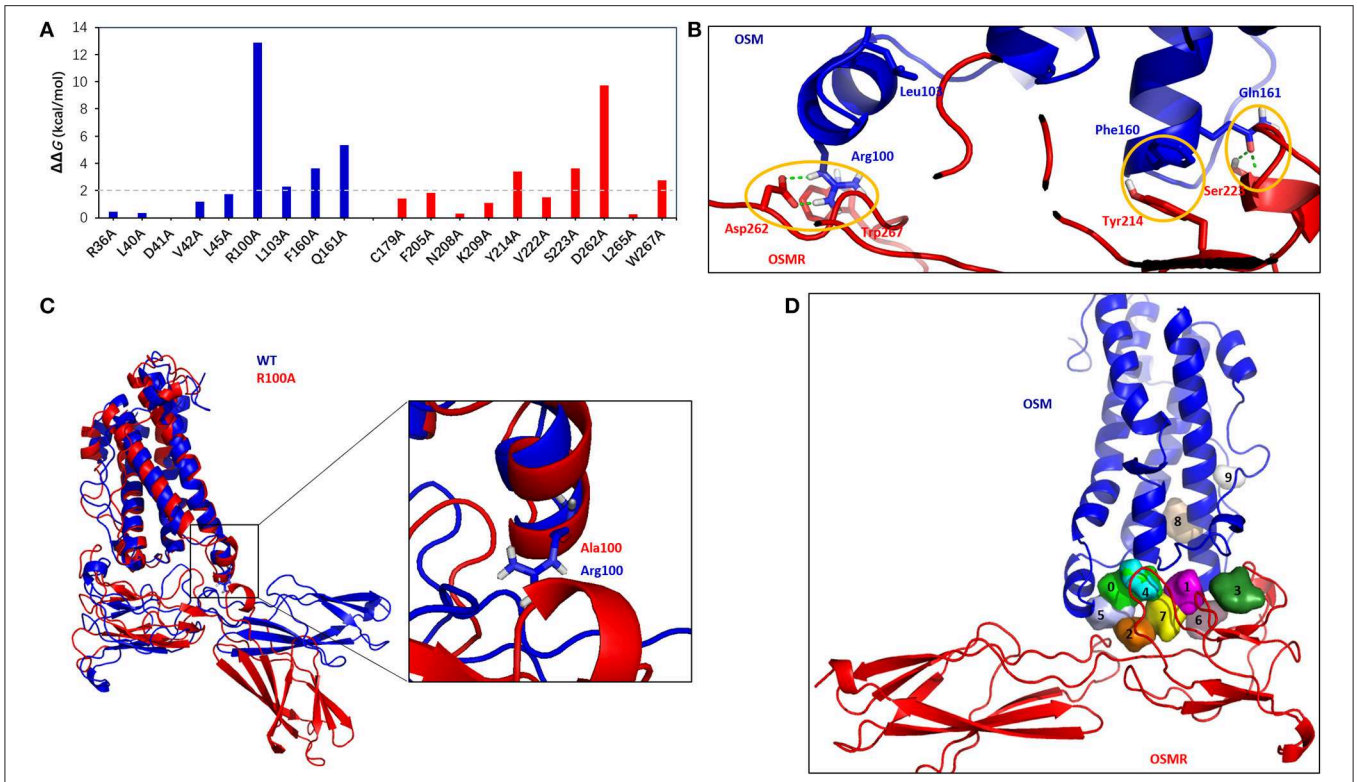


FIGURE 3 | "Hot spots" and potential binding sites located at the OSM–OSMR interface. **(A)** Computational alanine scanning calculation of the 19 residues with absolute energy contribution of more than 1 kcal/mol identified in per-residue energy decomposition analysis. **(B)** Interactions of the eight "hot spots" located at the OSM–OSMR interface. The hydrogen bonds are displayed as green dashes. **(C)** Comparison of the equilibrated state conformation of wild-type OSM–OSMR with the snapshot of R100A mutant after 470-ns molecular dynamics (MD) simulation. **(D)** Potential binding sites in the OSM–OSMR complex identified through FTMap analysis. The detected 10 sites are labeled (0–9) and shown as surface with different colors in the structure.

free energy (*G*) of more than 2 kcal/mol (**Figure 3A**) (Moreira et al., 2007; Tu et al., 2018). **Figure 3B** clearly shows that some important non-bond interactions formed among those "hot spots," such as the hydrogen bonds between Arg100 and Asp262, Gln161, and Ser223, and the π-π interaction between Phe160 and Tyr214. Among them, Phe160 was found to play an important role in OSM–OSMR recognition (Adrian-Segarra et al., 2018a,b). In addition, the other predicted "hot spots," especially R100 and D262, were predicted to have a ΔΔG larger than 8 kcal/mol, which might be very useful for further theoretical and experimental studies. To investigate the stability of the mutant, using R100A complex as an example, an additional MD simulation (500 ns) was performed starting from the representative snapshot of wild-type OSM–OSMR. The calculated RMSD values of the OSM–OSMR complex are shown in **Figure S4D**. It is noted that RMSD significantly increased by around 470 ns for the R100A (∼8 Å) complex. In addition, snapshots with the largest RMSD value during the simulation were extracted and shown in **Figure 3C**. Compared with the equilibrated state conformation of the wild-type OSM–OSMR, significant conformational change near the mutation site occurred in the R100A complex.

Moreover, using the crystal structure of the LIF–LIFR complex (PDB code 2Q7N), CAS analysis was performed on residues (Pro51, Phe52, Pro53, Leu56, Pro106, Leu109, Phe156, Gln157, Ile234, Val258, Asn261, Ser262, Ile267, Ile310, and Leu313) corresponding to the residues (Arg39, Asp41, Val42, Leu45, Arg100, Leu103, Phe160, Gln161, Phe205, Asn208, Lys209, Tyr214, Asp262, and Leu265) located at same position in the OSM–OSMR interface. It is found that Pro106, Phe156 in LIF, and Ile267 in LIFR (**Figure S6**), corresponding to Arg100, Phe160 in OSM, and Tyr214 in OSMR could be regarded as common "hot-spot" residues for both the OSM–OSMR and LIF-LIFR complexes. In the meanwhile, alanine mutations of other residues have little effect in the interaction energy of the LIF–LIFR complex, suggesting that the interface of the LIF–LIFR complex is significantly different from that of the predicted OSM–OSMR.

Detection of Druggable Sites in OSM–OSMR Interface

Through FTMap (Kozakov et al., 2011) analysis of the MD simulation-derived structure of the OSM–OSMR complex, 10 potential druggable binding sites were detected from fragment-based searching of the global protein surface (**Figure 3D**), indicating that the conformation of the residues in the recognition interface is very flexible. To further verify the feasibility of the predicted OSM–OSMR model for potential binding sites analysis, the crystal structure of the LIF–LIFR

TABLE 1 | List of key interacting residues within 4 Å of the bound probe molecules in the detected potential binding sites rendered as spheres in **Figure 3D**.

Site	OSM	OSMR
0	Arg36, Ile37, Gln38, Gly39, Pro93, Asp97, Leu98, Ser101, **Leu103**	Ile206, Arg207, Asn208, Lys209
1	Gln38, Gly39, Leu40, Leu45, **Phe160**, Lys163	Ser178, Cys179, Gly210, Thr211, Asn212, **Tyr214**,
2		Phe205, Ile206, Leu231, Phe232, Val233, Ser234, Ala264, Leu265, Gly266
3	Lys44, Leu45, His48, **Phe160**	Asn176, Val177, Ser178, **Tyr214**, Cys215, Glu216, Ser218, Gln219, Gly220, Val222
4	Arg36, Ile37, Gly39, **Leu103**	Ile206, Arg207, Asn208, Lys209, Gly210
5	Asp97, Leu98, **Arg100**, Ser101	Ile206, Ala264, Leu265, Gly266
6	Asp158, Ala159, **Phe160**	Gln146, Asn212, **Tyr214**, Val222, Lys227, Gly228, Val230
7		Ile206, Gly210, Thr211, Asn212, Leu231, Val233
8	Arg84, Asp87, Leu88, Arg91, Arg162, Glu165, Gly166	
9	Arg84, Pro151, Thr152, Pro153	

The identified "hot spots" by computational alanine mutagenesis were shown in bold.

complex was submitted for FTMap analysis using the same approach. The result showed that a total of 14 potential binding sites were detected, and the location of the position is similar with that of the OSM–OSMR complex (**Figure S7**). Therefore, it could be concluded that the predicted OSM–OSMR model is feasible for binding site prediction analysis. For the OSM–OSMR complex, the protein residues that interacted with bound fragments (within 4 Å) in the binding sites are summarized in **Table 1**. As shown in **Table 1**, six of the 10 sites (sites 0, 1, 3–6) were located at the interface of the OSM and OSMR interaction. Sites 2 and 7 were located in the OSMR (**Table 1**), and sites 8 and 9 were found in OSM (**Table 1** and **Figure 3D**).

To further evaluate the druggability of the detected binding sites in the OSM–OSMR complex, the identified eight "hot spots" were mapped to the protein residues summarized in **Table 1**. Interestingly, two common "hot spots" (OSM: Phe160, OSMR: Tyr214) were found in sites 1, 3, and 6. However, only one hot spot (OSM: Arg100 or Leu103) was found in sites 0, 4, and 5, and no "hot spot" was found in sites 2, 7, 8, and 9. This could be understand through the relationship between "hot spots" and ligand binding "hot spots" in the protein–protein interface, in which additional topological requirements were needed in a "hot spot" for small molecule binding (Zerbe et al., 2012). Therefore, sites 1, 3, and 6 were important target sites for designing inhibitors that may inhibit the protein–protein interaction between OSM and OSMR. In addition, as the binding site analysis was performed on the global protein surface, the predicted sites 2, 7 in OSMR and sites 8, 9 in OSM, especially the latter two sites (**Figure 3D**), which are located far away from the interface, could be regarded as potential allosteric sites.

CONCLUSION

Targeting the OSM and OSMR pathway represents a potential strategy for the treatment of IBD. In this work, the interaction between OSM and OSMR was investigated by employing computational simulation techniques including homology modeling, protein–protein docking, and long-time scale MD simulation. Post-analysis of the equilibrated simulation trajectory characterized seven new residues in OSM and 18 residues characterized in the OSMR as the important ones contributing to the protein–protein interaction. Based on these important residues, computational alanine scanning and FTMap analysis detected eight "hot spots" and six potential binding sites located at the OSM–OSMR interface. It is interesting to note that, compared with the equilibrated state conformation, significant conformational change near the mutation site occurred in the R100A (one of the identified "hot spots") complex during MD simulation. Further mapping of the eight "hot spots" in the detected binding sites suggested that sites 1, 3, and 6 were important target sites, which may be used for designing inhibitors to block OSM and OSMR interaction.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

QD and WX designed the experiments and performed computational simulations. QD, YQ, and WX analyzed the data and wrote the paper.

FUNDING

This research was funded by the Fundamental Research Funds for Central Universities (2019CDYGYB005), Technology Innovation and Application Demonstration Project of Chongqing (cstc2018jscx-msybX0287), and the Young Talents Nursery Foundation of the Second Affiliated Hospital of Chongqing Medical University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00029/full#supplementary-material

REFERENCES

Adrian-Segarra, J. M., Schindler, N., Gajawada, P., Lorchner, H., Braun, T., and Poling, J. (2018a). The AB loop and D-helix in binding site III of human Oncostatin M (OSM) are required for OSM receptor activation. *J. Biol. Chem.* 293, 7017–7029. doi: 10.1074/jbc.RA118.001920

Adrian-Segarra, J. M., Sreenivasan, K., Gajawada, P., Lorchner, H., Braun, T., and Poling, J. (2018b). The AB loop of oncostatin M (OSM) determines species-specific signaling in humans and mice. *J. Biol. Chem.* 293, 20181–20199. doi: 10.1074/jbc.RA118.004375

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., et al. (2017). The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* 13, 3031–3048. doi: 10.1021/acs.jctc.7b00125

Arkin, M. R., Tang, Y., and Wells, J. A. (2014). Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.* 21, 1102–1114. doi: 10.1016/j.chembiol.2014.09.001

Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T., et al. (2014). *AMBER, version 14.* California, SF: University of California.

Chaudhury, S., Berrondo, M., Weitzner, B. D., Muthu, P., Bergman, H., and Gray, J. J. (2011). Benchmarking, and analysis of protein docking performance in Rosetta v3.2. *PLoS ONE* 6:e22477. doi: 10.1371/journal.pone.0022477

Choi, S. Y., Kang, B., Lee, J. H., and Choe, Y. H. (2017). Clinical use of measuring trough levels and antibodies against infliximab in patients with pediatric inflammatory bowel disease. *Gut Liver* 11, 55–61. doi: 10.5009/gnl16041

Deller, M. C., Hudson, K. R., Ikemizu, S., Bravo, J., Jones, E. Y., and Heath, J. K. (2000). Crystal structure and functional dissection of the cytostatic cytokine oncostatin M. *Structure* 8, 863–874. doi: 10.1016/S0969-2126(00)00176-3

Du, Q., Qian, Y., Yao, X., and Xue, W. (2020). Elucidating the tight-binding mechanism of two oral anticoagulants to factor Xa by using induced-fit docking and molecular dynamics simulation. *J. Biomol. Struct. Dyn.* 38, 625–633. doi: 10.1080/07391102.2019.1583605

Fiser, A., Do, R. K., and Sali, A. (2000). Modeling of loops in protein structures. *Protein Sci.* 9, 1753–1773. doi: 10.1110/ps.9.9.1753

Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., et al. (2011). RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS ONE* 6:e20161. doi: 10.1371/journal.pone.0020161

Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., et al. (2003). Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* 331, 281–299. doi: 10.1016/S0022-2836(03)00670-3

Grosdidier, S., and Fernandez-Recio, J. (2008). Identification of hot-spot residues in protein-protein interactions by computational docking. *BMC Bioinformatics* 9:447. doi: 10.1186/1471-2105-9-447

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi: 10.1002/prot.21123

Huang, D., Qi, Y., Song, J., and Zhang, J. Z., H. (2019). Calculation of hot spots for protein-protein interaction in p53/PMI-MDM2/MDMX complexes. *J. Comput. Chem.* 40, 1045–1056. doi: 10.1002/jcc.25592

Huo, S., Massova, I., and Kollman, P. A. (2002). Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.* 23, 15–27. doi: 10.1002/jcc.1153

Huyton, T., Zhang, J. G., Luo, C. S., Lou, M. Z., Hilton, D. J., Nicola, N. A., et al. (2007). An unusual cytokine:Ig-domain interaction revealed in the crystal structure of leukemia inhibitory factor (LIF) in complex with the LIF receptor. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12737–12742. doi: 10.1073/pnas.0705577104

Kim, W. M., Kaser, A., and Blumberg, R. S. (2017). A role for oncostatin M in inflammatory bowel disease. *Nat. Med.* 23, 535–536. doi: 10.1038/nm.4338

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33, 889–897. doi: 10.1021/ar000033j

Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., et al. (2015). The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* 10, 733–755. doi: 10.1038/nprot.2015.043

Kozakov, D., Hall, D. R., Chuang, G. Y., Cencic, R., Brenke, R., Grove, L. E., et al. (2011). Structural conservation of druggable hot spots in protein-protein interfaces. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13528–13533. doi: 10.1073/pnas.1101835108

Laraia, L., McKenzie, G., Spring, D. R., Venkitaraman, A. R., and Huggins, D. J. (2015). Overcoming chemical, biological, and computational challenges in the development of inhibitors targeting protein-protein interactions. *Chem. Biol.* 22, 689–703. doi: 10.1016/j.chembiol.2015.04.019

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255

Monaco, C., Nanchahal, J., Taylor, P., and Feldmann, M. (2015). Anti-TNF therapy: past, present and future. *Int. Immunol.* 27, 55–62. doi: 10.1093/intimm/dxu102

Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2007). Computational alanine scanning mutagenesis–an improved methodological approach. *J. Comput. Chem.* 28, 644–654. doi: 10.1002/jcc.20566

Nero, T. L., Morton, C. J., Holien, J. K., Wielens, J., and Parker, M. W. (2014). Oncogenic protein interfaces: small molecules, big challenges. *Nat. Rev. Cancer* 14, 248–262. doi: 10.1038/nrc3690

Neurath, M. F. (2014). Cytokines in inflammatory bowel disease. *Nat. Rev. Immunol.* 14, 329–342. doi: 10.1038/nri3661

Nim, S., Jeon, J., Corbi-Verge, C., Seo, M. H., Ivarsson, Y., Moffat, J., et al. (2016). Pooled screening for antiproliferative inhibitors of protein-protein interactions. *Nat. Chem. Biol.* 12, 275–281. doi: 10.1038/nchembio.2026

Onufriev, A., Bashford, D., and Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55, 383–394. doi: 10.1002/prot.20033

Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., et al. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 29, 2994–3005. doi: 10.1093/nar/29.14.2994

Schrödinger, L. (2010). *The PyMOL Molecular Graphics System, version 1.3*, New York, NY: Schrödinger, LLC.

Shen, M. Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 15, 2507–2524. doi: 10.1110/ps.062416606

Sitkoff, D., Sharp, K. A., and Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98, 1978–1988. doi: 10.1021/j100058a043

Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., et al. (2002). Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16899–16903. doi: 10.1073/pnas.242603899

Tu, G., Fu, T., Yang, F., Yao, L., Xue, W., and Zhu, F. (2018). Prediction of GluN2B-CT1290-1310/DAPK1 interaction by protein(-)peptide docking and molecular dynamics simulation. *Molecules* 23:3018. doi: 10.3390/molecules23113018

Verstockt, S., Verstockt, B., and Vermeire, S. (2019). Oncostatin M as a new diagnostic, prognostic and therapeutic target in inflammatory bowel disease (IBD). *Expert Opin. Ther. Targets* 23, 943–954. doi: 10.1080/14728222.2019.1677608

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem. Rev.* 119, 9478–9508. doi: 10.1021/acs.chemrev.9b00055

Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247–260. doi: 10.1016/j.jmgm.2005.12.005

Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELER. *Curr. Protoc. Bioinformatics* 54, 5.6.1–5.6.37. doi: 10.1002/cpbi.3

Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D., et al. (2019). HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res.* 47, W322–W330. doi: 10.1093/nar/gkz397

West, N. R., Hegazy, A. N., Owens, B. M. J., Bullers, S. J., Linggi, B., Buonocore, S., et al. (2017). Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease. *Nat. Med.* 23, 579–589. doi: 10.1038/nm.4307

Yang, F., Zheng, G., Fu, T., Li, X., Tu, G., Li, Y. H., et al. (2018). Prediction of the binding mode and resistance profile for a dual-target pyrrolyl diketo acid scaffold against HIV-1 integrase and reverse-transcriptase-associated ribonuclease H. *Phys. Chem. Chem. Phys.* 20, 23873–23884. doi: 10.1039/C8CP01843J

Yu, Z., Li, Z., Wang, C., Pan, T., Chang, X., Wang, X., et al. (2019). Oncostatin M receptor, positively regulated by SP1, promotes gastric cancer growth and metastasis upon treatment with Oncostatin M. *Gastric Cancer* 22, 955–966. doi: 10.1007/s10120-019-00934-y

Zerbe, B. S., Hall, D. R., Vajda, S., Whitty, A., and Kozakov, D. (2012). Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces. *J. Chem. Inf. Model.* 52, 2236–2244. doi: 10.1021/ci300175u

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 11 March 2020
doi: 10.3389/fmolb.2020.00041

Selective Inhibition of HDAC1 by Macrocyclic Polypeptide for the Treatment of Glioblastoma: A Binding Mechanistic Analysis Based on Molecular Dynamics

Yang Zhang[1,2], *Tingting Fu*[2], *Yuxiang Ren*[1], *Fengcheng Li*[1], *Guoxun Zheng*[2], *Jiajun Hong*[1], *Xiaojun Yao*[3], *Weiwei Xue*[2*†] and *Feng Zhu*[1,2*†]

[1] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China, [2] School of Pharmaceutical Sciences, Chongqing University, Chongqing, China, [3] State Key Laboratory of Applied Organic Chemistry and Department of Chemistry, Lanzhou University, Lanzhou, China

Glioblastoma (GBM) is the most common and aggressive intracranial malignant brain tumor, and the abnormal expression of HDAC1 is closely correlated to the progression, recurrence and metastasis of GBM cells, making selective inhibition of HDAC1 a promising strategy for GBM treatments. Among all available selective HDAC1 inhibitors, the macrocyclic peptides have gained great attention due to their remarkable inhibitory selectivity on HDAC1. However, the binding mechanism underlying this selectivity is still elusive, which increases the difficulty of designing and synthesizing the macrocyclic peptide-based anti-GBM drug. Herein, multiple computational approaches were employed to explore the binding behaviors of a typical macrocyclic peptide FK228 in both HDAC1 and HDAC6. Starting from the docking conformations of FK228 in the binding pockets of HDAC1&6, relatively long MD simulation (500 ns) shown that the hydrophobic interaction and hydrogen bonding of E91 and D92 in the Loop2 of HDAC1 with the Cap had a certain traction effect on FK228, and the sub-pocket formed by Loop1 and Loop2 in HDAC1 could better accommodate the Cap group, which had a positive effect on maintaining the active conformation of FK228. While the weakening of the interactions between FK228 and the residues in the Loop2 of HDAC6 during the MD simulation led to the large deflection of FK228 in the binding site, which also resulted in the decrease in the interactions between the Linker region of FK228 and the previously identified key amino acids (H134, F143, H174, and F203). Therefore, the residues located in Loop1 and Loop2 contributed in maintaining the active conformation of FK228, which would provide valuable hints for the discovery and design of novel macrocyclic polypeptide HDAC inhibitors.

Keywords: HDAC, macrocyclic peptides, molecular docking, MD simulation, binding free energies, interaction fingerprints

OPEN ACCESS

Edited by:
Hongchun Li,
Shenzhen Institutes of Advanced Technology (CAS), China

Reviewed by:
Quan Zou,
University of Electronic Science and Technology of China, China
Sophie Sacquin-Mora,
UPR9080 Laboratoire de Biochimie Théorique (LBT), France
Xingcheng Lin,
Massachusetts Institute of Technology, United States

***Correspondence:**
Weiwei Xue
xueww@cqu.edu.cn
Feng Zhu
zhufeng@zju.edu.cn

†ORCID:
Weiwei Xue
orcid.org/0000-0002-3285-0574
Feng Zhu
orcid.org/0000-0001-8069-0053

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *20 November 2019*
Accepted: *21 February 2020*
Published: *11 March 2020*

Citation:
Zhang Y, Fu T, Ren Y, Li F, Zheng G, Hong J, Yao X, Xue W and Zhu F (2020) Selective Inhibition of HDAC1 by Macrocyclic Polypeptide for the Treatment of Glioblastoma: A Binding Mechanistic Analysis Based on Molecular Dynamics. Front. Mol. Biosci. 7:41. doi: 10.3389/fmolb.2020.00041

INTRODUCTION

Glioblastoma (GBM) is the most common and aggressive intracranial malignant brain tumor with the median survival duration <2 years in spite of chemotherapy, radiation or surgical resection (Natsume et al., 2019). In the current chemotherapies, such as *temozolomide*, drug resistance is the predominant obstacle (Zhang et al., 2012; Chen et al., 2019; Kim et al., 2019; Rahman et al., 2019; Su et al., 2019; Xingyi et al., 2019). On the basis of the latest experimental results obtained from the large-scale profiling which included the whole exome and RNA sequencing, it can be learnt that genetic and epigenetic mechanisms are involved in the occurrence and progress of glioma cells (Cancer Genome Atlas Research, 2008; Brennan et al., 2013), especially the aberrant epigenetic silencing of genes caused by histone deacetylation (Vaissiere et al., 2008; Cartron et al., 2013). A large number of researches have proven that significant nuclear expression of histone deacetylase 1 (HDAC1) occurred in GBM cells during the process of tumor progression, recurrence, and metastasis (Bhat et al., 2008; Kim et al., 2008; Campos et al., 2011; Li et al., 2016, 2018a; Zhang et al., 2016; Staberg et al., 2017; He et al., 2019; Natsume et al., 2019). In addition, the invasive and proliferative phenotype of GBM cells was found to be related to the overexpression of HDAC1 level (Han et al., 2013). Moreover, HDAC1 inhibitors developed for a variety of tumors have been extensively tested in clinical trials as a single drug or in combination with other chemotherapy agents (Lu et al., 2008; Tan et al., 2010; Campos et al., 2011; Dong et al., 2013; Tang et al., 2018). Currently, four HDAC inhibitors (HDACi) including Vorinostat, Romidepsin, Panobinostat, and Belinostat have been approved by FDA for anticancer therapeutics, and some other HDAC inhibitors (such as Ricolinostat) are still in the clinical trials to treat hematological and solid malignancies (Yang et al., 2016; Eckschlager et al., 2017; Li et al., 2018b).

Unfortunately, there are no clinical or approved cases of HDACi currently effective for the treatment of GBM. This is because targeting the key epigenetic enzymes, oncogenes, and pathways specific to glioblastoma cells by the drugs has proved to be of great challenges (Sturm et al., 2014), for example, the lower effective inhibitory concentrations within the tumor cells and adverse toxicological effects (Lee et al., 2015). In order to overcome the shortcoming caused by the limited stability and unacceptable pharmacokinetic properties of most existing drugs or molecules, various molecular skeletons were designed to improve the HDAC1-based drugs development, which conform the pharmacophore model of traditional HDACi, namely containing Cap group (Cap), Connect unit (CU), Linker region (Linker), and Zinc Binding Group (ZBG) (**Figure 1**; Dehmel et al., 2008; Varasi et al., 2011; Choi et al., 2012; Giannini et al., 2014; Krieger et al., 2017). Among these pharmacophores, the ZBG should penetrate deep into the bottom of the active pocket and chelate with zinc ion located in the catalytic center to compete with the protein for zinc ion, thereby inhibiting the catalytic activities of HDACs. And such binding pattern in the active pocket is the active conformation of the HDAC inhibitors (Krieger et al., 2019; Shen et al., 2019; Vergani et al., 2019).

Interestingly, the skeletons with macrocyclic Cap have better inhibitory activities against HDAC Class I than Class II, among which the macrocyclic peptide inhibitors account for a large proportion (Mwakwari et al., 2010; Rajak et al., 2013; Tapadar et al., 2015). As inhibiting class II HDACs (represented by HDAC6) can lead to unwanted toxic and side effects (especially serious cardiac toxicity) (Roche and Bertrand, 2016), targeting specific HDAC subtypes has shown great therapeutic potential. The macrocyclic HDACi targeting only Class I HDAC family or HDAC1 are regarded as lower toxicity and more tolerable than pan-HDAC inhibitors, which have shown great potential values of clinical therapeutic effects (Benelkebir et al., 2011; Bhansali et al., 2011; Mallinson and Collins, 2012; Salvador et al., 2014; Decroos et al., 2015; Pilon et al., 2015; Chen et al., 2017; Cheng et al., 2017; Kim et al., 2017). However, there are currently no crystal structures of HDAC1 and HDAC6 complexed with macrocyclic HDACi that have been resolved. Therefore, it is urgent and necessary to reveal the difference in the binding mechanism of macrocyclic HDACi in HDAC1&6 at the atomic level.

In this study, Romidepsin (FK228) was applied as a case study to investigate why macrocyclic polypeptide inhibitors tend to inhibit HDAC1, and various computational approaches were adopted to explore the binding modes of FK228 in HDAC1&6. First, the studied complexes of FK228 in HDAC1&6 were constructed via molecular docking approach. Second, the docked results were further verified by molecular dynamic (MD) simulation. Finally, the key residues responsible for the difference in binding energy of macrocyclic HDACi in HDAC1&6 were identified. In summary, the mechanism underlying why FK228 prefer to inhibit HDAC1 was elaborate through differential energy contributions and interaction fingerprints among the identified key amino acids, which could provide valuable information for the drug discovery on the basis of selective inhibition of HDAC1 in the future.

RESULTS AND DISCUSSION

Construction of FK228 Complexed HDAC1and6 Structures

On the basis of the resolved protein crystals of HDAC1&6 available in Protein Data Bank (PDB) (Hai and Christianson, 2016; Watson et al., 2016), there were 75 and 68 binding poses of FK228 in HDAC1 and HDAC6 generated by molecular docking, respectively. Except for the docking score, the spatial similarity of the docking pose to the largazole thiol in HDAC8 was considered in selecting the conformations of FK228 in HDAC1&6 (**Figure 2**). This was because there were no HDAC1&6 crystals resolved with macrocyclic inhibitors, and HDAC8 (HDAC Class I) protein crystal with similar active pocket

Abbreviations: Cap, capping group; CD, catalytic domain; CU, connect unit; GBM, Glioblastoma; HATs, histone acetyltransferases; HDAC6, Histone deacetylase 6; HDACi, HDAC inhibitors; MD, Molecular Dynamics; MM/GBSA, Molecular Mechanics Generalized Born Surface Area; sHDAC6Is, selectivity of HDAC6 inhibitors; ZBG, zinc binding group

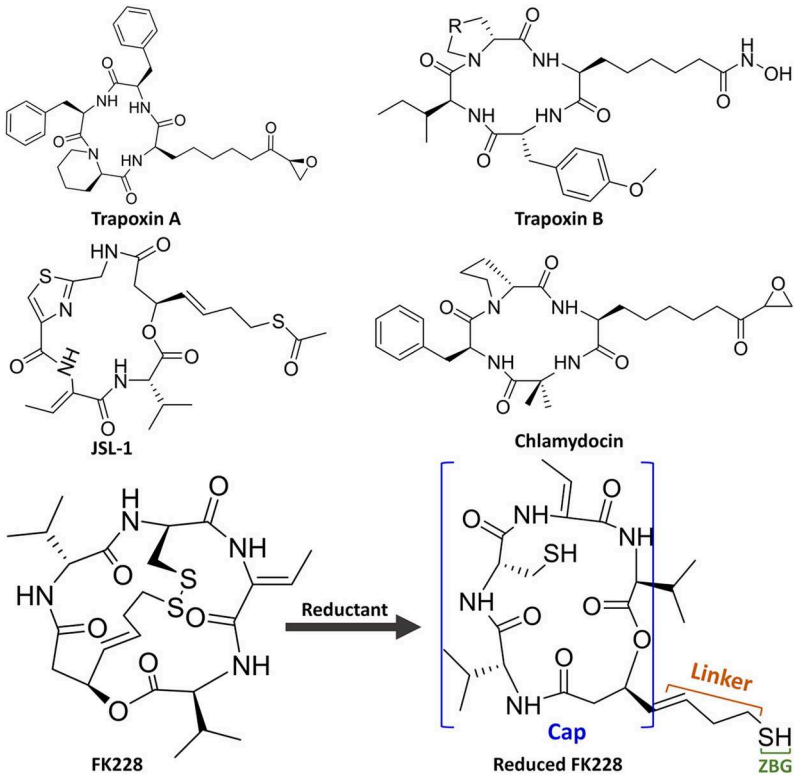


FIGURE 1 | Molecular skeletons of HDAC inhibitors with macrocyclic Cap group.

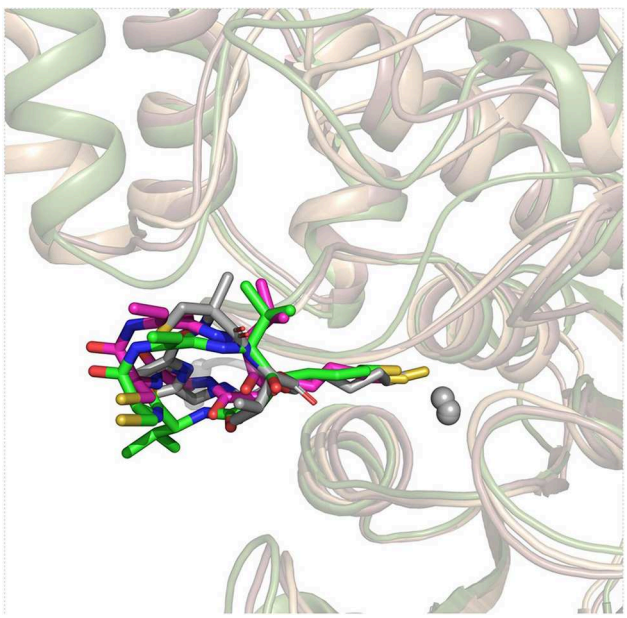


FIGURE 2 | Superimposition of FK228 in HDAC1&6 and HDAC8 complexed with depsipeptide inhibitor.

as HDAC1&6 complexed with largazole thiol could provide important clues for the choice of the initial conformations of FK228 in HDAC1&6 (**Figure 2**). According to **Figure S1**, it could be learnt that binding sites of HDAC1&6 were mainly composed of loop regions, namely loop 1–7. In addition, the selected docked poses suggested that the Cap group of FK228 had interactions with the residues at the rim of the active pocket of HDAC1&6, and the Linker coupled with ZBG penetrate the active pocket, which made the ZBG chelating with the zinc ion in catalytic center. Moreover, the orientation of FK228 in HDAC1&6 is highly coincident with the largazole thiol inhibitor in HDAC8 (**Figure 2**), which verified the reliability of the docking conformation to some extent. According to the **Tables S1, S2**, it could be found that the RMSD values were basically negatively correlated with the absolute value of the docking scores, and the smaller RMSD values could reflect the better binding of FK228 in the HDAC1&6 to some extent. In order to verify the reliability of the experiments, one additional initial conformation of the constructed system have been selected for the further molecular dynamic simulation (**Figure S2**).

Evaluating the Stability of MD Simulation via RMSD Analysis

The Complexes Stabilities Along the Simulation Monitored by RMSD

The selected docking conformations of HDAC1&6 in complex with FK228 were sampled by 500 ns MD simulation, and the dynamic trajectories of the studied complexes were supervised through the RMSD plots of the backbone-atoms of HDAC1&6, heavy-atoms of FK228, and the backbone-atoms of the amino acids in the binding pocket (within 5 Å of ligand) as the function

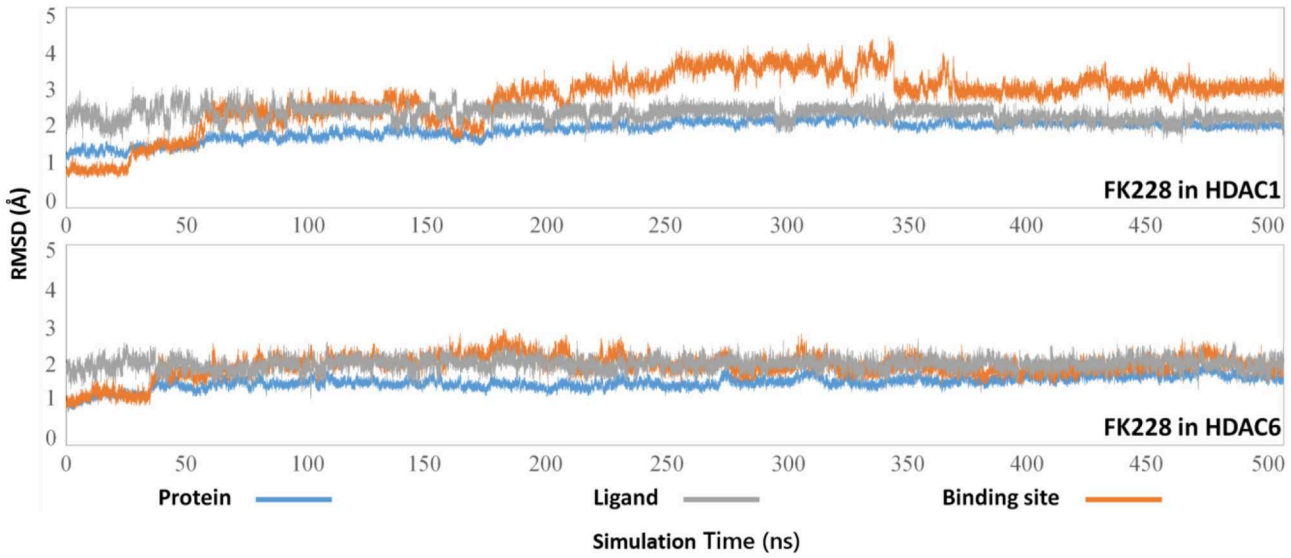


FIGURE 3 | Root mean square deviations of protein backbone atom, ligand heavy atoms, and the backbone atoms of the residues in the binding site as the function of time in MD simulations.

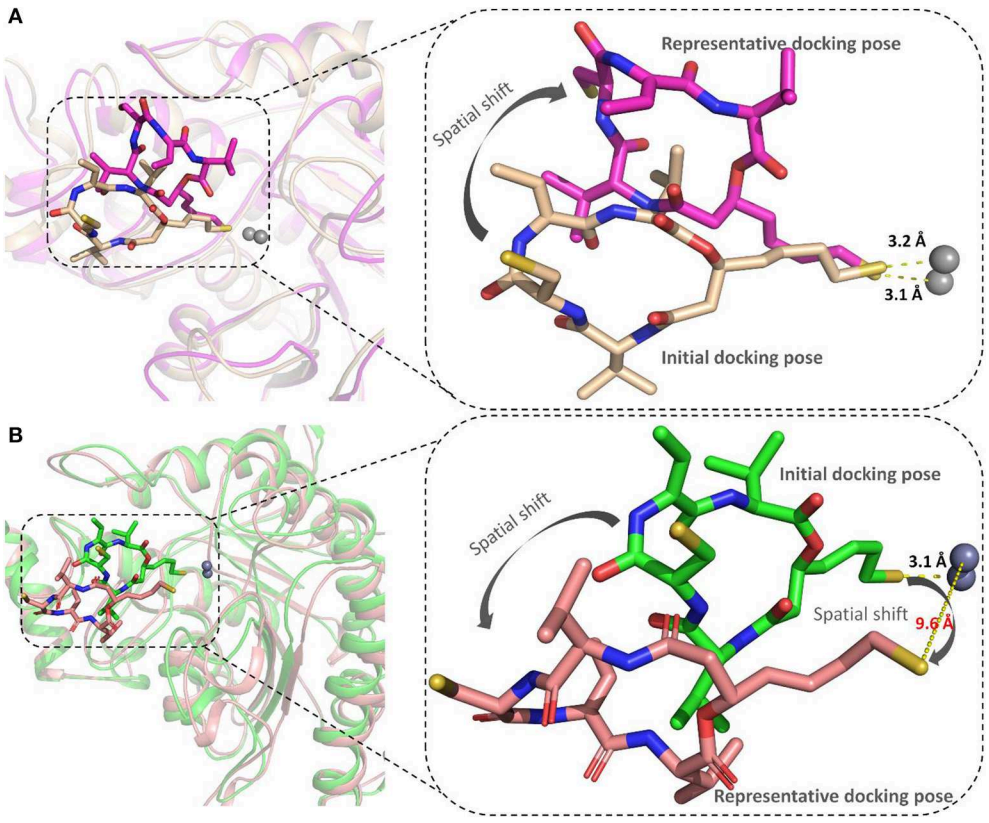


FIGURE 4 | Comparison of the initial conformation and the representative conformation of the FK228 in HDAC1&6: **(A)** FK228 in HDAC1 system; **(B)** FK228 in HDAC6 system.

of simulation time (**Figure 3**). Insight from the RMSD values in **Figure 3**, the FK228-HDAC1 and FK228-HDAC6 systems reached the equilibrium states around 350 and 50 ns, respectively. Moreover, according to the RMSD values of the additional independent simulations also showed that the constructed systems reached the equilibrium around 200 ns (**Figure S3**), and

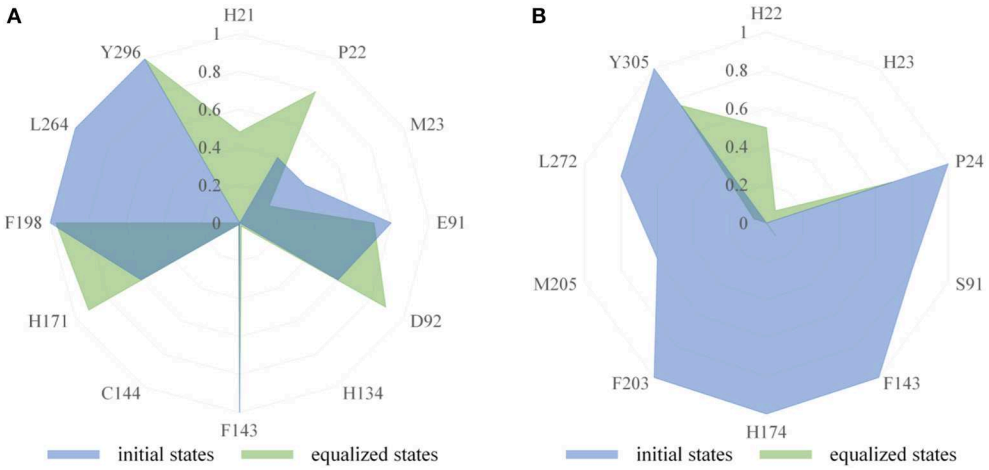


FIGURE 5 | Comparison of interaction fingerprints of FK228 in HDAC1&6 in the final 50 ns simulations with that of the optimized docking poses: **(A)** interaction fingerprints of FK228 in HDAC1; **(B)** interaction fingerprints of FK228 in HDAC6.

the difference in the fluctuation of the binding site of the two simulations were caused by the flexible loop domain.

The Conformational Rearrangements of FK228 in HDAC1and6

The representative structures of FK228 binding to HDAC1&6 were obtained from the equilibrated trajectories and were compared with their corresponding initial conformations (**Figure 4**). During the MD process, the protein conformational changes were calculated by *VMD* software, the values were 1.85 and 1.92 Å for the HDAC1-FK228 and HDAC6-FK228 systems, indicating the small change in the conformation of protein. According to **Figure 4A**, it can be learnt that slight spatial shift of FK228 occurred in HDAC1 active site and the binding conformation maintain the interaction of sulfhydryl group (ZBG) chelating with the zinc ion (∼3.2 Å) through inserting deeply into the active pocket. In contrast, for FK228 in HDAC6 (**Figure 4B**), there was a large deflection of the ZBG in the ligand from the initial conformation, namely straying from the catalytic center (∼9.6 Å). In order to verify the reliability of the experiment, the conformational rearrangement of FK228 in HDAC1&6 of the additional independent experiment was also analyzed, and based on **Figure S4**, it could learnt that FK228 could maintain the active conformation in HDAC1 but not in HDAC6 (ZBG was also far away from the zinc ion). The conformational rearrangements investigated by MD simulation imply that the protein-ligand binding modes is the leading cause of the significant difference of FK228 inhibitory activity to HDAC1&6 and need to be further explored.

Molecular Mechanism of FK228 Selectivity to HDAC1and6

Insights From the FK228-HDAC1and6 Interaction Fingerprints

The binding modes of FK228 in HDAC1&6 are related to the interactions between drugs and amino acids of the target proteins. Thus, the interaction fingerprints analysis was used to explore the difference of FK228-HDAC1&6 binding modes (**Figure 5**). **Figure 5A** indicates that FK228 can maintain its interactions with the P22, E91, and D92 located at Loop1 and Loop2 of HDAC1 before and after MD simulation. For HDAC6-FK228 complex, although FK228 can maintain the interaction with P24 of Loop1, the interaction with S91 of Loop2 in the initial conformation was disappeared after MD simulation (**Figure 5B**). In addition, according to the interaction fingerprints, E91 locating on the Loop2 of HDAC1 contributed to a strong hydrophobic interaction with FK228, and the corresponding site on HDAC6 has no interaction with FK228, leading to the weak interaction between FK228 and Loop2 of HDAC6, which is the main reason of the large spatial shift of FK228 in the binding site of HDAC6.

Insights From the Calculated Binding Free Energy of FK228-HDAC1and6 Complexes

The total binding free energies of HDAC1-FK228 and HDAC6-FK228 were −37.01 and −27.84 kcal/mol, which was consistent with the inhibitory gradient of FK228 toward HDAC1 and HDAC6 (**Table 1**). To qualify the energy contribution of each amino acid in HDAC1&6 for FK228's binding, the total binding free energies were decomposed at amino acid basis and the important ones with high contribution (≥0.1 kcal/mol) (Zheng et al., 2017) were identified. As shown in **Figure 6** the values of amino acids energy with high contribution in each complex varied significantly (taking FK228-HDAC1 as example, the contribution of F143 equaled to −2.39 kcal/mol, which was almost 22 times of C93's energy contribution). As expected, the contributions of the amino acids at the corresponding position on HDAC1&6 also varied greatly. Taking G295 in HDAC1 and N306 in HDAC6 as example, it contributed −0.18 and −1.94 kcal/mol to the binding of FK228 in HDAC1 and HDAC6, respectively.

In comparison with the residues D92 located at Loop2 of HDAC1, the reduction of the energy of corresponding residues

TABLE 1 | Calculated and experimental data of FK228 binding to HDAC1 and HDAC6 (ΔG is in kcal/mol and IC50 value is in nM).

Systems	ΔE_{ele}	ΔE_{vdW}	ΔG_{pol}	$\Delta G_{non\text{-}pol}$	$\Delta G^{a}_{MM/GBSA}$	IC^{b}_{50}
HDAC1–FK228	−12.47	−40.38	21.21	−5.37	−37.01	3.97
HDAC6–FK228	−8.18	−24.68	11.53	−4.51	−25.84	787

[a] *Calculated MM/GBSA binding free energies in this study.*
[b] *IC_{50} values obtained from previous study (Yurek-George et al., 2007).*

S91 of HDAC6 contributed to FK228's binding enhanced our understanding of the difficulty of FK228 to maintain the initial conformation in HDAC6 of during MD simulation. As a result, the large spatial shift of FK228 in the binding site of HDAC6 led to the decreased energy contribution of the amino acids in the active site of HDAC6, such as F143, H174, and F203 (**Figure 6**), consisted very well with the decrease in the interacting frequency with these residues when compared with HDAC1 (**Figure 7**). For the Zn^{2+}, the calculated energy contributing to FK228 binding in the active site of HDAC1 was −0.75 kcal/mol, while there was no energy contribution in HDAC6 system (**Figure 6**).

The Key Role of Residue D92 in FK228 Binding to HDAC1

Both interaction fingerprints and amino acid energy contribution analysis found that residue D92 plays a key role in FK228 binding to HDAC1. The representative conformation obtained from the MD simulation trajectory in **Figure 8** showed that the carbonyl group of D92 and nitrogen atom on the Cap group of FK228 could form a hydrogen bond. To further explore the huge difference in energy contribution of D92 in HDAC1 and its corresponding amino acid S91 in HDAC6, the distance between the two atoms forming the hydrogen bond during the equilibrium simulation (400–500 ns) was monitored, and the average distance between the two atoms forming hydrogen bonds was 3.15 Å (**Figure 8**). However, the side chain of S91 in HDAC6 lacked the hydrogen bond acceptor and its hydrophobic interaction with FK228 would gradually disappear with the deflection of its spatial position during the MD simulation (**Figures 4B**, **5B**).

The Active Site Radius of Gyration Confirmed the Trend of FK228 to HDAC1

Physical and structural properties of the active pockets are closely related to the binding affinities of the ligands (Narang et al., 2019; Thillainayagam et al., 2019), the calculated binding free energy (**Table 1**) has successfully predicted the higher binding affinity of FK228 to HDAC1. To further evaluate the interactions between the FK228 and HDAC1&6, the radius of gyration (Rg) for the Ca-atoms of HDAC1&6 active pockets, which could be applied as an important and effective parameter to evaluate the structural integrity and compactness of the studied systems. The time evolution plot of Rg was calculated and shown in **Figure 9**. It is noted that the average Rg value of HDAC1-FK228 system is lower than that of HDAC6-FK228. The lower value of Rg of HDAC1-FK228 system indicated that the binding pocket of HDAC1 much more compacted and that FK228 could stay stably at the active site, which provided a guarantee for stronger interaction between FK228 and amino acids in the active pocket of HDAC1.

Overall Comparison of the Binding Conformations of FK228 in HDAC1and6

According to previous studies, HDAC inhibitors could exert the inhibitory activities by chelating with zinc ion at the catalytic center via the ZBG group deep into the bottom of the active pocket. According to **Figure 10**, the distance between the zinc ion and the sulfur atom on the ZBG of FK228 varied greatly in the two studied systems. In the HDAC1-FK228 system, the distance between the zinc ion and the sulfur atom on the ZBG of FK228 was about 3.5 Å, but the relative positions of sulfur and zinc ion is relatively larger in HDAC6-FK228 system. Furthermore, it can be learnt that Loop1 and Loop2 of HDAC1 formed a sub-pocket during the MD simulation process that could well-accommodate the sulfhydryl group on the Cap group and anchored the Cap group (**Figure 11**). The anchoring effects of Loop1 and Loop2 played a vital role in maintaining the binding conformation of FK228, and the relatively small spatial biases ensured the interaction of FK228 with important amino acids in the HDAC1 active pocket. However, for the HDAC6-FK228 system, the ZBG group did not penetrate the active pocket bottom to compete with the protein for metal zinc ions, and the previously calculated Rg value also indicated that the HDAC6 active pocket is less compacted, reducing the potential for interaction with FK228, which was the main reason for the large deflection of the Cap group at the active pocket.

CONCLUSION

As the first approved macrocyclic HDAC inhibitor, FK228 was used as a molecular probe to compare its binding conformation in HDAC1 and HDAC6, and to explore the molecular mechanism of FK228' tendency to inhibit HDAC1 at the atomic level through a variety of *in silico* approaches. For HDAC6-FK228 system, the disappearing hydrophobic interaction of S91 (located in the HDAC6 Loop2 region) with FK228 during the MD simulation and the lack of the corresponding residue of E91 (located in the HDAC1 Loop2 region) together weakened the anchoring effects of HDAC6 Loop2 to the FK228 Cap during the MD process, leading to the large spatial conformational deviation of the docking conformation and resulting the reduced interaction between the FK228 Linker region and the conserved amino acids in the HDAC6 pocket. In the case of HDAC1-FK228 system, K228 could maintain the interactions with D92 (hydrogen bonding) and E91 (hydrophobic interaction) on Loop2 after the dynamic trajectory reached equilibrium, and the interactions with H21 and P22 on Loop1 were also strengthened. Moreover, in the process of molecular dynamics simulation, Loop1 and Loop2 on HDAC1 could form a sub-pocket that better accommodated the Cap group of FK228, maintaining the active conformation of FK228 at the binding pocket and ensuring ZBG chelating with the zinc ion and competing with the protein for the metal zinc ion, thereby exerting the inhibitory activity on HDAC1. The interaction of the Cap group with the Loop1 and Loop2

FIGURE 6 | The per-residue binding free energy decomposition of 31 residues with high energy contribution (≥0.1 kcal/mol) to the interaction in at least one studied complex: FK228 in HDAC1 (light green); FK228 in HDAC6 (light orange).

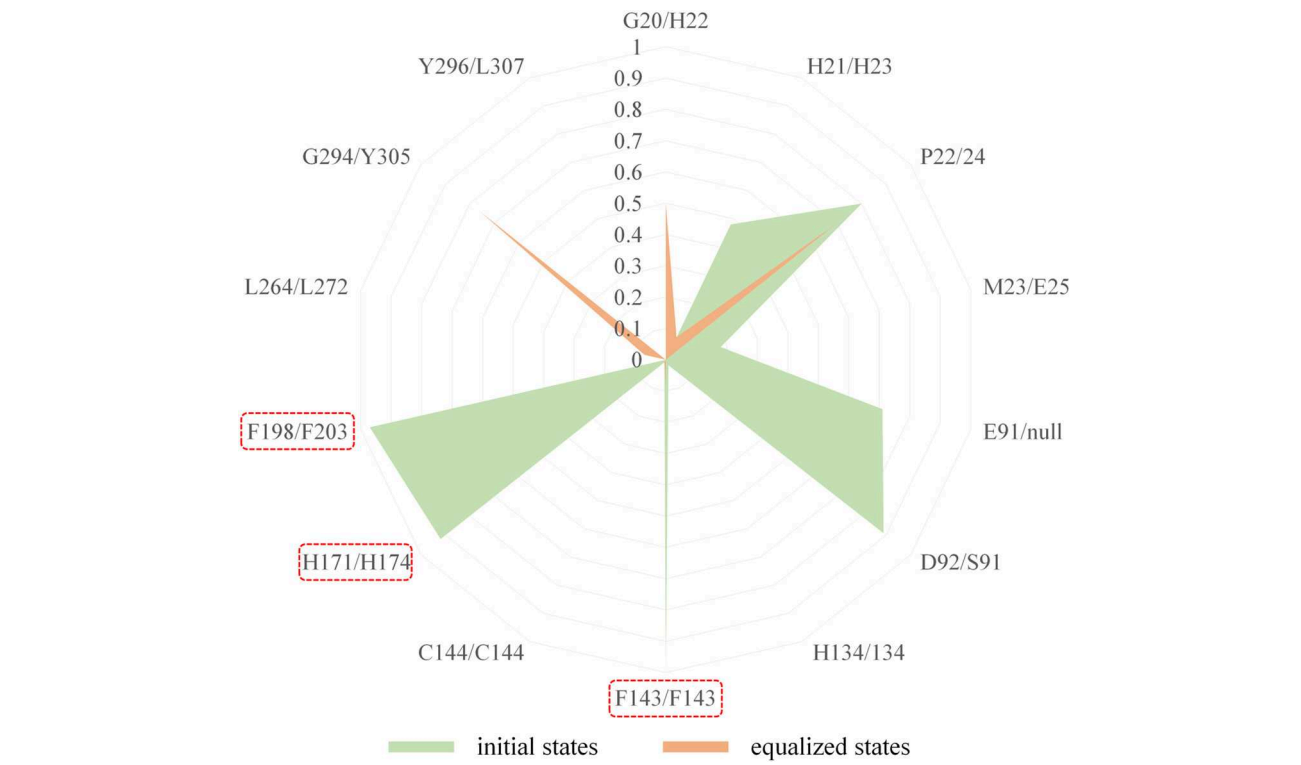


FIGURE 7 | Comparison of the interaction fingerprints of FK228 in HDAC1&6 under the equilibrium trajectories.

regions contributes to maintaining the active conformation of the HDACi and should be especially considered on subsequent drug design based on selective inhibition of HDAC1.

MATERIALS AND METHODS

The Construction of the Studied Systems

The studied systems FK228-HDAC1&6 were obtained by molecular docking using the Glide (2009) software embedded in Maestro (2009) with default parameters of standard precision. The 3D structure of FK228 was drawn by *ChemBioDraw* (Dickson et al., 2014) and saved in SDfile (*.sdf), then processed with *LigPrep* [*OPLS-2005* (Price and Brooks, 2005) force fields] to generate the low-energy stable conformation. Additionally, the 3D structure of FK228 was preprocessed by Epik (2009) ($pH = 7.0 \pm 2.0$) to generate the ionized state. After that, the protein structures of HDAC1&6 available in Protein Data Bank [PDB entry: 5ICN (Watson et al., 2016) and 5EDU (Hai and Christianson, 2016)] were processed by *Protein Preparation Wizard* (Maestro, 2009) module in Maestro (2009) to add the hydrogen atoms, assign protonation states and partial charges by *OPLS-2005* (Price and Brooks, 2005) force field, and minimize the whole protein crystal to prepare the receptor for molecular docking. The minimization process is completed when the RMSD value reached 0.30 Å. Furthermore, the spatial coordinates of largazole analog in HDAC8 were referred when defining the docking grid due to the similar binding pockets (Cole et al., 2011; Du et al., 2011; Decroos et al., 2015; Gantt et al., 2016).

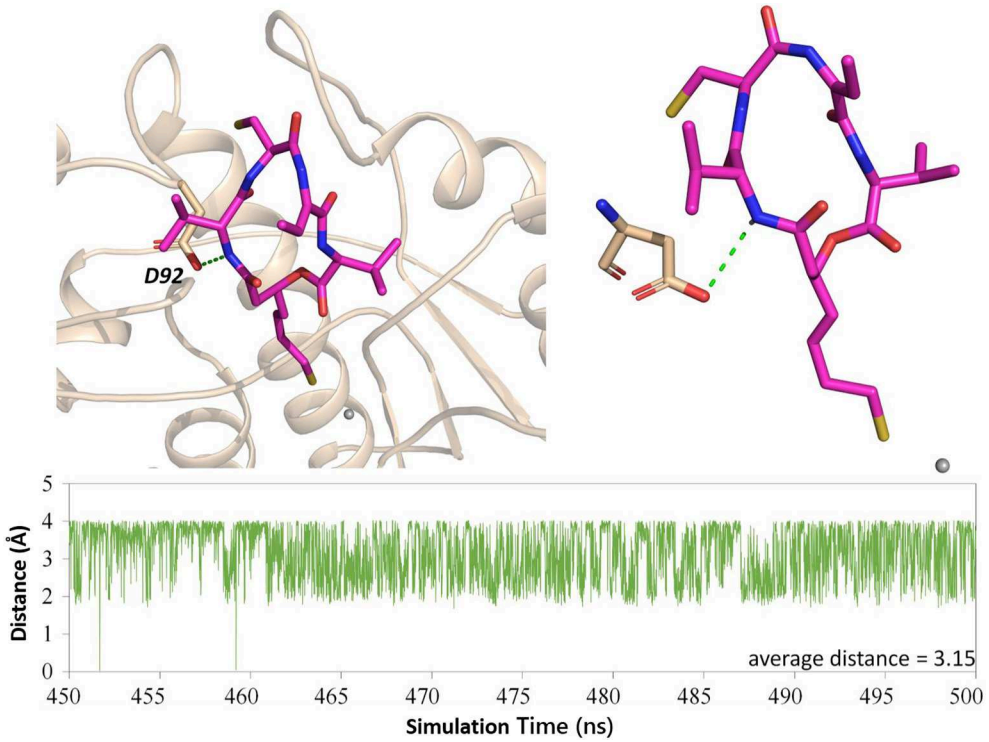


FIGURE 8 | H-bond analysis between D92 and FK228 in the two constructed systems.

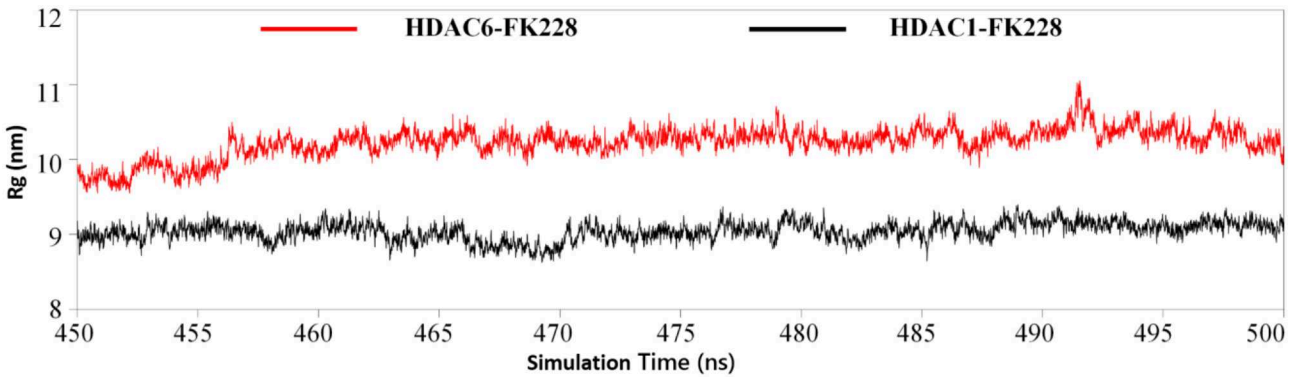


FIGURE 9 | Analysis of radius of gyration of the two studied systems.

In molecular docking, 5,000 poses were generated during the initial phase of the docking calculation, out of which best 400 poses were chosen for energy minimization by 100 steps of conjugate gradient minimizations (the details shown in **Supplementary Materials**).

Molecular Dynamics (MD) Simulation

MD simulation was performed within AMBER16 (2016) using GPU-accelerated *PMEMD* on 16 cores of an array of two 2.6 GHz Intel Xeon E5-2650v2 processors and 4 pieces of NVIDIA Tesla K40C graphic card. *AMBER* force field *ff14SB* (Dickson et al., 2014) and Li/Merz ion parameters (Li and Merz, 2014; Li et al., 2015a,b) were used for the protein and SPC/E water. *General AMBER force field* 2 (*gaff2*) was applied to assign the parameters of FK228 in each complex, and the atom types and partial charges of FK228 could be derived on the basis of RESP calculation through *antechamber* (Wang et al., 2006). In addition, the geometrical optimization and electrostatic potential calculation for FK228 were conducted at HF/6-31G* level through *Gaussian 09* software (Gaussian 09, 2009). The Zn^{2+} was processed by 12-6-4 model (Li et al., 2015a) imbedded in *Amber16*. When the constructed systems were processed by LEaP (AMBER16, 2016), it could be found that FK228-HDAC1 and FK228-HDAC6 systems were solvated with a cubic water box, and the vdw box

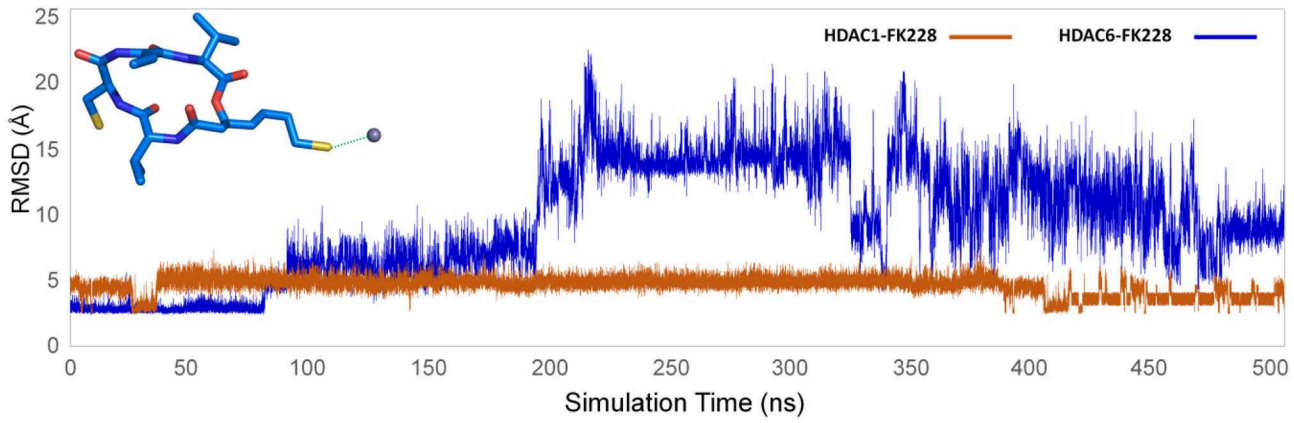


FIGURE 10 | Distance between the sulfur atom in the ZBG of FK228 and zinc ion in the studied systems.

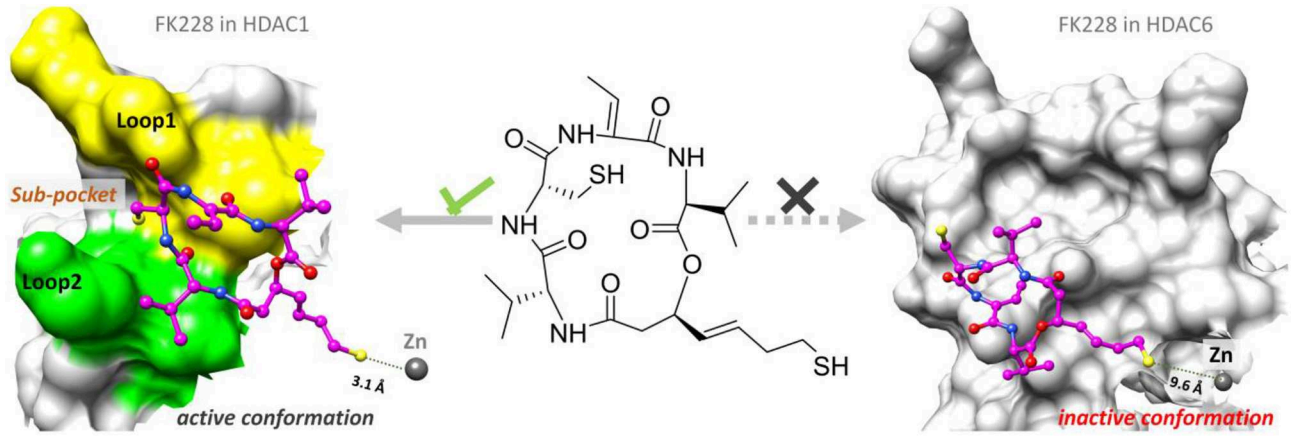


FIGURE 11 | Comparison of the binding pattern of FK228 in HDAC1&6.

sizes were 527706.39 $Å^3$ (12,126 water molecules) and 510534.33 $Å^3$ (11,910 water molecules), respectively. In addition, there were two sodium ions in HDAC1-FK228 system that were used to neutralize the negative charge, and six sodium ions were used to neutralize the negative grid charge in the HDAC6-FK228 system.

Before the MD simulation, the processed research systems were subjected to the initial energy minimization through two procedure (Xue et al., 2018a; Zhang et al., 2019). The first step was to apply harmonic restraint on solute atom (force constant = 10 $kcal \cdot mol^{-1} \cdot Å^{-2}$), and the second step was to release all atoms to move freely. In each step, energy minimization was conducted by the steepest descent method for the first 5,000 steps and the conjugated gradient method for the subsequent 5,000 steps. Then, each studied system was heated from 0 to 100 K and then gradually to 310 K with the protein restrained over 100 ps in the NVT ensembles. Subsequently, 10 times (5 ns) unrestrained equilibration at 310 K were performed to equilibrate system's periodic boundary condition. Finally, the unrestrained 500 ns production simulation was conducted for the prepared four systems in NPT ensembles under the temperature of 310 K and the pressure of 1 atm. Temperature was controlled by Langevin dynamics and the pressure was controlled using *Monte Carlo barostat* (2016). In all the simulations, *Particle-mesh Ewald* (PME) (Darden et al., 1993) was used to handle the long range electrostatic interaction, and *SHAKE algorithm* was exploited to keep all bonds rigid (Larini et al., 2007; Fu et al., 2018; Xue et al., 2018a; Yang et al., 2018). Time step of simulation was set 2.0 fs and a 10.0 Å cutoff was used for non-bonded interactions (Xue et al., 2018b; Du et al., 2019).

All the analysis of MD trajectories, including as root mean square deviation (RMSD), the representative structures from the trajectories, binding free energies, were analyzed and predicted via *cpptraj* and *mm_pbsa.pl* programs embedded in *AMBER16*. Structural visualization was performed in *PyMOL* software (PyMOL 1.3[1]).

Protein-Ligand Interaction Fingerprints Analysis

Interaction fingerprints between the FK228 and the HDAC1&6 were calculated via *Ichem* (Da Silva et al., 2018; Southan, 2018),

[1]PyMOL Molecular Graphics System, *v.* 1.3. *Schrödinger*, LLC, New York, NY.

and the calculation system mainly consisted of the ligand and the binding site (residues within 6 Å of the FK228's mass center). Firstly, conformation optimization and energy optimization were carried out for the docking poses of FK228 in HDAC1&6, and then interaction fingerprints was applied to carry out for the optimized conformations to calculate the interaction between FK228 and receptors in the initial conformation. Secondly, 500 snapshots were extracted from the equalized simulated trajectories (between 400 and 500 ns) to indicate the interacting effects between the FK228 and HDAC1&6, which was compared with interactions of the initial states of the studied systems. During the process of calculation, seven important interactions (hydrophobic interaction, aromatic, H-bond donor, H-bond acceptor, positively ionizable, negatively ionizable, and metal coordination) were applied to assess the interaction fingerprints between the ligand and receptor by parsing atoms and bond connectivity fields in the form of one-dimensional (1D) descriptors consisting of 1 and 0, and the results were shown by radar plots. In addition, detailed information about the rules of detecting the interactions between protein and ligand were shown in **Table S3**.

Calculation of the Binding Free Energy

MM/GBSA approach using a single molecular dynamic trajectory was adopted to calculate the binding free energy ($\Delta G_{MM/GBSA}$) regardless of entropic influence between the docked ligands and the receptor (Chen et al., 2017; Wang et al., 2017a, 2019; He et al., 2018), and in this study, 500 snapshots were extracted from the equilibrium trajectories (450–500 ns) for calculation via cpptraj. The calculation equation was as follows:

$$G_{MM/GBSA} = \Delta E_{vdW} + \Delta E_{ele} + \Delta G_{pol} + \Delta G_{nonpol} \quad (1)$$

Where, ΔE_{vdW} represented the van der Waals interactions contribution, ΔE_{ele} stood for the electrostatic energy contribution, ΔG_{pol} was the polar solvent interaction energy calculated with the GB model ($igb = 2$) and G_{nonpol} was the non-polar solvation free energy, which was evaluated using LCPO method (0.0072 × ΔSASA, SASA indicating the solvent accessible area with a probe radius of 1.4 Å) (Weiser et al., 1999; Zheng et al., 2016).

Calculating the Per-Residue Energy Contribution

The per-residue energy contribution $\Delta G^{per-residue}_{MM/GBSA}$ between the residues located in HDAC1&6 and the docked ligands was calculated using the following formula:

$$\Delta G^{per-residue}_{MM/GBSA} = \Delta E^{per-residue}_{vdW} + \Delta E^{per-residue}_{ele} + \Delta G^{per-residue}_{pol} + \Delta G^{per-residue}_{nonpol} \quad (2)$$

Where, the three terms, namely ($\Delta E^{per-residue}_{vdW}$), $\Delta E^{per-residue}_{ele}$ and $\Delta G^{per-residue}_{pol}$, were defined in the same way as the corresponding terms in formula 2, and $\Delta G^{per-residue}_{nonpol}$ was calculated using the *ICOSA* method (Wang et al., 2006, 2017b).

Radius of Gyration Calculation

The residues consisting of the binding site were selected to calculate the radius of gyration of the studied systems, and there were 30 residues in each systems. In this study, the equilibrium trajectories (450–500 ns) were used to calculate the Rg of the specified residues via *cpptraj*.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

FZ and WX conceived the work and directed the experiments. YZ performed the MD simulations. YZ, TF, YR, FL, GZ, JH, and XY collected and confirmed the data of protein and ligand structures and performed the analysis. FZ, WX, and YZ drafted the first and second version of the manuscript. All authors read, edited, and approved the final version of manuscript.

FUNDING

This study was supported by National Natural Science Foundation of China (81872798), National Key Research and Development Program of China (2018YFC0910500), Innovation Project on Industrial Generic Key Technologies of Chongqing (cstc2015zdcy-ztzx120003), and Fundamental Research Funds for the Central Universities (2018QNA7023, 10611CDJXZ238826, 2018CDQYSG0007, CDJZR14468801, and CDJKXB14011).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00041/full#supplementary-material

REFERENCES

AMBER16 (2016). i*AMBER, v. 16*. University of California, San Francisco.

Benelkebir, H., Marie, S., Hayden, A. L., Lyle, J., Loadman, P. M., Crabb, S. J., et al. (2011). Total synthesis of largazole and analogues: HDAC inhibition, antiproliferative activity and metabolic stability. *Bioorg. Med. Chem.* 19, 3650–3658. doi: 10.1016/j.bmc.2011.02.024

Bhansali, P., Hanigan, C. L., Casero, R. A., and Tillekeratne, L. M. (2011). Largazole and analogues with modified metal-binding motifs targeting histone deacetylases: synthesis and biological evaluation. *J. Med. Chem.* 54, 7453–7463. doi: 10.1021/jm200432a

Bhat, K. P., Pelloski, C. E., Zhang, Y., Kim, S. H., Delacruz, C., Rehli, M., et al. (2008). Selective repression of YKL-40 by NF-kappaB in glioma cell lines involves recruitment of histone deacetylase-1 and−2. *FEBS Lett.* 582, 3193–3200. doi: 10.1016/j.febslet.2008.08.010

Brennan, C. W., Verhaak, R. G., Mckenna, A., Campos, B., Noushmehr, H., Salama, S. R., et al. (2013). The somatic genomic landscape of glioblastoma. *Cell* 155, 462–477. doi: 10.1016/j.cell.2013.09.034

Campos, B., Bermejo, J. L., Han, L., Felsberg, J., Ahmadi, R., Grabe, N., et al. (2011). Expression of nuclear receptor corepressors and class I histone deacetylases in astrocytic gliomas. *Cancer Sci.* 102, 387–392. doi: 10.1111/j.1349-7006.2010.01792.x

Cancer Genome Atlas Research, N. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068. doi: 10.1038/nature07385

Cartron, P. F., Blanquart, C., Hervouet, E., Gregoire, M., and Vallette, F. M. (2013). HDAC1-mSin3a-NCOR1, Dnmt3b-HDAC1-Egr1 and Dnmt1-PCNA-UHRF1-G9a regulate the NY-ESO1 gene expression. *Mol. Oncol.* 7, 452–463. doi: 10.1016/j.molonc.2012.11.004

Chen, J. C., Lee, I. N., Huang, C., Wu, Y. P., Chung, C. Y., Lee, M. H., et al. (2019). Valproic acid-induced amphiregulin secretion confers resistance to temozolomide treatment in human glioma cells. *BMC Cancer* 19:756. doi: 10.1186/s12885-019-5843-6

Chen, X., Wang, C., Tang, S., Yu, C., and Zou, Q. (2017). CMSA: a heterogeneous CPU/GPU computing system for multiple similar RNA/DNA sequence alignment. *BMC Bioinformatics* 18:315. doi: 10.1186/s12859-017-1725-6

Cheng, K., Li, S., and Liao, C. (2017). Progress in the discovery of macrocyclic histone deacetylase inhibitors for the treatment of cancer. *Curr. Med. Chem.* 24, 4166–4179. doi: 10.2174/0929867324666170209105315

Choi, E., Lee, C., Cho, M., Seo, J. J., Yang, J. S., Oh, S. J., et al. (2012). Property-based optimization of hydroxamate-based gamma-lactam HDAC inhibitors to improve their metabolic stability and pharmacokinetic profiles. *J. Med. Chem.* 55, 10766–10770. doi: 10.1021/jm3009376

Cole, K. E., Dowling, D. P., Boone, M. A., Phillips, A. J., and Christianson, D. W. (2011). Structural basis of the antiproliferative activity of largazole, a depsipeptide inhibitor of the histone deacetylases. *J. Am. Chem. Soc.* 133, 12474–12477. doi: 10.1021/ja205972n

Da Silva, F., Desaphy, J., and Rognan, D. (2018). IChem: a versatile toolkit for detecting, comparing, and predicting protein-ligand interactions. *ChemMedChem* 13, 507–510. doi: 10.1002/cmdc.201700505

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald - an N.Log(N) method for ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397

Decroos, C., Clausen, D. J., Haines, B. E., Wiest, O., Williams, R. M., and Christianson, D. W. (2015). Variable active site loop conformations accommodate the binding of macrocyclic largazole analogues to HDAC8. *Biochemistry* 54, 2126–2135. doi: 10.1021/acs.biochem.5b00010

Dehmel, F., Weinbrenner, S., Julius, H., Ciossek, T., Maier, T., Stengel, T., et al. (2008). Trithiocarbonates as a novel class of HDAC inhibitors: SAR studies, isoenzyme selectivity, and pharmacological profiles. *J. Med. Chem.* 51, 3985–4001. doi: 10.1021/jm800093c

Dickson, C. J., Madej, B. D., Skjevik, A. A., Betz, R. M., Teigen, K., Gould, I. R., et al. (2014). Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* 10, 865–879. doi: 10.1021/ct4010307

Dong, L. H., Cheng, S., Zheng, Z., Wang, L., Shen, Y., Shen, Z. X., et al. (2013). Histone deacetylase inhibitor potentiated the ability of MTOR inhibitor to induce autophagic cell death in Burkitt leukemia/lymphoma. *J. Hematol. Oncol.* 6:53. doi: 10.1186/1756-8722-6-53

Du, J., Qiu, M., Guo, L., and Yao, X. (2019). Computational study of the binding mechanism between farnesoid X receptor alpha and antagonist N-benzyl-N-(3-(tertbutyl)-4-hydroxyphenyl)-2,6-dichloro-4-(dimethylamino) benzamide. *J. Biomol. Struct. Dyn.* 37, 1628–1640. doi: 10.1080/07391102.2018.1462735

Du, J., Sun, H., Xi, L., Li, J., Yang, Y., Liu, H., et al. (2011). Molecular modeling study of checkpoint kinase 1 inhibitors by multiple docking strategies and prime/MM-GBSA calculation. *J. Comput. Chem.* 32, 2800–2809. doi: 10.1002/jcc.21859

Eckschlager, T., Plch, J., Stiborova, M., and Hrabeta, J. (2017). Histone deacetylase inhibitors as anticancer drugs. *Int. J. Mol. Sci.* 18:E1414. doi: 10.3390/ijms18071414

Epik (2009). *Epik v. 2.0.* New York, NY: Schrödinger, LLC.

Fu, T., Zheng, G., Tu, G., Yang, F., Chen, Y., Yao, X., et al. (2018). Exploring the binding mechanism of metabotropic glutamate receptor 5 negative allosteric modulators in clinical trials by molecular dynamics simulations. *ACS Chem. Neurosci.* 9, 1492–1502. doi: 10.1021/acschemneuro.8b00059

Gantt, S. M., Decroos, C., Lee, M. S., Gullett, L. E., Bowman, C. M., Christianson, D. W., et al. (2016). General base-general acid catalysis in human histone deacetylase 8. *Biochemistry* 55, 820–832. doi: 10.1021/acs.biochem.5b01327

Gaussian 09 (2009). *Gaussian 09 v. D.01.* Wallingford CT: Gaussian, Inc.

Giannini, G., Vesci, L., Battistuzzi, G., Vignola, D., Milazzo, F. M., Guglielmi, M. B., et al. (2014). ST7612AA1, a thioacetate-omega(gamma-lactam carboxamide) derivative selected from a novel generation of oral HDAC inhibitors. *J. Med. Chem.* 57, 8358–8377. doi: 10.1021/jm5008209

Glide, v. 5.5. (2009). New York, NY: Schrödinger, LLC.

Hai, Y., and Christianson, D. W. (2016). Histone deacetylase 6 structure and molecular basis of catalysis and inhibition. *Nat. Chem. Biol.* 12, 741–747. doi: 10.1038/nchembio.2134

Han, S., Xia, J., Qin, X., Han, S., and Wu, A. (2013). Phosphorylated SATB1 is associated with the progression and prognosis of glioma. *Cell Death Dis.* 4:e901. doi: 10.1038/cddis.2013.433

He, M., Li, W., Zheng, Q., and Zhang, H. (2018). A molecular dynamics investigation into the mechanisms of alectinib resistance of three ALK mutants. *J. Cell. Biochem.* 119, 5332–5342. doi: 10.1002/jcb.26666

He, W., Jia, C., and Zou, Q. (2019). 4mCPred: machine learning methods for DNA N4-methylcytosine sites prediction. *Bioinformatics* 35, 593–601. doi: 10.1093/bioinformatics/bty668

Kim, B., Ratnayake, R., Lee, H., Shi, G., Zeller, S. L., Li, C., et al. (2017). Synthesis and biological evaluation of largazole zinc-binding group analogs. *Bioorg. Med. Chem.* 25, 3077–3086. doi: 10.1016/j.bmc.2017.03.071

Kim, B., Rincon Castro, L. M., Jawed, S., and Niles, L. P. (2008). Clinically relevant concentrations of valproic acid modulate melatonin MT(1) receptor, HDAC and MeCP2 mRNA expression in C6 glioma cells. *Eur. J. Pharmacol.* 589, 45–48. doi: 10.1016/j.ejphar.2008.04.058

Kim, H., Chong, K., Ryu, B. K., Park, K. J., Yu, M. O., Lee, J., et al. (2019). Repurposing penfluridol in combination with temozolomide for the treatment of glioblastoma. *Cancers (Basel)* 11:E1310. doi: 10.3390/cancers11091310

Krieger, V., Hamacher, A., Cao, F., Stenzel, K., Gertzen, C. G. W., Schaker-Hubner, L., et al. (2019). Synthesis of peptoid-based class I-selective histone deacetylase inhibitors with chemosensitizing properties. *J. Med. Chem.* 62, 11260–11279. doi: 10.1021/acs.jmedchem.9b01489

Krieger, V., Hamacher, A., Gertzen, C. G. W., Senger, J., Zwinderman, M. R. H., Marek, M., et al. (2017). Design, multicomponent synthesis, and anticancer activity of a focused histone deacetylase (HDAC) inhibitor library with peptoid-based cap groups. *J. Med. Chem.* 60, 5493–5506. doi: 10.1021/acs.jmedchem.7b00197

Larini, L., Mannella, R., and Leporini, D. (2007). Langevin stabilization of molecular-dynamics simulations of polymers by means of quasisymplectic algorithms. *J. Chem. Phys.* 126:104101. doi: 10.1063/1.2464095

Lee, P., Murphy, B., Miller, R., Menon, V., Banik, N. L., Giglio, P., et al. (2015). Mechanisms and clinical significance of histone deacetylase inhibitors: epigenetic glioblastoma therapy. *Anticancer Res.* 35, 615–625.

Li, P., and Merz, K. M. Jr. (2014). Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J. Chem. Theory Comput.* 10, 289–297. doi: 10.1021/ct400751u

Li, P., Song, L. F., and Merz, K. M. Jr. (2015a). Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water. *J. Phys. Chem. B* 119, 883–895. doi: 10.1021/jp505875v

Li, P., Song, L. F., and Merz, K. M. Jr. (2015b). Systematic parameterization of monovalent ions employing the nonbonded model. *J. Chem. Theory Comput.* 11, 1645–1657. doi: 10.1021/ct500918t

Li, S., Chen, X., Mao, L., Zahid, K. R., Wen, J., Zhang, L., et al. (2018a). Histone deacetylase 1 promotes glioblastoma cell proliferation and invasion via activation of PI3K/AKT and MEK/ERK signaling pathways. *Brain Res.* 1692, 154–162. doi: 10.1016/j.brainres.2018.05.023

Li, Y. H., Yu, C. Y., Li, X. X., Zhang, P., Tang, J., Yang, Q., et al. (2018b). Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.* 46, D1121–D1127. doi: 10.1093/nar/gkx1076

Li, Z. Y., Li, Q. Z., Chen, L., Chen, B. D., Wang, B., Zhang, X. J., et al. (2016). Histone deacetylase inhibitor RGFP109 overcomes temozolomide resistance

by blocking NF-kappaB-dependent transcription in glioblastoma cell lines. *Neurochem. Res.* 41, 3192–3205. doi: 10.1007/s11064-016-2043-5

Lu, Q., Lin, X., Feng, J., Zhao, X., Gallagher, R., Lee, M. Y., et al. (2008). Phenylhexyl isothiocyanate has dual function as histone deacetylase inhibitor and hypomethylating agent and can inhibit myeloma cell growth by targeting critical pathways. *J. Hematol. Oncol.* 9, 1–6. doi: 10.1186/1756-8722-1-6

Maestro, v. 9.0. (2009). New York, NY: Schrödinger, LLC.

Mallinson, J., and Collins, I. (2012). Macrocycles in new drug discovery. *Future Med. Chem.* 4, 1409–1438. doi: 10.4155/fmc.12.93

Mwakwari, S. C., Patil, V., Guerrant, W., and Oyelere, A. K. (2010). Macrocyclic histone deacetylase inhibitors. *Curr. Top. Med. Chem.* 10, 1423–1440. doi: 10.2174/156802610792232079

Narang, S. S., Goyal, D., and Goyal, B. (2019). Inhibition of Alzheimer's amyloid-beta42 peptide aggregation by a bi-functional bis-tryptoline triazole: key insights from molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 3, 1–14. doi: 10.1080/07391102.2019.1614093

Natsume, A., Hirano, M., Ranjit, M., Aoki, K., and Wakabayashi, T. (2019). Aberrant transcriptional regulation of super-enhancers by RET finger protein-histone deacetylase 1 complex in glioblastoma: chemoresistance to temozolomide. *Neurol. Med. Chir. (Tokyo).* 59, 293–298. doi: 10.2176/nmc.ra.2019-0049

Pilon, J. L., Clausen, D. J., Hansen, R. J., Lunghofer, P. J., Charles, B., Rose, B. J., et al. (2015). Comparative pharmacokinetic properties and antitumor activity of the marine HDACi Largazole and Largazole peptide isostere. *Cancer Chemother. Pharmacol.* 75, 671–682. doi: 10.1007/s00280-015-2675-1

Price, D. J., and Brooks, C. L. 3rd (2005). Detailed considerations for a balanced and broadly applicable force field: a study of substituted benzenes modeled with OPLS-AA. *J. Comput. Chem.* 26, 1529–1541. doi: 10.1002/jcc.20284

Rahman, M. A., Gras Navarro, A., Brekke, J., Engelsen, A., Bindesboll, C., Sarowar, S., et al. (2019). Bortezomib administered prior to temozolomide depletes MGMT, chemosensitizes glioblastoma with unmethylated MGMT promoter and prolongs animal survival. *Br. J. Cancer* 121, 545–555. doi: 10.1038/s41416-019-0551-1

Rajak, H., Singh, A., Dewangan, P. K., Patel, V., Jain, D. K., Tiwari, S. K., et al. (2013). Peptide based macrocycles: selective histone deacetylase inhibitors with antiproliferative activity. *Curr. Med. Chem.* 20, 1887–1903. doi: 10.2174/0929867311320140006

Roche, J., and Bertrand, P. (2016). Inside HDACs with more selective HDAC inhibitors. *Eur. J. Med. Chem.* 121, 451–483. doi: 10.1016/j.ejmech.2016.05.047

Salvador, L. A., Park, H., Al-Awadhi, F. H., Liu, Y., Kim, B., Zeller, S. L., et al. (2014). Modulation of activity profiles for largazole-based HDAC inhibitors through alteration of prodrug properties. *ACS Med. Chem. Lett.* 5, 905–910. doi: 10.1021/ml500170r

Shen, S., Hadley, M., Ustinova, K., Pavlicek, J., Knox, T., Noonepalle, S., et al. (2019). Discovery of a new isoxazole-3-hydroxamate-based histone deacetylase 6 inhibitor SS-208 with antitumor activity in syngeneic melanoma mouse models. *J. Med. Chem.* 62, 8557–8577. doi: 10.1021/acs.jmedchem.9b00946

Southan, C. (2018). Caveat usor: assessing differences between major chemistry databases. *ChemMedChem* 13, 470–481. doi: 10.1002/cmdc.201700724

Staberg, M., Michaelsen, S. R., Rasmussen, R. D., Villingshoj, M., Poulsen, H. S., and Hamerlik, P. (2017). Inhibition of histone deacetylases sensitizes glioblastoma cells to lomustine. *Cell. Oncol. (Dordr).* 40, 21–32. doi: 10.1007/s13402-016-0301-9

Sturm, D., Bender, S., Jones, D. T., Lichter, P., Grill, J., Becher, O., et al. (2014). Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge. *Nat. Rev. Cancer* 14, 92–107. doi: 10.1038/nrc3655

Su, R., Liu, X., Wei, L., and Zou, Q. (2019). Deep-resp-forest: a deep forest model to predict anti-cancer drug response. *Methods* 166, 91–102. doi: 10.1016/j.ymeth.2019.02.009

Tan, J., Cang, S., Ma, Y., Petrillo, R. L., and Liu, D. (2010). Novel histone deacetylase inhibitors in clinical trials as anti-cancer agents. *J. Hematol. Oncol.* 3, 5. doi: 10.1186/1756-8722-3-5

Tang, W., Wan, S., Yang, Z., Teschendorff, A. E., and Zou, Q. (2018). Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics* 34, 398–406. doi: 10.1093/bioinformatics/btx622

Tapadar, S., Fathi, S., Raji, I., Omesiete, W., Kornacki, J. R., Mwakwari, S. C., et al. (2015). A structure-activity relationship of non-peptide macrocyclic histone deacetylase inhibitors and their anti-proliferative and anti-inflammatory activities. *Bioorg. Med. Chem.* 23, 7543–7564. doi: 10.1016/j.bmc.2015.10.045

Thillainayagam, M., Ramaiah, S., and Anbarasu, A. (2019). Molecular docking and dynamics studies on novel benzene sulfonamide substituted pyrazole-pyrazoline analogues as potent inhibitors of *Plasmodium falciparum* Histo aspartic protease. *J. Biomol. Struct. Dyn.* 24, 1–11. doi: 10.1080/07391102.2019.1654923

Vaissiere, T., Sawan, C., and Herceg, Z. (2008). Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutat. Res.* 659, 40–48. doi: 10.1016/j.mrrev.2008.02.004

Varasi, M., Thaler, F., Abate, A., Bigogno, C., Boggio, R., Carenzi, G., et al. (2011). Discovery, synthesis, and pharmacological evaluation of spiropiperidine hydroxamic acid based derivatives as structurally novel histone deacetylase (HDAC) inhibitors. *J. Med. Chem.* 54, 3051–3064. doi: 10.1021/jm200146u

Vergani, B., Sandrone, G., Marchini, M., Ripamonti, C., Cellupica, E., Galbiati, E., et al. (2019). Novel benzohydroxamate-based potent and selective histone deacetylase 6 (HDAC6) inhibitors bearing a pentaheterocyclic scaffold: design, synthesis, and biological evaluation. *J. Med. Chem.* 62, 10711–10739. doi: 10.1021/acs.jmedchem.9b01194

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem Rev.* 119, 9478–9508. doi: 10.1021/acs.chemrev.9b00055

Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247–260. doi: 10.1016/j.jmgm.2005.12.005

Wang, P., Fu, T., Zhang, X., Yang, F., Zheng, G., Xue, W., et al. (2017a). Differentiating physicochemical properties between NDRIs and sNRIs clinically important for the treatment of ADHD. *Biochim. Biophys. Acta* 1861, 2766–2777. doi: 10.1016/j.bbagen.2017.07.022

Wang, P., Zhang, X., Fu, T., Li, S., Li, B., Xue, W., et al. (2017b). Differentiating physicochemical properties between addictive and nonaddictive ADHD drugs revealed by molecular dynamics simulation studies. *ACS Chem. Neurosci.* 8, 1416–1428. doi: 10.1021/acschemneuro.7b00173

Watson, P. J., Millard, C. J., Riley, A. M., Robertson, N. S., Wright, L. C., Godage, H. Y., et al. (2016). Insights into the activation mechanism of class I HDAC complexes by inositol phosphates. *Nat. Commun.* 7:11262. doi: 10.1038/ncomms11262

Weiser, J., Peter, S., Shenkin, and Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* 20, 217–230. doi: 10.1002/(SICI)1096-987X(19990130)20:2217::AID-JCC43.0.CO;2-A

Xingyi, J., Guonan, C., Xin, Z., and Naijie, L. (2019). AbCD133 modified alphaCT1 loaded target magnetic mesoporous silica nano-drugcarriers can sensitizes glioma cancer stem cells to TMZ and have therapeutic potential on TMZ resistant glioblastoma. *J. Biomed. Nanotechnol.* 15, 1468–1481. doi: 10.1166/jbn.2019.2795

Xue, W., Wang, P., Tu, G., Yang, F., Zheng, G., Li, X., et al. (2018a). Computational identification of the binding mechanism of a triple reuptake inhibitor amitifadine for the treatment of major depressive disorder. *Phys. Chem. Chem. Phys.* 20, 6606–6616. doi: 10.1039/C7CP07869B

Xue, W., Yang, F., Wang, P., Zheng, G., Chen, Y., Yao, X., et al. (2018b). What contributes to serotonin-norepinephrine reuptake inhibitors' dual-targeting mechanism? the key role of transmembrane domain 6 in human serotonin and norepinephrine transporters revealed by molecular dynamics simulation. *ACS Chem Neurosci.* 9, 1128–1140. doi: 10.1021/acschemneuro.7b00490

Yang, F., Zheng, G., Fu, T., Li, X., Tu, G., Li, Y. H., et al. (2018). Prediction of the binding mode and resistance profile for a dual-target pyrrolyl diketo acid scaffold against HIV-1 integrase and reverse-transcriptase-associated ribonuclease H. *Phys. Chem. Chem. Phys.* 20, 23873–23884. doi: 10.1039/C8CP01843J

Yang, H., Qin, C., Li, Y. H., Tao, L., Zhou, J., Yu, C. Y., et al. (2016). Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Res.* 44, D1069–1074. doi: 10.1093/nar/gkv1230

Yurek-George, A., Cecil, A. R., Mo, A. H., Wen, S., Rogers, H., Habens, F., et al. (2007). The first biologically active synthetic analogues of FK228, the

depsipeptide histone deacetylase inhibitor. *J. Med. Chem.* 50, 5720–5726. doi: 10.1021/jm0703800

Zhang, J., Stevens, M. F., and Bradshaw, T. D. (2012). Temozolomide: mechanisms of action, repair and resistance. *Curr. Mol. Pharmacol.* 5, 102–114. doi: 10.2174/1874467211205010102

Zhang, Y., Ying, J. B., Hong, J. J., Li, F. C., Fu, T. T., Yang, F. Y., et al. (2019). How does chirality determine the selective inhibition of histone deacetylase 6? A lesson from trichostatin a enantiomers based on molecular dynamics. *ACS Chem. Neurosci.* 10, 2467–2480. doi: 10.1021/acschemneuro.8b00729

Zhang, Z., Wang, Y., Chen, J., Tan, Q., Xie, C., Li, C., et al. (2016). Silencing of histone deacetylase 2 suppresses malignancy for proliferation, migration, and invasion of glioblastoma cells and enhances temozolomide sensitivity. *Cancer Chemother. Pharmacol.* 78, 1289–1296. doi: 10.1007/s00280-016-3188-2

Zheng, G., Xue, W., Wang, P., Yang, F., Li, B., Li, X., et al. (2016). Exploring the inhibitory mechanism of approved selective norepinephrine reuptake inhibitors and reboxetine enantiomers by molecular dynamics study. *Sci. Rep.* 6:26883. doi: 10.1038/srep26883

Zheng, G., Xue, W., Yang, F., Zhang, Y., Chen, Y., Yao, X., et al. (2017). Revealing vilazodone's binding mechanism underlying its partial agonism to the 5-HT1A receptor in the treatment of major depressive disorder. *Phys. Chem. Chem. Phys.* 19, 28885–28896. doi: 10.1039/C7CP05688E

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 24 March 2020
doi: 10.3389/fmolb.2020.00027

Conformational Changes Induced by S34Y and R98C Variants in the Death Domain of Myd88

Vijayakumar Gosu[1], *KyeongHye Won*[1], *Jae-Don Oh*[1]* *and Donghyun Shin*[2]*

[1] Department of Animal Biotechnology, Jeonbuk National University, Jeonju-si, South Korea, [2] The Animal Molecular Genetics and Breeding Center, Jeonbuk National University, Jeonju-si, South Korea

Myeloid differentiating factor 88 (Myd88) is a universal adaptor protein that plays a critical role in innate immunity by mediating TLR downstream signaling. Myd88 death domain (DD) forms an oligomeric complex by association with other DD-containing proteins such as IRAK4. Despite its universal role, polymorphisms in Myd88 can result in several diseases. Previous studies have suggested that, out of several non-synonymous single-nucleotide polymorphisms (nsSNPs), the variants S34Y and R98C in the DD of Myd88 disrupt the formation of the Myddosome complex. Therefore, we performed molecular dynamics (MD) simulations on wild-type ($Myd88^{WT}$) and mutant ($Myd88^{S34Y}$, $Myd88^{R98C}$) DDs to evaluate the subtle conformational changes induced by these mutations. Our results suggest that the S34Y variant induces large structural transitions compared to the R98C variant as evidenced by residual flexibility at the variable loop regions, particularly in the H1–H2 loop, and variations in the collective modes of motion observed for wild-type and mutant Myd88 DDs. The residue interaction network strongly suggests a distortion in the interaction pattern at the location of the mutated residue between the wild type and mutants. Moreover, betweenness centrality values indicate that variations in the distribution of functionally important residues may be reflected by distinct residue signal transductions in both wild-type and mutant Myd88 DDs, which may influence the interaction with other DDs in TLR downstream signaling.

Keywords: Myd88, polymorphism, molecular dynamics simulation, principal components, betweenness centrality

OPEN ACCESS

Edited by:
Ebru Demet Akten,
Kadir Has University, Turkey

Reviewed by:
Igor N. Berezovsky,
*Bioinformatics Institute (A*STAR), Singapore*
Richard Hrabal,
University of Chemistry and Technology in Prague, Czechia

***Correspondence:**
Jae-Don Oh
oh5ow@naver.com
Donghyun Shin
sdh1214@gmail.com

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *04 November 2019*
Accepted: *10 February 2020*
Published: *24 March 2020*

Citation:
Gosu V, Won K, Oh J-D and Shin D (2020) Conformational Changes Induced by S34Y and R98C Variants in the Death Domain of Myd88. Front. Mol. Biosci. 7:27. doi: 10.3389/fmolb.2020.00027

INTRODUCTION

Pattern recognition receptors (PRRs) play a crucial role in triggering the host innate immune response against harmful microbial organisms (Takeuchi and Akira, 2010; Thompson et al., 2011). Toll-like receptors (TLRs) are an important class of PRRs that are activated when pathogen-associated molecular patterns (PAMPs) are sensed. There are about 13 TLRs in mammals, and each TLR senses PAMPs with different specificities. Few TLRs, such as TLR3, TLR7/8, and TLR9, are localized in endosomes, whereas other TLRs are localized extracellularly. Intracellular TLRs recognize nucleic acids, whereas extracellular TLRs such as TLR2, TLR4, and TLR5 recognize lipopeptide (TLR2), lipopolysaccharide (LPS), and flagellin, respectively (Kawai and Akira, 2010; Kawasaki and Kawai, 2014). When PAMPs are detected, TLRs trigger the signaling mediators and activate NF-κB, thereby inducing proinflammatory genes encoding cytokines and chemokines (Gosu et al., 2012; Liu et al., 2017). TLR signaling is predominantly described as myeloid differentiating factor 88 (Myd88)-dependent and Myd88-independent pathways. Myd88 is a

universal adaptor protein that associates with all TLRs except TLR3. TLR3 is mediated through adaptor proteins such as TRIF. In humans, there are five adaptor proteins such as Myd88, TRIF, Mal, TRAM, and SARM in TLR signaling (Troutman et al., 2012). Out of these, MyD88 seems to be crucial because of its universal role in signaling and involvement in innate immunity. Myd88-deficient mice have been shown to be susceptible to leishmanial infection (Muraille et al., 2003; von Bernuth et al., 2012), and Myd88-deficient macrophages have been shown to be defective in the production of TNF and NO upon mycobacterial stimulation (Fremond et al., 2004). Moreover, Myd88 point mutations have been associated with several deadly bacterial infections (von Bernuth et al., 2008; Cervantes, 2017). In addition, Myd88 polymorphisms in chicken also increase their susceptibility to salmonella pullorum infection (Liu et al., 2015). The point mutations in Myd88 have been reported in a previous report (von Bernuth et al., 2008). Experimental reports suggest that out of several non-synonymous single-nucleotide polymorphisms (nsSNPs), S34Y and R98C variants interfere with the Myddosome complex (George et al., 2011). In particular, the S34Y variant is inactive in all Myd88-dependent signaling (Nagpal et al., 2011).

The Myd88 adaptor protein is 296 amino acids (aa) long and comprises a modular domain structure with N-terminal death domain (DD), intermediate domain (ID), and C-terminal TIR domain. The DD is crucial in forming the Myddosome complex via interactions with IRAK4 and IRAK1/IRAK2 DDs, whereas the TIR domain is important for initiating downstream signaling via interactions with the TLR-TIR domain. Extensive structural analyses have been performed on the TIR domains, which is largely conserved among adaptor proteins as well as TLRs (Ohnishi et al., 2009; Mahita and Sowdhamini, 2018). However, there have been very few studies on the structure of Myd88 DD and the variants of Myd88. The Myddosome complex has been resolved through protein crystallography, which suggested that the Myd88 DD is oligomeric in solution with six molecules of Myd88, four molecules of IRAK4, and four molecules of IRAK2. The DD is a small domain that is ∼90 aa, and is composed of six helices connected by loops (Lin et al., 2010) (**Figure 1A**). The rare point mutations S34Y and R98C have been suggested in a previous report to disrupt the formation of the Myddosome complex (George et al., 2011; Nagpal et al., 2011). Moreover, several studies have been reported that show that perturbations due to point mutations originate allosteric mechanism of protein functional activity (Guarnera and Berezovsky, 2019a,b; Tan et al., 2019; Tee et al., 2019). Hence, studies on conformational changes or allosteric mechanism induced by point mutations in the DD of Myd88 may provide structural insights to better understand how the single point mutations influence the DD–DD interactions.

In this study, we evaluated the wild-type Myd88 DD ($Myd88^{WT}$) and Myd88 mutants ($Myd88^{S34Y}$, $Myd88^{R98C}$) using molecular dynamics (MD) simulations of 100 ns with three replicates using different initial velocities. Furthermore, we performed principal component analysis (essential dynamics) and residue network analysis to better understand the subtle conformational changes induced by these point mutations in the DD of Myd88.

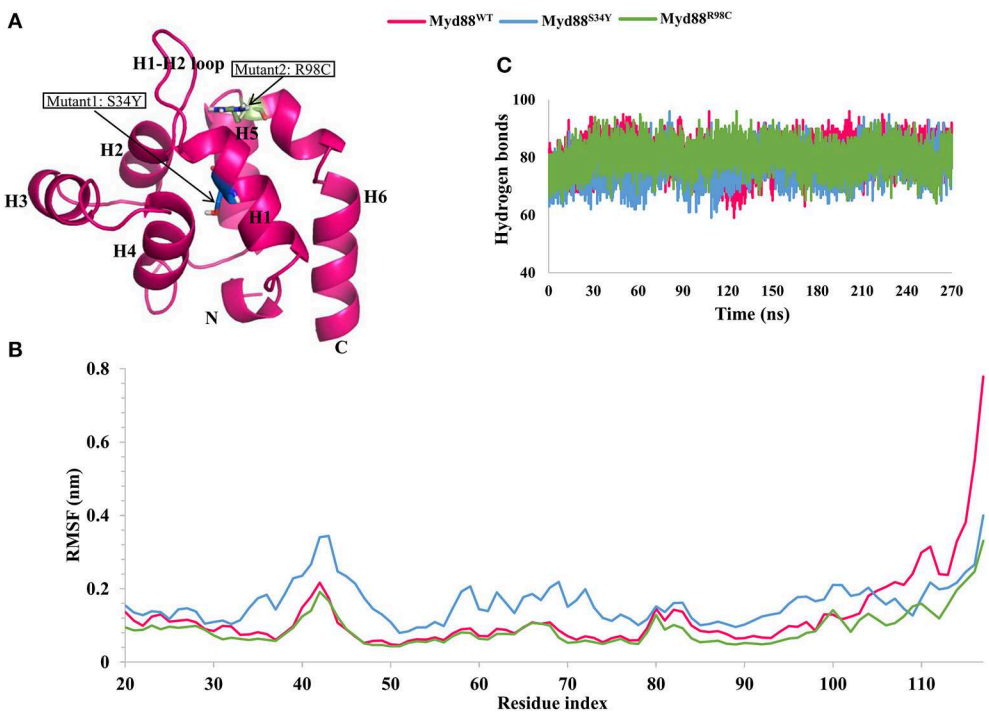


FIGURE 1 | Root mean square fluctuation (RMSF) and hydrogen bond analysis. **(A)** The structure of Myd88 DD. Mutant positions are highlighted. **(B)** The RMSF of backbone atoms of concatenated trajectory (270 ns) from all replicates of each system for wild-type and mutant Myd88. **(C)** The intramolecular hydrogen bonds for concatenated trajectory (270 ns) of wild-type and mutant Myd88.

MATERIALS AND METHODS

Preparation of Wild-Type and Mutant Structures of Myd88 DD

The Myd88 crystal structure has been previously deposited in the Protein Data Bank (PDB) as a Myddosome complex (Lin et al., 2010). We downloaded the Myddosome complex (PDB ID: 3MOP) and extracted the structure of the Myd88 DD. After initial minimization, we truncated both the N- and C-termini because of the long loop structures and considered the residues between 20 and 117 aa; this was considered the wild-type model ($Myd88^{WT}$). Furthermore, we changed serine to tyrosine at residue 34, and changed arginine to cysteine at residue 98 using the mutation option in PyMOL with probable rotamers to construct mutant models ($Myd88^{S34Y}$ and $Myd88^{R98C}$) of the Myd88 DD. Finally, we assessed the stereochemical properties using the ProQ webserver.

MD Simulation Protocol

We subjected all of the structures ($Myd88^{WT}$, $Myd88^{S34Y}$, and $Myd88^{R98C}$) to atomistic MD simulations using Gromacs 5.1.4 (Van Der Spoel et al., 2005; Pronk et al., 2013) with AMBER-ff99SB-ILDN force field (Lindorff-Larsen et al., 2010) similar to our previous report (Gosu et al., 2019). At first, we prepared the topology files using pdb2gmx, placed the structure in a dodecahedral box, and maintained a periodic distance of 12 nm from the protein to the box wall. Tip3p water molecules 6,947, 6,945, and 6,947 were included for $Myd88^{WT}$, $Myd88^{S34Y}$, and $Myd88^{R98C}$, respectively. To neutralize the system, four, four, and five sodium ions were included for $Myd88^{WT}$, $Myd88^{S34Y}$, and $Myd88^{R98C}$, respectively. The energy minimization was performed using the steepest descent method with 1,000 kJ/mol/nm as a maximum tolerance. Using the energy-minimized structure, NVT equilibration simulations were performed for 100 ps. Subsequently, an NPT equilibration simulation was performed for 500 ps using positional restraints. All of the bonds were restrained using the Lincs algorithm, and short-range electrostatics and Van der Waals interactions were accounted for with a 1.0-nm cutoff, and long-range electrostatics were maintained using the Particle Mesh Ewald (PME) method. Temperature (300 K) and pressure (1.0 bar) were maintained using v-rescale, a modified Berendsen thermostat, and Parrinello-Rahman barostat, respectively. Production simulations were performed without positional restraints for 100 ns. Three independent replicate simulations using different initial velocities were performed for all the systems.

Principal Components (PCs) and Free Energy Landscape (FEL) Analysis

PCs calculations were used to obtain concerted motions during simulations that are likely significant for biological function. We calculated PCs on the concatenated trajectory (last 90 ns of each replicate from three systems) as reported in previous reports (Gosu and Choi, 2014; Gosu et al., 2019). After removing the

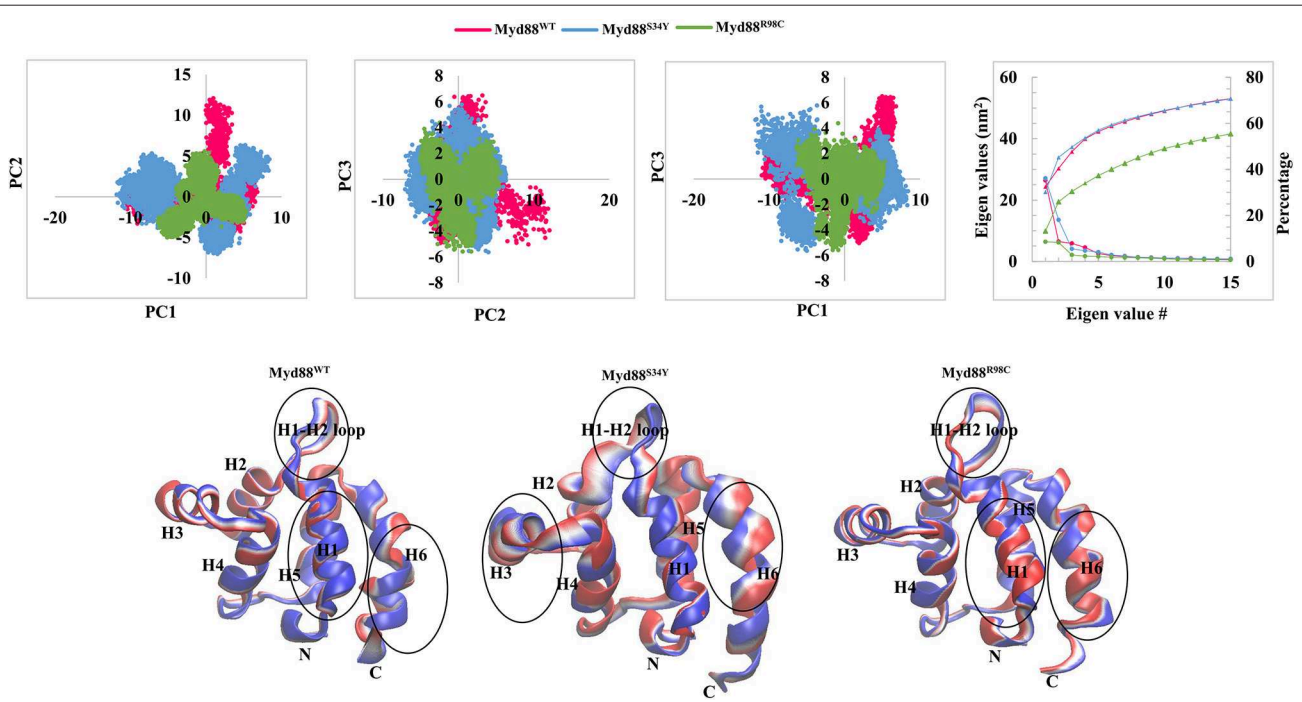


FIGURE 2 | Collective modes of motions. **(Top panel)** The projection of PC1 and PC2, PC2 and PC3, and PC1 and PC3 is shown. The covariance of each eigenvalue with respective eigenvector is shown. In addition, cumulative percentage is also shown. **(Bottom panel)** Two extreme conformations with 30 frames from the concatenated trajectory for PC1 are shown and sequentially superimposed. The minimum (red) to maximum (blue) conformations are shown. The circle represents the large conformational changes.

rotational and translational motions, the covariance matrix was constructed and diagonalized, which yielded a set of eigenvalues (amplitude) and eigenvectors (direction of motion). We obtained eigenvalues and eigenvectors using the gmx covar and analyzed the data using the gmx anaeig tool from the gromacs package (van Aalten et al., 1995; Yamaguchi et al., 1998). The gmx sham tool was used to construct the FEL, which is a combination of data points from the reaction coordinates of PC1 and PC2, and plots were drawn using the Mathematica version 12.

Residue Interaction Networks

Residue interaction networks (RINs) are widely used to understand the impact of mutations on proteins. We constructed representative structures of each system (both wild type and mutants) as a network, in which the nodes represent residues and the edge represents contacts between residues using RINalyzer (Doncheva et al., 2011) and the structure viz module implemented in cytoscape 3.14, similar to a previous report (Anwar and Choi, 2017). The contact distance between any two residues was considered at 5 Å. Furthermore, the network topological parameters were calculated as an undirected network using the network analyzer (Assenov et al., 2008). Betweenness centrality (C_B), closeness and node degree distribution were calculated. C_B is an important factor that suggests the residues crucial for the functional importance of the proteins (Lee et al., 2014; Basith et al., 2019). In addition, we also constructed RIN using the RING2.0 webserver, which was useful to inspect the various interactions such as hydrogen bonds, Van der Waals interactions, and ionic interactions in protein models (Piovesan et al., 2016).

Evaluation of Allosteric Effects

To quantify the allosteric effects under mutation for Myd88 DD, a statistical mechanical model implemented in the AlloSigMA server (Guarnera et al., 2017) was used. This method estimates the free energy of each residue used by the allosteric

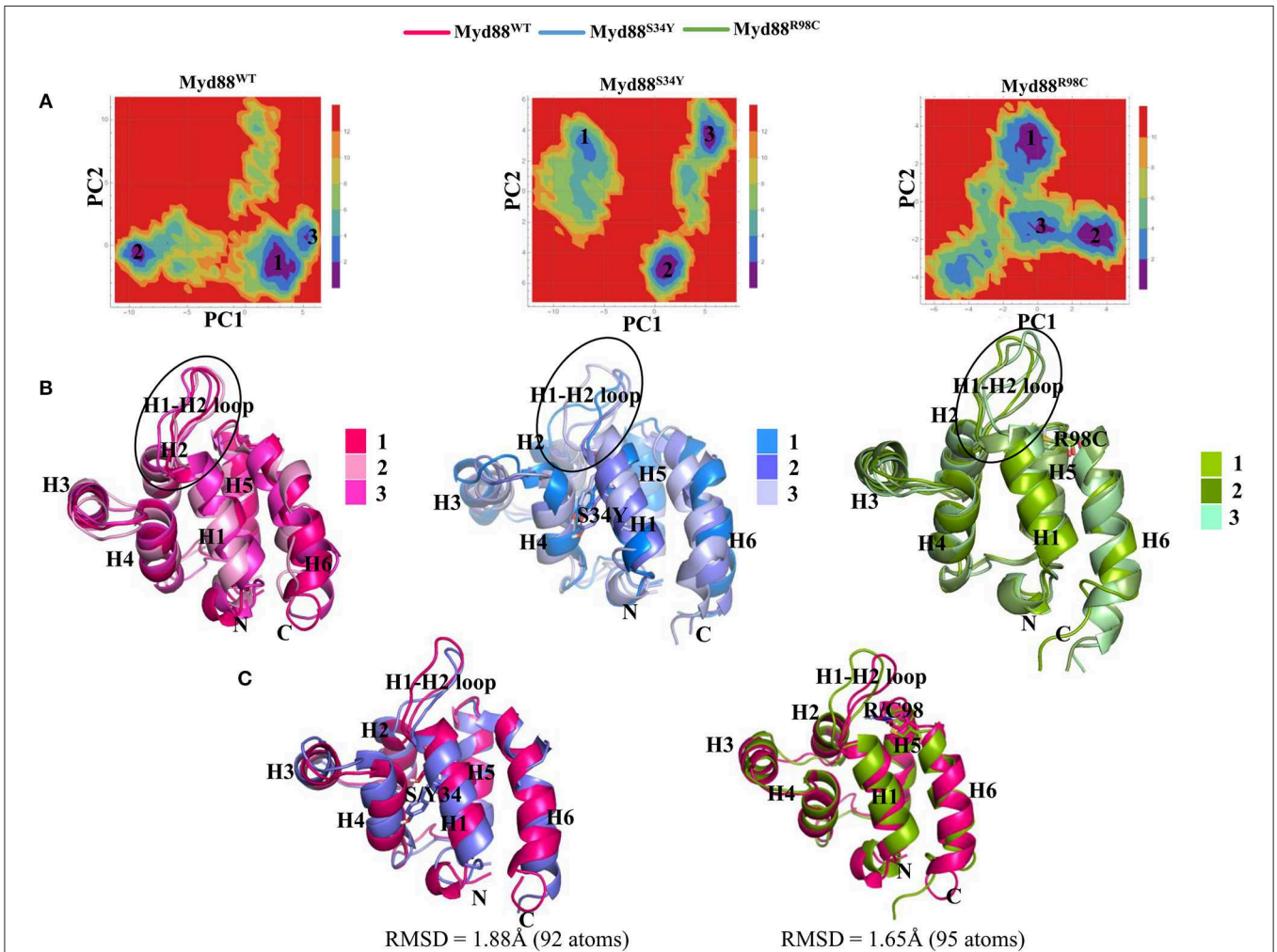


FIGURE 3 | Free energy landscape (FEL) and representative structures. **(A)** FEL of the wild-type and mutant Myd88 are shown using PC1 and PC2 as reaction coordinates. **(B)** The representative structures of Myd88 were considered and superimposed. The black circle represents the variation in the H1–H2 loop. **(C)** The superimpositions of the wild-type representative structure on both mutants are shown along with RMSD values.

communication under mutation or binding events, which is effectively verified in previous studies (Takeuchi and Akira, 2010; Guarnera and Berezovsky, 2016; Su et al., 2018a,b). We initiated perturbations at different positions (34 and 98) using UP mutation to simulate the bulky residues at these positions and subsequently calculate the free energies accountable for allosteric communication.

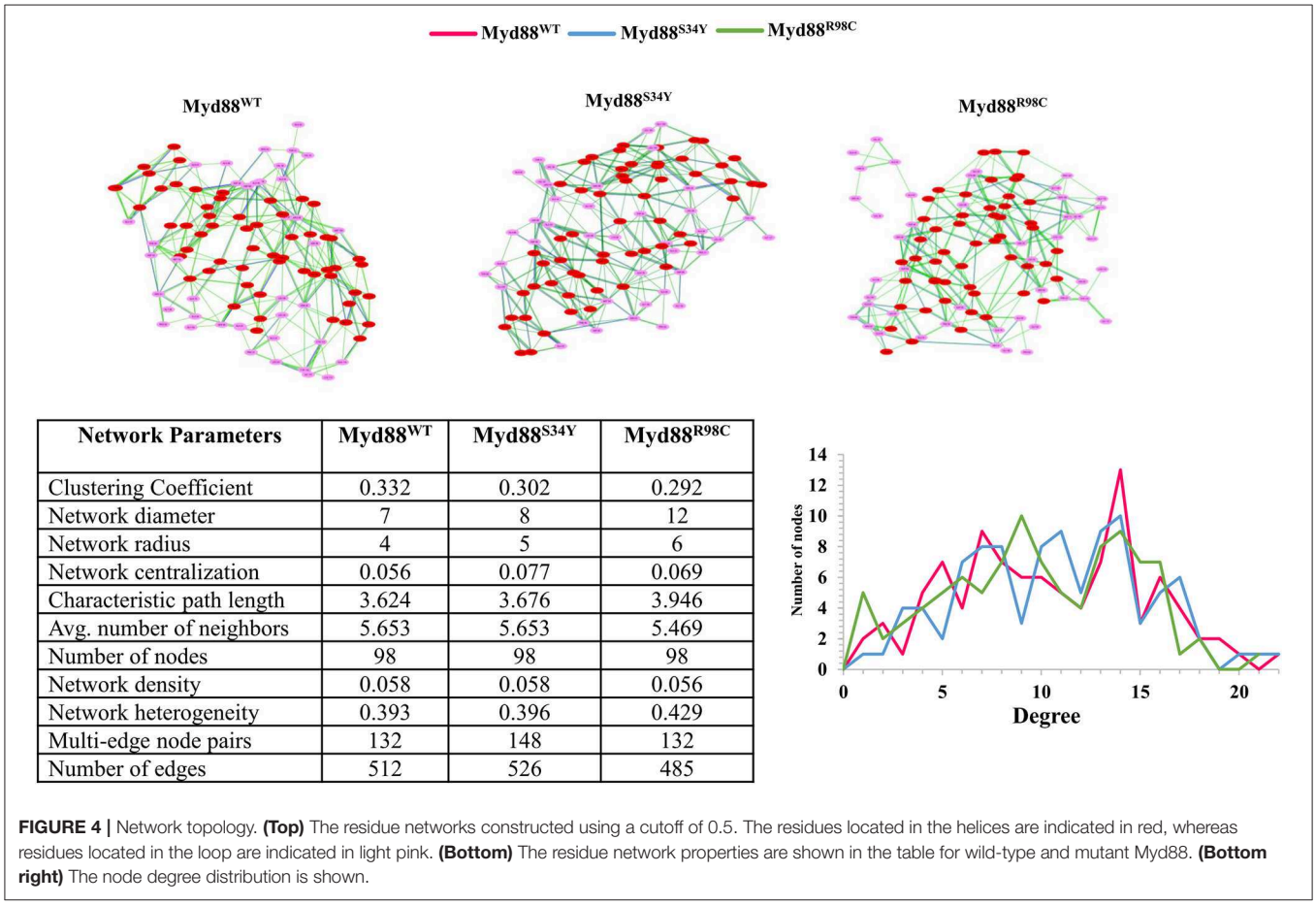
RESULTS AND DISCUSSION

Structural Dynamics of Wild-Type and Mutant Myd88 DDs

The Myd88 DD structure ($Myd88^{WT}$) was extracted from the Myddosome complex (PDB ID: 3MOP). After initial inspection, we constructed the mutant Myd88 DD models ($Myd88^{S34Y}$ and $Myd88^{R98C}$). Subsequently, we subjected all three models ($Myd88^{WT}$, $Myd88^{S34Y}$, and $Myd88^{R98C}$) to conventional MD simulations for 100 ns with three independent replicates (total 300 ns) using different initial velocities. To assess the stability of all of the systems, we calculated the root mean square deviation (RMSD) of backbone atoms with respect to the initial structure for the whole MD trajectories. RMSD suggested that all of the replicates showed a consistent increase in RMSD until ∼10 ns. Thereafter, RMSD values ranged from 0.1 to 0.3 nm for $Myd88^{WT}$, 0.1–0.4 nm for $Myd88^{S34Y}$, and 0.1–0.2 nm for $Myd88^{R98C}$ (**Figure S1**). However, the replicates of each system had slight variations in the RMSD values. In particular, $Myd88^{WT}$ showed more variations among replicates. $Myd88^{WT}$ was stable during the whole MD trajectory; however, it showed large flexibility at the end of the simulation (**Figure S1**). The observed RMSD values between wild-type and mutant models of the Myd88 DD indicated that the global structural deviation did not vary largely. Furthermore, the radius of gyration (Rg) of the backbone atoms revealed that all of the systems were compact during the simulations (**Figure S1**). However, the difference in Rg values indicate that S34Y mutant in Myd88 may exhibit large conformational changes compared to R98C mutant. In addition, we also observed that solvent accessible surface area (SASA) averaged from three replicates were 61.52, 63.03, and 61.8 for $Myd88^{WT}$, $Myd88^{S34Y}$, and $Myd88^{R98C}$, respectively. Despite the compact folding of the DD, the increase in SASA for mutant Myd88 (in particular $Myd88^{S34Y}$) compared to wild-type Myd88 indicated obvious conformational changes induced by mutations (**Figure S1**).

Residue Flexibility and Intra-Hydrogen Bonds During Simulations

From the three replicates of each system, we concatenated the last 90 ns trajectory (a total of 270 ns for each system) for further analysis. We checked residual flexibility using root



Network Parameters	$Myd88^{WT}$	$Myd88^{S34Y}$	$Myd88^{R98C}$
Clustering Coefficient	0.332	0.302	0.292
Network diameter	7	8	12
Network radius	4	5	6
Network centralization	0.056	0.077	0.069
Characteristic path length	3.624	3.676	3.946
Avg. number of neighbors	5.653	5.653	5.469
Number of nodes	98	98	98
Network density	0.058	0.058	0.056
Network heterogeneity	0.393	0.396	0.429
Multi-edge node pairs	132	148	132
Number of edges	512	526	485

FIGURE 4 | Network topology. **(Top)** The residue networks constructed using a cutoff of 0.5. The residues located in the helices are indicated in red, whereas residues located in the loop are indicated in light pink. **(Bottom)** The residue network properties are shown in the table for wild-type and mutant Myd88. **(Bottom right)** The node degree distribution is shown.

mean square fluctuations (RMSFs) of backbone atoms for the concatenated trajectory. The variation in the RMSF values suggested that the residual flexibility was similar between the wild-type and mutant systems except at the C-terminal residues and variable loop regions (**Figure 1B**). Deeper inspection of RMSF values indicated high residue flexibility at the H1–H2 loop region (39–48) in Myd88^{S34Y} compared to both Myd88^{WT} and Myd88^{R98C}. Moreover, the residue movements were larger in Myd88^{S34Y} compared to other systems. This was reasonable because the change in serine (small amino acid) to tyrosine (bulky amino acid) at residue 34 may induce conformational changes in the surrounding region (**Figure S2**). Additionally, Myd88^{R98C} and Myd88^{WT} show similar residue fluctuations; however, large fluctuations for wild type were observed in the region of helix 6 (100–117). Furthermore, we analyzed the intra-hydrogen bonds for all systems, which indicated their slight variations, particularly for the Myd88^{S34Y} mutant. On average, 80, 77, and 80 hydrogen bonds were observed for Myd88^{WT}, Myd88^{S34Y}, and Myd88^{R98C}, respectively (**Figure 1C**). The above analyses cumulatively suggested that mutations in Myd88 DD had an impact on the overall structural organization, which may influence the DD–DD interaction in the downstream signaling of TLRs.

Collective Motions of Wild-Type and Mutant Myd88 DD

In order to assess the dominant modes and conformational changes particularly induced by mutations, we performed a principal component analysis (PCA) on the concatenated trajectory (270 ns) of each system. The PCA indicated that the first few eigenvectors had eigenvalues >1 nm^2 as shown in **Figure 2**. The diagonalized co-variances of wild-type and mutant models were 81.19 for Myd88^{WT}, 89.9 for Myd88^{S34Y}, and 48.65 for Myd88^{R98C}. This indicated that Myd88^{S34Y} underwent a large fluctuation compared to Myd88^{WT} and Myd88^{R98C}. The cumulative percentage of mean square fluctuation for the first 15 eigenvectors were 68% for Myd88^{WT}, 69% for Myd88^{S34Y}, and 55% for Myd88^{R98C}. Furthermore, the first three eigenvectors contribute large motions, i.e., 60% for Myd88^{WT}, 61% for Myd88^{S34Y}, and 32% for Myd88^{R98C} (**Figure 2**). To assess the possible reason for less global dynamics (only 32% for first three PCs), of the Myd88^{R98C} mutant, we performed the RMSD calculations at residue 98 position for concatenated trajectory of all models, which show that Arg at the 98 position is largely flexible compared to Cys at position 98 in mutant (Myd88^{R98C}) model (**Figure S2A**). The minimum distance calculations between Arg98/Cys98 with Phe36 from helix 1, Ala45 from the H1–H2 loop, and Asp100 from helix 6 suggest possible local structural alterations; in particular, the minimum distance between Cys98 and Ala45 from the H1–H2 loop is higher in Myd88^{R98C} compared to wild type (**Figure S2C**). Moreover, the SASA (**Figure S2E**) and atomic fluctuations (**Figure S3A**) at the 98 position is largely varied compared to wild type. All the above analysis suggests that this mutant (R98C) may alter the interaction pattern particularly with the H1–H2 loop at the mutant surrounded region, thereby leading to

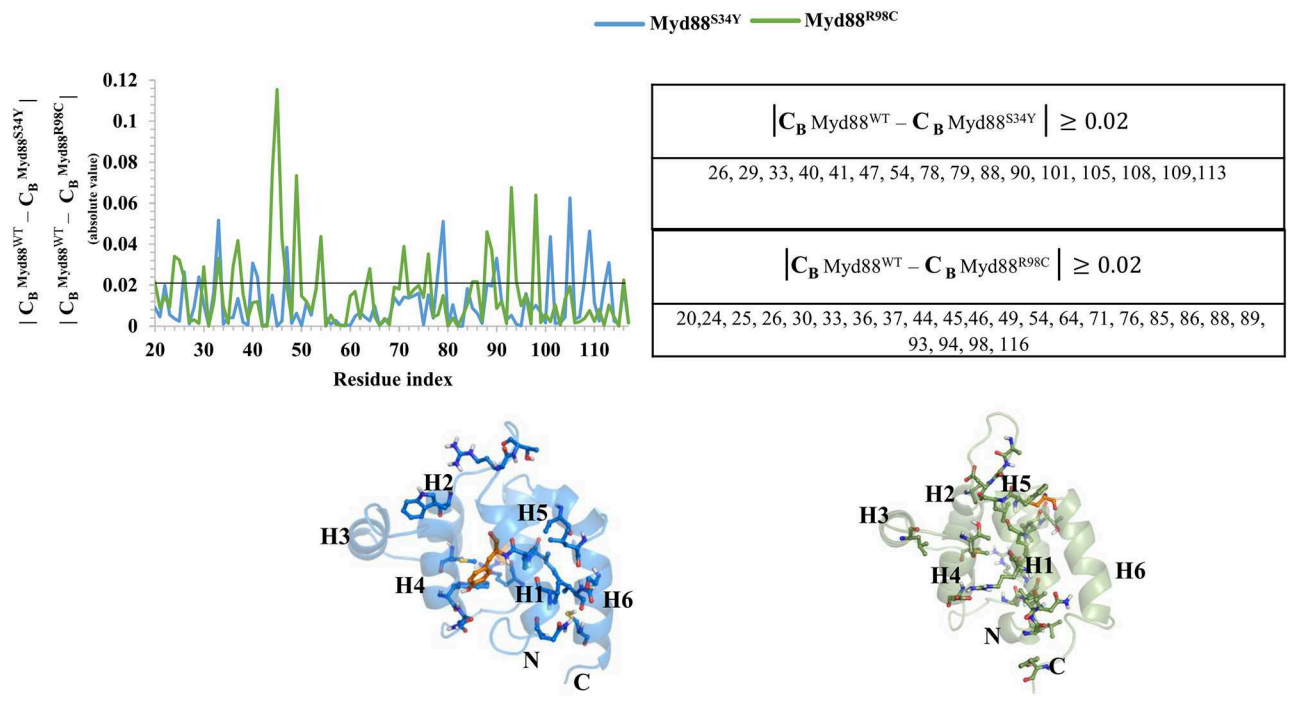


FIGURE 5 | Network centrality. The variation in the betweenness centrality (C_B) between wild-type and mutant Myd88 is shown. Residues showing $C_B \geq 0.02$ are presented in the table. The residues identified are shown on the corresponding structures of Myd88^{S34Y} and Myd88^{R98C}. The mutant residues Y34 and C98 are shown in orange color on the corresponding structures.

less dynamics globally compared to Myd88^{WT} and Myd88^{S34Y}. Additionally, the trajectories obtained for the first three PCs were projected on the phase space (**Figure 2**), which indicated that the Myd88^{S34Y} mutant was more dynamic compared to other systems. The spreading of the trajectory on the phase space clearly indicated few clusters for the Myd88^{S34Y} mutant, whereas Myd88^{R98C} shows less spreading of the trajectory compared to wild type. Hence, it was evident that the serine-to-tyrosine mutation at residue 34 had a large impact on the structure within the Myd88 DD.

In order to assess the variations in the collective motions of the wild-type and mutant systems, we considered two extreme positions with 30 frames form the concatenated trajectory of all the systems along the first three eigenvectors as shown in **Figure 2** and **Figure S3B**. In addition, **Movies S1–S9** show animations of the motions for all systems along the first three eigenvectors. The mutant Myd88, in particular Myd88^{S34Y}, shows overall large fluctuations. The first dominant mode (PC1) in Myd88^{WT} shows that the H1–H2 loop region (39–48) moves toward helix 3; in contrast, the mutants moved in the opposite direction (**Figure 2** and **Movies S1–S3**). However, large motions were observed in Myd88^{S34Y}, indicating that this mutant altered the structural organization of Myd88 DD, which may influence the symmetry required for Myddosome formation. Moreover, helix 6 was observed to undergo large motion in both the wild type and mutants. The second and third dominant modes (PC2 and PC3) show overall large motion in helix 6 as well as in the H1–H2 loop (**Figure S3B** and **Movies S4–S9**). Furthermore, in order to assess the low-energy structures, we performed a FEL analysis, which suggested that Myd88^{WT} may not undergo large conformations during simulations compared to mutants. Importantly, Myd88^{S34Y} underwent large conformations as shown in **Figures 3A,B**, particularly in the H1–H2 loop. Myd88^{R98C} exhibits less conformational difference; however, this mutant also shows variation in the H1–H2 loop region compared to Myd88^{WT} (**Figure 3C**).

RINs and Network Centrality

Recently, it has become common to consider graph theory to construct the RINs, thereby identifying the crucial residues for protein function. To understand the difference in the flow of information for wild-type and mutant Myd88, we constructed RINs on the representative structures using RING (default parameters) and RINlyzer with a contact threshold of 5 Å and then analyzed the topological parameters using the network analyzer. Each residue was considered as node and the contacts between the nodes are considered as edges. The RING analysis, which is useful to understand multiple interaction types involved in the residue networks, indicated that the hydrogen bond and van der Waals interactions largely varied at the location of the mutations, suggesting that this distinct conformation at the mutant position lead to the overall conformational changes in the DD (**Figure S4**). The RIN and RIN topological parameters are shown in **Figure 4**. The node degree distribution indicated large variations in RINs between wild-type and mutant Myd88. The network diameter and radius varied among wild-type and mutant Myd88; however, the clustering coefficient and characteristic path length indicated that the network belonged to small world topology. Moreover, the total number of edges was varied among the three systems, indicating that each one exhibits significant differences in signal transduction within the protein (**Figure 4**). Subsequently, we calculated betweenness centrality

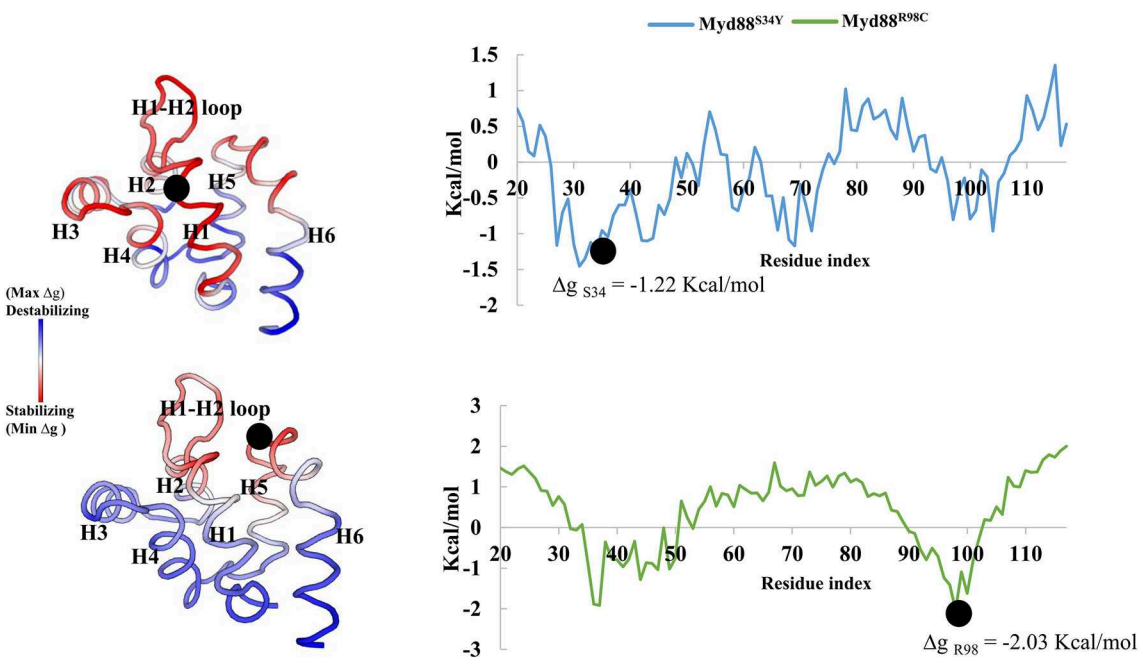


FIGURE 6 | Free energy of Myd88 mutants. Dot represents the mutant residue in the corresponding mutant model.

(C_B), closeness (C_C), and degree (C_D) for indications of large variations between wild-type and mutant Myd88 (**Figure S5**). In particular, betweenness centrality was useful for identifying the functionally important residues that are involved in transducing the signal within the protein. Hence, we considered the cutoff ($C_B \geq 0.04$) and illustrated the residues to better understand the differences in wild-type and mutant Myd88 (**Figure S5**). We observed that the distribution of functionally important residues varied in the wild type and mutants. For an in-depth understanding, using the condition ($|\mathbf{C}_B{}^{\text{Myd88WT}} - \mathbf{C}_B{}^{\text{Myd88S34Y}}| \geq 0.02$), ($|\mathbf{C}_B{}^{\text{Myd88WT}} - \mathbf{C}_B{}^{\text{Myd88S34Y}}| \geq 0.02$), we calculated the betweenness centrality variation between wild-type and mutant models (**Figure 5**). From this analysis, we observed that some of the residues in both the mutants were similar, which indicated that the mutation in Myd88 DD may lead to an allosteric mechanism. In order to assess the allosteric mechanism under mutation, free energy (Δg_{res}) of each residue was calculated using the AlloSigMA server. The prediction of free energy (Δg_{res}) indicated that the residues surrounded by the mutant at the 34 position is stabilized (negative free energy); however, the H2–H3 loop, the H4–H5 loop, and helix 5 were destabilized (positive free energy). Similarly, the residue free energy (Δg_{res}) calculated under mutation at position 98 indicated less stabilization (negative free energy) at the surrounding region of position 98 and destabilization (positive free energy) in the region of helices 1, 3, 4, and 6 (**Figure 6**). Compared to S34, R98 mutant shows a large destabilized region, indicating that R98 side chain is involved in interactions between the H1–H2 loop, helix 1, and helix 5; however, upon mutation, these interactions may be disturbed as shown from minimum distance calculations (**Figure S2**). Hence, it is possible that point mutation in Myd88 DD may exert an allosteric mechanism to regulate the function of protein, which is similar to previous reports suggesting the possible allosteric mechanism of protein function induced by point mutations through perturbations (Guarnera and Berezovsky, 2019a,b; Tee et al., 2019).

CONCLUSION

Myd88 is a crucial signaling adaptor in TLR signaling and is associated with the innate immune system. Myd88 deficiency in mice results in several immune diseases. Moreover, the polymorphisms in TLR signaling mediators are important because of the crucial role of these molecules in innate immunity. Hence, in this study, we performed MD simulations on wild-type and mutant Myd88 DDs to better understand the subtle conformational changes induced by the mutations in the DD of Myd88. First, we extracted the crystal structures of the Myd88 DD from PDB (ID: 3MOP) and then applied mutations by replacing serine with tyrosine at position 34 and arginine with cysteine at position 98. Finally, all three models ($Myd88^{WT}$, $Myd88^{S34Y}$, and $Myd88^{R98C}$) were subjected to MD simulations. Our results suggest that the high residue flexibility (in the mutant model $Myd88^{S34Y}$) in the H1–H2 loop region at residues 39–48 may affect Myddosome formation. Furthermore, we observed variations largely at the loop regions of $Myd88^{S34Y}$ compared to other systems ($Myd88^{WT}$ and $Myd88^{R98C}$), which strongly indicated that the S34Y mutation induced large conformational changes particularly at the loop regions and might affect the symmetry required for Myddosome formation. The variation in the betweenness centrality (C_B) as well as changes in the free energies (Δg_{res}) of each residue upon mutation indicate that point mutation in Myd88 DD affects the dynamics, which may lead to allosteric regulation of Myd88 functional activity. MDs and residue network analysis are powerful tools to better understand the structure–function relationship of proteins. We hope that the results obtained from this study may help in understanding allosteric synergism induced by mutations in the DD of Myd88, which may have an influence in the formation of the Myddosome complex.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

VG designed the work and performed the simulations. VG and KW analyzed the data. VG, J-DO, and DS wrote the paper. J-DO and DS supervised the work. All authors read and approved the manuscript.

FUNDING

This work was supported by the Next-Generation BioGreen 21 Program (No. PJ0131512020), Rural Development Administration, Republic of Korea, and was also supported by a grant from the National Research Foundation of Korea (NRF-2018R1C1B6008141).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00027/full#supplementary-material

REFERENCES

Anwar, M. A., and Choi, S. (2017). Structure-activity relationship in TLR4 mutations: atomistic molecular dynamics simulations and residue interaction network analysis. *Sci. Rep.* 7:43807. doi: 10.1038/srep43807

Assenov, Y., Ramirez, F., Schelhorn, S. E., Lengauer, T., and Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics* 24, 282–284. doi: 10.1093/bioinformatics/btm554

Basith, S., Manavalan, B., Shin, T. H., and Lee, G. (2019). A molecular dynamics approach to explore the intramolecular signal transduction of PPAR-alpha. *Int. J. Mol. Sci.* 20:1666. doi: 10.3390/ijms20071666

Cervantes, J. L. (2017). MyD88 in *Mycobacterium tuberculosis* infection. *Med. Microbiol. Immunol.* 206, 187–193. doi: 10.1007/s00430-017-0495-0

Doncheva, N. T., Klein, K., Domingues, F. S., and Albrecht, M. (2011). Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* 36, 179–182. doi: 10.1016/j.tibs.2011.01.002

Fremond, C. M., Yeremeev, V., Nicolle, D. M., Jacobs, M., Quesniaux, V. F., and Ryffel, B. (2004). Fatal *Mycobacterium tuberculosis* infection despite adaptive immune response in the absence of MyD88. *J. Clin. Invest.* 114, 1790–1799. doi: 10.1172/JCI21027

George, J., Motshwene, P. G., Wang, H., Kubarenko, A. V., Rautanen, A., Mills, T. C., et al. (2011). Two human MYD88 variants, S34Y and R98C, interfere with MyD88-IRAK4-myddosome assembly. *J. Bbiol. Chem.* 286, 1341–1353. doi: 10.1074/jbc.M110.159996

Gosu, V., Basith, S., Kwon, O. P., and Choi, S. (2012). Therapeutic applications of nucleic acids and their analogues in Toll-like receptor signaling. *Molecules* 17, 13503–13529. doi: 10.3390/molecules171113503

Gosu, V., and Choi, S. (2014). Structural dynamic analysis of apo and ATP-bound IRAK4 kinase. *Sci. Rep.* 4:5748. doi: 10.1038/srep05748

Gosu, V., Son, S., Shin, D., and Song, K. D. (2019). Insights into the dynamic nature of the dsRNA-bound TLR3 complex. *Sci. Rep.* 9:3652. doi: 10.1038/s41598-019-39984-8

Guarnera, E., and Berezovsky, I. N. (2016). Structure-based statistical mechanical model accounts for the causality and energetics of allosteric communication. *PLoS Comput. Biol.* 12:e1004678. doi: 10.1371/journal.pcbi.1004678

Guarnera, E., and Berezovsky, I. N. (2019a). On the perturbation nature of allostery: sites, mutations, and signal modulation. *Curr. Opin. Struct. Biol.* 56, 18–27. doi: 10.1016/j.sbi.2018.10.008

Guarnera, E., and Berezovsky, I. N. (2019b). Toward comprehensive allosteric control over protein activity. *Structure 27*, 866–878.e861. doi: 10.1016/j.str.2019.01.014

Guarnera, E., Tan, Z. W., Zheng, Z., and Berezovsky, I. N. (2017). AlloSigMA: allosteric signaling and mutation analysis server. *Bioinformatics* 33, 3996–3998. doi: 10.1093/bioinformatics/btx430

Kawai, T., and Akira, S. (2010). The role of pattern-recognition receptors in innate immunity: update on toll-like receptors. *Nat. Immunol.* 11, 373–384. doi: 10.1038/ni.1863

Kawasaki, T., and Kawai, T. (2014). Toll-like receptor signaling pathways. *Front. Immunol.* 5:461. doi: 10.3389/fimmu.2014.00461

Lee, Y., Choi, S., and Hyeon, C. (2014). Mapping the intramolecular signal transduction of G-protein coupled receptors. *Proteins* 82, 727–743. doi: 10.1002/prot.24451

Lin, S. C., Lo, Y. C., and Wu, H. (2010). Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling. *Nature* 465, 885–890. doi: 10.1038/nature09121

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–1958. doi: 10.1002/prot.22711

Liu, T., Zhang, L., Joo, D., and Sun, S.-C. (2017). NF-κB signaling in inflammation. *Signal Transduct. Target. Ther.* 2:17023. doi: 10.1038/sigtrans.2017.23

Liu, X. Q., Wang, F., Jin, J., Zhou, Y. G., Ran, J. S., Feng, Z. Q., et al. (2015). MyD88 polymorphisms and association with susceptibility to Salmonella Pullorum. *Biomed. Res. Int.* 2015:692973. doi: 10.1155/2015/692973

Mahita, J., and Sowdhamini, R. (2018). Probing subtle conformational changes induced by phosphorylation and point mutations in the TIR domains of TLR2 and TLR3. *Proteins* 86, 524–535. doi: 10.1002/prot.25471

Muraille, E., De Trez, C., Brait, M., De Baetselier, P., Leo, O., and Carlier, Y. (2003). Genetically resistant mice lacking MyD88-adapter protein display a high susceptibility to Leishmania major infection associated with a polarized Th2 response. *J. Immunol.* 170, 4237–4241. doi: 10.4049/jimmunol.170.8.4237

Nagpal, K., Plantinga, T. S., Sirois, C. M., Monks, B. G., Latz, E., Netea, M. G., et al. (2011). Natural loss-of-function mutation of myeloid differentiation protein 88 disrupts its ability to form Myddosomes. *J. Biol. Chem.* 286, 11875–11882. doi: 10.1074/jbc.M110.199653

Ohnishi, H., Tochio, H., Kato, Z., Orii, K. E., Li, A., Kimura, T., et al. (2009). Structural basis for the multiple interactions of the MyD88 TIR domain in TLR4 signaling. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10260–10265. doi: 10.1073/pnas.0812956106

Piovesan, D., Minervini, G., and Tosatto, S. C. (2016). The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* 44, W367–W374. doi: 10.1093/nar/gkw315

Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854. doi: 10.1093/bioinformatics/btt055

Su, C. T., Kwoh, C. K., Verma, C. S., and Gan, S. K. (2018b). Modeling the full length HIV-1 Gag polyprotein reveals the role of its p6 subunit in viral maturation and the effect of non-cleavage site mutations in protease drug resistance. *J. Biomol. Struct. Dyn.* 36, 4366–4377. doi: 10.1080/07391102.2017.1417160

Su, C. T.-T., Lua, W.-H., Ling, W.-L., and Gan, S. K.-E. (2018a). Allosteric effects between the antibody constant and variable regions: a study of iga fc mutations on antigen binding. *Antibodies* 7:20. doi: 10.3390/antib7020020

Takeuchi, O., and Akira, S. (2010). Pattern recognition receptors and inflammation. *Cell* 140, 805–820. doi: 10.1016/j.cell.2010.01.022

Tan, Z. W., Tee, W. V., Guarnera, E., Booth, L., and Berezovsky, I. N. (2019). AlloMAPS: allosteric mutation analysis and polymorphism of signaling database. *Nucleic Acids Res.* 47, D265–D270. doi: 10.1093/nar/gky1028

Tee, W. V., Guarnera, E., and Berezovsky, I. N. (2019). On the allosteric effect of NSSNPS and the emerging importance of allosteric polymorphism. *J. Mol. Biol.* 431, 3933–3942. doi: 10.1016/j.jmb.2019.07.012

Thompson, M. R., Kaminski, J. J., Kurt-Jones, E. A., and Fitzgerald, K. A. (2011). Pattern recognition receptors and the innate immune response to viral infection. *Viruses* 3, 920–940. doi: 10.3390/v3060920

Troutman, T. D., Bazan, J. F., and Pasare, C. (2012). Toll-like receptors, signaling adapters and regulation of the pro-inflammatory response by PI3K. *Cell Cycle* 11, 3559–3567. doi: 10.4161/cc.21572

van Aalten, D. M., Findlay, J. B., Amadei, A., and Berendsen, H. J. (1995). Essential dynamics of the cellular retinol-binding protein–evidence for ligand-induced conformational changes. *Protein Eng.* 8, 1129–1135. doi: 10.1093/protein/8.11.1129

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi: 10.1002/jcc.20291

von Bernuth, H., Picard, C., Jin, Z., Pankla, R., Xiao, H., Ku, C. L., et al. (2008). Pyogenic bacterial infections in humans with MyD88 deficiency. *Science* 321, 691–696. doi: 10.1126/science.1158298

von Bernuth, H., Picard, C., Puel, A., and Casanova, J. L. (2012). Experimental and natural infections in MyD88- and IRAK-4-deficient mice and humans. *Eur. J. Immunol.* 42, 3126–3135. doi: 10.1002/eji.201242683

Yamaguchi, H., van Aalten, D. M., Pinak, M., Furukawa, A., and Osman, R. (1998). Essential dynamics of DNA containing a cis.syn cyclobutane thymine dimer lesion. *Nucleic Acids Res.* 26, 1939–1946. doi: 10.1093/nar/26.8.1939

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 16 April 2020
doi: 10.3389/fmolb.2020.00051

Understanding Thermostability Factors of Barley Limit Dextrinase by Molecular Dynamics Simulations

*Juan Du[1,2], Jianjun Dong[1], Songjie Du[2], Kun Zhang[2], Junhong Yu[1], Shumin Hu[1] and Hua Yin[1]**

[1] State Key Laboratory of Biological Fermentation Engineering of Beer, Tsingtao Brewery, Qingdao, China, [2] Shandong Province Key Laboratory of Applied Mycology, College of Life Sciences, Qingdao Agricultural University, Qingdao, China

Limit dextrinase (LD) is the only endogenous starch-debranching enzyme in barley (*Hordeum vulgare*, *Hv*), which is the key factor affecting the production of a high degree of fermentation. Free LD will lose its activity in the mashing process at high temperature in beer production. However, there remains a lack of understanding on the factor affecting the themostability of *Hv*LD at the atomic level. In this work, the molecular dynamics simulations were carried out for *Hv*LD to explore the key factors affecting the thermal stability of LD. The higher value of root mean square deviation (RMSD), radius of gyration (R_g), and surface accessibility (SASA) suggests the instability of *Hv*LD at high temperatures. Intra-protein hydrogen bonds and hydrogen bonds between protein and water decrease at high temperature. Long-lived hydrogen bonds, salt bridges, and hydrophobic contacts are lost at high temperature. The salt bridge interaction analysis suggests that these salt bridges are important for the thermostability of *Hv*LD, including E568–R875, D317–R378, D803–R884, D457–R214, D468–R395, D456–R452, D399–R471, and D541–R542. Root mean square fluctuation (RMSF) analysis identified the thermal-sensitive regions of *Hv*LD, which will facilitate enzyme engineering of *Hv*LD for enhanced themostability.

Keywords: barley limit dextrinase, thermostability, molecular dynamics simulation, hydrogen bond, salt-bridge

OPEN ACCESS

Edited by:
Guang Hu,
Soochow University, China

Reviewed by:
Jian Tian,
Chinese Academy of Agricultural Sciences, China
Hao Dong,
Nanjing University, China

***Correspondence:**
Hua Yin
yinhua@tsingtao.com.cn

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *02 December 2019*
Accepted: *16 March 2020*
Published: *16 April 2020*

Citation:
Du J, Dong J, Du S, Zhang K, Yu J, Hu S and Yin H (2020) Understanding Thermostability Factors of Barley Limit Dextrinase by Molecular Dynamics Simulations. Front. Mol. Biosci. 7:51. doi: 10.3389/fmolb.2020.00051

INTRODUCTION

Limit dextrinase (LD), also termed R-enzyme, pullulanase, isoamylase, or amylopectin 6-glucanohydrolase, is the only endogenous starch-debranching enzyme in barley (*Hordeum vulgare*, *Hv*) that digests amylopectin and dextrins (Manners et al., 1970; Yang et al., 2008). *Hv*LD belongs to the glycoside hydrolase family 13 subfamily 13 (GH13_13) and can cleave α-1,6-glucosidic bonds in limit dextrins derived from amylopectin (Stam et al., 2006).

Barley is a major raw material in beer production. The major biochemical process in brewing is to degrade barley starch into fermentable sugars, which are further converted into alcohol by yeast metabolism. *Hv*LD as a specific enzyme to digest amylopectin and dextrins is the key factor affecting the production of a high degree of fermentation (Wang et al., 2015). There are three different forms of LD existing in barley: insoluble bound, soluble inactive, and active free. Only the free form is capable for degrading amylopectin (Sissons et al., 1994). The essential industrial process

of brewing includes three steps: malting, mashing, and fermentation. Mashing is usually performed at 60–70°C and at a pH of below 4.5 (Moshi et al., 2015). The heat resistance of LD in free form is poorer than the bound or latent form (Sissons et al., 1995). Free LD will lose its activity in the mashing process at a temperature higher than 63°C (Sissons et al., 1995). High themostability and activity of LD is desirable for the beer production.

Several crystal structures of *Hv*LD have been reported (Vester-Christensen et al., 2010; Møller et al., 2012a, Møller et al., 2015a,b). The *Hv*LD structure contains four domains (Vester-Christensen et al., 2010) (shown in **Figure 1**): the N-terminal domain, a carbohydrate binding module 48 (CBM48), a catalytic (β/α) 8 domain containing the two catalytic residues (Asp473, nucleophile; Glu510, general acid/base) and the transition-state stabilizer (Asp642), and a C-terminal domain. The N-terminal domain includes residues 2–124 resembling carbohydrate binding module 21. CBM48 includes residues 125–230. The catalytic domain contains residues 231–774 and the C-terminal domain contains residues 775–885.

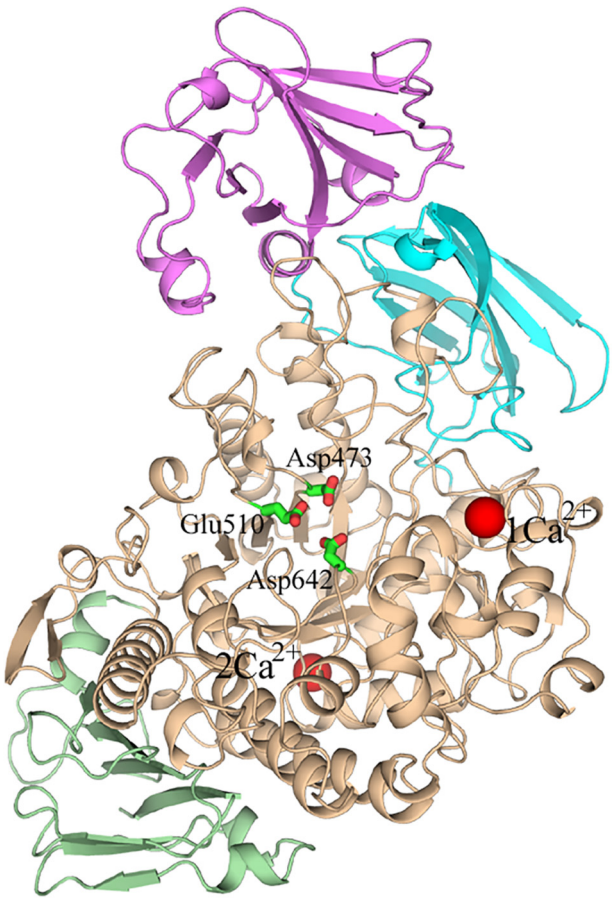


FIGURE 1 | The overall structure of Barley limit dextrinase. N-domain, Magenta; CBM48, cyan; catalytic domain, wheat; C-domain, light green; Ca^{2+} ions, red spheres. The catalytic residues (Asp473, Glu510, and Asp642) are represented in sticks.

There are several works focused on improvement of the themostability of pullulanase derived from bacteria (Chen et al., 2015; Li et al., 2015; Chang et al., 2016; Wang et al., 2016). However, it is still lacking the understanding on the factor affecting the themostability of *Hv*LD at the atomic level. It is suggested that enzymes keep their structural stability by various kinds of non-covalent interactions, such as hydrogen bonds, salt bridges, disulfide bonds, and hydrophobic interaction (Nick Pace et al., 2014; Nilofer et al., 2017). Recently, molecular dynamics (MD) simulation, as a useful tool, has been widely applied to find important characteristics of protein stability (Alizadeh-Rahrovi et al., 2015; Sharma and Sastry, 2015; Jiang et al., 2016; Idrees et al., 2017; Gu et al., 2019).

In this work, MD simulations were carried out for barley limit dextrinase (*Hv*LD) to explore the key factors affecting the thermal stability of LD. The root mean square deviation (RMSD), radius of gyration (R_g), and surface accessibility (SASA) were calculated to explore the dynamics of *Hv*LD. Intra-protein hydrogen bonds, protein–water hydrogen bonds, salt bridges, and hydrophobic interaction were analyzed to find the factors about the thermal stability of *Hv*LD. Finally, root mean square fluctuation (RMSF) analysis was performed to identify the thermal-sensitive regions of LD. The structural and dynamic details will help to understand the driving forces that lead to the stability of *Hv*LD at different temperatures, which will facilitate enzyme engineering of *Hv*LD.

MATERIALS AND METHODS

Systems Preparation

The X-ray structure of barley LD (PDB ID: 4CVW) (Møller et al., 2015a) was obtained from the RCSB Protein Data Bank. The LD inhibitor was removed from this structure. The structure of barley LD (PDB ID: 4CVW) was superimposed on the free form of HvLD (PDB ID: 4AIO) (Møller et al., 2012b). The missing residues (43-PSN-45, 102-FGADGK-107) were also built based on the coordinate of the corresponding residues in the free form of HvLD (PDB ID: 4AIO). A mutant of LD^{D317A} was constructed to evaluate the effect of salt bridge between Asp317 and Arg378.

MD Simulations

Standard AMBER ff03 force field (Wang et al., 2004; Hornak et al., 2006) was assigned to the protein. The force field parameter developed by Bradbrook et al. (1998) was assigned for the Ca^{2+}. The protonation state of ionizable residues was set under pH 5.5 based on the pKa values calculated by the H++ server (Anandakrishnan et al., 2012). Na^{+} ions were added to neutralize the overall system. Each system was embedded in a rectangular box of the TIP3P water molecule (Jorgensen et al., 1983), maintaining a distance of 10 Å from any solute atom to the boundary.

The MD simulations were performed using AMBER12. Energy minimization was carried out with a decreasing harmonic force constraint on the protein. The minimized system was gradually heated from 0 K to the desired temperature within 200 ps under the NVT ensemble condition. The temperature was set as 298 K, 318 K (optimum temperature), and 343 K (the

highest mashing temperature), respectively. To investigate the effect of calcium ions for the structural stability, three systems without Ca^{2+} at 298, 318, and 343 K were also constructed. The temperature was set as 298 and 343 K for LD^{D317A}. Then, the system was relaxed within 1.55 ns under the *NPT* ensemble condition. Finally, a total of 100 ns was simulated to produce trajectories under the *NPT* ensemble condition for each system. A 50-ns MD simulation was conducted for LD^{D317A} at both temperatures. The covalent bonds to hydrogen atoms were constrained using the SHAKE algorithm (Coleman et al., 1977) and the Particle Mesh Ewald (PME) method (Darden et al., 1993) was employed to calculate long-range electrostatic interactions. The real space cutoff was set at 10.0 Å, the same as that for van der Waals interactions. The grid-spacing and convergence criteria of PME calculation was set to 1 Å and 1.0E-05, respectively. The time step used for the simulations was set to 2 fs. The atom coordinates were saved every 10 ps for subsequent analysis.

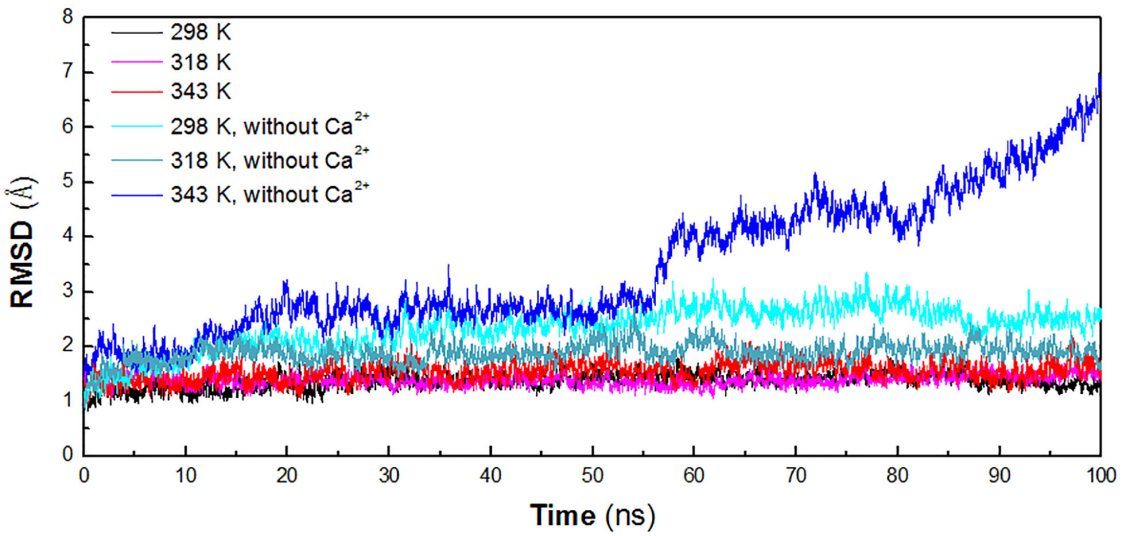


FIGURE 2 | Time evolution of the backbone RMSD versus the starting structure of *Hv*LD. The simulation data obtained at 298, 318, 343, 298 K, without Ca^{2+}, 318 K without Ca^{2+}, and 343 K without Ca^{2+} are shown in black, magenta, red, cyan, deep teal, and blue, respectively.

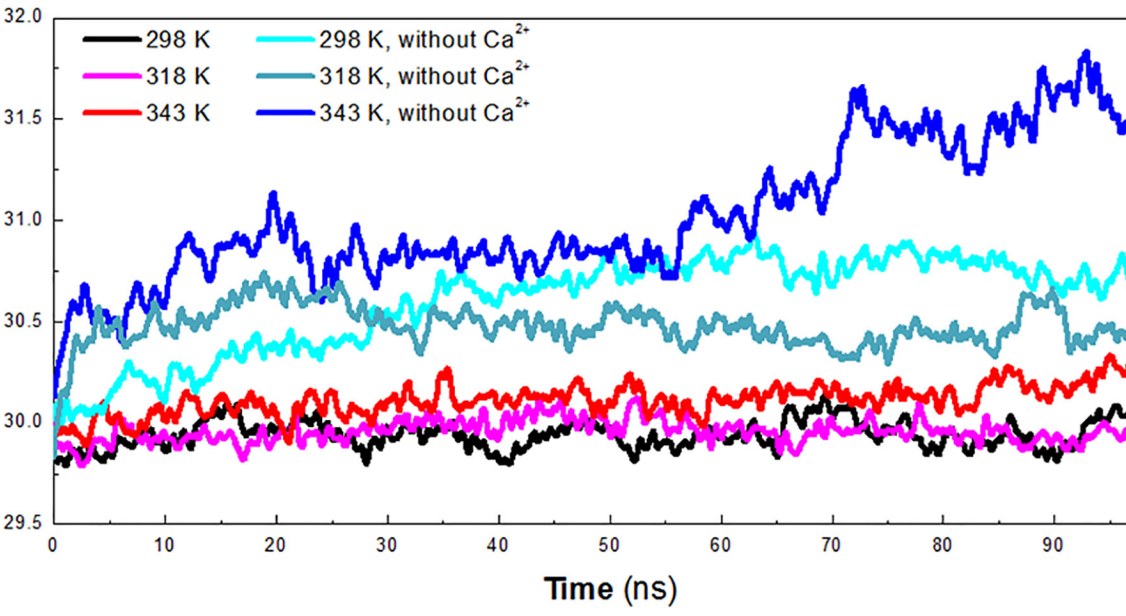


FIGURE 3 | Radius of gyration (R_g) plot. The simulation data obtained at 298 K, 318 K, 343 K, 298 K without Ca^{2+}, 318 K without Ca^{2+}, and 343 K without Ca^{2+} is shown in black, magenta, red, cyan, deep teal, and blue, respectively.

Analysis

All of the analyses were performed using the analytical tools cpptraj (Roe and Cheatham, 2013) module in AMBER tools and VMD (Humphrey et al., 1996). The RMSD was calculated as the deviation between backbone atoms of the protein with respect to the initial structure's backbone atoms, averaged over the backbone atoms. For the RMSF and SASA, the average value was calculated over time. SASA was calculated for all of the residues, hydrophobic residues, hydrophilic residues, and catalytic residues (Asp473, Glu510, and Asp642), respectively. The R_g was calculated by VMD.

The hydrogen bond was calculated based on a maximum cutoff distance between the donor and the acceptor at 3.5 Å and the angle of donor–hydrogen–acceptor larger than 120°. The average value of the number of hydrogen bond was calculated as the ratio of the sum of the total number of HBs in each frame to the total number of frames. The redundant hydrogen bonds between the same donor and acceptor but with different hydrogen atoms were removed, saving the one with the highest occupancy. The HBs were analyzed by considering the chemical properties of different residues, including charged residues (Arg, Lys, Asp, and Glu), polar residues (Gln, Asn, Ser, Thr, Tyr, and Cys), and hydrophobic residues (Ala, Ile, Leu, Phe, Val, Pro, Gly, Met, and Trp).

The salt bridges (SBs) were considered to be formed if the distance between an oxygen atom of an acidic residue (Oδ1 and Oδ2 of ASP and Oδ1 and Oδ2 of Glu) and the nitrogen atom of a basic residue (Nε, Nη1 and Nη2 of Arg and Nζ of Lys) was less than 4 Å. The average value of the number of SBs was calculated as the ratio of the sum of the total number of SBs in each frame to the total number of frames. The SBs between the same two residues but different atoms were regarded as unique, keeping the one with the highest occupancy.

It is suggested that all C atoms within 3.9 Å interacts through hydrophobic contacts (Stojanovic and Zaric, 2009). We calculated the hydrophobic contacts between all the hydrophobic atoms (C, Cα, Cβ, Cδ, Cδ1, Cδ2, Cε, Cε1, Cε2, Cε3, Cγ, Cγ1, Cγ2, Cζ, Cζ2, Cζ3, and Cη2) with a cutoff of 4 Å, without redundancy. The structures were visualized by VMD and PyMOL (Schrödinger, 2010).

A one-way ANOVA was conducted here to evaluate whether the differences are significant for systems containing Ca^{2+}, 318 K/298 K and 343 K/298 K, and systems without Ca^{2+} (318 K/298 K and 343 K/298 K) for the RMSD, R_g, and SASA. The difference is considered significant in the case of $P < 0.05$.

RESULTS AND DISCUSSION

Dynamics of Barley LD

In order to identify the key factors responsible for instability of *Hv*LD at high temperature, MD simulations were performed at different temperature conditions (298, 318, and 343 K) to predict the molecular behavior over the period of time using AMBER. To

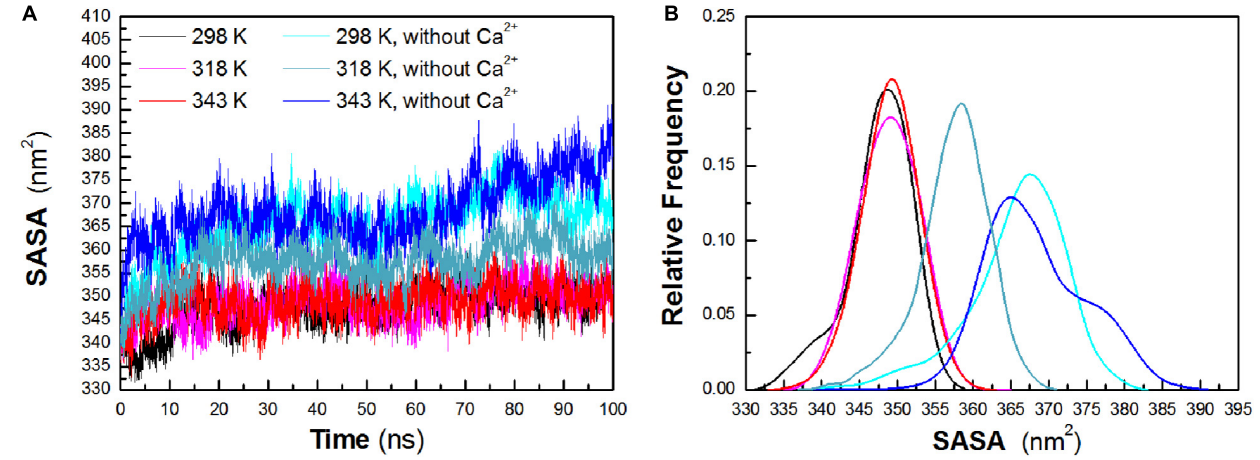


FIGURE 4 | Total solvent-accessible surface area (SASA). **(A)** Time evolution of total SASA. **(B)** The distribution of total SASA. The simulation data obtained at 298, 318, 343, 298 K without Ca^{2+}, 318 K without Ca^{2+}, and 343 K without Ca^{2+} are shown in black, magenta, red, cyan, deep teal, and blue, respectively.

TABLE 1 | Average value of SASA in different systems.

Systems	$SASA_{pho}$	$SASA_{phil}$	$SASA_{total}$	$SASA_{catalytic}$
298 K	110.48 ± 2.87	237.15 ± 2.52	347.63 ± 4.29	60.32 ± 7.83
318 K	111.56 ± 1.98	237.34 ± 2.81	348.90 ± 3.95	61.81 ± 6.90
343 K	110.05 ± 2.18	239.00 ± 2.81	349.05 ± 3.78	65.57 ± 7.34
298 K, no Ca^{2+}	114.29 ± 3.16	251.51 ± 4.05	365.80 ± 6.52	67.34 ± 8.36
318 K, no Ca^{2+}	111.15 ± 2.05	246.28 ± 3.30	357.43 ± 4.75	71.56 ± 9.25
343 K, no Ca^{2+}	116.19 ± 2.67	252.08 ± 4.91	368.26 ± 6.79	71.88 ± 8.14

investigate the effect of calcium ions for the structural stability, MD simulations were also conducted for the three systems without Ca^{2+} (298, 318, and 343 K). The structural stability of six systems was examined by calculation of the RMSD of the backbone atoms relative to the initial structure. **Figure 2** shows the RMSD variations of the six systems with respect to simulation time. It is observed that the systems are equilibrated and thus suitable for exploring the dynamics of *Hv*LD. The first three systems achieved equilibrium at 20, 10, and 25 ns, respectively. The RMSD values for the backbone atoms of *Hv*LD converge at 1.40 ± 0.11 and 1.39 ± 0.10 Å at 298 and 318 K and 1.59 ± 0.16 Å at 343 K. The statistical analysis suggests that the differences are significant ($P < 0.05$) (**Supplementary Table S1**). *Hv*LD in the systems without Ca^{2+} exhibits large variation compared with the initial structure at 298 and 343 K. The backbone RMSD increases rapidly and major structural distortion occurs at 343 K.

The radius of gyration (R_g) reflects the compactness of protein structure. To detect the compactness of the overall structure, the

TABLE 2 | Average numbers of hydrogen bonds in different systems.

Systems	Total HB	MM HB[a]	MS HB[b]	SS HB[c]	chr-chr HB[d]	pho-pho HB[e]	phi-phi HB[f]	pho-phi HB[g]
298 K	756.05	283.61	178.40	171.83	99.54	130.50	74.05	207.83
318 K	776.17	297.84	175.45	169.51	104.83	136.48	72.03	207.73
343 K	720.27	265.37	175.74	166.44	91.80	127.46	73.69	206.42
298 K, no Ca^{2+}	754.33	295.91	178.82	150.73	81.55	140.52	74.21	204.34
318 K, no Ca^{2+}	732.57	283.64	171.22	154.77	82.36	133.26	73.05	205.53
343 K, no Ca^{2+}	745.39	287.05	174.38	155.38	90.66	138.89	66.77	195.73

[a]Main chain–main chain hydrogen bonds. [b]Main chain–side chain hydrogen bonds. [c]Side chain–side chain hydrogen bonds. [d]Hydrogen bonds between charged residues. [e]Hydrogen bonds between hydrophilic residues. [f]Hydrogen bonds between hydrophobic residues. [g]Hydrogen bonds between hydrophilic residues and hydrophobic residues.

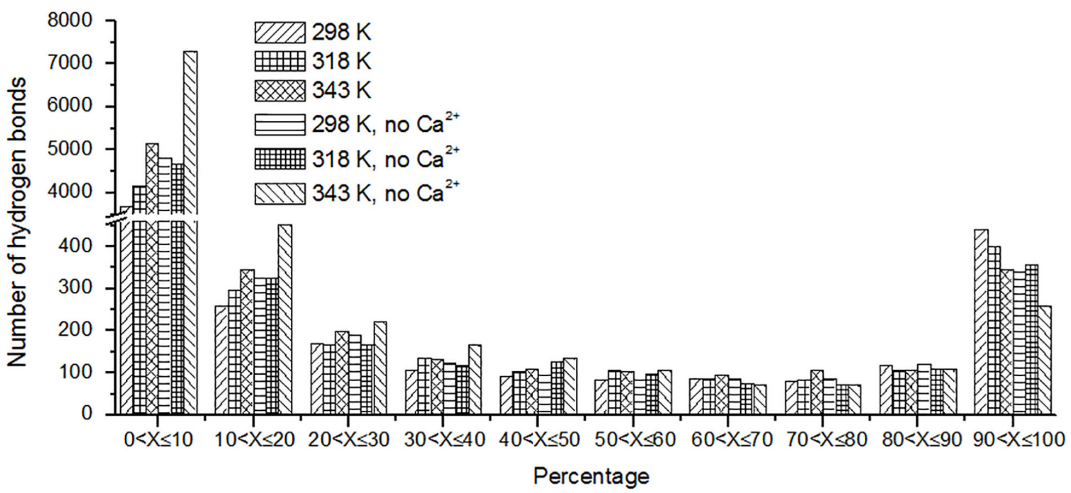


FIGURE 5 | The number of hydrogen bonds (*y*-axis) based on percentage existence time (*x*-axis).

TABLE 3 | Average numbers of protein–water hydrogen bonds in different systems.

Systems	Pro-water Total HB[a]	M-water HB[b]	S-water HB[c]	chr-water HB[d]	pho-water HB[e]	phi-water HB[f]
298 K	928.69	273.55	655.14	358.89	243.81	323.87
318 K	925.18	270.69	654.49	356.18	239.94	326.95
343 K	894.48	257.36	637.11	349.38	227.40	315.66
298 K, no Ca^{2+}	967.29	275.80	691.49	390.55	245.21	329.83
318 K, no Ca^{2+}	954.70	277.43	677.27	378.16	245.75	328.70
343 K, no Ca^{2+}	924.28	262.96	661.31	363.69	237.24	321.18

[a]Protein–water total hydrogen bond. [b]Main chain–water hydrogen bond. [c]Side chain–water hydrogen bond. [d]Charge residues–water hydrogen bond. [e]Hydrophilic residues–water hydrogen bond. [f]Hydrophobic residues–water hydrogen bond.

radius of gyration (R_g) of the protein in six systems was also calculated, as shown in **Figure 3**. According to RMSD plot, all the systems converged after 25 ns. We calculated the average R_g ranging from 25 to 100 ns in each system. The averaged R_g value indicated that the compactness of *Hv*LD increases when the temperature rises. *Hv*LD has the most compact structure at the lowest temperature (298 K, R_g: 29.95 ± 0.09 Å). Moreover, *Hv*LD exhibits similar compactness at 318 K (R_g: 29.97 ± 0.08 Å) and 343 K (R_g: 30.14 ± 0.09 Å). The structure of *Hv*LD is less compact at the higher temperature than it at the lower temperature, indicating expansion of protein structure at higher temperature. The *Hv*LD exhibited higher R_g values at three systems without Ca^{2+}, with the value of 30.69 ± 0.15 Å (298 K), 30.47 ± 0.10 (318 K), and 31.13 ± 0.34 Å (343 K), respectively. This result indicated that the structure of protein in the system without Ca^{2+} was less compact than those systems with Ca^{2+}. In addition, with the temperature rising, the structure of *Hv*LD becomes less compact.

To evaluate the exposure of protein atoms to solvent, SASA was also obtained by calculating the surface area of atom in contact with solvent molecules. From **Figure 4A**, it is found that the total SASA values show a slight increase with a rise of temperature. The averaged SASA is 347.63 ± 4.29 nm^2, 348.90 ± 3.95 nm^2, and 349.05 ± 3.78 nm^2 at 298 K, 318 K, and 343 K, respectively, while in the systems without Ca^{2+}, the total SASA values exhibit significant increase at both temperatures. The averaged SASA is 365.80 ± 6.52 nm^2, 357.43 ± 4.75 nm^2, and 368.26 ± 6.79 nm^2 at 298 K, 318 K, and 343 K, respectively. Total SASA, SASA of hydrophilic residues and catalytic residues, increases relatively from 298 K to 343 K (**Table 1**). The distribution of SASA is displayed in **Figure 4B**. It can be observed that the total SASA is from 331 to 359 nm^2 at 298 K, with the major peak at 349 nm^2 (20.73%). The same major peak appears at 318 K (18.94%) and 343 K (22.00%), respectively. In the systems without Ca^{2+}, the total SASA increases dramatically. The range of SASA is from 337 to 383 nm^2, with the major peak at 367 nm^2 (14.91%) at 298 K. The range of SASA is from 337 to 371 nm^2, with the major peak at 359 nm^2 (18.95%) at 318 K. The SASA at 343 K distributes at a range of 339–391 nm^2, with the major peak at 365 nm^2 (13.31%). This profile is consistent with the trend of R_g values, which indicates that *Hv*LD become less compact with more solvent penetration into the core of the enzyme at high temperatures.

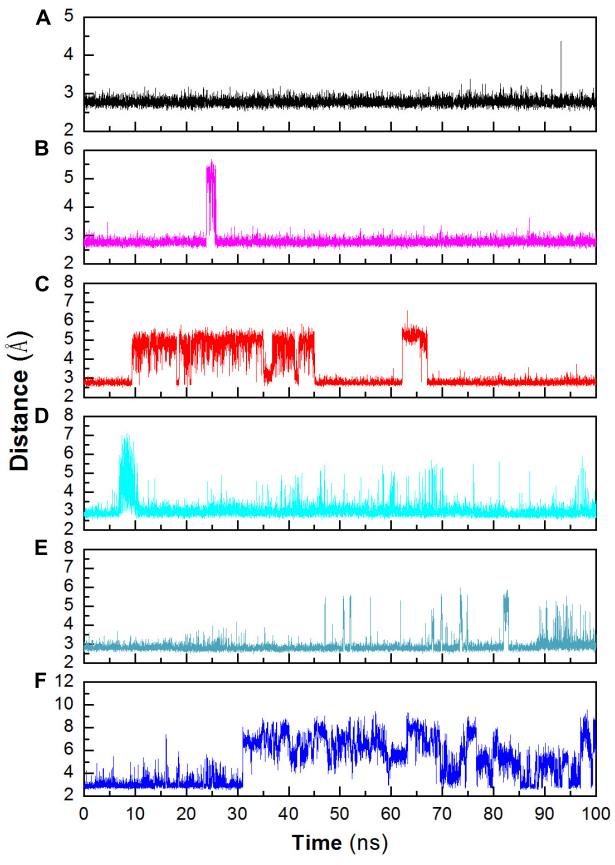


FIGURE 6 | Monitoring the hydrogen bonds between the OD1 atom of Asp403 and the ND1 atom of His404 during MD simulation. **(A)** 298 K; **(B)** 318 K; **(C)** 343 K; **(D)** 298 K without Ca^{2+}, **(E)** 318 K without Ca^{2+}, **(F)** 343 K without Ca^{2+}.

The hydrogen bond interaction is considered important in protein folding, stability, and function. It can be seen that *Hv*LD lost 55 hydrogen bonds at 343 K with respect to 298 K (**Table 2**). The number of hydrogen bonds also decreases in the systems without Ca^{2+}. Based on the percentage existence time of HBs (**Figure 5**), short-lived HBs ($0 < X \leq 10\%$) increase significantly at 343 K, indicating that most of them appear transiently at high temperature. In contrast, the number of long-lived HBs decreases at 343 K, indicating that these interactions are unable to maintain at high temperature. In addition, the number of substantially live HBs ($10 < X \leq 90\%$) at 343 K is more than the corresponding value at 298 K, indicating that these HBs contribute to the stability of *Hv*LD. We analyzed the HBs with occupancy higher than 50% at 298, 318, and 343 K. There are 56 HBs becoming weak and their occupancy decreases with the increase in temperature (**Supplementary Table S2**). These HBs are very sensitive to temperature changes and they affect the thermostability of *Hv*LD.

Besides, we also analyzed the effect of temperature on main chain–main chain HBs, main chain–side chain HBs, and side chain–side chain HBs. There are more MM HBs compared with two other types. There are 18 MM HBs broken at 343 K compared with those at 298 K. MM HBs are important in secondary structure formation. The decrease of the number of MM HBs indicates that the stability of the secondary structure of the enzyme would be impaired. MS HBs and SS HBs exhibit slight decrease at 343 K with respect to 298 K. Moreover, HBs among the residues having similar/different chemical properties (charged–charged residues, hydrophobic–hydrophobic residues, hydrophilic–hydrophilic residues, and hydrophobic–hydrophilic residues) were analyzed. Among these, the number of HBs between hydrophilic–hydrophilic residues and hydrophobic–hydrophilic residues does not exhibit a significant difference at different temperatures, while the amount

of HBs between charged–charged residues and hydrophobic–hydrophobic residues reduces at 343 K compared with that at 298 K. In the systems without Ca^{2+}, the amount of the total HBs, MM HBs, MS HBs, and HBs for hydrophobic–hydrophobic residues, hydrophilic–hydrophilic residues, and hydrophobic–hydrophilic residues also decreases at 343 K compared with 298 K. The amount of SS HBs and HBs for charged–charged residues increases at 343 K compared with 298 K, indicating that there are new hydrogen bonds formed in the distorted structure.

The changes of HBs between protein and water were also explored in the temperature range of 298–343 K, which is displayed in **Table 3**. There is a decrease in protein–water HBs from 298 to 343 K, due to the loss of both M-water HBs and S-water HBs from 298 to 343 K. Besides, the number of HBs for hydrophobic residues to water, hydrophilic residues to water, and charged residues to water reduces at higher temperatures, indicating that the network of HBs between *Hv*LD and water molecules is broken. In the systems without Ca^{2+}, the amount of total protein–water HBs and other types of protein–water HBs decreases at 343 K compared with 298 K, indicating the large change of the network of HBs between *Hv*LD and water molecules in these two systems.

The Stability of the Hydrogen Bond Between D403 and H404

Structural stability may also affect the catalytic activity of this enzyme. The hydrogen bond between D403 and H404 is favorable for the stability of the catalytic triad, which is suggested in a previous work (Møller et al., 2015b). The distance between these two residues was monitored during the whole trajectory. It can be seen that the interaction is more stable at 298 and 318 K (**Figure 6**). When the temperature rises, the interaction becomes unstable (at 343 K). In addition, most of this interaction disappeared in the system without Ca^{2+} at 343 K. The unstable interaction would be unfavorable for the stability for the catalytic triad of Asp473–Glu510–Asp642.

Salt Bridge Interaction Analysis

Besides hydrogen bonds, salt bridges are also very important for the stability of protein (Horovitz et al., 1990; Strop and Mayo, 2000; Jelesarov and Karshikoff, 2009). Salt bridges in enzyme may contribute to its stability at high temperature (Vieille and Zeikus, 2001; Kundu and Roy, 2010). Recently, Guo et al. (2018) summarized factors may contribute to the thermostability for pullulan-hydrolyzing enzymes. They found that there are more

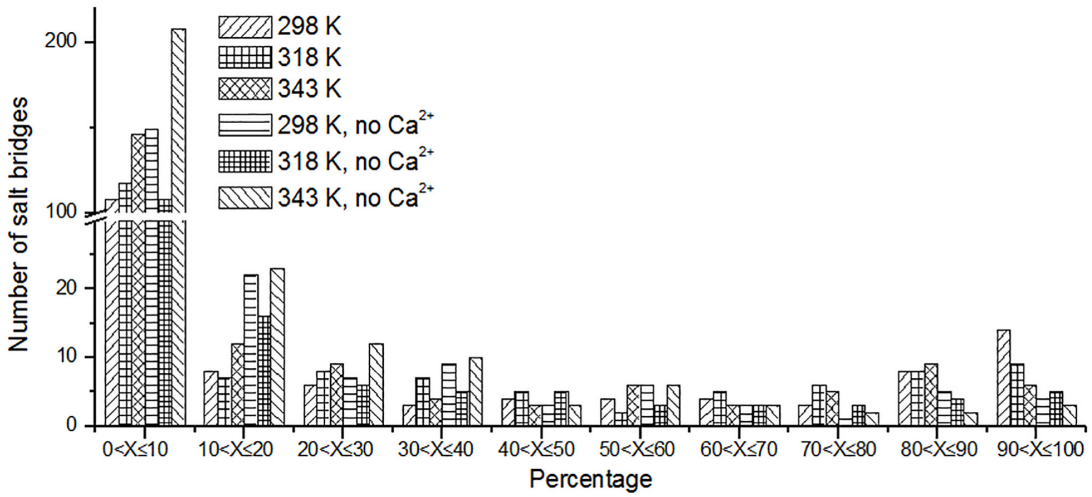


FIGURE 7 | The number of unique salt bridges based on the percentage existence time at cutoff of 4 Å.

TABLE 4 | Average length and occupancy of important salt bridge interactions of *Hv*LD at 298, 318, and 343 K.

Salt bridge	298 K		318 K		343 K	
	Occupancy (%)	Distance (Å)	Occupancy (%)	Distance (Å)	Occupancy (%)	Distance (Å)
E568–R875	99.75	2.85 ± 0.17	81.06	3.28 ± 0.93	64.94	3.79 ± 1.12
D317–R378	99.19	2.84 ± 0.16	52.92	3.70 ± 0.62	59.23	3.55 ± 0.83
D803–R884	98.19	2.80 ± 0.12	92.80	2.82 ± 0.14	86.48	2.86 ± 0.20
D457–R214	97.66	2.88 ± 0.26	98.04	2.89 ± 0.26	73.57	3.35 ± 0.72
D468–R395	94.19	2.93 ± 0.25	91.04	2.96 ± 0.34	88.36	2.97 ± 0.32
D456–R452	94.10	3.06 ± 0.31	77.22	3.41 ± 0.92	57.88	3.75 ± 1.01
D399–R471	89.25	2.88 ± 0.24	88.84	2.85 ± 0.30	75.38	2.98 ± 0.34
D541–R542	84.31	3.11 ± 0.53	86.82	3.13 ± 0.44	67.38	3.63 ± 1.15

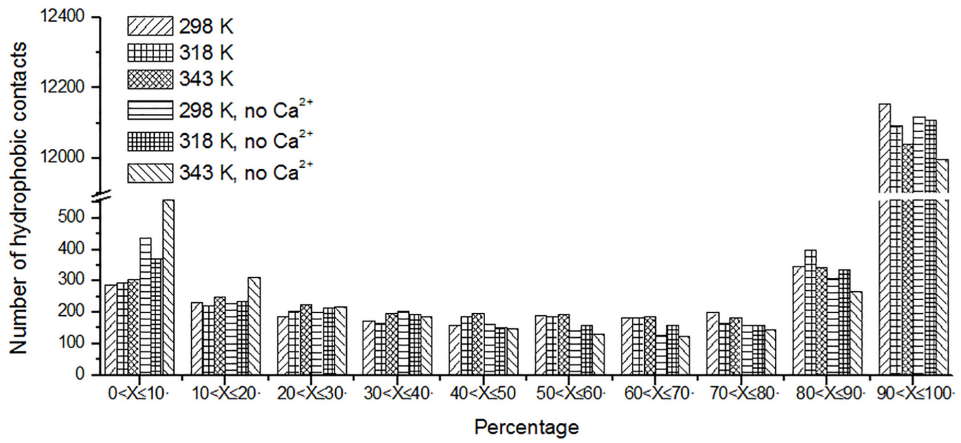


FIGURE 8 | The number of unique hydrophobic contacts based on the percentage existence time at cutoff of 4 Å.

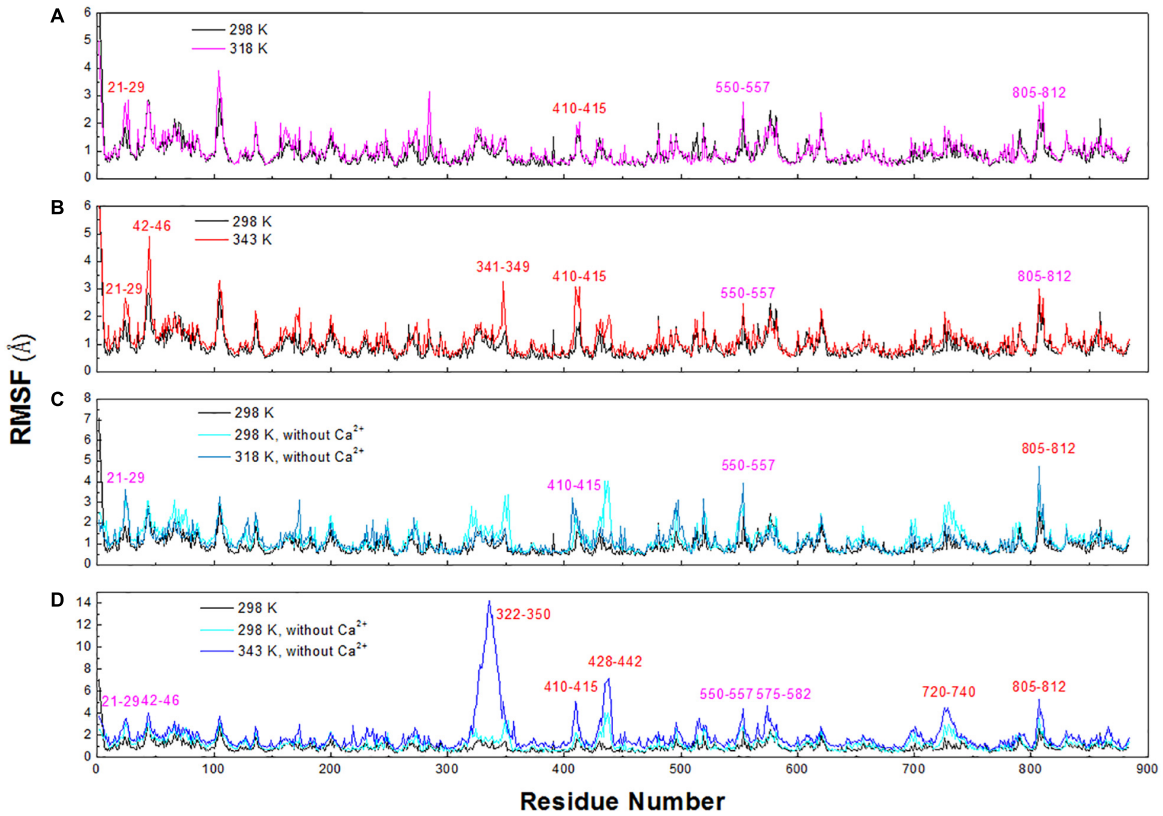


FIGURE 9 | RMSF plots of the backbone atoms of six systems. **(A)** 298 and 318 K. **(B)** 298 and 343 K. **(C)** Simulation at 298 and 318 K without Ca^{2+}. **(D)** Simulation at 298 and 343 K without Ca^{2+}.

salt bridges in thermophilic pullulan-hydrolyzing enzymes than mesophilic ones, suggesting the importance of salt bridge for the enzyme pullulan-hydrolyzing thermostability. In this work, the salt bridges were identified using a 4-Å distance cutoff. We can observe that the number of short-lived salt bridges increases from 298 to 343 K (**Figure 7**), while the number of long-lived salt bridges decreases from 298 to 343 K. It is indicated that the short-lived salt bridges form transiently at high temperature. Some long-lived salt bridges are broken at high temperature. The number of long-lived salt bridges also decreases in the systems without Ca^{2+}, which suggests that ions Ca^{2+} contribute to the stability of salt bridges.

We also analyzed the salt bridges with occupancy higher than 50%. The length of each salt bridge was averaged over

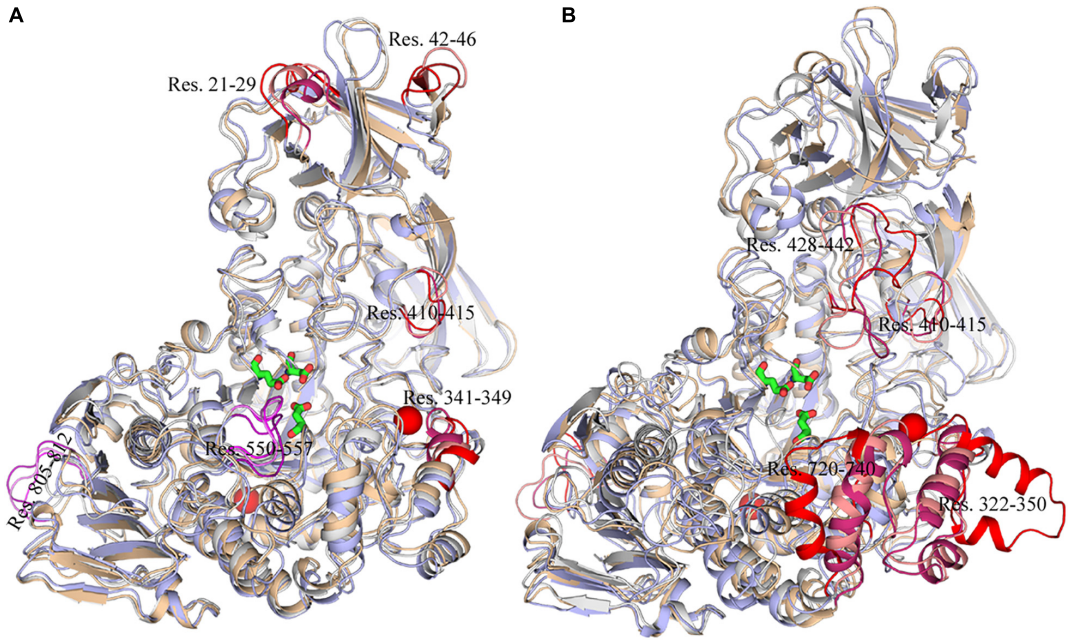


FIGURE 10 | Snapshots of *Hv*LD structures from MD trajectories. **(A)** Comparison of the structure extracted from the system at 298, 318, and 343 K, which are colored by white, wheat, and light blue, respectively. **(B)** Comparison of the structure extracted from the system at 298, 298, and 343 K without Ca^{2+}, which are colored by white, wheat, and light blue, respectively. The high flexible region was colored by red.

the whole trajectory. There are 33, 30, and 29 salt bridges with occupancy higher than 50% during the simulations at 298, 318, and 343 K and 19 and 16 salt bridges in two systems without Ca^{2+} at 298 and 343 K. The salt bridges are noted to decrease in number with an increase in temperature. There are 21 common salt bridges identified at 298, 318, and 343 K. Among these salt bridges, five could still be maintained well even at high temperatures, suggesting an essential role in stabilizing this enzyme. There are eight salt bridges exhibiting comparable occupancy at three temperatures. These salt bridges were also not affected by increases in temperature. These 13 salt bridges are correlated with the partial structural stability of *Hv*LD, while 8 salt bridges become weak and their occupancy decreases with an increase in temperature (**Table 4**), including E568–R875, D317–R378, D803–R884, D457–R214, D468–R395, D456–R452, D399–R471, and D541–R542. These salt bridges are very sensitive to temperature changes and they affect the thermostability of *Hv*LD. All of the salt bridges are located on the surface of *Hv*LD.

To evaluate the function of salt bridge, the salt bridge between Asp317 and Arg378 was selected. The MD simulations on LD^{D317A} at 298 and 343 K were conducted to evaluate the effect of salt bridge D317–R378 on the thermal stability of *Hv*LD. It is indicated that the break of this salt bridge destabilizes the protein. The LD^{D317A} exhibited higher RMSD, R_g, and SASA than LD^{WT} at 298 K and 343 K (**Supplementary Figure S1**).

Hydrophobic Contacts

Protein stabilization is maintained by various interactions. Hydrophobic interaction is one of the important part of them (Van Dan Burg et al., 1994). Unique hydrophobic contacts were calculated as described in the "Materials and Methods" section. The percentage of time is displayed in **Figure 8**. It can be seen that the short-lived unique hydrophobic contacts ($0 < X \leq 10\%$) increase from 298 to 343 K. Long-lived hydrophobic contacts ($90 < X \leq 100\%$) decrease from 298 to 343 K. It exhibits the same trend with hydrogen bonds and salt bridges. A similar situation also occurs in systems without Ca^{2+}.

Identification of Thermal-Sensitive Regions

To explore the structural and dynamic changes, we identified the thermal-sensitive regions of *Hv*LD by analyzing the structural mobility based on the RMSF of the backbone atoms with respect to the initial structure. **Figure 9** shows that the RMSF values of most regions of *Hv*LD fluctuate slightly at a temperature of up to 343 K, suggesting that these regions are relatively thermostable. Some regions showed steep RMSF fluctuations at high temperature, such as 318 and 343 K, indicating that those are thermal-sensitive regions. It can be observed in **Figure 9** that the highest fluctuations occur at the N-terminal because it was not restrained. Also, some loops that always exhibit high fluctuation at different temperatures may be due to their intrinsic flexibility, including residues 102–107, 135–138, and 575–582. Besides, regions that fluctuated higher than 0.5 and 1.0 Å are highlighted by magenta and red in **Figure 9**. Larger fluctuations are observed at 318 K or 343 K, such as residues 21–29, 42–46, 341–349, and 410–415. These regions are predicted to be thermal-sensitive regions. In a previous work, it was suggested that residues 23–27, 42–48, 102–109, and 806–810 exhibited high

flexibility, with low level electron density (Vester-Christensen et al., 2010), which is consistent with our results. To compare with those systems without Ca^{2+}, it can be observed that the flexibility of LD is all higher in these systems than that at 298 K. In particular, these loops exhibit high fluctuations, including residues 21–29, 42–46, 322–350, 410–415, 428–442, 550–557, 575–582, 720–740, and 805–812.

To exhibit these regions in *Hv*LD, the higher fluctuation regions were mapped onto the tertiary structure. As shown in **Figure 10**, it is observed that some of them are located near the catalytic crevice, including residues 550–557 (**Figure 10A**). The high flexibility of these regions would decrease the stability of the catalytic triad. Residues 21–29, 42–46, 341–349, and 410–415 are located in surface loops (**Figure 10A**). In the systems without Ca^{2+}, there are large conformational change of residues 322–350, 428–442, 720–740, and 805–812 at high temperature (343 K) (**Figure 10B**). The loss of Ca^{2+} made the significant conformational change of a long loop in catalytic (β/α) eight domain, previously described as loop 2 (Jespersen et al., 1991; Mikami et al., 2006; Turkenburg et al., 2009) (the site for Ca1) (**Supplementary Figure S2**), indicating that this Ca^{2+} are important for maintaining the stable conformation of *Hv*LD at high temperature. The conformational change of loop 2 affected the conformation of residue 322–350 and 720–740. The removal of another Ca^{2+}, which is located between α1 and α2, had a weaker effect on the conformation of *Hv*LD at different temperatures. This Ca^{2+} contributes less to the stability of *Hv*LD.

CONCLUSION

In the present study, thermostability factors of barley LD were investigated by MD simulations. The higher value of RMSD, R_g, and SASA suggests the instability of *Hv*LD at high temperatures. Intra-protein hydrogen bonds and hydrogen bonds between protein and water decrease at high temperature. Long-lived hydrogen bonds, salt bridges, and hydrophobic contacts are lost at high temperature. The salt bridge interaction analysis suggests that these salt bridges are important for the thermostability of *Hv*LD, including E568–R875, D317–R378, D803–R884, D457–R214, D468–R395, D456–R452, D399–R471, and D541–R542, which are located on the surface of *Hv*LD. Based on RMSF calculations for *Hv*LD at various temperatures, several thermally sensitive regions of *Hv*LD were identified, such as residues 21–29, 42–46, 341–349, and 410–415. The structural and dynamic details will help to understand the driving forces that lead to the stability of *Hv*LD at different temperatures, which will facilitate enzyme engineering of *Hv*LD for enhanced thermostability.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

JDu conceived and supervised the experiments. JDo, SD, and KZ performed the MD simulations. JDo, JDu, JY, and SH analyzed the data. JDu and HY wrote the manuscript.

FUNDING

This work was supported by the Open Research Fund of State Key Laboratory of Biological Fermentation Engineering of Beer (grant no. K201804), the Natural Science Foundation of Shandong Province, China (grant no. ZR2018QB004) and the Off-site collaborative calculation of marine medicinal biological resources (grant no. 2018ASKJ01-05).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00051/full#supplementary-material

FIGURE S1 | RMSD, R_g, and SASA for $HvLD^{D317A}$. **(A)** RMSD plot. **(B)** R_g plot. **(C)** SASA plot.

FIGURE S2 | The conformation of Ca1 site. Initial structure and the extracted structure form system without Ca^{2+} at 343 K are colored by cyan and purple, respectively.

TABLE S1 | Statistical analysis results of RMSD, R_g and SASA for HvLD at different simulation temperatures.

TABLE S2 | Length and occupancy of important hydrogen bond interactions of HvLD at 298, 318, and 343 K.

REFERENCES

Alizadeh-Rahrovi, J., Shayesteh, A., and Ebrahim-Habibi, A. (2015). Structural stability of myoglobin and glycomyoglobin: a comparative molecular dynamics simulation study. *J. Biol. Phys.* 41, 349–366. doi: 10.1007/s10867-015-9383-2

Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40, W537–W541. doi: 10.1093/nar/gks375

Bradbrook, G. M., Gleichmann, T., Harrop, S. J., Habash, J., Raftery, J., Kalb, J., et al. (1998). X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding. *J. Chem. Soc. Faraday Trans.* 94, 1603–1611. doi: 10.1039/A800429C

Chang, M., Chu, X., Lv, J., Li, Q., Tian, J., and Wu, N. (2016). Improving the thermostability of acidic pullulanase from bacillus naganoensis by rational design. *PLoS One* 11:e0165006. doi: 10.1371/journal.pone.0165006

Chen, A., Li, Y., Nie, J., McNeil, B., Jeffrey, L., Yang, Y., et al. (2015). Protein engineering of Bacillus acidopullulyticus pullulanase for enhanced thermostability using in silico data driven rational design methods. *Enzyme Microb. Technol.* 78, 74–83. doi: 10.1016/j.enzmictec.2015.06.013

Coleman, T. G., Mesick, H. C., and Darby, R. L. (1977). Numerical integration: a method for improving solution stability in models of the circulation. *Ann. Biomed. Eng.* 5, 322–328. doi: 10.1007/BF02367312

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem.Phys.* 98, 10089–10092. doi: 10.1063/1.464397

Gu, J., Tong, H., Sun, L., and Lin, Z. (2019). Molecular dynamics perspective on the thermal stability of mandelate racemase. *J. Biomol. Struct. Dyn.* 37, 383–393. doi: 10.1080/07391102.2018.1427631

Guo, J., Coker, A. R., Wood, S. P., Cooper, J. B., Keegan, R. M., Ahmad, N., et al. (2018). Structure and function of the type III pullulan hydrolase from Thermococcus kodakarensis. *Acta Crystallogr. Sec. D* 74, 305–314. doi: 10.1107/S2059798318001754

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi: 10.1002/prot.21123

Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990). Strength and co-operativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* 216, 1031–1044. doi: 10.1016/S0022-2836(99)80018-7

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Idrees, D., Rahman, S., Shahbaaz, M., Haque, M. A., Islam, A., Ahmad, F., et al. (2017). Estimation of thermodynamic stability of human carbonic anhydrase IX from urea-induced denaturation and MD simulation studies. *Int. J. Biol. Macromol.* 105, 183–189. doi: 10.1016/j.ijbiomac.2017.07.010

Jelesarov, I., and Karshikoff, A. (2009). "Defining the role of salt bridges in protein stability," in *Protein Structure, Stability, and Interactions*, ed. J. W. Shriver (Totowa, NJ: Humana Press), doi: 10.1007/978-1-59745-367-7_10

Jespersen, H. M., MacGregor, E. A., Sierks, M. R., and Svensson, B. (1991). Comparison of the domain-level organization of starch hydrolases and related enzymes. *Biochem. J.* 280, 51–55. doi: 10.1042/bj2800051

Jiang, X., Chen, G., and Wang, L. (2016). Structural and dynamic evolution of the amphipathic N-terminus diversifies enzyme thermostability in the glycoside hydrolase family 12. *Phys. Chem. Chem. Phys.* 18, 21340–21350. doi: 10.1039/C6CP02998A

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Kundu, S., and Roy, D. (2010). Structural study of carboxylesterase from hyperthermophilic bacteria Geobacillus stearothermophilus by molecular dynamics simulation. *J. Mol. Graph. Model.* 28, 820–827. doi: 10.1016/j.jmgm.2010.03.001

Li, S., Xu, J., Bao, Y., Zheng, H., and Song, H. (2015). Structure and sequence analysis-based engineering of pullulanase from Anoxybacillus sp. LM18-11 for improved thermostability. *J. Biotechnol.* 210, 8–14. doi: 10.1016/j.jbiotec.2015.06.406

Manners, D. J., Marshall, J. J., and Yellowlees, D. (1970). The specificity of cereal limit dextrinases. *Biochem. J.* 116, 539–541. doi: 10.1042/bj1160539

Mikami, B., Iwamoto, H., Malle, D., Yoon, H.-J., Demirkan-Sarikaya, E., Mezaki, Y., et al. (2006). Crystal structure of pullulanase: evidence for parallel binding of oligosaccharides in the active site. *J. Mol. Biol.* 359, 690–707. doi: 10.1016/j.jmb.2006.03.058

Møller, M., Hachem, M. Abou, Svensson, B., and Henriksen, A. (2012a). Structure of the starch-debranching enzyme barley limit dextrinase reveals homology of the N-terminal domain to CBM21. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 68(Pt 9), 1008–1012. doi: 10.1107/S1744309112031004

Møller, M. S., Abou Hachem, M., Svensson, B., and Henriksen, A. (2012b). Structure of the starch-debranching enzyme barley limit dextrinase reveals homology of the N-terminal domain to CBM21. *Acta Crystallogr.* 68(Pt 9), 1008–1012.

Møller, M. S., Vester-Christensen, M. B., Jensen, J. M., Hachem, M. A., Henriksen, A., and Svensson, B. (2015a). Crystal Structure of Barley Limit Dextrinase-Limit Dextrinase Inhibitor (LD-LDI) Complex Reveals Insights into Mechanism and Diversity of Cereal Type Inhibitors. *J. Biol. Chem.* 290, 12614–12629. doi: 10.1074/jbc.M115.642777

Møller, M. S., Windahl, M. S., Sim, L., Bøjstrup, M., Abou Hachem, M., Hindsgaul, O., et al. (2015b). Oligosaccharide and Substrate Binding in the Starch Debranching Enzyme Barley Limit Dextrinase. *J. Mol. Biol.* 427(6 Pt B), 1263–1277. doi: 10.1016/j.jmb.2014.12.019

Moshi, A. P., Hosea, K. M. M., Elisante, E., Mamo, G., and Mattiasson, B. (2015). High temperature simultaneous saccharification and fermentation of starch from inedible wild cassava (Manihot glaziovii) to bioethanol using *Caloramator boliviensis*. *Bioresour. Technol.* 180, 128–136. doi: 10.1016/j.biortech.2014.12.087

Nick Pace, C., Scholtz, J. M., and Grimsley, G. R. (2014). Forces stabilizing proteins. *FEBS Lett.* 588, 2177–2184. doi: 10.1016/j.febslet.2014.05.006

Nilofer, C., Sukhwal, A., Mohanapriya, A., and Kangueane, P. (2017). Protein-protein interfaces are vdW dominant with selective H-bonds and (or) electrostatics towards broad functional specificity. *Bioinformation* 13, 164–173. doi: 10.6026/97320630013164

Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theor. Comput.* 9, 3084–3095. doi: 10.1021/ct400341p

Schrödinger, L. L. C. (2010). *The PyMOL Molecular Graphics System, Version 1.3.1. The PyMOL Molecular Graphics System, Version 1.3.1.*

Sharma, R., and Sastry, G. N. (2015). Deciphering the dynamics of non-covalent interactions affecting thermal stability of a protein: molecular dynamics study on point mutant of thermus thermophilus isopropylmalate dehydrogenase. *PLoS One* 10:e0144294. doi: 10.1371/journal.pone.0144294

Sissons, M., Taylor, M., and Proudlove, M. (1995). Barley malt limit dextrinase: Its extraction, heat stability, and activity during malting and mashing. *Am. Soc. Brew. Chem.* 21, S356.

Sissons, M. J., Lance, R. C. M., and Wallace, W. (1994). Bound and free forms of barley limit dextrinase. *Cereal Chem.* 71, 520–521.

Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M., and Henrissat, B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Eng. Design Select.* 19, 555–562. doi: 10.1093/protein/gzl044

Stojanovic, S. D., and Zaric, S. D. (2009). Hydrogen bonds and hydrophobic interactions of porphyrins in porphyrin-containing proteins. *Open Struct. Biol. J.* 3, 34–41. doi: 10.2174/1874199100903010034

Strop, P., and Mayo, S. L. (2000). Contribution of surface salt bridges to protein stability. *Biochemistry* 39, 1251–1255. doi: 10.1021/bi992257j

Turkenburg, J. P., Brzozowski, A. M., Svendsen, A., Borchert, T. V., Davies, G. J., and Wilson, K. S. (2009). Structure of a pullulanase from Bacillus acidopullulyticus. *Proteins* 76, 516–519. doi: 10.1002/prot.22416

Van Dan Burg, B., Dijkstra, B. W., Vriend, G., Van Dar Vinne, B., Venema, G., and Eijsink, V. G. H. (1994). Protein stabilization by hydrophobic interactions at the surface. *Eur. J. Biochem.* 220, 981–985. doi: 10.1111/j.1432-1033.1994.tb18702.x

Vester-Christensen, M. B., Hachem, M. A., Svensson, B., and Henriksen, A. (2010). Crystal structure of an essential enzyme in seed starch degradation: barley limit dextrinase in complex with cyclodextrins. *J. Mol. Biol.* 403, 739–750. doi: 10.1016/j.jmb.2010.09.031

Vieille, C., and Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* 65, 1–43. doi: 10.1128/mmbr.65.1.1-43.2001

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035

Wang, X., Nie, Y., Mu, X., Xu, Y., and Xiao, R. (2016). Disorder prediction-based construct optimization improves activity and catalytic efficiency of Bacillus naganoensis pullulanase. *Sci. Rep.* 6:24574. doi: 10.1038/srep24574

Wang, X., Zhang, X., Cai, S., Ye, L., Zhou, M., Chen, Z., et al. (2015). Genetic diversity and QTL mapping of thermostability of limit dextrinase in barley. *J. Agric. Food Chem.* 63, 3778–3783. doi: 10.1021/acs.jafc.5b00190

Yang, X., Westcott, S., Gong, X., Evans, E., Zhang, X.-Q., Lance, R. C. M., et al. (2008). Amino acid substitutions of the limit dextrinase gene in barley are associated with enzyme thermostability. *Mol. Breed.* 23, 61. doi: 10.1007/s11032-008-9214-2

Conflict of Interest: JDu, JDo, JY, SH, and HY were employed by company Tsingtao Brewery.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 14 May 2020
doi: 10.3389/fmolb.2020.00088

Identification of Alternative Allosteric Sites in Glycolytic Enzymes for Potential Use as Species-Specific Drug Targets

*Merve Ayyildiz[1], Serkan Celiker[1], Fatih Ozhelvaci[2] and E. Demet Akten[3]**

[1] Graduate Program of Computational Biology and Bioinformatics, Graduate School of Science and Engineering, Kadir Has University, Istanbul, Turkey, [2] Graduate Program of Computational Science and Engineering, Graduate School of Science and Engineering, Bogazici University, Istanbul, Turkey, [3] Department of Bioinformatics and Genetics, Faculty of Engineering and Natural Sciences, Kadir Has University, Istanbul, Turkey

Three allosteric glycolytic enzymes, phosphofructokinase, glyceraldehyde-3 phosphate dehydrogenase and pyruvate kinase, associated with bacterial, parasitic and human species, were explored to identify potential allosteric sites that would be used as prime targets for species-specific drug design purposes using a newly developed approach which incorporates solvent mapping, elastic network modeling, sequence and structural alignments. The majority of binding sites detected by solvent mapping overlapped with the interface regions connecting the subunits, thus appeared as promising target sites for allosteric regulation. Each binding site was then evaluated by its ability to alter the global dynamics of the receptor defined by the percentage change in the frequencies of the lowest-frequency modes most significantly and as anticipated, the most effective ones were detected in the vicinity of the well-reported catalytic and allosteric sites. Furthermore, some of our proposed regions intersected with experimentally resolved sites which are known to be critical for activity regulation, which further validated our approach. Despite the high degree of structural conservation encountered between bacterial/parasitic and human glycolytic enzymes, the majority of the newly presented allosteric sites exhibited a low degree of sequence conservation which further increased their likelihood to be used as species-specific target regions for drug design studies.

Keywords: allosteric regulation, glycolytic enzyme, elastic network modeling, species-specific, drug discovery

OPEN ACCESS

Edited by:
Gennady Verkhivker, Chapman University, United States

Reviewed by:
*Igor N. Berezovsky, Bioinformatics Institute (A*STAR), Singapore*
Natalia Kulik, Academy of Sciences of the Czech Republic (ASCR), Czechia

***Correspondence:**
E. Demet Akten
demet.akten@khas.edu.tr

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *12 February 2020*
Accepted: *16 April 2020*
Published: *14 May 2020*

Citation:
Ayyildiz M, Celiker S, Ozhelvaci F and Akten ED (2020) Identification of Alternative Allosteric Sites in Glycolytic Enzymes for Potential Use as Species-Specific Drug Targets. Front. Mol. Biosci. 7:88. doi: 10.3389/fmolb.2020.00088

INTRODUCTION

Glycolysis is the most essential metabolic sequence of enzymatic reactions in all living cells that converts glucose into pyruvate to produce energy in the form of adenosine triphosphate (ATP) and reduced nicotinamide adenine dinucleotide (NADH). The process has a dual effect in the sense that while it metabolizes six-carbon sugars into smaller three-carbon compounds which are later used for a large amount of ATP production or fat synthesis, it also generates a small amount of ATP (Meyerhof and Junowicz-Kocholaty, 1943; Barnett, 2003). Thus, it is nearly ubiquitous in all living cells and essential for the survival of biological organisms. Glycolytic pathway is a sequence of ten consecutive reactions catalyzed by ten different enzymes, three of which are known to be

allosteric; phosphofructokinase, glyceraldehyde-3 phosphate dehydrogenase and pyruvate kinase which appear on the third, the sixth and the last reaction, respectively.

As glycolysis is essential for living cells, allostery is equally crucial for regulating protein's activity (Monod et al., 1963; Perutz, 1989; Koshland and Hamadani, 2002). Allostery is defined as the correspondence of conformational changes between two distant sites in the protein which usually incorporate a catalytic region and another so-called effector site. The functional state of the enzyme becomes under the regulation of a ligand or the so-called effector binding, since the catalytic region consequently becomes either accessible or inaccessible to substrates. After the first allosteric model (MWC model) proposed by Monod et al. (1965) which defined allosteric proteins as symmetric oligomers with identical protomers found in at least two conformational states (T and R) with different ligand-binding affinities (Monod et al., 1965), Weber put forward a powerful concept for allosteric regulation which is the population shift or re-distribution of protein's conformational states (Weber, 1972). Accordingly, all proteins have a repertoire of conformational states from which they select to adopt in a particular functional state, and the ligand binding merely alters the selection of these conformations (Elber and Karplus, 1987; Pan et al., 2000). Hence, if that repertoire or the dynamic ensemble of conformations underlies the allosteric behavior, apparently one can suggest that all proteins are potentially allosteric (Gunasekaran et al., 2004). In fact, two decades old experiments demonstrated that allostery can be introduced into proteins of which their functional state do not rely on allostery, either by site-directed mutagenesis or a strong binding molecule (Falcon and Matthews, 2001; Wang and Kemp, 2001; Santamaría et al., 2002).

Allostery is merely a redistribution of conformational states as a consequence of a structural perturbation which is merely the binding of a ligand at a distal site. The same type manifestation is also recognized as a result of mutation, changes in pH, temperature, ionic strength and covalent modification such as phosphorylation and acetylation as the population shift is an intrinsically embedded dynamic feature of proteins. As previously reported for HIV protease and reverse transcriptase, the apo and ligand-bound forms of an enzyme represent two different conditions under which the receptor display distinct dynamics or communication networks (Temiz and Bahar, 2002; Kurt et al., 2003).

The general acceptance of allostery as an intrinsic feature of all proteins revolutionized the drug design efforts in an unprecedented way (Ellis, 1998; Christopoulos, 2002). One of the major advantages of targeting allosteric sites rather than catalytic or so-called orthosteric regions was the low degree of sequence conservation which enables the design of species-specific drug molecules. The first step of allosteric drug design thus involves the identification of these distinct sites away from the catalytic region which would display a high degree of sequence variability among species. For allosteric proteins, the so-called allosteric regions are usually well-established through experimental studies, yet alternative sites might exist for the same protein which will enrich the likelihood of effective drugs with greater specificity. Furthermore, for non-allosteric proteins, these "secret" allosteric sites can be exposed and used as target in drug design studies with unprecedented success.

Several well-established approaches exist to detect alternative allosteric sites. Some relies on static structures of proteins acquired from NMR or X-ray experimental studies, while others investigate large scale motions such as hinge bending via normal mode analysis (NMA) using coarse-grained elastic network model (Bahar and Rader, 2005; Tama and Brooks, 2006) or molecular dynamics simulations (Hornak et al., 2006; Lou and Cukier, 2006; Dilcan et al., 2019), since large scale motions involving large domains can be correlated with protein function. Moreover, large scale motions defined by the slowest frequency modes present an intrinsic feature of the protein (Tobi and Bahar, 2005) and also defines the distant couplings which is the nature of allostery. Therefore, it is crucial to identify potential sites in the protein that will perturb this communication and eventually the dynamic equilibrium which might lead to a functional disorder. Besides low-frequency modes, local disturbances in the conformation represented by high-frequency modes also play a critical role in transmitting signals between distant sites (Hammes-Schiffer and Benkovic, 2006; Hawkins and McLeish, 2006).

Allosteric communication in a protein is evolutionarily encoded in a protein structure and conducted via a well-defined network comprising a limited amount of conserved residues which is strongly coupled (Lockless and Ranganathan, 1999). This well-defined communication channel is evolutionized, i.e., optimized to fulfill the functional requirements with minimal energy requirement. There exist several theoretical studies which highlight the existence of functional key residues which persistently appear in pathways of allosteric signal propagation (Süel et al., 2003; Ming and Wall, 2005). Perturbations on these residues strongly affect the cooperative network within proteins and thus it is of paramount importance to develop novel approaches to effectively identify these residues. A computational study conducted by Liu and his coworkers used an ensemble-based model and suggested that functional sites may be uniquely coupled to structural fluctuations and can be identified by the way a bound ligand to these sites effect the conformational manifold (Liu et al., 2007). Another noteworthy computational algorithm developed by Flechsig makes use of *in silico* designed synthetic structures which are represented by elastic networks and a strategy of evolutionary optimization to iteratively improve allosteric coupling or signal propagation along simple pathways incorporating a set of interacting residues (Flechsig, 2017). According to the model, allostery is considered as a consequence of optimized communication between distant functional sites. Another pioneering work by Guarnera and Berezovsky emphasizes the importance of the causality and energetics of allosteric communication (Guarnera and Berezovsky, 2019). They used ligand binding and mutations as a source of perturbations and hypothesized that perturbation of functional sites can identify latent allosteric sites based on the fact that allosteric communication is symmetric in nature (Guarnera and Berezovsky, 2016a).

Our procedure in this study uses the well-known normal mode analysis using a coarse-grained elastic network model

which predicts the change in the frequencies of lowest-frequency modes as a result of a ligand binding (Kaynak et al., 2018). The approach is based on the fact that as the lowest-frequency modes consist of global motions that control the protein function, the sites which would display the highest frequency shift would correspond to either active catalytic sites or potential allosteric sites. Combining this structure-based approach with an energy-based algorithm for detecting "hot spots" that are likely to be druggable sites, a powerful prediction tool was obtained. Each one of the catalytic sites was identified as strongly druggable in addition to well-recognized allosteric sites. Besides, our procedure suggested unique alternative allosteric locations observed at the interface of monomeric subunits. Interface regions in oligomeric proteins usually accommodate potential allosteric sites as the global dynamics in complex systems is most often described by the relative rearrangement of these subunits (Kurkcuoglu et al., 2011, 2015). Thus, a structural perturbation at the interface such as ligand binding most often disrupts the dynamic character and eventually the catalytic site. Moreover, proposed allosteric sites were investigated based on sequence and structural similarity between bacterial/parasitic enzyme and its human counterpart. In all these sites, a satisfactory amount of sequence variation was observed despite a high degree of structural similarity. Thus, our future drug design efforts which will target these slightly conserved sites will potentially yield species-specific drug molecules. Furthermore, our results were compared to a well-established algorithm which predict binding sites (DoGSiteScorer) using a Difference of Gaussian filter solely based on 3D structure of the protein and assess their druggability using a support vector machine which is a linear combination of three descriptors describing volume, hydrophobicity and enclosure (Volkamer et al., 2012a). The binding pockets with highest scores successfully agreed with our predictions of druggable binding sites. Despite the lack of experimental support, the observation of all well-known catalytic and allosteric sites as druggable provided a powerful critical assessment of our approach. Finally, the allosteric effect of our top druggable sites in each enzyme was confirmed via a powerful tool AlloSigMA (Guarnera and Berezovsky, 2016b; Guarnera et al., 2017), which demonstrated a decrease in the dynamics of several catalytic regions as a result of a ligand binding.

MATERIALS AND METHODS

System Preparation

Several X-ray crystallographic structures deposited at the Protein Data Bank for three glycolytic enzymes phosphofructokinase (PFK), glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and pyruvate kinase (PK) were extracted for species of Homo sapiens (*H. sapiens*) (Kung et al., 2012; Kloos et al., 2015; White et al., 2015), Staphylococcus aureus (*S. aureus*) (Mukherjee et al., 2010; Axerio-Cilies et al., 2012; Tian et al., 2018) and three parasites, Trypanosoma cruzi for GADPH (*T. cruzi*) (Guido et al., 2009), Trypanosoma brucei (*T. brucei*) for PFK (McNae et al., 2009) and Leishmania mexicana (*L. mexicana*) for PK (Rigden et al., 1999) and the selected ones were listed in **Table 1** along with the details of each structure provided at the footnote section.

TABLE 1 | Tetrameric structures of glycolytic enzymes extracted from PDB databank.

Enzyme	Human (*H. sapiens*)	Bacterium (*S. aureus*)	Parasite
PFK	4RH3[a]	5XZ7[b]	3F5M[c] (*T. brucei*)
GAPDH	4WNI[d]	3HQ4[e]	3DMT[f] (*Cruzi*)
PK	4G1N[g]	3T0T[h]	1PKL[i] (*L. mexicana*)

[a] *4RH3; the structure with the least amount of missing residues with a resolution of 3.02 Å.*
[b] *5XZ7; the only available X-ray structure reported so far with a resolution of 1.6 Å.*
[c] *3F5M; the ATP-bound form of PFK enzyme with a resolution of 2.7 Å*
[d] *4WNI; the T229K mutant form at 2.3 Å resolution.*
[e] *3HQ4; the C151S mutant of GADPH1 complexed with NAD from S. aureus MRSA252.*
[f] *3DMT; the glycosomal GADPH at 2.3 Å resolution in complex with irreversible iodoacetate inhibitor.*
[g] *4G1N; M2 isoform in complex with an activator and at 2.3 Å resolution.*
[h] *3T0T; MRSA PK structure at 3.1 Å resolution in complex with an inhibitor.*
[i] *1PKL: the structure of Leishmania Pyruvate Kinase at 2.35 Å resolution.*

Sequence and Structural Alignment

To identify similarities and differences between human and bacterial/parasitic species at the level of primary structure, pairwise amino acid sequence alignment was performed using Needleman-Wunsch global alignment algorithm (Needleman and Wunsch, 1970) via EMBOSS-Needle (Rice et al., 2000) web server using the following default parameters; Blosum62 as similarity matrix (Henikoff and Henikoff, 1992), gap penalty as 10 for opening and 0.5 for extension, and no end gap penalty. As for displaying the structural differences, the *super* module of PyMOL graphics visualization tool was used (Schrödinger, 2015). *Super* module superposes two structures based on the positions of backbone α-Carbon atoms regardless of their amino acid identity. It uses a dynamic programming algorithm which incorporates a series of refinement cycles to eliminate unfit pairing and thus minimizing the root mean square deviation (RMSD) between two aligned structures. Finally, each receptor structure was colored based on sequence identity, similarity and differences as well as RMSD value, to identify variations emerging at both primary and tertiary level.

Computational Solvent Mapping (CS-Map)

Computational solvent-mapping was used to identify all possible ligand binding sites via docking small drug-like organic molecules over the entire receptor surface. For that purpose, the widely used FTMap (Brenke et al., 2009; Kozakov et al., 2015) tool was employed. As for all CS-Map algorithms, FTMap was constructed based on the assumption that binding pockets incorporating the "hot spots" provide major contributions to the free energy of binding, and thus are likely to bind drug-like ligands with high affinity (DeLano, 2002; Ciulli et al., 2006; Metz et al., 2012; Hall et al., 2015). The algorithm uses fast Fourier transform (FFT) correlation approach which effectively and quickly samples billions of probe's poses and

calculate their energies based on a detailed energy function which is CHARMM27 (Brooks et al., 1983). A total of sixteen organic probe molecules (isopropanol, acetaldehyde, phenol, benzaldehyde, urea, dimethyl ether, acetonitrile, ethane, acetamide, benzene, methylamine, cyclohexane, ethanol, N,N-dimetylformamide, isobutanol and acetone) varying in size and chemical compositions were used for docking. Initially, each probe was docked using rigid body algorithm, and a total of 2,000 generated poses were energy-minimized and clustered based on proximity. Clusters were then ranked by their Boltzmann-averaged energy values. Overlapping clusters of different probe types were assembled into consensus sites (CS) identified as "hot spots." When several CS were found to be near each other on the surface of the protein, there is a strong indication of a potential "druggable" binding region. In a sense, FTMap mimics the experimental NMR or X-ray crystallographic studies which attempt to solve the protein structure using a variety of organic solvents which often form clusters in active sites of the protein.

In addition to solvent-mapping the overall tetrameric structure, each monomeric subunit was solvent-mapped individually. This approach increases the number of alternative solutions by enabling regions that would not be accessible in a tetrameric arrangement. Considering the fact that an X-ray structure only represents an instantaneous state of the receptor in time, the monomeric decomposition and mapping approach attempts to alleviate that drawback, and provides alternative binding sites that would not be detected otherwise. However, this approach may give rise to clusters that would be inaccessible from outside if they happen to be located at the interface of monomeric subunits and thus should be discarded.

While all parasitic/bacterial species of PFK are tetrameric structures, *H. sapiens* PFK is dimeric where each monomer consists of two domains. As depicted in **Figure 1A**, each domain is the counterpart of one chain in tetrameric structure of parasitic/bacterial PFK. Thus, when *H. sapiens* PFK was decomposed into its monomeric subunits for solvent-mapping, bacterial PFK was also decomposed into its two-chain units corresponding to one monomeric unit in human and then solvent-mapped for compatibility, in addition to chain-by-chain decomposition. For GADPH and PK, two-chain solvent mapping was not necessary, as they were tetrameric in all species (see **Figure 1B**).

ENM-Based Residue Scanning

Elastic network model (ENM) is a powerful theoretical approach used to predict the global or essential dynamics of biomolecular structures which is then used to establish the relationship between the structure and the functional mechanism (Tirion, 1996; Haliloglu et al., 1997; Doruker et al., 2000; Atilgan et al., 2001). In this model, the protein was represented as a collection of beads connected by Hookean springs corresponding to a collection of atoms connected by fluctuating bonds. Furthermore, the springs connected the atoms only if they were closer than a predefined cutoff distance of 15 Å in the native structure. In our study, we used a residue scanning method that was developed based on this coarse-grained standard ENM (Kurkcuoglu et al., 2015). In this new approach, each residue represented by its backbone α-Carbon as a single node was redefined such that side-chain heavy atoms will be included as extra nodes. It was proposed that these new additions will mimic the presence of a bound ligand interacting with that residue. The effect was then quantified by the change in the *ith* collective mode's eigenvalue λ_i upon adding the extra nodes to the selected residue,

$$\%\ shift\ for\ mode\ i\ (\%s_i) = \frac{\lambda_i\left(modified\right) - \lambda_i\left(original\right)}{\lambda_i\left(original\right)} \times 100$$

The percentage shift for each residue was determined as an average over the 20 slowest modes as 20 slowest essential modes dominated more than 90% of the global dynamics of all three receptors. The average value $\%s = \sum_{i=1}^{20}(\%s_i)/20$ was then represented using a color gradient on the protein's X-ray structure. The regions which incorporate residues with highest $\%s_i$ values were simply proposed as potential allosteric sites. Furthermore, another theoretical method DoGSiteScorer (Volkamer et al., 2012b) incorporating physicochemical pocket features and perturbation based on normal-mode analysis (NMA) has been employed to support our findings.

Merging FTMap and ENM-Based Residue Scanning Results

Clusters identified from FTMap were further explored to identify all proximal residues situated within 5 Å of the bound solvent molecule observed in that cluster. Then, a mean percentage frequency shift value for each cluster was determined as the average over all n residues neighboring all the bound solvent molecules in that cluster as $\hat{S} = \sum_{j=1}^{n}(\%s_j)/n$. If a cluster's $\hat{S}$ value was smaller than 50%, that cluster was simply discarded from analysis as its interaction with a ligand would have a negligible impact on the global dynamics of the receptor. In case the number of alternative solutions is scarce, the threshold value was decreased to 25%.

Determination of Interface Regions Using Relative Solvent Accessible Surface Area (rSASA)

Interface regions are known to incorporate conserved "hot spot" residues which majorly contribute to the free energy of binding to another subunit or partner protein, thus are frequently targeted in species-specific drug design studies (Clackson and Wells, 1995; Bogan and Thorn, 1998). In addition, binding of a ligand at the interface is suggested to disrupt protein's global dynamics which is most often governed by the close correspondence between monomeric units. In this study, the interface regions were determined based on relative solvent accessibility surface area (rSASA) which is a widely used metric to identify buried and exposed residues in the structure. rSASA was defined as a residue's solvent accessibility (ASA) normalized by its maximum ASA value. Maximum ASA for each residue X previously reported in Tien's work (Tien et al., 2013) was derived as the highest ASA achieved in a Gly-X-Gly tripeptide construction evaluated for

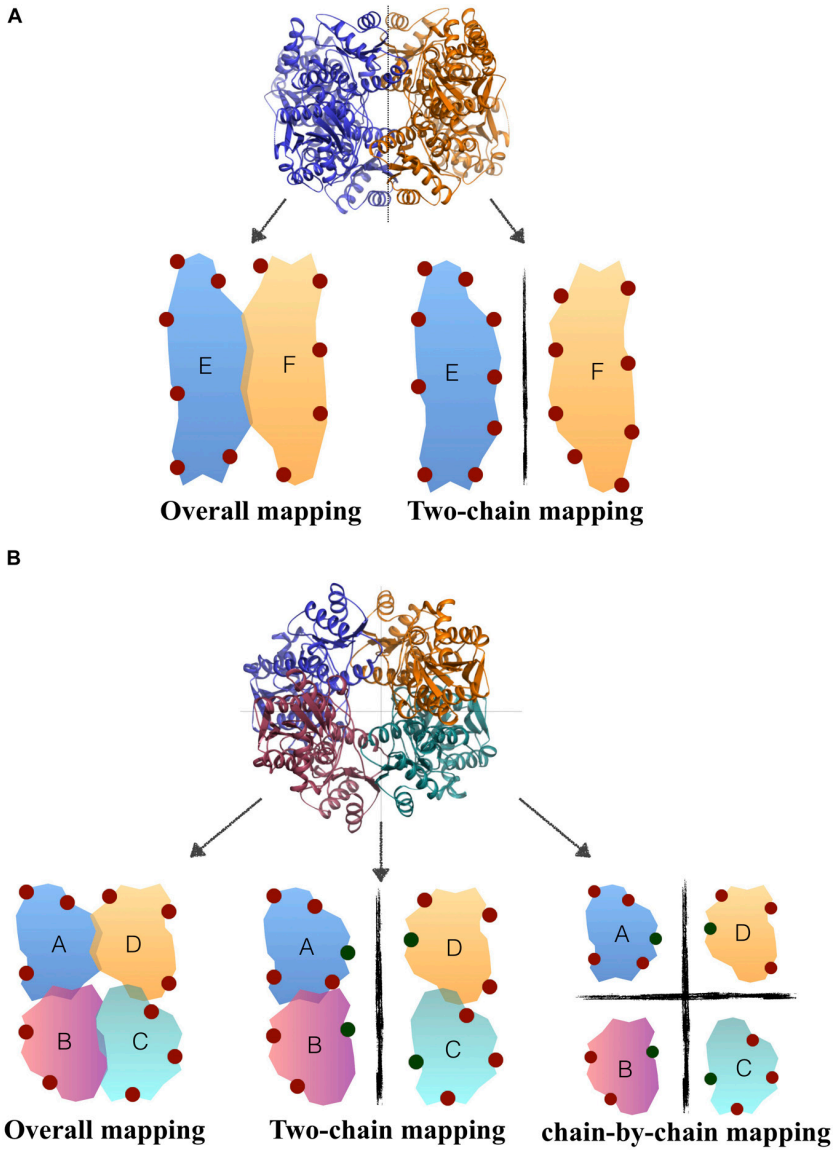


FIGURE 1 | Solvent mapping strategy in **(A)** *H. sapiens* PFK, **(B)** *T. brucei*/*S. aureus* PFK where binding sites proposed by FTMap were illustrated with circles.

all sterically possible conformations. Accordingly, a residue was found at the interface if its rASA value in the monomeric form is greater than its rASA value in the complex form (Levy, 2010).

RESULTS AND DISCUSSION

Solvent Mapping and ENM Analysis Detected Several Druggable and Potential Allosteric Sites At/Near Interface Regions

Consensus sites (CS sites) or hot spots were determined for the overall tetramer, and also for each chain separately, to increase the number of alternative binding sites. In addition, for PFK enzyme only, solvent mapping was also employed on an assembly of two chains, as the corresponding PFK structure in human species existed as a dimer with each monomeric unit corresponding to two chains in bacterial/parasitic species tetrameric structure (see **Figure 1A** in Materials and Methods section). As listed on the third column of **Table 2**, for tetramer mapping, the highest number of CS sites was 18 in *S. aureus* of PK (*Sa*PK), and the lowest number was 8 observed in human GADPH (*h*GADPH). The number of CS sites in chain-by-chain mapping was comparable to that found in tetramer mapping. Overall, GADPH demonstrated the lowest amount of CS sites in all three species.

Several CS sites obtained from chain-by-chain mapping had to be discarded as they either coincided with CS sites obtained from tetramer mapping or became solvent inaccessible in tetrameric

TABLE 2 | Number of clusters determined before and after filtering protocols for three glycolytic enzymes (PFK, GADPH, and PK) in different species (*H. sapiens*, *S. aureus*, *T. brucei*, *T. cruzi,* and *L. mexicana*).

		(1) Total Number of Clusters/ (2) Non-Overlapping Solvent-Accessible Clusters/ (3) After ENM filtering (frequency shift > 25%/50%)							
Enzyme	**Species**	**Tetramer**	**Chain A**	**Chain B**	**Chain C**	**Chain D**	**Chain AB***	**Chain CD***	**TOTAL**
PFK	*H. sapiens*	13	11	12	13	9	10	11	79
		13	9	11	12	8	2	2	66
		13/13	7/3	9/5	9/5	8/5	2/2	2/2	50/35
	S. aureus	17	12	11	11	11	11	12	85
		17	9	8	8	8	4	4	58
		17/17	5/4	4/3	4/3	5/4	4/4	4/4	43/39
	T. brucei	13	10	11	9	10	13	12	78
		13	10	10	9	10	4	5	61
		13/8	8/7	8/3	6/3	7/3	4/3	5/5	51/32
GADPH	*H. sapiens*	8	7	7	9	8	–	–	39
		8	6	7	9	7	–	–	37
		8/8	3/3	3/2	4/2	3/1	–	–	21/16
	S. aureus	14	6	9	8	7	–	–	44
		12	3	6	6	4	–	–	31
		12/12	3/3	6/6	6/5	4/3	–	–	31/29
	T. cruzi	15	7	9	7	9	–	–	47
		15	5	5	5	7	–	–	37
		15/2	5/5	5/2	5/2	7/7	–	–	37/18
PK	*H. sapiens*	12	12	10	10	9	–	–	53
		12	11	9	9	8	–	–	49
		12/5	9/7	8/4	7/3	8/4	–	–	44/23
	S. aureus	18	8	12	10	10	–	–	58
		18	8	11	9	9	–	–	55
		18/10	6/4	7/4	7/4	7/4	–	–	45/26
	L. mexicana	15	9	11	9	9	–	–	53
		15	6	9	5	5	–	–	40
		15/6	3/3	7/6	5/5	2/1	–	–	32/21

**Chains AB and CD correspond to one monomeric subunit in H. Sapiens (see Materials and Methods section).*

arrangement. Numbers listed in the second row of each cell in **Table 2** indicate the number of non-overlapping and solvent accessible clusters. Lastly, each site in the remaining list was evaluated based on its percent frequency shift value averaged for all the residues in the immediate vicinity (%*s*), as mentioned in Materials and Methods section. Accordingly, CS sites displaying an average %*s* lower than 50% was eliminated in the first run. To increase the number of alternative binding sites, a second threshold of 25% was also used in case the number of solutions is limited. As listed in the third row of each cell in **Table 2**, the total number of CS sites was found to be significantly higher in PFK enzyme for all three species than either GADPH or PK. Another unexpected outcome was *S. aureus* displaying the highest number of hot spots among species for all three enzymes, with the highest number being 39 observed for PFK and all satisfying 50% frequency threshold.

The location of all consensus sites listed in **Table 2** was presented extensively in **Supplementary Figures S1–S3** for all three enzymes. It was noticeable that the majority of CS sites was detected at/near interface regions as indicated in blue color. Furthermore, the existence of more than one CS site situated nearby further emphasized the existence of a druggable site. **Supplementary Tables S1–S3** list all these druggable sites with two or more CS sites. Some of these clusters were marked with either a single or a double star which indicate those that did not fulfill 25 and 50% frequency shift criterion, i.e., ineffective sites. CS sites with double stars were those having a frequency shift between 25 and 50%, and were used in case of limited number of alternative solutions. Isolated CS sites were those with no close proximity to any other CS sites. They were only observed for PFK and PK enzymes and listed separately in the footnote section of the corresponding table.

To further highlight the most probable target regions, all druggable sites with two or more effective CS which gave rise to a frequency shift above 50% in global dynamics were listed in the following **Table 3**. In addition, residues constituting each site were determined based on their proximity to the clusters (<5Å) and listed in **Supplementary Tables S4–S6** for each enzyme separately. The top druggable site on the list given in bold character incorporates the highest amount of CS and

TABLE 3 | Druggable sites incorporating several consensus sites labeled with an ID composed of a number and a letter.

Enzyme	Druggable site ID	*S. aureus*	Parasite	*H. sapiens*
PFK	1	4-5-7-8-12-15-17	1-9-13-6AB-7AB	4-7-8-11CD
	2	1-2-3-6-11-16	7-11-5CD-12CD	5D-6D-7D
	3	2A-10A-11A#	5-6CD-9CD-11CD	6B-7AB-1
	4	2D-9D-11D	2A-3A-5A-8A	3-5-9
	5	2B-10B	3-12-13AB	5B-10B
	6		6A-7A	4C-8C
	7		10B-11B	6-11
	8		7D-10D	10-12
	9			4B-9B
	10			6C-9C
	11			2-5C
GADPH	1	1A-2A-5A-7A-8A-9A-2	1D-3D-4D-6D-7D-9D-3	1-2-3-4-5-6
	2	2B-3B-4B-7B-8B-5	2A-3A-4A-5A-7A-12	2A-4A-7
	3	3C-4C-6C-7-8-12	1B-6B-7B	2B-4B
	4	3D-4D-5D-3-14	2C-4C	2C-5C
	5	1-4-6-10-11		2D-8
PK	1	2-3-16-18	4C-5C-6C	2-4-5-6
	2	4-5-13-15	1B-9B-11B	3A-8A-9A-12A
	3	1A-2A-6A	9C-5-14	1B-9B
	4	1B-2B-8B	8-8A	1C-5C
	5	2C-3C-5C	7-13	7D-9D
	6	2D-3D-4D		

 # The ID is composed of a single number which is sometimes followed by a letter. The number indicates the rank of that cluster; the smaller the number, the most populated the consensus site is, which increases the likelihood of that cluster. The letter indicates the mapping type, i.e., "A" indicates chain-by-chain mapping result coming from chain A, etc.

was illustrated in **Figure 2** for each three enzymes of each species. PFK exhibited the highest amount of distinct druggable sites among three enzymes, varying from 5 for *S. aureus* up to 11 for *H. sapiens*. In GADPH, 2 –5 druggable sites were detected only, yet each site was crowded with several CS. Pyruvate kinase displayed a total of 5–6 druggable sites for each species, each holding 2–4 effective consensus sites. It is important to note that all druggable sites shown in **Figure 2** also have symmetric counterparts which are shown in detail in **Supplementary Figures S1–S3**.

S. aureus Phosphofructokinase (*Sa*PFK) Indicated an Alternative Allosteric Region in Addition to Well-Known Allosteric and Catalytic Regions

For phosphofructokinase enzyme, all druggable sites listed in **Table 3** were observed at the interface region as depicted in detail in **Supplementary Figure S1**. Seven CS located at the top druggable site were picked up from solvent mapping of the tetrameric structure, as illustrated in green at the top left figure of **Figure 2**. In the vicinity of this region, there exist isolated consensus sites obtained from chain-by-chain solvent mapping and are distinguishable by their magenta color, reinforcing the promise of this site for allosteric regulation. The second top druggable site on the list with six CS is the symmetric counterpart of the first and is located on the exact opposite face of the enzyme (see **Supplementary Figure S1**A). Either one of these sites can be safely proposed as an allosteric target region. Furthermore, a computational study conducted by Mitternacht et al. recognized the same exact region via Monte Carlo simulations as a possible binding site as it showed characteristics of being coupled to the intrinsic motion of the protein (Mitternacht and Berezovsky, 2011). Furthermore, this proposed region has an equivalence in human species which also incorporates top druggable site as depicted in **Figure 2** (top right corner). On the other hand, as the human PFK is composed of two dimers where each monomeric unit is equivalent to two dimers in bacterial PFK, there is no interface in this proposed allosteric site. Recently, drug discovery studies aim the interface regions for identifying new allosteric drug candidates that would likely inhibit enzymatic activity through changing the global dynamics and thus preventing large dynamics subunit motion required for forming the active site (Pommier and Cherfils, 2005; Rahimova et al., 2018). Hence, the absence of any interface at the correspondin garea in human PFK might offer some advantage when designing drug molecules specific for *S. aureus* PFK.

Moreover, our computational approach was employed for the dimeric form of human PFK which represents the inactive state, hence our conclusion would not be complete without investigating the active state of human PFK which is a tetramer. The tetrameric active form of human PFK incorporates two dimers and as each dimer corresponds to one tetrameric structure in bacteria, the human PFK becomes the equivalence of two bacterial tetramers. Consequently, an additional solvent mapping and elastic network modeling was employed using the human

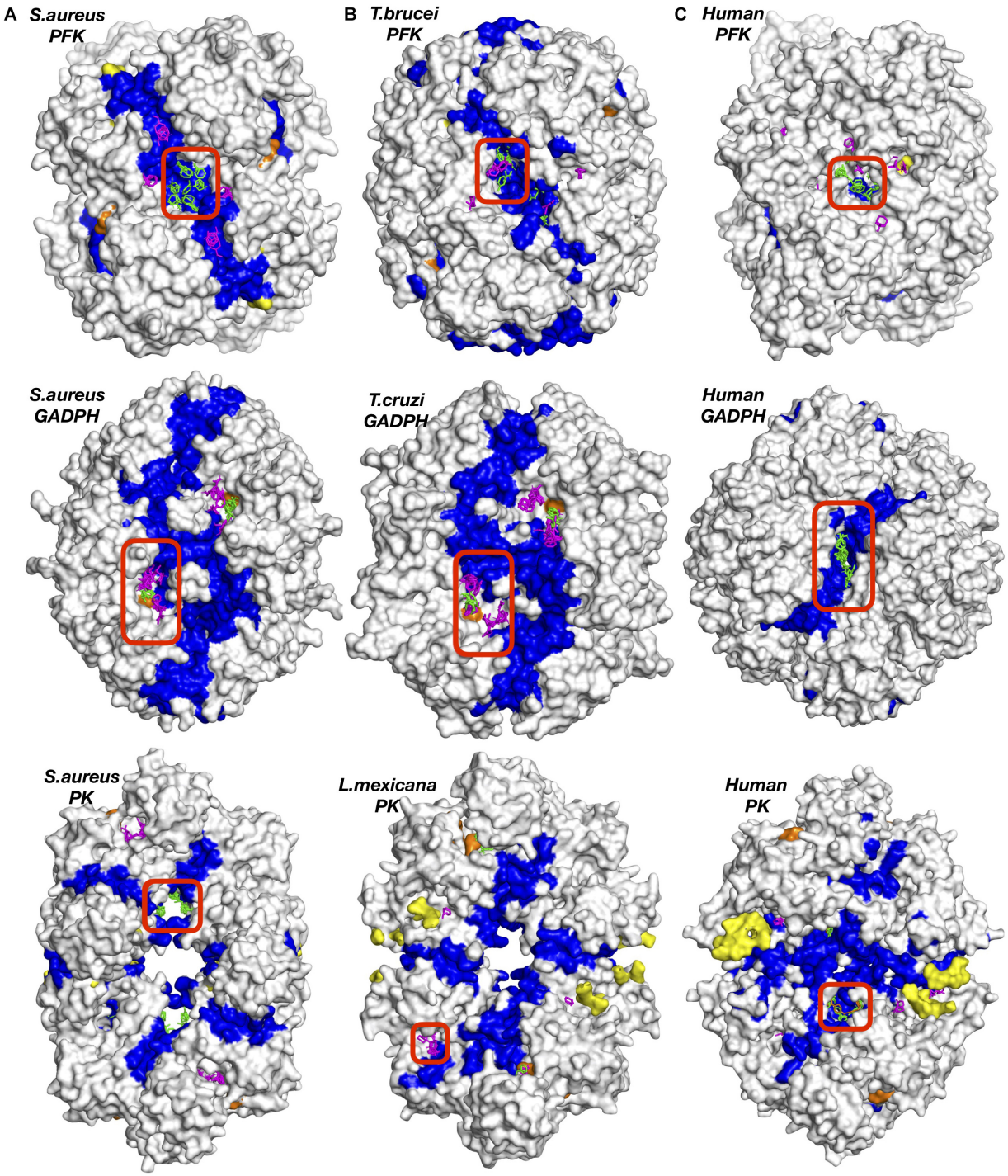


FIGURE 2 | Potential druggable sites proposed for three enzymes of different species, **(A)** bacteria (*S. aureus*), **(B)** parasite (*T. brucei*, *Cruzi,* or *L. mexicana*) and **(C)** human (*H. sapiens*) using solvent mapping (FTMap). Interface regions between subunits indicated in blue color. Experimentally reported allosteric and catalytic regions were highlighted in yellow and orange, respectively. Clusters colored in green and magenta correspond to results for tetrameric and chain-by-chain solvent mapping, respectively.

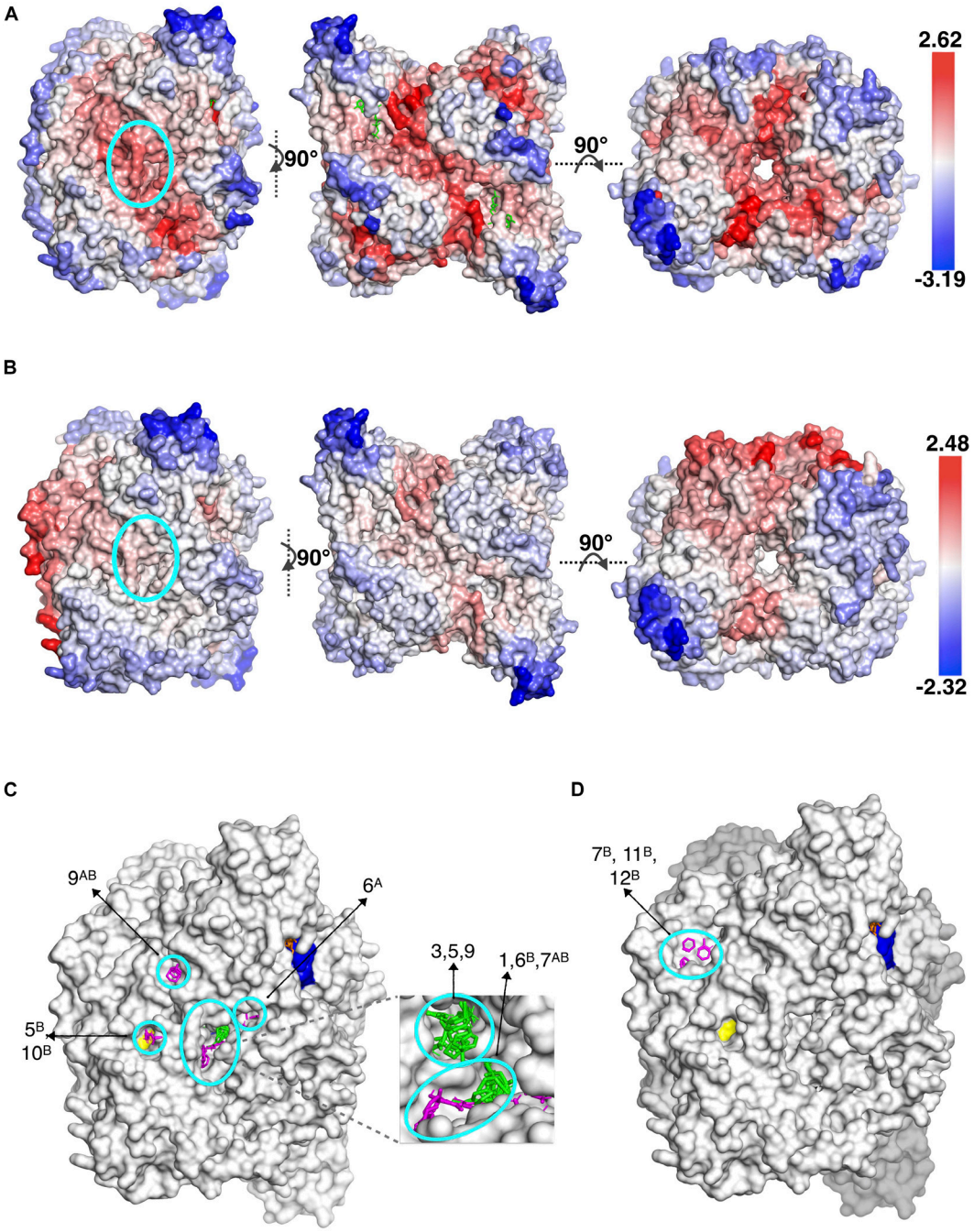


FIGURE 3 | Different views of the same snapshot of human PFK colored based on frequency shift in **(A)** dimeric (inactive) and **(B)** tetrameric (active) forms, for which the top druggable site is highlighted with a cyan circle. Consensus sites at the top druggable site of human PFK in **(C)** dimeric (inactive) and **(D)** tetrameric (active) forms. Clusters colored in green and magenta correspond to results for tetrameric and chain-by-chain solvent mapping, respectively.

tetramer and the clusters with frequency shifts above 50% were collected together with clusters obtained for the human dimer only. As indicated with a color gradient in **Figures 3A,B**, the intensity of the frequency shifts in human tetramer in 3b was significantly lower than those in dimer form in 3a. On the other hand, as anticipated, the highest intensity of frequency shift in tetramer form was observed at the interface region through which the second subunit bind.

The proposed druggable site encircled in the left figures clearly indicate the active tetramer form displaying a lower degree of frequency shift compared to dimer form. Consequently, the number of druggable sites which satisfied the frequency

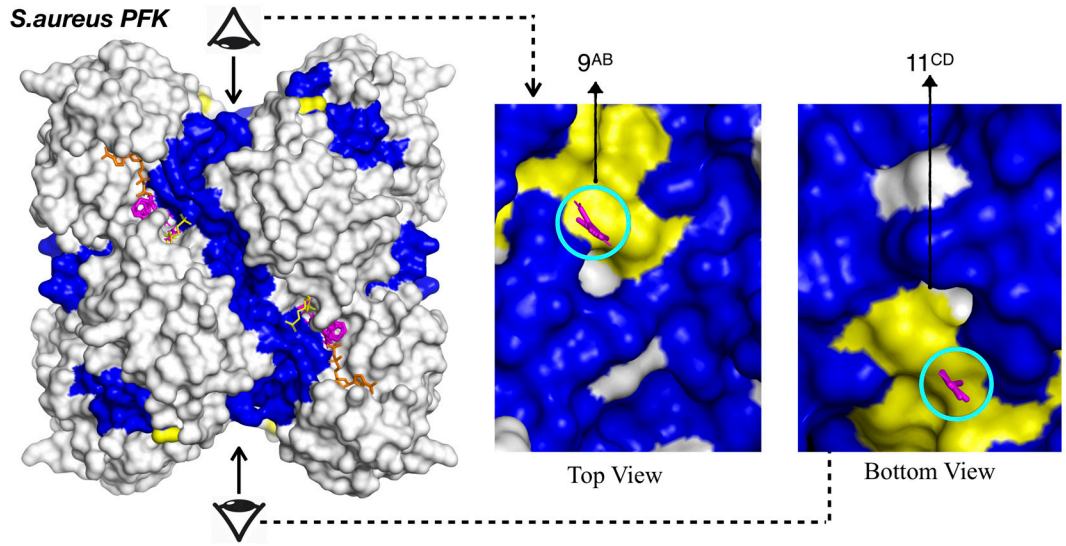


FIGURE 4 | Alternative allosteric regions proposed in *S. aureus* and PFK indicated by circled consensus sites. Yellow patches indicate the experimentally observed allosteric sites (Kloos et al., 2015; Tian et al., 2018). ATP and F6P were illustrated with orange and yellow sticks. Clusters colored in magenta correspond to results for chain-by-chain solvent mapping.

shift threshold of 50% in tetramer was significantly reduced (see **Figures 3C,D**). This further increased the potential of our proposed site to be the most suitable target region for designing species-specific drug molecules, as the same region in active form of human PFK would not favorably accommodate any drug molecule or if that happens, the receptor's global dynamics would not be affected by its binding as much as its bacterial counterpart would.

The three remaining druggable sites listed for *S. aureus* in **Table 3**, were observed in the vicinity of the active site (depicted in orange in **Supplementary Figure S1**), thus they are far from functioning allosterically. Still, it clearly demonstrates the power of our computational approach to detect catalytic sites as well as allosteric sites which both require a coupling between ligand binding and protein's intrinsic dynamics. Furthermore, there exist a second region in *Sa*PFK which incorporates one isolated CS visible at the top and its symmetric counterpart at the bottom view of the receptor as depicted in **Figure 4**, thus creating a region for possible allosteric regulation. No such cluster was observed in the same region in human PFK. Besides, this second alternative site is passing through an interface region, further accentuating its potential role in allostery. However, these consensus sites were detected within reach to a well-known binding site for two allosteric effectors which are the activator ADP-Mg and the inhibitor phosphoenolpyruvate (PEP), as depicted with yellow surface in **Figure 4** (Schirmer and Evans, 1990). Although this site cannot be introduced as novel, the findings are supportive of our procedure's prediction power.

For proposing a potential target region for designing species-specific drug molecules that would bind more strongly to *Sa*PFK than its human equivalence, we need to make sure that either structural or sequence conservation is minimum at the region of interest. As illustrated in **Figure 5A**, a snapshot of *Sa*PFK colored based on sequence similarity/identity between human and bacteria clearly displays a low degree of sequence conservation in the proposed site, as highlighted with an abundancy of white spaces corresponding to dissimilar residues. Furthermore, the overall structural RMSD between two species was determined as 1.56 Å, and this value is even lower for the confined region at the top druggable site. As the human counterpart of this proposed allosteric region in *Sa*PFK also incorporates the top druggable site with four CS as depicted in **Supplementary Figure S1**, the degree of variation at sequence level is satisfactorily low for proposing this site as a target in the design of species-specific drug molecules.

T. brucei PFK (*Tb*PFK) Suggested an Alternative Allosteric Region in Addition to a Site Within Reach to a Catalytic Region

Similar to *S. aureus*, the top druggable site incorporating five CS was observed in a region passing through an interface and was the counterpart of the allosteric region in *S. aureus*, as illustrated in **Figure 2B** (top middle). In the vicinity of this region, there exist several other druggable and consensus sites strengthening its likelihood to be allosteric. Moreover, there exist two alternative sites represented by encircled areas located at close proximity to each other and to an interface region as illustrated in **Figure 5B**. The symmetric counterparts of these regions also exist at the opposite site of the receptor as illustrated in detail in **Supplementary Figure S1**. However, each of these sites coincide with the well-known binding area of the substrate F6P shown with yellow sticks, therefore unlikely to be suggested as an allosteric site. In human counterpart, a similar observation was made, i.e., two distinct druggable sites were detected as

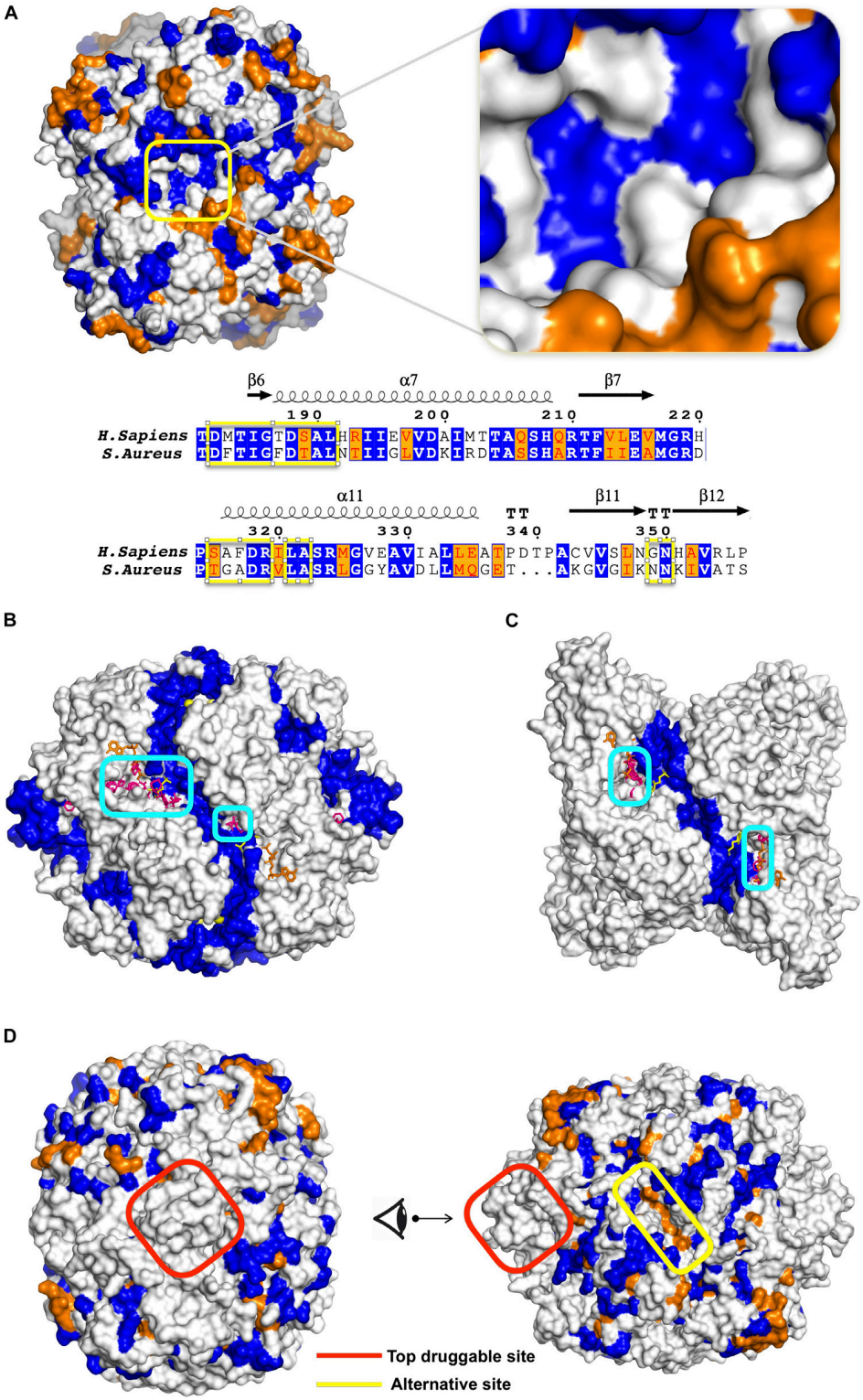


FIGURE 5 | (A) Sequence similarity between human and *S. aureus* PFK illustrated on a snapshot and a sequence alignment with top druggable site encircled in yellow. ESPript 3.0 tool (Robert and Gouet, 2014) used for graphical illustration of sequence alignment. Potential allosteric sites represented by clusters encircled in blue for **(B)** *T. brucei* and **(C)** human PFK. ATP and F6P colored in orange and yellow, respectively. **(D)** Sequence similarity illustrated on a snapshot of *Tb*PFK at two different angles. Similar, identical and dissimilar residues colored in orange, blue and white, respectively.

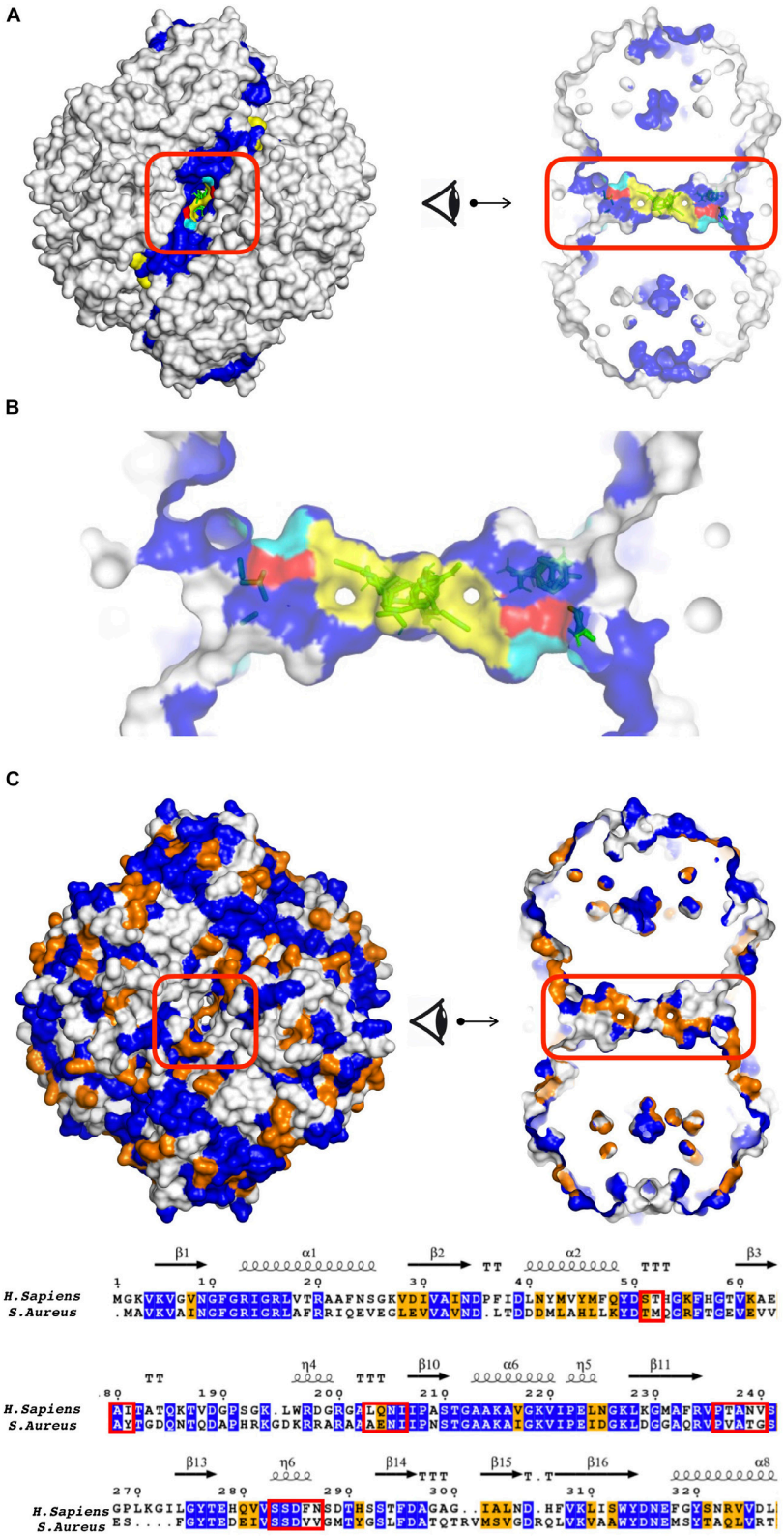


FIGURE 6 | (A) Tunnel like region as a potential allosteric site in *S. aureus* GADPH using different perspectives. **(B)** S-loop was depicted with yellow patches, key residues S50 and S287 in the tunnel region, colored in red and cyan, respectively. Sequence similarity between *S. aureus* and *human* GADPH, illustrated on **(B)** a snapshot in two different angles and **(C)** a sequence alignment. See caption of **Figure 5** for color coding. Tunnel region encircled in red.

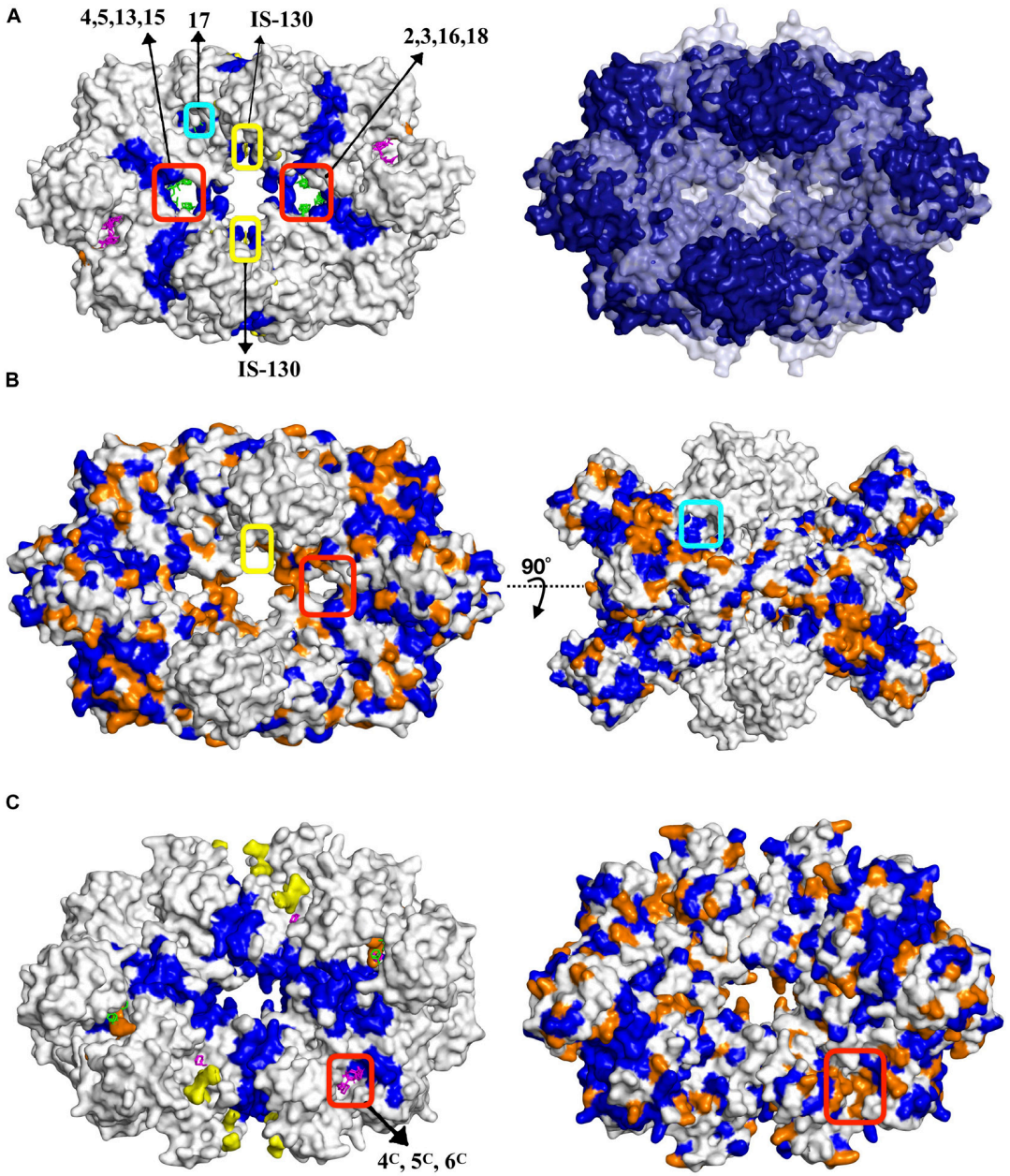


FIGURE 7 | (A) Potential allosteric sites in *S. aureus* pyruvate kinase along with structural alignment of *S. aureus* and human PKs. Dark blue represents regions similar in both species, whereas white and pale blue regions correspond to unmatched regions, respectively. Sequence similarity illustrated on **(B)** a snapshot with two different views where proposed site encircled in red, well-reported IS–130 bound allosteric site encircled in yellow, proposed site with cluster ID 17 encircled in cyan. See caption of **Figure 5** for color coding. **(C)** Top druggable site strongly proposed as a potential allosteric site in *L. mexicana* pyruvate kinase.

depicted in **Figure 5C**, each close to ATP and F6P binding area, but also near the interface region. On the other hand, the area in between these two druggable sites might be proposed as a druggable target site for allosteric drug candidates as it is passing through an interface which might perturb the global dynamics of the receptor essential for its activity. In addition, it displays a low level of sequence conservation represented by mostly white and orange spaces as in **Figure 5D**. Furthermore, the top druggable site displays a significantly low level of sequence conservation as depicted with a nearly white area encircled as in **Figure 5D**, thus would be an ideal location to be targeted for species-specific drug discovery.

Tunnel Region Observed in GADPH Can Be a Potential Allosteric Site

GADPH displayed a distinct profile of druggable sites which were well packed with several CS in both bacteria and parasite. However, top sites on the list were observed near the catalytic

TABLE 4 | Top druggable sites for each enzyme from each species and their corresponding DogSite binding pockets with score and rank.

Enzyme	*S. aureus*	Region	Score*/rank	Parasite	Score/rank	Region
PFK	4-5-7-8-12-15-17	Allosteric	0.81/3	1-9-13-6AB-7AB	0.80/4	Allosteric
	1-2-3-6-11-16	Allosteric	0.81/3	7-11-5CD-12CD	0.80/4	Allosteric
	2A-10A-11A	Catalytic	0.49/13	5-6CD-9CD-11CD	0.80/4	Allosteric
	2D-9D-11D	Catalytic	0.49/13	2A-3A-5A-8A	0.81/3	Catalytic
	2B-10B	Catalytic	0.78/6	3-12-13AB	0.80/4	Allosteric
				6A-7A	N/A	Catalytic
				10B-11B	0.53/16	Allosteric
				7D-10D	0.52/17	Allosteric
GADPH	1A-2A-5A-7A-8A-9A-2	Catalytic	0.80/2	1D-3D-4D-6D-7D-9D-3	0.48/6	Catalytic
	2B-3B-4B-7B-8B-5	Catalytic	0.76/3	2A-3A-4A-5A-7A-12	0.44/8	Catalytic
	3C-4C-6C-7-8-12	Catalytic	0.73/4	1B-6B-7B	0.72/3	Catalytic
	3D-4D-5D-3-14	Catalytic	0.64/8	2C-4C	0.81/1	Catalytic
	1-4-6-10-11	Allosteric	0.80/2			
PK	2-3-16-18	Allosteric	0.76/6	4C-5C-6C	0.42/9	Allosteric
	4-5-13-15	Allosteric	0.71/8	1B-9B-11B	0.81/4	Catalytic
	1A-2A-6A	Catalytic	0.58/18	9C-5-14	0.83/3	Catalytic
	1B-2B-8B	Catalytic	0.57/19	8-8A	0.81/4	Catalytic
	2C-3C-5C	Catalytic	0.63/16	7-13	0.81/4	Catalytic
	2D-3D-4D	Catalytic	0.53/21			

**Minimum-maximum range for score values:*
S. aureus PFK: [0.14-0.88]; T. brucei PFK:[0.16-0.86];
S. aureus GADPH: [0.14-0.83]; T. cruzi GADPH:[0.15-0.81];
S. aureus PK: [0.15-0.84]; L. mexicana PK:[0.11-0.86].

region and thus cannot be proposed as allosteric (see **Figure 2**). On the other hand, our procedure accurately detects all catalytic sites in addition to allosteric ones. An additional druggable site which appeared in **Table 3** with five CS for *S. aureus*, was detected in a tunnel like region passing through the center of the receptor as depicted in **Figure 6A**. In *T. Cruzi* GADPH, consensus sites which appeared in the same tunnel region only displayed a moderate amount of frequency shift which was determined between 25 and 50%, whereas those in human and *S. aureus* GADPH, the 50% criterion was fulfilled.

Agreeably, the tunnel like location coincided with a well-known dynamic S-loop which is known to be modulated by phosphorylation of Ser50, Ser203, and Tyr41 in regulating the enzymatic activity through NAD-binding pocket and oligomer assembly (Dubey et al., 2017). The regulatory effect of GADPH S-loop via its phosphorylation is a universal feature as the phosphorylated sites consist of well conserved residues. Dephosphorylated Ser50 and Tyr41 both play a part in homodimerization by hydrogen bonding across the dimer interface with S287 and stabilizes the neighboring S-loop, whereas dephosphorylated Ser203 induces the fit of S-loop into the neighboring NAD-binding pocket by forming atomic interactions with three other S-loop residues. Among these residues, S50 and S287 were visible in the tunnel region, as illustrated in **Figure 6B** in red and cyan color, respectively. In addition, S-loop was depicted with yellow patches.

The tunnel region was further investigated for the amino acid sequence similarity between human and bacterium/parasite in order to guarantee that the proposed site incorporates distinctive features for identifying drug molecules that would specifically inhibit the enzyme of the infecting organism which is *S. aureus*. Colored based on sequence similarity between human and bacterium/parasite, the snapshots and the sequence alignment in **Figure 6C** clearly demonstrate the low degree of sequence conservation at the tunnel region. On the other hand, the structural RMSD value for the same region was exceptionally low at around 0.34 Å.

S. aureus Pyruvate Kinase Displayed One Allosteric Site at the Center Cutting Across the Interface Region and Another at the Junction of A/C Domain

As listed in **Table 3**, the top druggable site corresponds to a region which is located at an opening in the center of the receptor and crossing an interface region. Its symmetric counterpart can also be observed at the other side of the orifice and both of these clusters were listed as top two druggable sites in **Table 3**. Moreover, a well-known allosteric site exists in the same orifice which accommodates the inhibitor IS-130 (N'-[(1E)-1-(1H-benzimidazol-2-yl)ethylidene]-5-bromo-2-hydroxybenzohydrazide) which was previously identified by Cilies and his coworkers as a potential allosteric inhibitor targeting methicillin-resistant Staphylococcus aureus (MRSA) (Axerio-Cilies et al., 2012). As illustrated in **Figure 7A** with yellow sticks at the top and bottom of the orifice, it is located at the so-called small C-C interface separating the subunits of the receptor, thus by disrupting the essential salt bridges which help to stabilize the small C-C interface and lock the tetramer in the active R-state, it may prevent the conformational transition

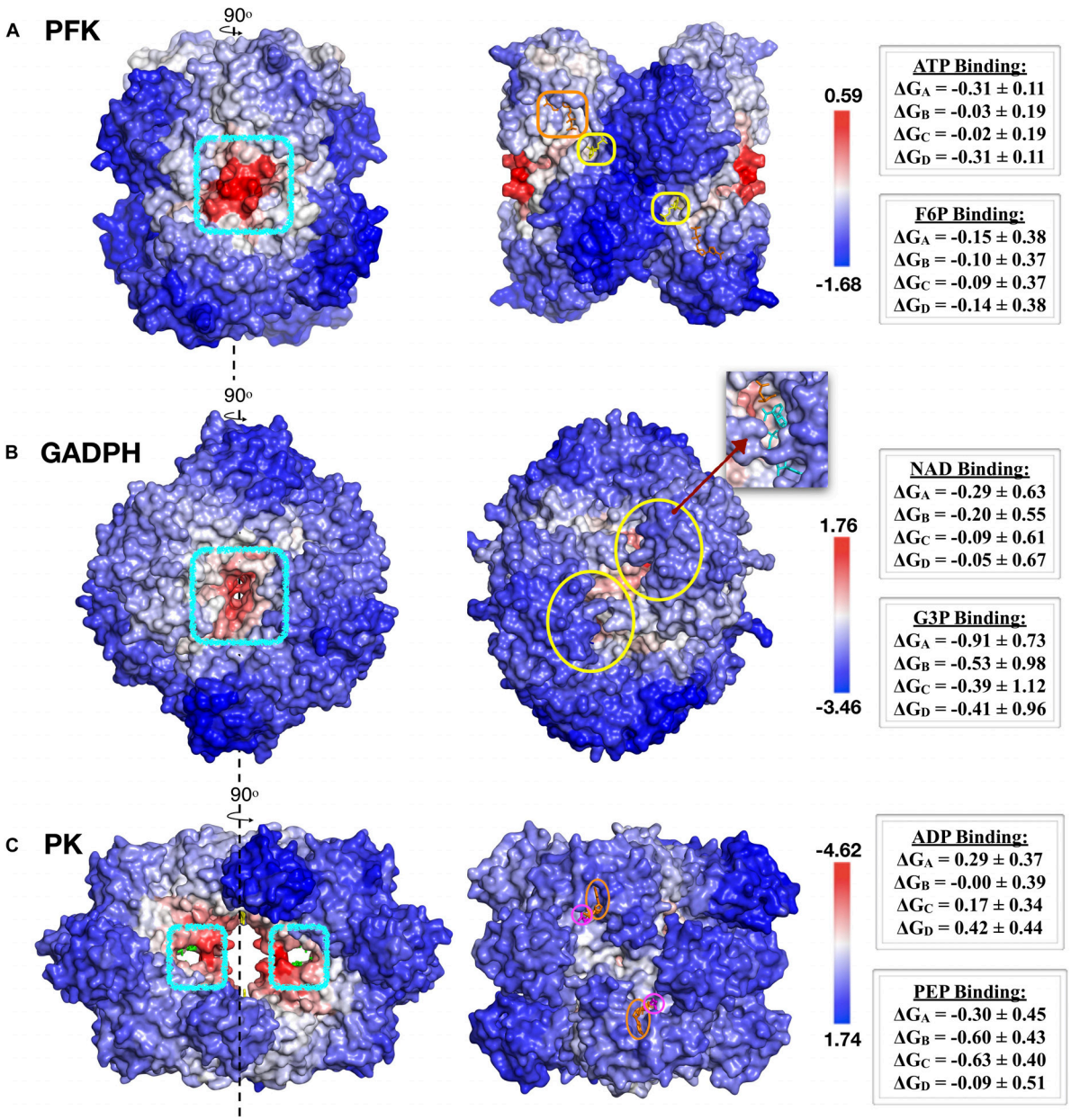


FIGURE 8 | AlloSigMA results for top druggable sites and their symmetric counterparts in **(A)** PFK, **(B)** GADPH and **(C)** PK enzymes of *S. aureus* species. Proposed sites depicted with blue circles on the left, while catalytic regions indicated on the right side. Red and blue regions correspond to regions with decreased ($\Delta G < 0$) and increased ($\Delta G > 0$) dynamics, respectively.

to an active state (Morgan et al., 2010). Our newly proposed target site indicated by green sticks on the right and the left side of the orifice (encircled in red) is passing directly through the so-called large interface region, thus might eventually affect the rocking motion of the subunits necessary for activation. The human pyruvate kinase has a potential druggable site in the same corresponding region, however, the center of the receptor has a distinct shape with a nearly closed orifice almost inaccessible to the other side (see **Figure 7A**). Furthermore, this predicted druggable site coincide with the same pocket where quinolone sulfonamide activators bind (Kung et al., 2012). Besides, sequence alignment indicates this area with high amounts of variations which further emphasizes our proposed site as an ideal target for species-specific drug design (see **Figures 7B** and **Supplementary Figure S4A**). The remaining four clusters appeared in the vicinity of each of the four catalytic sites, as illustrated in **Supplementary Figure S3**, hence do not suggest an allosteric feature. A second alternative allosteric site in *Sa*PK appeared at the junction of A and C domain of one subunit as depicted by cluster with id 17 encircled with cyan in **Figures 7A,B**. No cluster was observed

in human PK at the same corresponding site. Furthermore, sequence similarity analysis displayed a high degree of variation which indicated a likelihood of species specificity.

L. mexicana Pyruvate Kinase Displayed a Distinct Allosteric Site Nearby an Interface Region

As illustrated in **Figure 7C**, a distinct druggable site on *Lm*PK was observed in the vicinity of an interface region, far away from both catalytic sites and the central region. Based on a low degree of sequence conservation, it is likely to provide a distinct binding site for specific drug candidates (see **Supplementary Figure S4B**). The remaining four druggable sites listed in **Table 3** were detected at each of the four catalytic sites. Furthermore, there was no druggable site at the center which satisfied 50% frequency shift threshold as in human or bacteria. On the other hand, four isolated consensus sites have been detected at the center at the same exact locations as in human or bacteria, but in terms of effecting/shifting the frequency of the normal modes, they remained moderately within the range of 25–50%. Still, they can be proposed as possible target sites for species-specific drug design studies for *L. mexicana* PK.

Critical Assessment of Binding Pockets With DoGSiteScorer

Our findings were compared to potential binding pockets predicted by the algorithm DogSiteScorer which is a grid-based method solely based on protein's tertiary structure divided into subpockets, each assigned to a score value. DogSite scores appear between 0 and 1 with the most probable binding pockets displaying score values closer to 1. Each one of our druggable sites previously listed in **Table 3** was re-evaluated based on the scores of DogSite pockets to which they overlapped. As shown in **Table 4**, high-score DogSite pockets coincided successfully with our predicted top druggable sites.

For *S. aureus* PFK, the predicted top druggable sites overlapped with the DogSite pocket ranked in third with 0.81 score value. The top two DogSite pockets with only slightly higher scores, 0.87 and 0.88, corresponded to catalytic regions where ATP binds (see **Supplementary Figures S5A,B**). The top druggable site in *T. brucei* which corresponded to the same location as in *S. aureus* and proposed to be allosteric, successfully coincided with a DogSite pocket of 0.8 score value which was the fourth highest. Interestingly, a second alternative allosteric site which was observed at the interface and proposed for *T. brucei* PFK as outlined with a yellow rectangle in **Figure 6C**, displayed a favorable DogSite pocket with 0.81 score value as also depicted in **Supplementary Figure S5C**. Moreover, the top score DogSite pocket in *T. brucei* PFK was detected in the interior region of the receptor unlike other binding cavities reported so far (see **Supplementary Figure S5D**).

For *S. aureus* GADPH, the tunnel region proposed to be an allosteric site displayed a pocket with 0.8 score value ranked in second (see **Supplementary Figure S6A**). Almost all catalytic regions overlapped with high-score pockets (see **Supplementary Figures 6B,D**). Interestingly, the corresponding tunnel region in *Cruzi* GADPH which did not appear among druggable sites due to its moderate frequency shift coincided with a favorable DogSite binding pocket with 0.80 score value ranked in second (see **Supplementary Figure S6C**). This finding increases the likelihood of the same tunnel region to be an allosteric site in parasite species as well, despite its relatively low frequency shift.

The new allosteric region proposed for *S. aureus* PK at the center of the structure neighboring the large interface coincided with the DogSite pocket ranked in sixth with a value of 0.76 which is not far from the highest score of 0.84 obtained for this structure (see **Supplementary Figure S7A**). On the other hand, catalytic regions appeared as druggable sites in our list were not strongly selected by DogSite. For *L. mexicana* PK, the proposed allosteric site located far from the origin and nearby an interface was not a highly favorable pocket for DogSite with only 0.42 score value ranked in the ninth position. On the other hand, the remaining four catalytic sites coincided well with highly scored DogSite pockets (see **Supplementary Figure S7B**).

Support From AlloSigMA Server

Finally, our proposed allosteric sites were evaluated with AlloSigMA tool (Guarnera and Berezovsky, 2016b; Guarnera et al., 2017) which quantifies the allosteric effect of a ligand binding and/or mutation at a site on the basis of a per-residue free energy which is obtained by solving all possible protein local configurations. For our three allosteric enzymes, we investigated the effect of a ligand binding to our top druggable sites in *S. aureus* only. Other druggable sites and species will be considered in a future work.

Accordingly, the ligand binding to the proposed top druggable site and its symmetric counterpart in each of three enzymes caused a fair amount of decrease in residue dynamics in all catalytic regions. In phosphofructokinase, the highest decrease in allosteric effect was quantified by a negative mean free energy of -0.31 ± 0.11 and -0.15 ± 0.38 for ATP and F6P binding sites, respectively. Mean ΔG values of all four catalytic sites were listed as in **Figure 8A**. All four catalytic regions encircled in yellow for F6P displayed a comparable degree of mean ΔG which was around −0.1, whereas ATP binding site encircled in orange displayed two different values, one nearby −0.3 and the other −0.02.

Similarly analysis was conducted for the known allosteric site of *Sa*PFK illustrated in **Figure 4**, for comparison only. Binding of an effector molecule at the allosteric site is known to increase the activity of the receptor. Ligand binding with AlloSigMA exhibited a moderate amount of increase in the dynamics of F6P binding site with mean ΔG values varying between 0.12 and 0.25, whereas ATP binding site in two of the four monomeric units displayed a decrease in dynamics with a mean ΔG value of −0.5.

In GADPH, there exist four catalytic sites in which the substrate glyceraldehyde 3-phosphate (G3P) as well as the cofactor NAD binds. Two of these sites were illustrated with yellow circles as in **Figure 8B**, while the remaining two are on the opposite face of the receptor. The probe ligand was bound on both sides of the tunnel simultaneously. Accordingly, all four G3P sites displayed negative ΔG values between -0.66 ± 0.84 and -0.24 ± 1.1, whereas only two NAD binding sites showed

unaltered dynamics with low positive ΔG values, 0.05 ± 0.67 and 0.09 ± 0.61. The proposed tunnel region clearly demonstrated a fair amount of allosteric effect on all four catalytic regions.

Finally, for pyruvate kinase, the allosteric effect via ligand binding to two symmetric proposed sites at the central region as depicted in **Figure 8C**, manifested itself as a moderate amount of decrease in the dynamics of each of the four catalytic regions where the substrate PEP would bind. On the other hand, all four ADP binding sites displayed only a slight increase in their dynamics. Furthermore, a similar analysis was conducted for the known allosteric site, which was occupied by the allosteric inhibitor IS-130 at the central region as illustrated in **Figure 7A**. Surprisingly, the allosteric effect was the opposite of that observed for our proposed site, with an increase in dynamics in the majority of PEP and ADP binding sites with mean ΔG values as high as 0.78 ± 0.21.

CONCLUDING REMARKS

Our new approach consisting of a combination of well-established algorithms such as normal mode analysis using elastic network model and solvent-molecule binding site detection algorithm along with sequence and structural alignments demonstrated an exceptional prediction power for discovering alternative allosteric sites in the protein which were proposed as potential target sites for species-specific drug design efforts. The fact that nearly all well-reported catalytic and allosteric sites for three glycolytic enzymes have been identified undoubtedly supports the accuracy of our findings. Besides, several alternative allosteric sites have been identified for each one of three enzymes. *Sa*PFK presented a novel allosteric site which had one of the highest DogSiteScore value in addition to an allosteric effect perturbing the dynamics of all four catalytic regions. The second glycolytic enzyme, GADPH, presented the tunnel region as a potential allosteric site. Notably, this tunnel region incorporates the critical S-loop which owns the universal regulatory effect of the enzyme activity via its phosphorylation. The ligand binding to two symmetric sites at the tunnel region induced a fair amount of decrease in all four catalytic regions of the receptor. Finally, the two symmetric binding sites proposed for pyruvate kinase at the central region, exhibited allosteric features which were stronger than the known allosteric inhibitor sites nearby.

Although our current work was focused on allosteric enzymes only, the remaining seven glycolytic enzymes that do not display any allosteric feature in their functioning can be investigated using the same approach to identify potential allosteric sites that might be used to regulate the enzymatic activity of these enzymes. As our current strategy is solely based on the intrinsic nature of allostery supposedly owned by all proteins, there is always a likelihood of encountering a novel allosteric site that will be proposed as a target region for developing effective allosteric drugs.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work has been partially supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK, Project # 218M320). MA acknowledges Kadir has University for her MS. Scholarship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00088/full#supplementary-material

REFERENCES

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80, 505–515. doi: 10.1016/S0006-3495(01)76033-X

Axerio-Cilies, P., See, R. H., Zoraghi, R., Worral, L., Lian, T., Stoynov, N., et al. (2012). Cheminformatics-driven discovery of selective, nanomolar inhibitors for staphylococcal pyruvate kinase. *ACS Chem. Biol.* 7, 350–359. doi: 10.1021/cb2003576

Bahar, I., and Rader, A. J. (2005). Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* 15, 586–592. doi: 10.1016/j.sbi.2005.08.007

Barnett, J. A. (2003). A history of research on yeasts 5: the fermentation pathway. *Yeast* 20, 509–543. doi: 10.1002/yea.986

Bogan, A. A., and Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 280, 1–9. doi: 10.1006/jmbi.1998.1843

Brenke, R., Kozakov, D., Chuang, G. Y., Beglov, D., Hall, D., Landon, M. R., et al. (2009). Fragment-based identification of druggable 'hot spots' of proteins using fourier domain correlation techniques. *Bioinformatics* 25, 621–627. doi: 10.1093/bioinformatics/btp036

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M., et al. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4, 187–217. doi: 10.1002/jcc.540040211

Christopoulos, A. (2002). Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat. Rev. Drug Discov.* 1, 198–210. doi: 10.1038/nrd746

Ciulli, A., Williams, G., Smith, A. G., Blundell, T. L., and Abell, C. (2006). Probing hot spots at protein-ligand binding sites: a fragment-based approach using biophysical methods. *J. Med. Chem.* 49, 4992–5000. doi: 10.1021/jm060490r

Clackson, T., and Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science* 267, 383–386. doi: 10.1126/science.7529940

DeLano, W. L. (2002). Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* 12, 14–20. doi: 10.1016/S0959-440X(02)00283-X

Dilcan, G., Doruker, P., and Akten, E. D. (2019). Ligand-binding affinity of alternative conformers of human B2-adrenergic receptor in the presence of intracellular loop 3 (ICL3) and their potential use in virtual screening studies. *Chem. Biol. Drug Des.* 93, 883–899. doi: 10.1111/cbdd.13478

Doruker, P., Atilgan, A. R., and Bahar, I. (2000). Dynamics of proteins predicted by molecular simulations and analytical approaches: application to α-amylase inhibitor. *Proteins Struct. Funct. Genet.* 40, 512–524.

Dubey, R., Staker, B. L., Foe, I. T., Bogyo, M., Myler, P. J., Ngô, H. M., et al. (2017). Membrane skeletal association and post-translational allosteric regulation of *Toxoplasma gondii* GAPDH1. *Mol. Microbiol.* 103, 618–634. doi: 10.1111/mmi.13577

Elber, R., and Karplus, M. (1987). Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science* 235, 318–321. doi: 10.1126/science.3798113

Ellis, J. (1998). Allosteric binding sites on muscarinic receptors. *Drug Dev. Res.* 40, 193–204. doi: 10.1124/mol.105.019141

Falcon, C. M., and Matthews, K. S. (2001). Engineered disulfide linking the hinge regions within lactose repressor dimer increases operator affinity, decreases sequence selectivity, and alters allostery. *Biochemistry* 40, 15650–15659. doi: 10.1021/bi0114067

Flechsig, H. (2017). Design of elastic networks with evolutionary optimized long-range communication as mechanical models of allosteric proteins. *Biophys. J.* 113, 558–571. doi: 10.1016/j.bpj.2017.06.043

Guarnera, E., and Berezovsky, I. N. (2016a). Allosteric sites: remote control in regulation of protein activity. *Curr. Opin. Struct. Biol.* 37, 1–8. doi: 10.1016/j.sbi.2015.10.004

Guarnera, E., and Berezovsky, I. N. (2016b). Structure-based statistical mechanical model accounts for the causality and energetics of allosteric communication. *PLoS Comput. Biol.* 12:e1004678. doi: 10.1371/journal.pcbi.1004678

Guarnera, E., and Berezovsky, I. N. (2019). On the perturbation nature of allostery: sites, mutations, and signal modulation. *Curr. Opin. Struct. Biol.* 56, 18–27. doi: 10.1016/j.sbi.2018.10.008

Guarnera, E., Tan, Z. W., Zheng, Z., and Berezovsky, I. N. (2017). AlloSigMA: allosteric signaling and mutation analysis server. *Bioinformatics* 33, 3996–3998. doi: 10.1093/bioinformatics/btx430

Guido, R., Balliano, T., Andricopulo, A., and Oliva, G. (2009). Kinetic and crystallographic studies on glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma cruzi* in complex with iodoacetate. *Lett. Drug Des. Discov.* 6, 210–214. doi: 10.2174/157018009787847774

Gunasekaran, K., Ma, B., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? *Proteins Struct. Funct. Genet.* 57, 433–443. doi: 10.1002/prot.20232

Haliloglu, T., Bahar, I., and Erman, B. (1997). Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79, 3090–3093. doi: 10.1103/PhysRevLett.79.3090

Hall, D. R., Kozakov, D., Whitty, A., and Vajda, S. (2015). Lessons from hot spot analysis for fragment-based drug discovery. *Trends Pharmacol. Sci.* 36, 724–736. doi: 10.1016/j.tips.2015.08.003

Hammes-Schiffer, S., and Benkovic, S. J. (2006). Relating protein motion to catalysis. *Annu. Rev. Biochem.* 75, 519–541. doi: 10.1146/annurev.biochem.75.103004.142800

Hawkins, R. J., and McLeish, T. C. B. (2006). Coupling of global and local vibrational modes in dynamic allostery of proteins. *Biophys. J.* 91, 2055–2062. doi: 10.1529/biophysj.106.082180

Henikoff, S., and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* 89, 10915–10919. doi: 10.1073/pnas.89.22.10915

Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006). HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. *J. Am. Chem. Soc.* 128, 2812–2813. doi: 10.1021/ja058211x

Kaynak, B. T., Findik, D., and Doruker, P. (2018). RESPEC incorporates residue specificity and the ligand effect into the elastic network model. *J. Phys. Chem. B* 122, 5347–5355. doi: 10.1021/acs.jpcb.7b10325

Kloos, M., Brüser, A., Kirchberger, J., Schöneberg, T., and Sträter, N. (2015). Crystal structure of human platelet phosphofructokinase-1 locked in an activated conformation. *Biochem. J.* 469, 421–432. doi: 10.1042/BJ20150251

Koshland, D. E., and Hamadani, K. (2002). Proteomics and models for enzyme cooperativity. *J. Biol. Chem.* 277, 46841–46844. doi: 10.1074/jbc.R200014200

Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., et al. (2015). The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* 10, 733–755. doi: 10.1038/nprot.2015.043

Kung, C., Hixon, J., Choe, S., Marks, K., Gross, S., Murphy, E., et al. (2012). Small molecule activation of Pkm2 in cancer cells induces serine auxotrophy. *Chem. Biol.* 19, 1187–1198. doi: 10.1016/j.chembiol.2012.07.021

Kurkcuoglu, Z., Findik, D., Akten, E. D., and Doruker, P. (2015). How an inhibitor bound to subunit interface alters triosephosphate isomerase dynamics. *Biophys. J.* 109, 1169–1178. doi: 10.1016/j.bpj.2015.06.031

Kurkcuoglu, Z., Ural, G., Akten, E. E. D., and Doruker, P. (2011). Blind dockings of benzothiazoles to multiple receptor conformations of triosephosphate isomerase from *Trypanosoma cruzi* and human. *Mol. Inform.* 30, 986–995. doi: 10.1002/minf.201100109

Kurt, N., Scott, W. R. P., Schiffer, C. A., and Haliloglu, T. (2003). Cooperative fluctuations of unliganded and substrate-bound HIV-1 protease: a structure-based analysis on a variety of conformations from crystallography and molecular dynamics simulations. *Proteins Struct. Funct. Genet.* 51, 409–422. doi: 10.1002/prot.10350

Levy, E. D. (2010). A simple definition of structural regions in proteins and its use in analyzing interface evolution. *J. Mol. Biol.* 403, 660–670. doi: 10.1016/j.jmb.2010.09.028

Liu, T., Whitten, S. T., and Hilser, V. J. (2007). Functional residues serve a dominant role in mediating the cooperativity of the protein ensemble. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4347–4352. doi: 10.1073/pnas.0607132104

Lockless, S. W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286, 295–299. doi: 10.1126/science.286.5438.295

Lou, H., and Cukier, R. I. (2006). Molecular dynamics of apo-adenylate kinase: a distance replica exchange method for the free energy of conformational fluctuations. *J. Phys. Chem. B* 110, 24121–24137. doi: 10.1021/jp064303c

McNae, I. W., Martinez-Oyanedel, J., Keillor, J. W., Michels, P. A. M., Fothergill-Gilmore, L. A., and Walkinshaw, M. D. (2009). The crystal structure of ATP-bound phosphofructokinase from trypanosoma brucei reveals conformational transitions different from those of other phosphofructokinases. *J. Mol. Biol.* 385, 1519–1533. doi: 10.1016/j.jmb.2008.11.047

Metz, A., Pfleger, C., Kopitz, H., Pfeiffer-Marek, S., Baringhaus, K. H., and Gohlke, H. (2012). Hot spots and transient pockets: predicting the determinants of small-molecule binding to a protein-protein interface. *J. Chem. Inform. Model.* 52, 120–133. doi: 10.1021/ci200322s

Meyerhof, O., and Junowicz-Kocholaty, R. (1943). The equilibria of isomerase and aldolase, and the problem of the phosphorylation of glyceraldehyde phosphate. *J. Biol. Chem.* 149, 71–92.

Ming, D., and Wall, M. E. (2005). Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.* 96:159902. doi: 10.1103/PhysRevLett.95.198103

Mitternacht, S., and Berezovsky, I. N. (2011). Binding leverage as a molecular basis for allosteric regulation. *PLoS Comput. Biol.* 7:e1002148. doi: 10.1371/journal.pcbi.1002148

Monod, J., Changeux, J. P., and Jacob, F. (1963). Allosteric proteins and cellular control systems. *J. Mol. Biol.* 6, 306–329. doi: 10.1016/s0022-2836(63)80091-1

Monod, J., Wyman, J., and Changeux, J. P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12, 88–118. doi: 10.1016/s0022-2836(65)80285-6

Morgan, H. P., McNae, I. W., Nowicki, M. W., Hannaert, V., Michels, P. A. M., Fothergill-Gilmore, L. A., et al. (2010). Allosteric mechanism of pyruvate kinase from *Leishmania mexicana* uses a rock and lock model. *J. Biol. Chem.* 285, 12892–12898. doi: 10.1074/jbc.M109.079905

Mukherjee, S., Dutta, D., Saha, B., and Das, A. (2010). Crystal structure of glyceraldehyde-3-phosphate dehydrogenase 1 from methicillin-resistant *Staphylococcus aureus* MRSA252 provides novel insights into substrate binding and catalytic mechanism. *J. Mol. Biol.* 401, 949–968. doi: 10.1016/j.jmb.2010.07.002

Needleman, S. B., and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443–453. doi: 10.1016/0022-2836(70)90057-4

Pan, H., Lee, J. C., and Hilser, V. J. (2000). Binding sites in *Escherichia coli* dihydrofolate reductase communicate by modulating the conformational

ensemble. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12020–12025. doi: 10.1073/pnas.220240297

Perutz, M. F. (1989). Mechanisms of cooperativity and allosteric regulation in proteins. *Q. Rev. Biophys.* 22, 139–237. doi: 10.1017/s0033583500003826

Pommier, Y., and Cherfils, J. (2005). Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery. *Trends Pharmacol. Sci.* 26, 138–145. doi: 10.1016/j.tips.2005.01.008

Rahimova, R., Fontanel, S., Lionne, C., Jordheim, L. P., Peyrottes, S., and Chaloin, L. (2018). Identification of allosteric inhibitors of the ecto-5'-nucleotidase (CD73) targeting the dimer interface. *PLoS Comput. Biol.* 14:e1005943. doi: 10.1371/journal.pcbi.1005943

Rice, P., Longden, L., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1093/bib/3.1.92

Rigden, D. J., Phillips, S. E. V., Michels, P. A. M., and Fothergill-Gilmore, L. A. (1999). The structure of pyruvate kinase from *Leishmania mexicana* reveals details of the allosteric transition and unusual effector specificity. *J. Mol. Biol.* 291, 615–635. doi: 10.1006/jmbi.1999.2918

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316

Santamaría, B., Estévez, A. M., Martínez-Costa, O. H., and Aragón, J. J. (2002). Creation of an allosteric phosphofructokinase starting with a nonallosteric enzyme: the case of dictyostelium discoideum phosphofructokinase. *J. Biol. Chem.* 277, 1210–1216. doi: 10.1074/jbc.M109480200

Schirmer, T., and Evans, P. R. (1990). Structural basis of the allosteric behaviour of phosphofructokinase. *Nature* 343, 140–145. doi: 10.1038/343140a0

Schrödinger, L. L. C. (2015). *The PyMol Molecular Graphics System, Versión 1.8.*

Süel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* 10, 59–69. doi: 10.1038/nsb881

Tama, F., and Brooks, C. L. (2006). Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu. Rev. Biophys. Biomol. Struct.* 35, 115–133. doi: 10.1146/annurev.biophys.35.040405.102010

Temiz, N. A., and Bahar, I. (2002). Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. *Proteins Struct. Funct. Genet.* 49, 61–70. doi: 10.1002/prot.10183

Tian, T., Wang, C., Wu, M., Zhang, X., and Zang, J. (2018). Structural insights into the regulation of *Staphylococcus aureus* phosphofructokinase by tetramer-dimer conversion. *Biochemistry* 57, 4252–4262. doi: 10.1021/acs.biochem.8b00028

Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J., and Wilke, C. O. (2013). Maximum allowed solvent accessibilites of residues in proteins. *PLoS One* 8:e80635. doi: 10.1371/journal.pone.0080635

Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77, 1905–1908. doi: 10.1103/PhysRevLett.77.1905

Tobi, D., and Bahar, I. (2005). Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18908–18913. doi: 10.1073/pnas.0507603102

Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., and Rarey, M. (2012a). Combining global and local measures for structure-based druggability predictions. *J. Chem. Inform. Model.* 52, 360–372. doi: 10.1021/ci200454v

Volkamer, A., Kuhn, D., Rippmann, F., and Rarey, M. (2012b). Dogsitescorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* 28, 2074–2075. doi: 10.1093/bioinformatics/bts310

Wang, X., and Kemp, R. G. (2001). Reaction path of phosphofructo-1-kinase is altered by mutagenesis and alternative substrates. *Biochemistry* 40, 3938–3942. doi: 10.1021/bi002709o

Weber, G. (1972). Ligand binding and internal equilibria in proteins. *Biochemistry* 11, 864–878. doi: 10.1021/bi00755a028

White, M. R., Khan, M. M., Deredge, D., Ross, C. R., Quintyn, R., Zucconi, B. E., et al. (2015). A dimer interface mutation in glyceraldehyde-3-phosphate dehydrogenase regulates its binding to AU-rich RNA. *J. Biol. Chem.* 290, 1770–1785. doi: 10.1074/jbc.A114.618165

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

REVIEW
published: 09 July 2020
doi: 10.3389/fmolb.2020.00136

Allosteric Regulation at the Crossroads of New Technologies: Multiscale Modeling, Networks, and Machine Learning

Gennady M. Verkhivker[1,2]*, *Steve Agajanian*[1], *Guang Hu*[3] *and Peng Tao*[4]

[1] *Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA, United States,* [2] *Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, CA, United States,* [3] *Center for Systems Biology, Department of Bioinformatics, School of Biology and Basic Medical Sciences, Soochow University, Suzhou, China,* [4] *Department of Chemistry, Center for Drug Discovery, Design, and Delivery (CD4), Center for Scientific Computation, Southern Methodist University, Dallas, TX, United States*

Allosteric regulation is a common mechanism employed by complex biomolecular systems for regulation of activity and adaptability in the cellular environment, serving as an effective molecular tool for cellular communication. As an intrinsic but elusive property, allostery is a ubiquitous phenomenon where binding or disturbing of a distal site in a protein can functionally control its activity and is considered as the "second secret of life." The fundamental biological importance and complexity of these processes require a multi-faceted platform of synergistically integrated approaches for prediction and characterization of allosteric functional states, atomistic reconstruction of allosteric regulatory mechanisms and discovery of allosteric modulators. The unifying theme and overarching goal of allosteric regulation studies in recent years have been integration between emerging experiment and computational approaches and technologies to advance quantitative characterization of allosteric mechanisms in proteins. Despite significant advances, the quantitative characterization and reliable prediction of functional allosteric states, interactions, and mechanisms continue to present highly challenging problems in the field. In this review, we discuss simulation-based multiscale approaches, experiment-informed Markovian models, and network modeling of allostery and information-theoretical approaches that can describe the thermodynamics and hierarchy allosteric states and the molecular basis of allosteric mechanisms. The wealth of structural and functional information along with diversity and complexity of allosteric mechanisms in therapeutically important protein families have provided a well-suited platform for development of data-driven research strategies. Data-centric integration of chemistry, biology and computer science using artificial intelligence technologies has gained a significant momentum and at the forefront of many cross-disciplinary efforts. We discuss new developments in the machine learning field and the emergence of deep learning and deep reinforcement learning applications in modeling of molecular mechanisms and allosteric proteins. The experiment-guided integrated approaches empowered by recent advances in multiscale modeling, network science, and machine

OPEN ACCESS

Edited by:
Shozeb Haider,
UCL School of Pharmacy,
United Kingdom

Reviewed by:
Sarath Chandra Dantu,
Brunel University London,
United Kingdom
Serena H. Chen,
Oak Ridge National Laboratory (DOE),
United States

****Correspondence:***
Gennady M. Verkhivker
verkhivk@chapman.edu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *01 May 2020*
Accepted: *08 June 2020*
Published: *09 July 2020*

Citation:
Verkhivker GM, Agajanian S, Hu G and Tao P (2020) Allosteric Regulation at the Crossroads of New Technologies: Multiscale Modeling, Networks, and Machine Learning. Front. Mol. Biosci. 7:136. doi: 10.3389/fmolb.2020.00136

learning can lead to more reliable prediction of allosteric regulatory mechanisms and discovery of allosteric modulators for therapeutically important protein targets.

Keywords: allosteric regulation, multiscale modeling, Markov state models, network analysis, deep learning, reinforcement learning, drug discovery

INTRODUCTION

Allosteric regulation is an efficient and robust mechanism for molecular communication and signaling in the cell employed by proteins for regulation of activity and adaptability during processes of signal transduction, catalysis, and gene regulation (Monod et al., 1965; Koshland, 1998; Changeux and Edelstein, 2005; Popovych et al., 2006; Changeux, 2012). The recent breakthroughs in nuclear magnetic resonance (NMR) technologies have enabled dynamic studies of large biomolecules at atomic resolution, and are now frequently employed as powerful diagnostic tools of allosteric communications in proteins (Boehr et al., 2006; Jarymowycz and Stone, 2006; Mittermaier and Kay, 2006, 2009; Sprangers et al., 2007; Korzhnev and Kay, 2008; Kalodimos, 2011; Kay, 2011, 2016; Rosenzweig and Kay, 2014; Lisi and Loria, 2016, 2017; Huang and Kalodimos, 2017; Jiang and Kalodimos, 2017). Allosteric molecular events can involve complex cascades of thermodynamic and rapid dynamic changes that occur on different spatial and temporal scales. The thermodynamic-centric energy landscape concepts and conformational selection models of allosteric regulation have gained a considerable prominence in recent years, rooted in the assumption that statistical ensembles of preexisting conformational states and communication pathways are intrinsic to a given protein system (Astl et al., 2019) and allow for modulation and redistribution induced by external perturbations, ligand binding, and mutations (Gunasekaran et al., 2004; Tsai et al., 2008, 2009; del Sol et al., 2009; Csermely et al., 2010, Zhuravlev and Papoian, 2010; Ma et al., 2011; Wrabl et al., 2011; Hilser et al., 2012; Nussinov, 2012; Motlagh et al., 2014; Tsai and Nussinov, 2014; Nussinov and Tsai, 2015; Guo and Zhou, 2016; Liu and Nussinov, 2016; Astl et al., 2019). Conformational dynamics redistributions in the absence of appreciable structural transformations are the hallmark of the "entropy-driven" allosteric mechanisms in which allosteric interactions can be mediated through alterations of functional motions and rebalancing of rigid and flexible protein regions (Cooper and Dryden, 1984; Stevens et al., 2001; Dam et al., 2002; Kern and Zuiderweg, 2003; Frederick et al., 2007; Tzeng and Kalodimos, 2009, Nesmelova et al., 2010; Kalodimos, 2011, 2012; McLeish et al., 2013; Li et al., 2014; Buchenberg et al., 2017; Stock and Hamm, 2018; Wodak et al., 2019). The quantitative elucidation of these highly dynamic and often elusive processes continues to present formidable technical and conceptual challenges. Despite significant advances, the quantitative characterization and prediction of functional allosteric states, interactions and mechanisms continue to present highly challenging problems in the field. The fundamental biological importance and complexity of these processes require innovative computational and experimental approaches that can advance current understanding of allosteric regulatory processes. A systematic interdisciplinary effort is needed to leverage the burgeoning knowledge about allosterically regulated proteins to develop robust experiment-informed computational tools for atomistic prediction of allosteric mechanisms. In this review we discuss and analyze how recent advances in biophysical simulations and network science can be integrated with NMR spectroscopy experiments and leverage the rising power of machine learning (ML) approaches to enable the reliable quantitative characterization of allosteric regulation mechanisms and facilitate allosteric drug discovery. We discuss in details computational strategies that leverage biophysical and network-based modeling with NMR experiments for characterization and probing of allosteric regulatory mechanisms. The review also critically discusses advantages and limitations of emerging approaches including Markovian modeling and the information-theoretical analysis of dynamic flows in allosteric networks in addressing present challenges and open questions of allosteric regulation mechanisms.

NETWORK-BASED APPROACHES IN STUDIES OF ALLOSTERIC REGULATION MECHANISMS

It has been recognized that allosteric regulation is a global property of protein systems that can be described by the residue interaction networks in which the effector binding initiates a cascade of coupled fluctuations propagating through the network and eliciting long-range functional responses at distal sites (Atilgan et al., 2004; Brinda and Vishveshwara, 2005, 2010; del Sol and O'Meara, 2005; Bode et al., 2007; Sethi et al., 2009; Vijayabaskar and Vishveshwara, 2010; Csermely et al., 2013; Di Paola and Giuliani, 2015; Dokholyan, 2016). The graph-based network approaches have offered a simple and effective formalism for describing allosteric interactions, where the dynamic fluctuations are mapped onto a graph with nodes representing residues and edges representing weights of the measured dynamic properties. The network-centric methods have represented a powerful complementary strategy to physics-based landscape models of protein dynamics by quantifying global functional changes (Vendruscolo et al., 2002; Atilgan et al., 2004; Brinda and Vishveshwara, 2005, 2010; Ghosh and Vishveshwara, 2007, 2008; Hansia et al., 2009; Bhattacharyya and Vishveshwara, 2011; Ghosh et al., 2011; Csermely et al., 2012; Gasper et al., 2012; Bhattacharya and Vaidehi, 2014; General et al., 2014; Dokholyan, 2016; Adhireksan et al., 2017), identifying key functional centers and allosteric communication pathways (Verkhivker et al., 2002; del Sol and O'Meara, 2005; del Sol et al., 2006; Sethi et al., 2009, 2013; Vijayabaskar and

Vishveshwara, 2010; Rivalta et al., 2012; Vanwart et al., 2012; Farabella et al., 2014; Di Paola and Giuliani, 2015; Kalescky et al., 2015, 2016; Hertig et al., 2016; Ricci et al., 2016; Stolzenberg et al., 2016; Palermo et al., 2017; Zhou et al., 2017, 2019a,b; Liang et al., 2019; Li et al., 2019). Recent years have witnessed the proliferation of numerous computational tools for predicting allosteric pathways and communications in proteins (Ming and Wall, 2005, 2006; McClendon et al., 2009; Tehver et al., 2009; Mitternacht and Berezovsky, 2011; Bowman and Geissler, 2012; Panjkovich and Daura, 2012, 2014; Goncearenco et al., 2013; Kaya et al., 2013; Stetz and Verkhivker, 2017). The network studies have also suggested that rapid signal transmission of allosteric interactions through small-world networks encoded in protein folds may be a universal signature encoded in protein families (Tsai et al., 2009; Di Paola and Giuliani, 2015). Significant bodies of computational and experimental studies have shown that integration of network-based approaches with structural and biochemical studies can provide a robust platform for further exploration and atomistic characterization of allosteric states and regulatory mechanisms controlled by allostery.

Functional residues in residue networks are often connected via strong evolutionary relationships (Lockless and Ranganathan, 1999; Suel et al., 2003; Halabi et al., 2009; Aguilar et al., 2012; McLaughlin et al., 2012; Simonetti et al., 2013). Coevolution of protein residues can reflect correlated functional dynamics of these sites in mediating residue-residue contacts (Socolich et al., 2005), protein folding transitions (Morcos et al., 2011), and allosteric signaling in protein complexes (Wang et al., 2019). Coevolving residues could also form direct communication paths in the interaction networks with connections weighted according to dynamic couplings and coevolutionary interaction strengths between nodes (Chakrabarti and Panchenko, 2009, 2010; Nishi et al., 2011). Dynamic and coevolutionary residue correlations may also act as synchronizing forces that determine modular organization of allosteric interaction networks and enable efficient allosteric regulation (Stetz and Verkhivker, 2017). These results have motivated the development of novel community-based methods for modeling ensembles of allosteric communication pathways in protein structures (Tse and Verkhivker, 2015a,b; Verkhivker et al., 2016; Stetz and Verkhivker, 2017). Using this computational framework, it was found that efficient allosteric communications in various signaling proteins could be controlled by structurally stable functional centers that exploit dynamically coupled residues in their local communities to propagate cooperative structural changes. The important revelation of these studies was that dynamic and evolutionary residue correlations may act as synchronizing forces to enable efficient and robust allosteric regulation.

Examining proteins as dynamic regulatory machineries that fluctuate between functional allosteric states and modulated by ligand binding or mutations is critical to understanding the molecular principles of signaling in the cell. Computational studies of allosteric regulation in signaling proteins have led to important mechanistic insights, better atomistic understanding of complex regulatory processes and continuous integration with structural and functional experiments. A variety of computational approaches have been extensively explored in investigations of allosteric mechanisms in protein kinases. These studies included experiment-guided structural modeling and protein folding analysis (Levinson et al., 2006; Zhang et al., 2006; Kornev et al., 2008; Han et al., 2011; Jura et al., 2011; Shan et al., 2011, 2012, 2013; Taylor and Kornev, 2011; Tzeng and Kalodimos, 2011; Levinson and Boxer, 2012, 2014; Taylor et al., 2012a,b; Meharena et al., 2013; Shaw et al., 2014; Shukla et al., 2014; Kornev and Taylor, 2015; Schulze et al., 2016; Narayanan et al., 2017; Levinson, 2018; Ruff et al., 2018), molecular simulations and free energy computations (Yang and Roux, 2008; Dixit and Verkhivker, 2009, 2011a,b; Yang et al., 2009; Arkhipov et al., 2013; Lin and Roux, 2013; Lin et al., 2013, 2014; Dixit and Verkhivker, 2014; Meng and Roux, 2014; Fajer et al., 2017; Kim et al., 2017; Meng et al., 2017), and network modeling (James and Verkhivker, 2014; Tse and Verkhivker, 2015a,b,c; Czemeres et al., 2017; Stetz et al., 2017; Astl and Verkhivker, 2019a,b). By examining residue interaction networks in protein kinases a unifying mechanistic model of allosteric coupling between the ATP-binding and substrate binding sites conserved among kinases was proposed (Tse and Verkhivker, 2015a,b,c; Stetz et al., 2017). A theoretical framework for rationalizing binding preferences of the kinase inhibitors was developed establishing the relationships between ligand binding and modulation of the residue interaction networks (Tse and Verkhivker, 2015a,b,c). Atomistic modeling of the ABL kinase regulation using a combination molecular dynamics (MD) simulations, structural perturbation methods and network-centric analysis (Astl and Verkhivker, 2019a,b) has provided evidence of allosteric interactions and communication pathways in the ABL interaction networks that supported and explained the underlying mechanisms proposed in the pioneering NMR studies (Saleh et al., 2017).

Computational studies of allosteric regulation in molecular chaperones Hsp90 and Hsp70 have also been instrumental to the progress in the field by complementing biochemical experiments and providing a detailed dynamic view of the functional cycle and mechanisms (Colombo et al., 2008; Morra et al., 2009, Verkhivker et al., 2009; Morra et al., 2010, 2012; Matts et al., 2011a,b; Chiappori et al., 2012, 2016; Dixit and Verkhivker, 2012; Lawless et al., 2013; Verkhivker, 2014, 2018a,b; Paladino et al., 2015; Stetz and Verkhivker, 2015, 2016, 2017, 2018; Czemeres et al., 2017; Stetz et al., 2017). Using a network-based formalism of allostery, computational studies have captured NMR-observed direction-specific nature of signal propagation pathways in the Hsp70 chaperone (Stetz and Verkhivker, 2015, 2017).

Studies of allosteric mechanisms have indicated that integration of experiment-informed molecular simulations with network-based formalisms of allostery may provide a convenient and powerful platform for atomistic characterization of allosteric states and regulatory mechanisms. The lessons from studies of signaling proteins including protein kinases and molecular chaperones have suggested that allosteric regulation mechanisms can proceed via a non-trivial and often elusive combination of the three classical models of allostery: induced fit, conformational selection, and dynamic allostery. Computational modeling and atomistic simulations of protein systems and functional assemblies have shown that allosteric mechanisms may not necessarily imply a simple switching between the crystal

structures of the inactive and active states, but often represent a complex regulatory machinery in which binding and external perturbations could give rise to a spectrum of functionally relevant and yet often hidden allosteric conformations exhibiting a range of activity levels.

ALLOSTERIC REGULATION AND DETECTING ALLOSTERIC STATES THROUGH INTEGRATION OF NMR EXPERIMENTS AND COMPUTATIONAL MODELING

The growing number of high-resolution crystal structures and wealth of structural information about protein systems have had an enormous impact on computational and simulation approaches, facilitating development of knowledge-based methods and advanced sampling techniques. However, allosteric functional states in proteins are often highly dynamic and short-lived representing low populated, high energy states that are rarely directly observed in X-ray crystallography experiment. A large amount of conformational sampling is typically needed to uncover and isolate high-energy functional states simulations. For instance, cryptic (or hidden) allosteric sites sporadically appear during conformational transitions of a protein in the presence of a bound ligand. These hidden allosteric sites are invisible in crystal structures and can be detected due to the stabilization of the low-populated, higher-energy conformation by certain compounds. Even with the advanced sampling techniques and enormous computer power that is now available, the experimental validation and confirmation of allosteric states represents the key component to ensure robust quantitative modeling and analysis of allosteric mechanisms.

NMR spectroscopy is a powerful method for studying protein dynamics and allosteric mechanisms by probing multiple time scales and detecting residue-specific conformational changes associated with ligand binding (Boehr et al., 2006; Jarymowycz and Stone, 2006; Mittermaier and Kay, 2006, 2009; Sprangers et al., 2007; Korzhnev and Kay, 2008; Kalodimos, 2011; Kay, 2011, 2016; Rosenzweig and Kay, 2014; Lisi and Loria, 2016, 2017; Huang and Kalodimos, 2017; Jiang and Kalodimos, 2017). The micro- to milli-second time scale protein motions measured in relaxation-dispersion experiments can provide information about the distribution of conformational states and thermodynamics and kinetics of allosteric protein changes. Protein dynamics can also be investigated by NMR methods other than traditional relaxation experiments. Residual dipolar couplings are sensitive to motions occurring across a vast time scale, ranging from seconds to faster than nanoseconds. Conformational changes in isotopically labeled proteins upon ligand binding can be detected by two-dimensional heteronuclear single quantum coherence (HSQC) spectroscopy for large protein systems (Sprangers et al., 2007; Korzhnev and Kay, 2008). Chemical shift mapping of protein residues upon ligand binding provides a specific and precise fingerprint of allosteric propagation effects that allows to detect site-specific binding responses, characterize pathways of allosteric communication and differentiate between competitive and allosteric inhibitor binding (Grutsch et al., 2016; Berjanskii and Wishart, 2017; Krivdin, 2017; Nerli et al., 2018). The NMR technologies have enabled structural studies of conformational dynamic processes at atomic resolution and are used to identify coupled networks and communication pathways in allosteric proteins (Swain and Gierasch, 2006; Smock and Gierasch, 2009; Shi and Kay, 2014; Grutsch et al., 2016). Relaxation dispersion NMR methods have enabled detection and characterization of rare and energetically excited conformational states that play significant role in dynamic activation of protein function and allosteric mechanisms (Vallurupalli et al., 2012; Kalbitzer et al., 2013; Munte et al., 2013; Sekhar and Kay, 2013, 2019; Williamson and Kitahara, 2019). Characterization of low-lying excited states of proteins by high-pressure NMR under equilibrium conditions can allow for detection of reversible transitions that are functionally relevant, providing means for probing dynamic energy landscapes of allosteric mechanisms (Kalbitzer et al., 2013; Williamson and Kitahara, 2019). High-pressure NMR can help to identify these conformations, including low populated functional states, and characterize their energies and kinetics of conformational changes (Williamson and Kitahara, 2019). By measuring redistributions in the conformational entropy, pressure-dependent chemical shifts can help to sequester low-populated functional states (Kalbitzer et al., 2013; Munte et al., 2013; Williamson and Kitahara, 2019).

Recent years have witnessed the development of various approaches that investigate NMR chemical shift perturbations to identify potential allosteric networks and structural dynamics in proteins (Selvaratnam et al., 2011, 2012; Robustelli et al., 2012; Cembran et al., 2014). NMR chemical exchange saturation transfer (CEST) experiments can provide adequate characterization of slower exchange processes, identify invisible states, and slow conformational exchange (Long et al., 2014; Anthis and Clore, 2015; Yuwen et al., 2017). NMR chemical shift covariance (CHESCA) and projection (CHESPA) analyses can identify blocks of dynamically coupled residues collectively forming allosteric interaction networks (Selvaratnam et al., 2011, 2012; Boulton et al., 2014, 2018; Boulton and Melacini, 2016). Allosteric proteins subjected to specific perturbations (ligand binding, mutations) cause residues that belong to the same effector-dependent allosteric network to exhibit a concerted response signal. CHESCA approach can detect patterns of correlated changes in the chemical shifts between apo and holo states due to perturbations and isolate allosterically coupled regions (**Figure 1**). This method is particularly effective in detecting allosteric networks within dynamic and partially unstructured regions (Boulton and Melacini, 2016; Boulton et al., 2018). NMR chemical shift perturbations have been recently used in combination with Markov model network analysis to reveal the dynamic flow of communication between allosteric communities in the protein kinases (Aoto et al., 2016). NMR-guided computational modeling can leverage CHESCA approach for computation of the chemical shift correlation matrices in the known allosteric states obtained using crystal structures of complexes with allosteric ligands. The experimental NMR chemical shifts can guide molecular

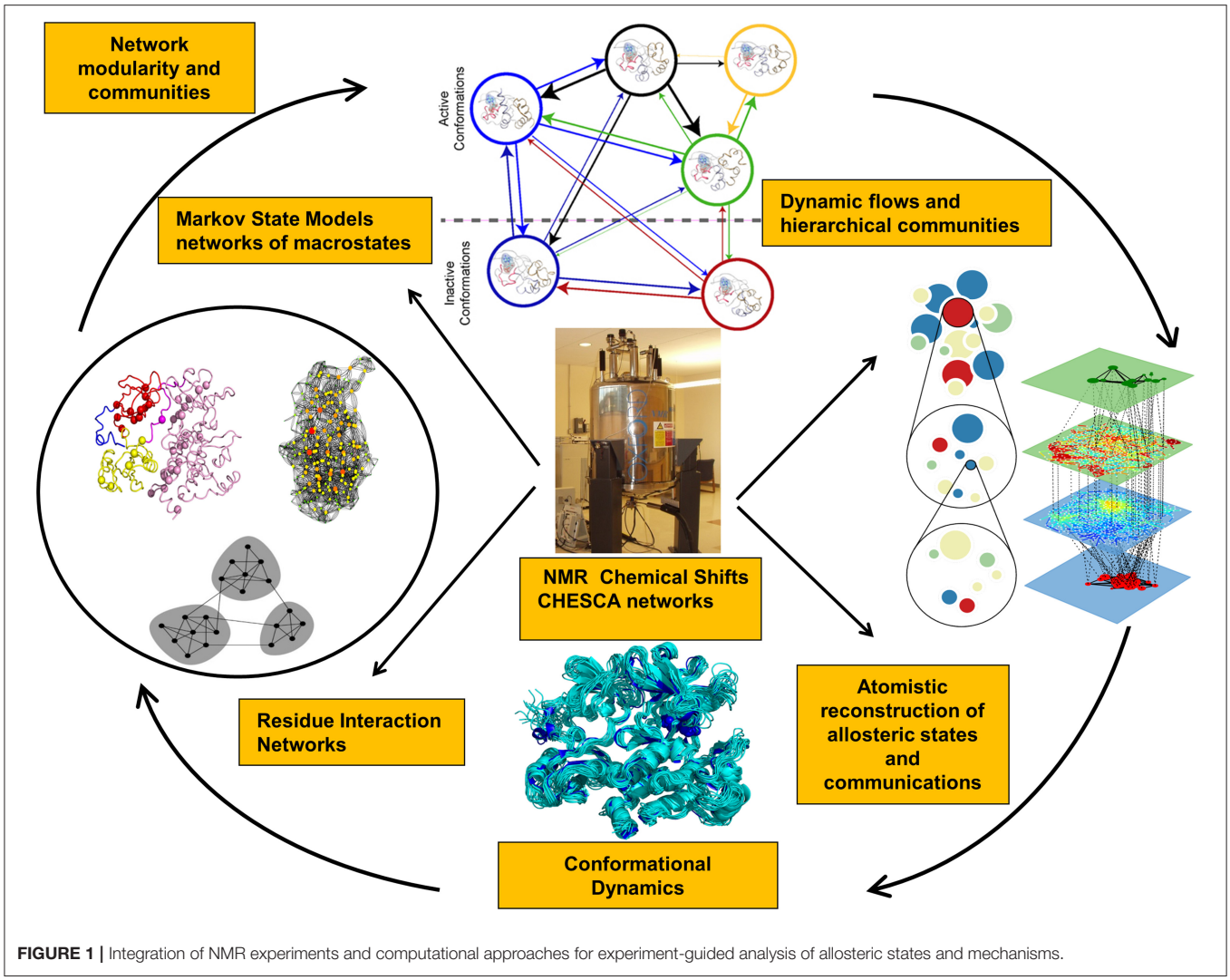


FIGURE 1 | Integration of NMR experiments and computational approaches for experiment-guided analysis of allosteric states and mechanisms.

simulations and network analysis by reporting on blocks of dynamically coupled residues forming allosteric interaction networks. Through integration of these experimental data into accelerated atomistic simulations, a more detailed mapping of the functional landscapes and relevant allosteric states can be achieved.

Protein systems can be efficiently simulated on longer time scales by accelerated meta-dynamics approaches (Limongelli et al., 2013; Palazzesi et al., 2013, 2017; Sutto and Gervasio, 2013; Bonomi and Camilloni, 2017; Kuzmanic et al., 2017; Yang et al., 2018; Brotzakis and Parrinello, 2019) where the experimental and computed NMR chemical shifts (Shen and Bax, 2010; Han et al., 2011) are often used to determine collective variables (Granata et al., 2013; Xia et al., 2013; Palazzesi et al., 2017). NMR chemical shifts can be also evaluated using structure-based CamShift approach (Kohlhoff et al., 2009) with collective variables defined as the difference between experimental and calculated chemical shifts. These NMR-guided simulation techniques have enabled adequate sampling of the conformational space and robust structure reconstruction using experimental constraints (Robustelli et al., 2010; Cavalli et al., 2011; Granata et al., 2013). NMR chemical shift observables can be also used in combination with other collective variables in meta-dynamics simulations to guide the efficient exploration of allosteric states and functional transitions (Kimanius et al., 2015; Ansari et al., 2016).

A combination of powerful and expensive NMR spectroscopy equipment, biophysical techniques and protein expression platforms is often required to obtain structures of allosteric states for protein systems and experimental validation of short-lived hidden functional conformations. Despite unique abilities of NMR spectroscopy to detect highly dynamic events and examine conformational landscapes of allosteric proteins, the NMR applications for high-resolution reconstruction of allosteric states are still fairly limited owing to complexity and cost of these experiments. Hence, development of novel research strategies based on innovative integration of NMR spectroscopy and experiment-guided simulation approaches become especially

important and clearly represent the most promising avenue for further explorations going forward.

MARKOV STATE MODELS IN STUDIES OF ALLOSTERIC REGULATION

Given the complexity of thermodynamic and kinetic factors underlying allosteric regulatory events, the information-based theory of signal propagation (Chennubhotla and Bahar, 2006, 2007; Chennubhotla et al., 2008) and stochastic Markov state models (MSMs) (Prinz et al., 2011; McGibbon et al., 2014; Pande, 2014; Shukla et al., 2015, 2017; Wu et al., 2016; Husic and Pande, 2018) have become increasingly useful states-and-rates network models with the continuously developing open source software infrastructure (Cronkite-Ratcliff and Pande, 2013; Bowman, 2014; Bowman and Noe, 2014; Harrigan et al., 2017). The MSMs have been successfully adopted for describing the transitions between functional states during allosteric events (Bowman et al., 2015; Hart et al., 2016; Sengupta and Strodel, 2018). Combined with MD simulations, MSM approaches can provide connectivity maps of states on the free energy landscape, estimate the effect of allosteric perturbations on the conformational equilibrium, and rigorously describe kinetics of allosteric transitions. Recent advances in the field have highlighted how MSM tools can help to recognize structural and dynamic patterns of conformational ensembles, identify functional allosteric states hidden in the conformational ensembles and reconstruct allosteric mechanisms (Sengupta and Strodel, 2018). Markov models have been employed for understanding of the reaction mechanisms, thermodynamics and free-energy landscape population shifts, the hierarchy of timescales and the structural basis of allosteric events (Prinz et al., 2011; Pande, 2014; Shukla et al., 2015, 2017; Zhou et al., 2017, 2019a,b).

When combined with appropriate general coordinates, MSM could be a very powerful tool to reveal intrinsic states of the proteins (Malmstrom et al., 2015). The important component of the MSM approach in studies of allosteric systems is the employment of robust dimensionality reduction techniques to identify experimentally-informed collective variables that can enhance sampling and provide efficient detection and separation of functional allosteric states. Dimension reduction is often performed using time-lagged independent component analysis (TICA) (Schwantes and Pande, 2014; Perez-Hernandez and Noe, 2016; Noe and Clementi, 2017; Olsson et al., 2017). In these approaches, the simulation samples can be divided into substates assuming that conformations within each substate share kinetic similarity and could interconvert rapidly (Bowman et al., 2009; Zhou and Tao, 2018; Zhou et al., 2018a,b). t-SNE method was recently developed as a dimensionality reduction method with minimum structural information loss revealing that both one-dimensional (1D) and two-dimensional (2D) models of t-SNE method are superior to other tools in distinguishing functional states of allosteric proteins (Zhou et al., 2018a,b). MSMs and transition network models are widely applied to extract kinetic descriptors from equilibrium simulations. Directed Kinetic Transition Network (DKTN) which is a graph representation of a master equation was developed for describing non-equilibrium kinetics in allosteric proteins (Zhou et al., 2019a,b). Markov modeling studies have also examined the timescales and intramolecular pathways implicated in allostery by introducing master equation-based approach for allostery by population shift (Long and Bruschweiler, 2011). Another study employed a graph-theoretic approach and Markov stability analysis for modeling of signaling pathways and characterization of allosteric sites (Amor et al., 2014).

Current allosteric models have suggested that conformational and dynamical distribution phase space accessible for allosteric interactions in proteins is much larger than the experimentally visible landscapes provided through crystallographic and NMR experiments. As a result, external perturbations, such as mutations and/or ligand binding that could significantly affect conformational space and dynamic distribution of allosteric proteins and can be employed as probes to explore functional consequences of allosteric phenomena. The recently developed Rigid Residue Scan (RRS) simulation method has been shown as effective tool to perturb protein dynamics and assess both conformational and dynamical redistributions in allosteric systems (Kalescky et al., 2015, 2016). Using the RRS method, the predictive models for light-oxygen-voltage-sensing (LOV) domains allostery have been developed that identified the experimentally verified mutants with distinctive allosteric regulatory effects. The results of this analysis have suggested how manipulating functional regions with light in LOV proteins could link chemistry and allostery, providing a path for rational engineering of LOV ontogenetic tools.

EXPLOITING ALLOSTERIC MECHANISMS AND CRYPTIC BINDING SITES FOR DISCOVERY OF ALLOSTERIC MODULATORS

Multiscale simulations and MSM approaches have shown that allosteric mechanisms may not necessarily imply a simple two-state switch between the major inactive and active states, but often represent a dynamic multilayered regulatory machine in which binding and external perturbations could give rise to a discrete spectrum of functionally relevant and yet often hidden allosteric conformations exhibiting a range of activity levels. Experiment-informed Markovian modeling studies have shown a promise in adequately describing the hierarchy of allosteric states by recognizing structural and dynamic patterns of conformational ensembles and identifying functional allosteric states that are hidden in the crystal structures of allosteric proteins. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments has been advanced and applied for antibiotic target TEM-1 β-lactamase (Bowman et al., 2015). Bowman et al. used MSM approach of a ligand-free protein to identify allosteric sites based on several signatures of collective dynamics, namely the presence of a pocket in a significant fraction of the population and the presence of correlated motions between the newly

discovered pocket and the active site which provides means for allosteric communication between distant sites. The central to this pioneering work is a close integration with labeling experiments on TEM-1 β-lactamase that were performed to test the existence of hidden allosteric sites as feasible targets for allosteric drug design (Bowman et al., 2015). These illuminating studies have shown for the first time the power of integrated tools to identify, characterize and exploit hidden allosteric sites, highlighting the robust nature of Markov modeling tools in guiding the experiments. It has been argued that the wealth of thermodynamic, kinetic and structural data derived from MSMs can guide further development of experimental tools for discovery of hidden allosteric states and invisible cryptic allosteric binding sites.

The results suggest there are many undiscovered hidden allosteric sites that can be characterized and targeted with rational drug design (Cimermancic et al., 2016; Oleinikovas et al., 2016; Beglov et al., 2018; Kuzmanic et al., 2020). The hidden allosteric sites are invisible in crystal structures and cryptic sites can emerge as a result of stabilization of rare, high-energy states by small fragment probes. The allosteric mechanisms of cryptic site formation may involve a delicate interplay between induced-fit and conformational selection that can be modeled using elaborate replica-exchange sampling techniques (Oleinikovas et al., 2016). Collectively, experiment-informed multiscale simulation studies have shown that these tools can adequately describe complexity and stochasticity that underlies the thermodynamics and hierarchy of allosteric states and the molecular basis of allosteric mechanisms.

Recent advances in understanding allosteric regulation and activation mechanisms of therapeutic signaling proteins such as protein kinases have fueled unprecedented efforts to discover targeted allosteric inhibitors. Allosteric kinase inhibitors do not compete with ATP and could be more selective by binding to the regulatory sites outside of the ATP binding site (Dar and Shokat, 2011). Allosteric kinase inhibitors can improve target specificity and play an important role in the precision medicine initiative in oncology. NMR and X-ray crystallography studies have revealed a detailed atomistic picture of allosteric regulation in many protein kinases, showing how interacting signaling modules form a multilayered regulatory mechanism that exploits various allosteric switch points powered by binding or phosphorylation at different sites of the regulatory kinase complexes (Saleh et al., 2017). Recently discovered allosteric inhibitors of the ABL kinase GNF-2, GNF-5, and ABL001 (Asciminib) bind to the allosteric pocket near the C terminus of the ABL kinase domain stabilizes the inactive conformation of the kinase (Adrian et al., 2006; Zhang et al., 2010). Using solution NMR, X-ray crystallography, mutagenesis and hydrogen exchange mass spectrometry, it was shown that allosteric inhibitors can induce long-range structural and dynamic changes in the remote ATP-binding site (Adrian et al., 2006; Zhang et al., 2010; Wylie et al., 2017; Schoepfer et al., 2018). While the field of kinase inhibitors has enjoyed unprecedented success manifested in multiple FDA approved drugs, the development of allosteric kinase activators has been lagging behind. The mechanisms underpinning allosteric action of kinase activators can proceed by destabilization of the inactive state, stabilization of the active state, facilitating of the active state, and dynamic responses to phosphorylation in regulatory sites (Dar and Shokat, 2011; Fang et al., 2013; Hu et al., 2013; Cowan-Jacob et al., 2014). Structural and biochemical studies of allosteric inhibitors and activators of ABL kinase have indicated that structural environment near the allosteric pocket can serve as a sensor of ligand binding, triggering either stabilization of the inactive state or large conformational shift and activation. Furthermore, synergistic actions of allosteric and ATP competitive inhibitors have provided evidence that binding can perturb dynamics at distal regions and elicit ligand-specific communication between binding sites. Computational studies have detailed how allosteric inhibitors and activators may exert a differential control on allosteric signaling between binding sites (Astl and Verkhivker, 2019a). It was found that while inhibitor binding can strengthen the inhibitory ABL state and induce allosteric communications directed from the allosteric pocket to the ATP binding site, DPH activator may induce a more dynamic kinase state and preferentially activate allosteric couplings between the ATP and substrate binding sites (Astl and Verkhivker, 2019a).

By combining computational and experimental approaches a significant progress has been made in discovery of allosteric modulators of Hsp90 and Hsp70 chaperones. Recent studies have demonstrated that the C-terminal domain (CTD) of Hsp90 is important for dimerization of the chaperone and contains a second nucleotide binding site (Marcu et al., 2000a,b). The bacterial gyrase inhibitor novobiocin, a member of the coumeromycin family of antibiotics, is an Hsp90 antagonist that was found to inhibit a second ATP binding site at the C-terminus. Novobiocin binds the C-terminal nucleotide pocket and displaces both ATP and geldanamycin, and inhibits Hsp90's function (Marcu et al., 2000a,b). The principal advantage of C-terminal inhibition over N-terminal inhibition is the lack of a heat shock response upon ligand binding at the C-terminus of Hsp90. The first compounds that clearly differentiated between the C-terminus of Hsp90 and DNA gyrase, and converted a well-established gyrase inhibitor into a selective Hsp90 inhibitor were initially reported by Donnelly and Blagg (2008), Matts et al. (2011a), Matts et al. (2011b), Garg et al. (2016, 2017a,b), Hall et al. (2016), Khandelwal et al. (2016), and Kumar MV et al. (2018). The first experimental-guided computational prediction and mapping of hidden allosteric sites in Hsp90 combined NMR analysis, proteolytic fingerprinting and photoaffinity labeling with multiscale modeling of Hsp90 interactions and docking (Matts et al., 2011a,b). Computational predictions provided the first atomic resolution model of Hsp90 binding with the C-terminal modulator that fully satisfies the available experimental data and provide key insight for the structure-based design of allosteric Hsp90 inhibitors. In the subsequent study, a novel, computational approach for the discovery of allosteric inhibitors based on the physical characterization of signal propagation mechanisms was applied to the Hsp90 chaperone (Morra et al., 2010). By characterizing the allosteric hotspots of the inter-domain communication pathways, dynamic pharmacophore models to screen small molecules were developed. The computational predictions were combined with experimental validation showing that the selected molecules bind the allosteric sites of Hsp90,

exhibit anti-proliferative activity in different tumor cell lines, and destabilize Hsp90 client proteins (Morra et al., 2010). The recent series of studies by Colombo and colleague have reported results of computer-aided design and synthesis of new allosteric ligands with micromolar to nanomolar anticancer activities, demonstrating the power of computational approaches in discovering allosteric modulators that can probe the relationships between structure dynamics and function of the Hsp90 proteins and regulatory complexes with client proteins (Sattin et al., 2015; D'Annessa et al., 2017; Masgras et al., 2017; Ferraro et al., 2019; Hu et al., 2020; Sanchez-Martin et al., 2020). Computational targeting of the Hsp90 client proteins based on the prediction of locally unstable substructures in proteins was used to develop potent probes and peptides blocking Hsp90-client interactions (Colombo et al., 2020). Recent efforts have also produced small molecules that can inhibit the inter-chaperone protein-protein interactions for Hsp70 chaperone (Gestwicki and Shao, 2019). These chemical probes have shown a considerable promise in interrogating chaperone networks in a range of models. Design, synthesis, and biological evaluation of small molecules that regulate the interaction between two Hsp70 and HOP chaperones reported the first class of compounds that specifically modulate these protein-protein interactions and inhibit protein folding events (Zaiter et al., 2019). An integrated computational and experimental approach probed allosteric mechanisms of Hsp70 binding, showing that symbiotic employment of different research tools in dissecting allosteric events in signaling proteins can be instrumental to discover selective allosteric modulators of protein functions (Rinaldi et al., 2018).

NEW DEVELOPMENTS IN MODELING OF ALLOSTERIC REGULATION: INFORMATION-THEORETICAL ANALYSIS OF DYNAMIC FLOWS AND ENTROPY TRANSFER IN PROTEINS

The emerging computational approaches that are now employed for studies of allosteric states and mechanisms include experiment-informed network approaches, Markovian modeling and also the information-theoretical methods that model dynamic flows and entropy transfer in complex systems. By describing protein dynamics as a dynamically evolving network of interconnected modules, the topological regularities of the network structure can be identified, while filtering out the relatively unimportant details. A modular description of a network can be viewed as a compression of that network topology, and the problem of community identification can be viewed as a problem of finding an efficient compression of the network structure and topology. Using this premise, the challenge of identifying the community structure of complex networks describing dynamic energy landscapes of allosteric proteins can be reformulated as an information-theoretic approach. Flow-based model methods operate through a stochastic walk on the dynamics of the network rather than on its topological structure, where communities consist of dynamically interconverting conformations among which the dynamic flow can persist for a long time and define functionally significant states (Rosvall and Bergstrom, 2007, 2008, 2010, 2011; Lancichinetti and Fortunato, 2009; Schaub et al., 2012; Rosvall et al., 2014; Kawamoto and Rosvall, 2015). This information-theoretical analysis can quantify the structure and dynamics of the proteins from a unified perspective in which short term dynamics is integrated into a long term behavior of the system through a modular description of dynamic flows occurring on a given network (**Figure 2**). In this approach, a random walk is used as a proxy for the dynamic flow on the network. The map equation method implemented by the Infomap algorithm (Edler et al., 2017) can find the optimal community partition of the dynamic conformational ensembles on different time scales (derived from MD simulations or MSM maps of macrostates) and identify dynamically persistent (as opposed to topology-derived) communities of functional macrostates. This dynamic flows method compresses the flows by aggregating nodes (states) with rapid stochastic movements, revealing network regularities as distinct dynamic modules in which flows are contained on a given time scale. The map equation has been also extended to the higher-order Markov dynamics (Lancichinetti and Fortunato, 2009; Lambiotte et al., 2011, 2019; Schaub et al., 2012; Rosvall et al., 2014; Delvenne et al., 2015; Salnikov et al., 2016). NMR chemical shift perturbations have been combined with Markov modeling and information-theoretical analysis to reveal the dynamic flow of communication between allosteric communities and identify critical residue nodes within the communication pathways in protein kinases (Aoto et al., 2016).

This information-theoretical approach can also explore the dynamic evolution of the hierarchical multi-layered interaction networks and has a potential to uncover hidden allosteric states associated with the different dynamic time scales (**Figure 2**). Synchronization and causality are basic non-linear phenomenon observed in diverse complex systems, including allosterically regulated proteins. When studying allosteric mechanisms and communications in proteins, it is important not only to detect synchronized allosteric states, but also to identify causal relationships between them. The knowledge of information-theoretic measures is essential for the analysis of information flow between allosteric states and presents a challenging problem (Hlaváčková-Schindler et al., 2007). The problem of finding a measure that is sensitive to the directionality of the flow of information has been explored using non-linear Granger causality of time series (Ancona et al., 2004). An asymmetric quantity termed Transfer Entropy (TE), has been proposed to estimate the directionality of the coupling and flow of information (Schreiber, 2000). The information-theoretic approaches measuring causal influences in multivariate time series (Hlaváčková-Schindler et al., 2007; Ito, 2016; Darmon and Rapp, 2017) can be also applied to studies of allosteric protein states and mechanisms. The quantitative models of information flow between two correlated processes (Schreiber, 2000) were adopted to quantify time delayed correlations and entropy transfer between residue pairs as a measure of allosteric coupling in proteins (Hacisuleyman and Erman, 2017a,b). Through analysis of entropy transfer, one can determine residues that act as drivers of the fluctuations of other residues, thereby

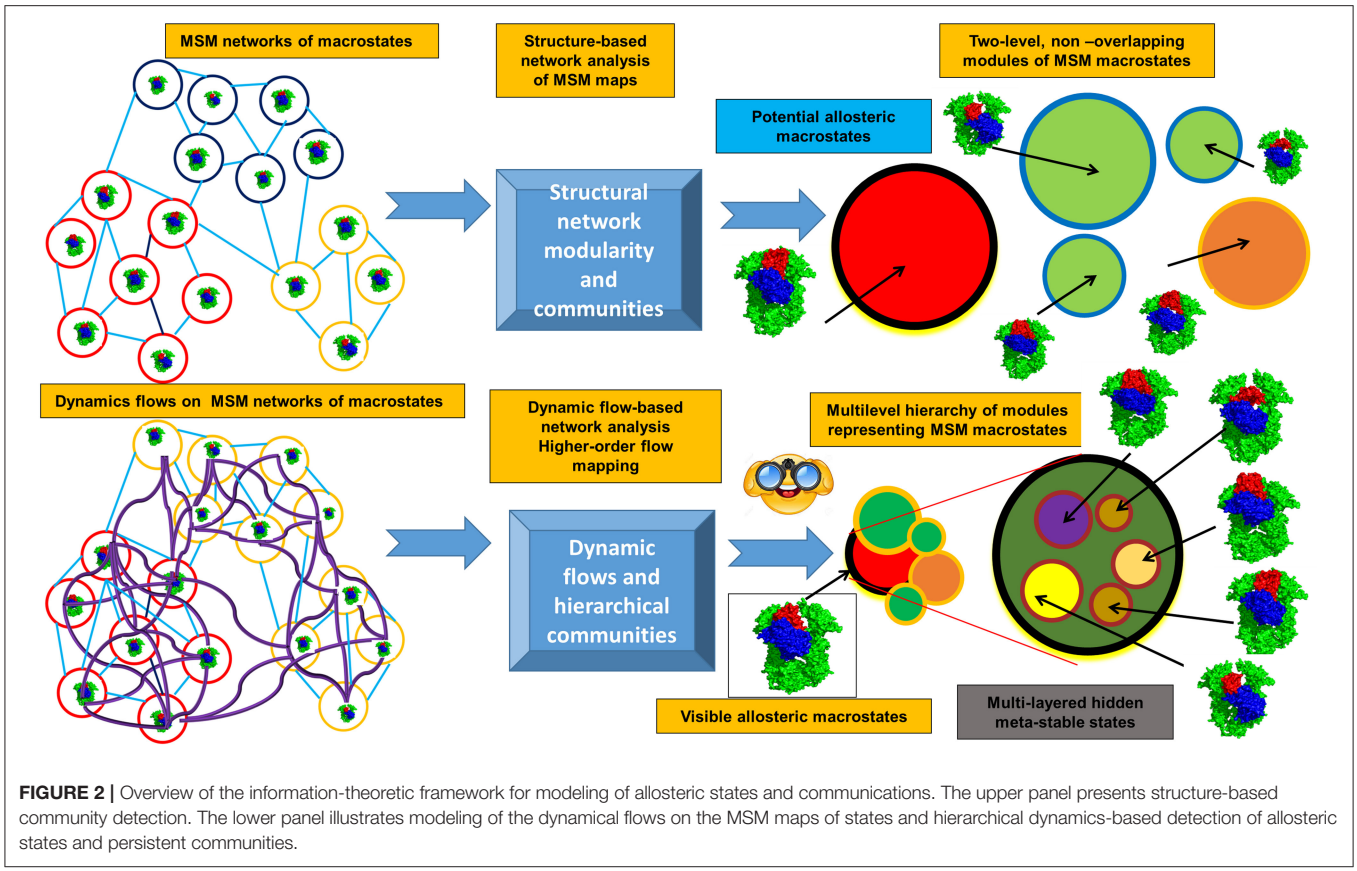


FIGURE 2 | Overview of the information-theoretic framework for modeling of allosteric states and communications. The upper panel presents structure-based community detection. The lower panel illustrates modeling of the dynamical flows on the MSM maps of states and hierarchical dynamics-based detection of allosteric states and persistent communities.

determining causality in the correlations and identifying residues that act as drivers of allosteric communication in proteins (Hacisuleyman and Erman, 2017a,b). The relative entropy concept from information theory was used as a quantitative metric to develop a method for measurement of the population shift during allosteric transitions (Zhou and Tao, 2019). The developed relative entropy-based dynamical allosteric network (REDAN) model was sucessfully applied to the second PDZ domain (PDZ2) in the human PTP1E protein, providing an accurate assessment of allosteic pathways (Zhou and Tao, 2019). A rigorous mathematical framework offered by the information-theoretical formalism of dynamic network flows combined with biophysical simulations may prove to be useful for finding modular patterns and dynamically persistent communities of macrostates. The integration of this methodology with NMR experiments can aid in the better identification of functional allosteric states by matching evolution of dynamic communities against the NMR chemical shift patterns and biophysical experiments.

THE RISE OF THE MACHINES: ALLOSTERIC MECHANISMS THROUGH THE LENS OF MACHINE LEARNING

Over the last few years, advances in the ML field have driven the design of new computational systems that improve with experience and are able to model increasingly complex chemical and biological phenomena (Goh et al., 2017; Korotcov et al., 2017; Chen et al., 2018a; Popova et al., 2018; Dimitrov et al., 2019; Mater and Coote, 2019). ML techniques have been successfully applied to various computational chemistry challenges (Husic and Pande, 2018), pharmaceutical data analysis (Burbidge et al., 2001), protein–ligand binding affinity prediction problems (Ballester and Mitchell, 2010, Decherchi et al., 2015), dissecting molecular determinants of protein mechanisms and biochemical reactions (Li et al., 2015, Cortina and Kasson, 2018, Shcherbinin et al., 2019). Data-intensive ML modeling can be also applied for detection and classification of allosteric protein states. The integration of Markov modeling, simulations and ML approaches into robust and reproducible computational pipelines with the experimental feedback can be explored for atomistic modeling and classification of allosteric states (**Figure 3**). Several ML algorithms including decision tree and artificial neural networks were employed in combination with MSM approaches to develop classification models of functionally relevant allosteric conformations that exhibit very similar tertiary structures (Zhou et al., 2018a,b). Despite the lack of significant conformational change between allosteric states of the second PDZ domain (PDZ2) in the human PTP1E protein, which is a prototypical example of dynamics-driven allostery, it was demonstrated that both algorithms could build effective prediction models and provide reliable quantitative evaluation of the contributions from

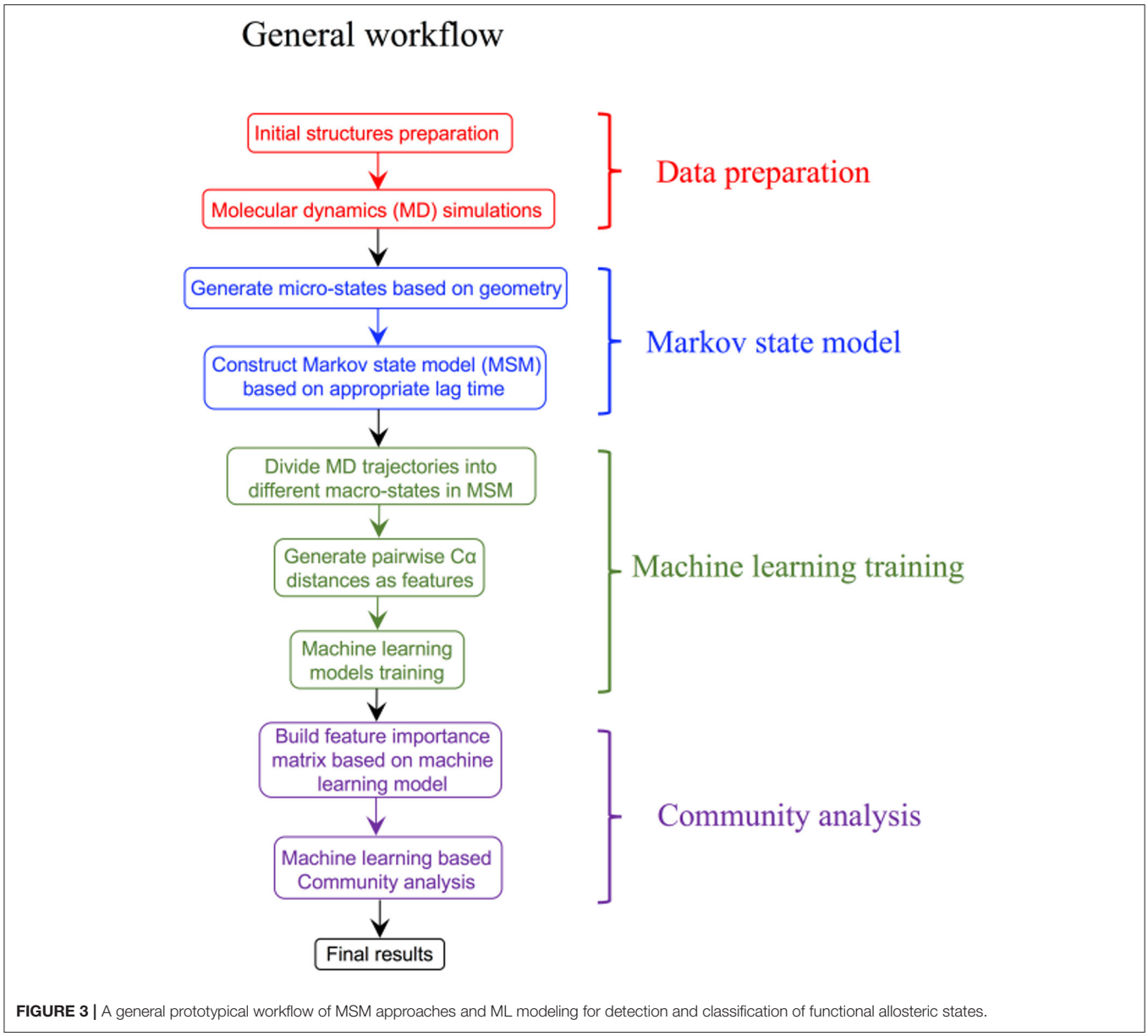


FIGURE 3 | A general prototypical workflow of MSM approaches and ML modeling for detection and classification of functional allosteric states.

individual residues to the difference between the two allosteric states (Zhou et al., 2018a,b). A high prediction accuracy and sensitivity of the ML models to small structural and dynamic changes have demonstrated the utility of these approaches in probing subtle allosteric changes. Deep neural networks were used in combination with MD simulations of the PDZ3 domain of PDS-95 revealing that allosteric effects can be quantified as residue-specific properties through two-dimensional property-residue maps (Hayatshahi et al., 2019). These residue response maps could accurately describe how different protein residues are affected by allosteric perturbations exerted on the protein system. Another ML-based analysis of protein dynamics was employed to compare the binding modes of TEM-1 β-lactamase in different catalytic states (Wang et al., 2019). While conventional analysis methods including principal components analysis (PCA) could not differentiate TEM-1 in different binding modes, neural network models resulted in an excellent classification scheme that differentiated between closely related functional states (Wang et al., 2019). This study has provided a unique insight into the role and specific function of individual residues, highlighting their contributions to the delicate thermodynamic balance between allosteric states.

The remarkable rise of deep learning (DL) relying on the robust function approximations and representation properties of deep neural networks has provided us with new tools to automatically find compact low-dimensional representations (features) of high-dimensional data (LeCun et al., 2015). DL models have achieved outstanding predictive performance making dramatic breakthroughs in a wide range of applications, including automatic speech processing and image recognition

(Toledano et al., 2018; Kim et al., 2019; Hey et al., 2020; Wu et al., 2020). In the words of Geoffrey Hinton who is the founder of DL technologies "Deep Learning is an algorithm which has no theoretical limitations on what it can learn; the more data you give and the more computational time you provide the better it is" (LeCun et al., 2015). Deep neural network methods were successfully applied to predict intrinsic molecular properties such as atomization energy based on simple molecular geometry and element types (Rupp et al., 2012). DL models were recently used for structure-functional prediction of cancer mutations and functional hotspots of ligand binding in cancer-associated genes (Agajanian et al., 2018). The developed models can capture ∼90% of experimentally validated mutational hotspots and yield novel information about molecular signatures of driver mutations. In the recent studies, we have proposed novel DL architectures capable of predicting functional protein hotspots directly from raw nucleotide sequence information (Agajanian et al., 2019). These studies have shown that DL models can learn high importance features from raw genomic information and produce reliable recognition and classification of functionally significant cancer mutation hotspots. Moreover, these DL models can often outperform computational predictors of cancer mutations that are based on protein sequence and structure features (Agajanian et al., 2019). The success of DL tools in deciphering important functional phenotypes directly from primary sequence information is encouraging as these models can bypass the need for a large number of empirically-derived functional and structural features. However, ML methods often result in "black box" models with limited interpretability. There has been an explosion of interest in transparent and interpretable ML models to enable more efficient data mining and scientific knowledge discovery (Holzinger et al., 2014). Our investigations have also provided a roadmap how to combine DL predictions of functional sites with subsequent biophysical analysis to aid in the interpretability of ML models and facilitate their applications in biological problems (Agajanian et al., 2018, 2019).

One of the primary goals of artificial intelligence (AI) is to produce fully autonomous agents that interact with their environments to learn optimal behavior, improving over time through trial and error. An important mathematical framework for experience-driven autonomous learning through interactions with the environment is reinforcement learning (RL) (Sutton and Barto, 1981; Barto, 1994; Botvinick, 2012; Hassabis et al., 2017). While previous RL approaches lacked scalability and were limited to fairly low-dimensional problems, a marriage between deep neural networks and RL resulted in the new rapidly evolving field of deep reinforcement learning (DRL) that has achieved remarkable success in game-oriented and various scientific applications, attaining a wide popularity and celebrity-like following among researchers (Mnih et al., 2015; Silver et al., 2017; Botvinick et al., 2019; Jaderberg et al., 2019; Senior et al., 2019). DRL concepts leverage and symbiotically combine neural network modeling with reinforcement learning, in which optimization strategies are crafted based on the trade-offs and competition between rewards and punishments rather than conventional deterministic or stochastic exploration methods. After years of serving as a mere inspiration rather than a practical tool, DRL techniques have taken off overcoming scalability and data limitation issues, and exploding into one of the most intense areas of AI research. Recent years have witnessed the expansion of DRL applications into biomedical research including but not limited to biomedical informatics, drug discovery (Baskin, 2020; Grebner et al., 2020), and toxicology (Chary et al., 2020).

The rationale for employing DRL technologies in studies of allosteric regulation is to capitalize on conceptual and algorithmic similarity between Markov decisions processes (MDPs) which are at the core of RL methods and Markovian modeling of allosteric states in proteins. Several methods adopted RL-based conceptualization to develop MDP algorithms for conformational mapping of the protein landscapes and detection of functional allosteric states. REinforcement learning based Adaptive samPling (REAP) algorithm has shown a considerable promise by adopting RL principles in which an agent (or learning algorithm) takes actions in an environment (conformational protein landscape) to maximize a reward function (Shamsi et al., 2018). In this study, the action is associated with launching a pool of simulations along different collective variables (reaction coordinates), with the reward function proportional to the efficiency of a reaction coordinate to sample space and detect unknown states, and the agent selecting the directions which are most rewarding ultimately leading to the optimal adaptive strategy (Shamsi et al., 2018). Similar concepts were used to develop a goal-oriented sampling method, termed fluctuation amplification of specific traits (FAST) for rapid search of conformational space and identification of distinct functional states by balancing search near promising solutions (exploitation) and attempts to find novel solutions (exploration). Inspired by the RL ideas, this methods runs pools of simulations from starting points chosen based on the reward functions that encourages discovery of new conformations with selected physical properties (Zimmerman and Bowman, 2015; Zimmerman et al., 2018). Generative neural networks have been recently proposed as a tool for the discovery of efficient collective variables that are fundamental for adaptive exploration of the conformational landscapes and finding functional states and hidden allosteric states by guiding sampling toward poorly explored regions (Chiavazzo et al., 2017; Chen et al., 2018b; Hernandez et al., 2018; Mardt et al., 2018). MD simulations were combined with DL approach to train an autoencoder (Hinton and Salakhutdinov, 2006) in order to generate new protein conformations and mine conformational space of the bound state from an ensemble of unbound protein structures (Degiacomi, 2019). Another interesting study employed autoencoder-based detection algorithm to explore dynamic allostery induced by ligand binding based on the comparison of time fluctuations of distance matrices obtained from MD simulations in ligand-bound and unbound forms (Tsuchiya et al., 2019). In this method, the autoencoder neural network is first trained on the time fluctuations of protein motions in the apo form, and the trained autoencoder is then applied to analyze patterns of fluctuations in the holo form. Using this elegant implementation of RL approach, the authors mapped allosteric communication networks of the dynamically coupled residues and ligand pockets in the PDZ2 domain induced by binding (Tsuchiya et al.,

2019). Allosteric pocket crosstalk defined as a temporal exchange of atoms between adjacent pockets in the MD trajectories can produce a fingerprint of hidden allosteric communication networks (La Sala et al., 2017). The recent RL-inspired studies of allosteric systems suggested that simulation-driven ML modeling and analysis of conformational landscapes may uncover rarely-populated functional states and shed the light on the key features of allosteric communications between visible and hidden binding pockets in proteins.

DRL is a continuous trial-and-error based sampling-learning process where the agent tries to apply different combination of actions on a state to find the highest cumulative reward. Although DRL methods can tackle a wide range of learning problems with a rigorous mathematical formulation, the challenges posed by the properly crafted interplay between rich data acquisition and delayed rewards remains a significant impediment to the widespread of RL methods in many application domains, including prediction of allosteric protein states and mechanisms. The challenges of DRL approaches often lie in the art of designing robust reward function. The hybrid reward functions with a weighted combination of topological, dynamic, and network-based rewards describing different characteristics of allosteric states may represent a potentially interesting strategy to mitigate the inherent deficiencies of RL and DRL methods. For this, the rewards may be treated as neural networks trained on the set of known allosteric states. A new saga in the rapidly evolving DRL field was the development of episodic-based DRL algorithms that estimate the value of actions and states using episodic memories where the agent stores each encountered state along with the sum of rewards obtained during the n time steps (Botvinick et al., 2019). The episodic memory-based models can be extended to develop curiosity reward bonus functions for efficient exploration of the environment and detecting states in the poorly accessible regions (Han et al., 2020). In this context, DRL framework that iterates episodes of DRL and community decomposition of the dynamic network flows on the conformational landscapes may enhance the interplay between sampling and learning, thus facilitating identification of hidden allosteric states. Different from traditional DRL approaches, this strategy can consider communities of states as intermediate stages in the learning process, rather than unique states, which could potentially lead to a more robust and versatile learning procedure (**Figure 4**).

Deep neural network (DNN) models, most notably variational autoencoder (VAE) (Gomez-Bombarelli et al., 2018) and generative adversarial networks (GANs) (Sorin et al., 2020; Zhong et al., 2020) have proven fruitful in chemical discovery and molecular design of novel synthesizable chemical probes. Automated chemical design approaches employed VAE neural networks for a data-driven continuous representation of molecules (Gomez-Bombarelli et al., 2018). GAN models are often considered as one of the most significant advances in the field of machine learning, and their success has generated a considerable momentum with growing number of applications including molecular design of novel chemical probes and materials (Guimaraes et al., 2017; Kadurin et al., 2017a,b; Olivecrona et al., 2017; Yu et al., 2017; Gupta et al., 2018; Polykovskiy et al., 2018; Putin et al., 2018a,b) (**Figure 4**). By leveraging sequence data generation (SeqGAN) approach (Yu et al., 2017); Objective-Reinforced Generative Adversarial Networks (ORGAN) (Guimaraes et al., 2017) combines GANs and RL to apply the GAN framework to molecular design with domain-specific rewards and feedback. Of particular importance is MolGAN, an implicit, generative model for small molecular graphs that circumvents the need for expensive graph matching procedures and adapts GAN approach to operate directly on graph-structured data (Cao and Kipf, 2018). CycleGAN provides unpaired image-to-image translation using Cycle-Consistent Adversarial Networks (Zhu et al., 2017). MolCycleGAN, which extended the CycleGAN framework with an added loss and extra encoding network, maps from distribution to distribution on unpaired samples, so it can amplify the size of our dataset in the process by taking all of the pairing combinations rather than relying on a training dataset of predefined molecule-inhibitor pairs (Maziarka et al., 2020). The advantage of MolCycleGAN is the ability to learn transformation rules from the sets of compounds with desired and undesired values of the considered property. The methodological and algorithmic progress in GAN applications to molecular discovery has been further catalyzed by the development of several comprehensive benchmarking sets embedded into a sophisticated cheminformatics infrastructure supporting open-source implementations of molecular features and learning algorithms (Olson et al., 2017; Polykovskiy et al., 2018; Racz et al., 2019). Despite recent developments in GANs models, the applicability of these tools for molecular design continues to present a promise rather than a validated strategy, lacking systematic and comprehensive tools to target specific protein families and interrogate molecular mechanisms. There is also growing interest in generative models which can incorporate both chemical and structural information about small molecules and their interactions with protein targets.

SYNERGIES AND LIMITATIONS OF COMPUTATIONAL APPROACHES FOR QUANTITATIVE MODELING OF ALLOSTERIC REGULATION

A systematic interdisciplinary effort is needed to leverage the burgeoning knowledge about allosterically regulated proteins in the development of experiment-informed data-oriented computational tools for prediction of allosteric mechanisms and allosteric drug discovery. The main advantage of experiment-informed Markovian modeling is the ability of this technique to adequately describe hierarchy of allosteric states and the molecular basis of allosteric mechanisms. Using a combination of NMR-guided simulations and MSM approach, we can determine structural and dynamic patterns of conformational ensembles and identify functional allosteric states that are hidden in the conformational ensembles. The critical challenges of these methodologies for modeling allosteric regulation phenomenon is selecting a set of experimentally-informed collective variables defined by the intrinsic dynamics to provide the optimal projection of the landscape into functional allosteric states. In this context, the newly emerging

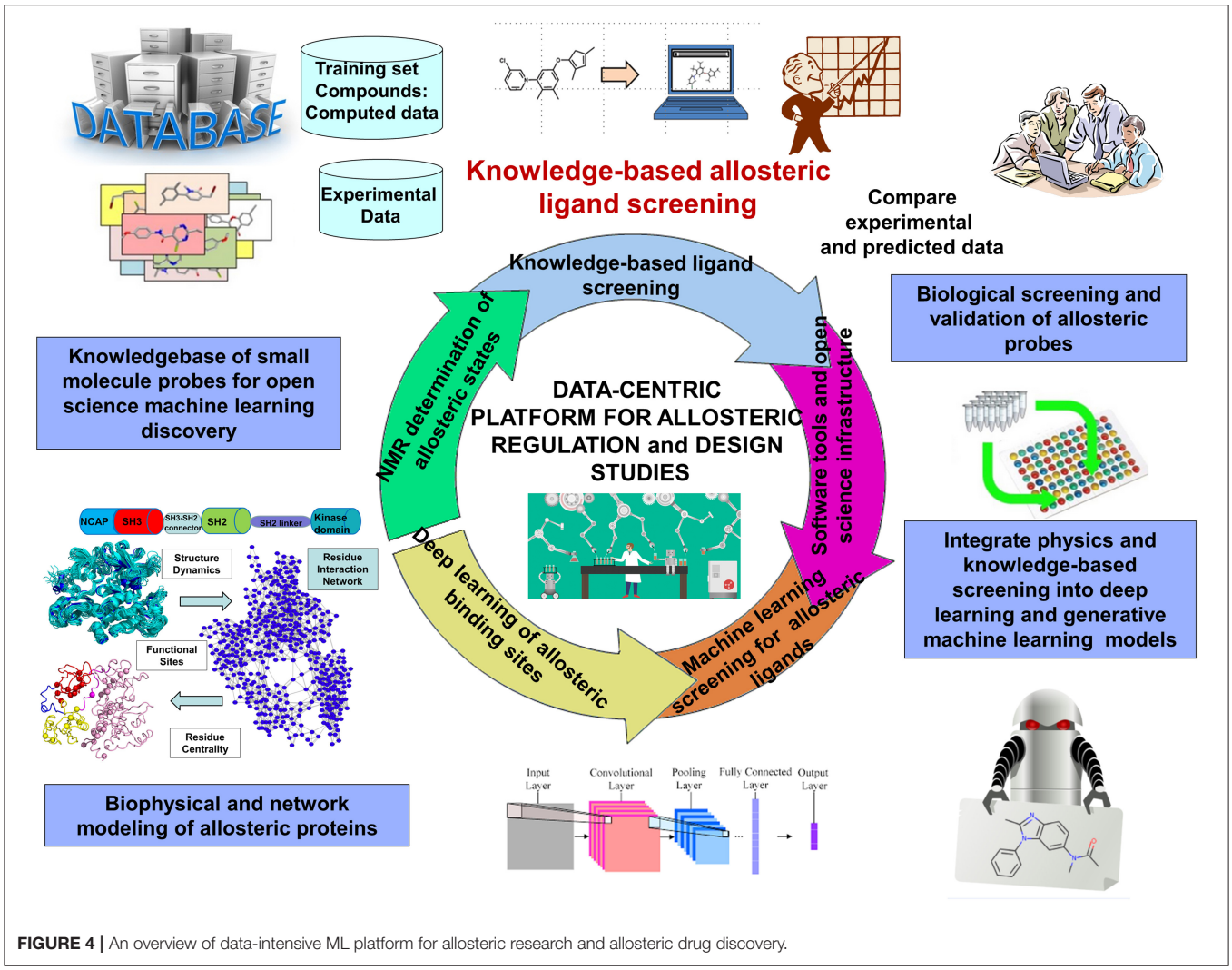


FIGURE 4 | An overview of data-intensive ML platform for allosteric research and allosteric drug discovery.

information-theoretical flow approaches and modeling of entropy transfer in proteins can represent viable complementary tools for adequate reconstruction of functional conformational landscapes in proteins. The proposed integration of biomolecular simulations and NMR experiments with machine learning into a comprehensive research platform is expected to produce a toolkit of approaches for prediction of allosteric states and mapping of allosteric mechanisms.

Network algorithms, information-theoretical approaches and DL models may be time-consuming and require a systematic exploration and engineering of features and neural net architectures with a constant and evolving feedback from NMR experiments to validate and confirm predictions. Several different ML architectures can be further explored to address potential efficiency and convergence problems including transfer learning, imitation learning, episodic control and dueling networks. To achieve synergies and robust integration of emerging technologies for predicting allosteric regulation mechanisms, a new open science infrastructure development is required which implies extensive sharing of experimental and computational data, software and knowledge across many discipline. Through integrative studies of allosteric mechanisms empowered by biophysical and data science approaches we can expand the toolkit of to dissect and interrogate allosteric mechanisms and functions in the therapeutically important protein families.

CONCLUSION

The growing body of computational and experimental studies has shown that integration of data-driven biophysical and ML approaches can bring about new drug discovery paradigms, opening up unexplored venues for further scientific innovation and unique biological insights. The integration of computational and NMR approaches into a novel research platform that explores experiment-informed physical simulations, Markov state modeling, information-theoretical formalism of dynamic allosteric networks under the unified umbrella of machine learning will key to dissect

molecular rules of allosteric regulation. The innovative cross-disciplinary approaches that expand the knowledge, resources and tools for studies of allosteric regulation can promote a broader usage of new technologies to understand and exploit allosteric phenomenon through the lens of chemical biology, material science, synthetic biology and bioengineering. By developing an open science infrastructure for machine learning studies of allosteric regulation and validating computational approaches using integrative studies of allosteric mechanisms, the scientific community can expand the toolkit of approaches and chemical probes for dissecting and interrogation allosteric mechanisms in many therapeutically important proteins. The development of community-accessible tools that uniquely leverage the existing experimental and simulation knowledgebase to enable interrogation of the allosteric functions can provide much needed impetus to further experimental technologies and enable steady progress.

AUTHOR CONTRIBUTIONS

GV, PT, and GH conceived and designed the research, analyzed the results, and wrote the manuscript. GV, SA, PT, and GH performed the research. GV wrote the final version of the manuscript and supervised the project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partly supported by institutional funding from Chapman University.

ACKNOWLEDGMENTS

The authors acknowledge the technical assistance of Schmid College Grand Challenge Initiative Postdoctoral Fellow Dr. Anne Sonnenschein.

REFERENCES

Adhireksan, Z., Palermo, G., Riedel, T., Ma, Z., Muhammad, R., Rothlisberger, U., et al. (2017). Allosteric cross-talk in chromatin can mediate drug-drug synergy. *Nat. Commun.* 8:14860. doi: 10.1038/ncomms14860

Adrian, F. J., Ding, Q., Sim, T., Velentza, A., Sloan, C., Liu, Y., et al. (2006). Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nat. Chem. Biol.* 2, 95–102. doi: 10.1038/nchembio760

Agajanian, S., Odeyemi, O., Bischoff, N., Ratra, S., and Verkhivker, G. M. (2018). Machine learning classification and structure-functional analysis of cancer mutations reveal unique dynamic and network signatures of driver sites in oncogenes and tumor suppressor genes. *J. Chem. Inf. Model.* 58, 2131–2150. doi: 10.1021/acs.jcim.8b00414

Agajanian, S., Oluyemi, O., and Verkhivker, G. M. (2019). Integration of random forest classifiers and deep convolutional neural networks for classification and biomolecular modeling of cancer driver mutations. *Front. Mol. Biosci.* 6:44. doi: 10.3389/fmolb.2019.00044

Aguilar, D., Oliva, B., and Marino Buslje, C. (2012). Mapping the mutual information network of enzymatic families in the protein structure to unveil functional features. *PLoS ONE* 7:e41430. doi: 10.1371/journal.pone.0041430

Amor, B., Yaliraki, S. N., Woscholski, R., and Barahona, M. (2014). Uncovering allosteric pathways in caspase-1 using Markov transient analysis and multiscale community detection. *Mol. Biosyst.* 10, 2247–2258. doi: 10.1039/C4MB00088A

Ancona, N., Marinazzo, D., and Stramaglia, S. (2004). Radial basis function approach to nonlinear granger causality of time series. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys. 70(5 Pt 2)*:056221. doi: 10.1103/PhysRevE.70.056221

Ansari, S. M., Coletta, A., Kirkeby Skeby, K., Sorensen, J., Schiott, B., and Palmer, D. S. (2016). Allosteric-activation mechanism of bovine chymosin revealed by bias-exchange metadynamics and molecular dynamics simulations. *J. Phys. Chem. B* 120, 10453–10462. doi: 10.1021/acs.jpcb.6b07491

Anthis, N. J., and Clore, G. M. (2015). Visualizing transient dark states by NMR spectroscopy. *Q Rev. Biophys.* 48, 35–116. doi: 10.1017/S0033583514000122

Aoto, P. C., Martin, B. T., and Wright, P. E. (2016). NMR characterization of information flow and allosteric communities in the MAP kinase p38gamma. *Sci. Rep.* 6:28655. doi: 10.1038/srep28655

Arkhipov, A., Shan, Y., Das, R., Endres, N. F., Eastwood, M. P., Wemmer, D. E., et al. (2013). Architecture and membrane interactions of the EGF receptor. *Cell* 152, 557–569. doi: 10.1016/j.cell.2012.12.030

Astl, L., Tse, A., and Verkhivker, G. M. (2019). Interrogating regulatory mechanisms in signaling proteins by allosteric inhibitors and activators: a dynamic view through the lens of residue interaction networks. *Adv. Exp. Med. Biol.* 1163, 187–223. doi: 10.1007/978-981-13-8719-7_9

Astl, L., and Verkhivker, G. M. (2019a). Atomistic modeling of the abl kinase regulation by allosteric modulators using structural perturbation analysis and community-based network reconstruction of allosteric communications. *J. Chem. Theory Comput.* 15, 3362–3380. doi: 10.1021/acs.jctc.9b00119

Astl, L., and Verkhivker, G. M. (2019b). Data-driven computational analysis of allosteric proteins by exploring protein dynamics, residue coevolution and residue interaction networks. *Biochim. Biophys. Acta Gen. Subj.* doi: 10.1016/j.bbagen.2019.07.008. [Epub ahead of print].

Atilgan, A. R., Akan, P., and Baysal, C. (2004). Small-world communication of residues and significance for protein dynamics. *Biophys. J.* 86(Pt. 1), 85–91. doi: 10.1016/S0006-3495(04)74086-2

Ballester, P. J., and Mitchell, J. B. (2010). A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* 26, 1169–1175. doi: 10.1093/bioinformatics/btq112

Barto, A. G. (1994). Reinforcement learning control. *Curr. Opin. Neurobiol.* 4, 888–893. doi: 10.1016/0959-4388(94)90138-4

Baskin, I. I. (2020). The power of deep learning to ligand-based novel drug discovery. *Expert. Opin. Drug Discov.* doi: 10.1080/17460441.2020.1745183. [Epub ahead of print].

Beglov, D., Hall, D. R., Wakefield, A. E., Luo, L., Allen, K. N., Kozakov, D., et al. (2018). Exploring the structural origins of cryptic sites on proteins. *Proc. Natl. Acad. Sci. U.S.A.* 115, E3416–E3425. doi: 10.1073/pnas.1711490115

Berjanskii, M. V., and Wishart, D. S. (2017). Unraveling the meaning of chemical shifts in protein NMR. *Biochim. Biophys. Acta Proteins Proteom.* 1865(Pt. B), 1564–1576. doi: 10.1016/j.bbapap.2017.07.005

Bhattacharya, S., and Vaidehi, N. (2014). Differences in allosteric communication pipelines in the inactive and active states of a GPCR. *Biophys. J.* 107, 422–434. doi: 10.1016/j.bpj.2014.06.015

Bhattacharyya, M., and Vishveshwara, S. (2011). Probing the allosteric mechanism in pyrrolysyl-tRNA synthetase using energy-weighted network formalism. *Biochemistry* 50, 6225–6236. doi: 10.1021/bi200306u

Bode, C., Kovacs, I. A., Szalay, M. S., Palotai, R., Korcsmaros, T., and Csermely, P. (2007). Network analysis of protein dynamics. *FEBS Lett.* 581, 2776–2782. doi: 10.1016/j.febslet.2007.05.021

Boehr, D. D., Dyson, H. J., and Wright, P. E. (2006). An NMR perspective on enzyme dynamics. *Chem. Rev.* 106, 3055–3079. doi: 10.1021/cr050312q

Bonomi, M., and Camilloni, C. (2017). Integrative structural and dynamical biology with PLUMED-ISDB. *Bioinformatics* 33, 3999–4000. doi: 10.1093/bioinformatics/btx529

Botvinick, M., Ritter, S., Wang, J. X., Kurth-Nelson, Z., Blundell, C., and Hassabis, D. (2019). Reinforcement learning, fast and slow. *Trends Cogn. Sci.* 23, 408–422. doi: 10.1016/j.tics.2019.02.006

Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Curr. Opin. Neurobiol.* 22, 956–962. doi: 10.1016/j.conb.2012.05.008

Boulton, S., Akimoto, M., Selvaratnam, R., Bashiri, A., and Melacini, G. (2014). A tool set to map allosteric networks through the NMR chemical shift covariance analysis. *Sci. Rep.* 4:7306. doi: 10.1038/srep07306

Boulton, S., and Melacini, G. (2016). Advances in NMR methods to map allosteric sites: from models to translation. *Chem. Rev.* 116, 6267–6304. doi: 10.1021/acs.chemrev.5b00718

Boulton, S., Selvaratnam, R., Ahmed, R., and Melacini, G. (2018). Implementation of the NMR CHEmical shift covariance analysis (CHESCA): a chemical biologist's approach to allostery. *Methods Mol. Biol.* 1688, 391–405. doi: 10.1007/978-1-4939-7386-6_18

Bowman, G. R. (2014). A tutorial on building markov state models with MSMBuilder and coarse-graining them with BACE. *Methods Mol. Biol.* 1084, 141–158. doi: 10.1007/978-1-62703-658-0_8

Bowman, G. R., Bolin, E. R., Hart, K. M., Maguire, B. C., and Marqusee, S. (2015). Discovery of multiple hidden allosteric sites by combining markov state models and experiments. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2734–2739. doi: 10.1073/pnas.1417811112

Bowman, G. R., and Geissler, P. L. (2012). Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11681–11686. doi: 10.1073/pnas.1209309109

Bowman, G. R., Huang, X., and Pande, V. S. (2009). Using generalized ensemble simulations and markov state models to identify conformational states. *Methods* 49, 197–201. doi: 10.1016/j.ymeth.2009.04.013

Bowman, G. R., and Noe, F. (2014). Software for building markov state models. *Adv. Exp. Med. Biol.* 797:139. doi: 10.1007/978-94-007-7606-7_11

Brinda, K. V., and Vishveshwara, S. (2005). A network representation of protein structures: implications for protein stability. *Biophys. J.* 89, 4159–4170. doi: 10.1529/biophysj.105.064485

Brinda, K. V., and Vishveshwara, S. (2010). Random network behaviour of protein structures. *Mol. Biosyst.* 6, 391–398. doi: 10.1039/B903019K

Brotzakis, Z. F., and Parrinello, M. (2019). Enhanced sampling of protein conformational transitions via dynamically optimized collective variables. *J. Chem. Theory Comput.* 15, 1393–1398. doi: 10.1021/acs.jctc.8b00827

Buchenberg, S., Sittel, F., and Stock, G. (2017). Time-resolved observation of protein allosteric communication. *Proc. Natl. Acad. Sci. U.S.A.* 114, E6804–E6811. doi: 10.1073/pnas.1707694114

Burbidge, R., Trotter, M., Buxton, B., and Holden, S. (2001). Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput. Chem.* 26, 5–14. doi: 10.1016/S0097-8485(01)00094-8

Cao, N. D., and Kipf, T. (2018). MolGAN: An implicit generative model for small molecular graphs. *arXiv [Preprint]*. arXiv: 1805.11973.

Cavalli, A., Montalvao, R. W., and Vendruscolo, M. (2011). Using chemical shifts to determine structural changes in proteins upon complex formation. *J. Phys. Chem. B* 115, 9491–9494. doi: 10.1021/jp202647q

Cembran, A., Kim, J., Gao, J., and Veglia, G. (2014). NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts upon ligand binding. *Phys. Chem. Chem. Phys.* 16, 6508–6518. doi: 10.1039/C4CP00110A

Chakrabarti, S., and Panchenko, A. R. (2009). Coevolution in defining the functional specificity. *Proteins* 75, 231–240. doi: 10.1002/prot.22239

Chakrabarti, S., and Panchenko, A. R. (2010). Structural and functional roles of coevolved sites in proteins. *PLoS ONE* 5:e8591. doi: 10.1371/journal.pone.0008591

Changeux, J. P. (2012). Allostery and the monod-wyman-changeux model after 50 years. *Annu. Rev. Biophys.* 41, 103–133. doi: 10.1146/annurev-biophys-050511-102222

Changeux, J. P., and Edelstein, S. J. (2005). Allosteric mechanisms of signal transduction. *Science* 308, 1424–1428. doi: 10.1126/science.1108595

Chary, M. A., Manini, A. F., Boyer, E. W., and Burns, M. (2020). The role and promise of artificial intelligence in medical toxicology. *J. Med. Toxicol.* doi: 10.1007/s13181-020-00769-5. [Epub ahead of print].

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., and Blaschke, T. (2018a). The rise of deep learning in drug discovery. *Drug Discov. Today* 23, 1241–1250. doi: 10.1016/j.drudis.2018.01.039

Chen, W., Tan, A. R., and Ferguson, A. L. (2018b). Collective variable discovery and enhanced sampling using autoencoders: innovations in network architecture and error function design. *J. Chem. Phys.* 149:072312. doi: 10.1063/1.5023804

Chennubhotla, C., and Bahar, I. (2006). Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. *Mol. Syst. Biol.* 2:36. doi: 10.1038/msb4100075

Chennubhotla, C., and Bahar, I. (2007). Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput. Biol.* 3, 1716–1726. doi: 10.1371/journal.pcbi.0030172

Chennubhotla, C., Yang, Z., and Bahar, I. (2008). Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol. Biosyst.* 4, 287–292. doi: 10.1039/b717819k

Chiappori, F., Merelli, I., Colombo, G., Milanesi, L., and Morra, G. (2012). Molecular mechanism of allosteric communication in Hsp70 revealed by molecular dynamics simulations. *PLoS Comput. Biol.* 8:e1002844. doi: 10.1371/journal.pcbi.1002844

Chiappori, F., Merelli, I., Milanesi, L., Colombo, G., and Morra, G. (2016). An atomistic view of Hsp70 allosteric crosstalk: from the nucleotide to the substrate binding domain and back. *Sci. Rep.* 6:23474. doi: 10.1038/srep23474

Chiavazzo, E., Covino, R., Coifman, R. R., Gear, C. W., Georgiou, A. S., Hummer, G., et al. (2017). Intrinsic map dynamics exploration for uncharted effective free-energy landscapes. *Proc. Natl. Acad. Sci. U.S.A.* 114, E5494–E5503. doi: 10.1073/pnas.1621481114

Cimermancic, P., Weinkam, P., Rettenmaier, T. J., Bichmann, L., Keedy, D. A., Woldeyes, R. A., et al. (2016). CryptoSite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. *J. Mol. Biol.* 428, 709–719. doi: 10.1016/j.jmb.2016.01.029

Colombo, G., Morra, G., Meli, M., and Verkhivker, G. (2008). Understanding ligand-based modulation of the Hsp90 molecular chaperone dynamics at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7976–7981. doi: 10.1073/pnas.0802879105

Colombo, G., Paladino, A., Woodford, M. R., Backe, S. J., Sager, R. A., Kancherla, P., et al. (2020). Chemical perturbation of oncogenic protein folding: from the prediction of locally unstable structures to the design of disruptors of Hsp90-client interactions. *Chemistry*. doi: 10.1002/chem.202000615. [Epub ahead of print].

Cooper, A., and Dryden, D. T. (1984). Allostery without conformational change. A plausible model. *Eur. Biophys. J.* 11, 103–109. doi: 10.1007/BF00276625

Cortina, G. A., and Kasson, P. M. (2018). Predicting allostery and microbial drug resistance with molecular simulations. *Curr. Opin. Struct. Biol.* 52, 80–86. doi: 10.1016/j.sbi.2018.09.001

Cowan-Jacob, S. W., Jahnke, W., and Knapp, S. (2014). Novel approaches for targeting kinases: allosteric inhibition, allosteric activation and pseudokinases. *Fut. Med. Chem.* 6, 541–561. doi: 10.4155/fmc.13.216

Cronkite-Ratcliff, B., and Pande, V. (2013). MSMExplorer: visualizing markov state models for biomolecule folding simulations. *Bioinformatics* 29, 950–952. doi: 10.1093/bioinformatics/btt051

Csermely, P., Korcsmaros, T., Kiss, H. J., London, G., and Nussinov, R. (2013). Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* 138, 333–408. doi: 10.1016/j.pharmthera.2013.01.016

Csermely, P., Palotai, R., and Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* 35, 539–546. doi: 10.1016/j.tibs.2010.04.009

Csermely, P., Sandhu, K. S., Hazai, E., Hoksza, Z., Kiss, H. J., Miozzo, F., et al. (2012). Disordered proteins and network disorder in network descriptions of protein structure, dynamics and function: hypotheses and a comprehensive review. *Curr. Protein. Pept. Sci.* 13, 19–33. doi: 10.2174/138920312799277992

Czemeres, J., Buse, K., and Verkhivker, G. M. (2017). Atomistic simulations and network-based modeling of the Hsp90-Cdc37 chaperone binding with Cdk4 client protein: a mechanism of chaperoning kinase clients by exploiting weak spots of intrinsically dynamic kinase domains. *PLoS ONE* 12:34. doi: 10.1371/journal.pone.0190267

Dam, T. K., Roy, R., Page, D., and Brewer, C. F. (2002). Negative cooperativity associated with binding of multivalent carbohydrates to lectins. Thermodynamic analysis of the "multivalency effect". *Biochemistry* 41, 1351–1358. doi: 10.1021/bi015830j

D'Annessa, I., Sattin, S., Tao, J., Pennati, M., Sanchez-Martin, C., Moroni, E., et al. (2017). Design of allosteric stimulators of the Hsp90 ATPase as new anticancer leads. *Chemistry* 23, 5188–5192. doi: 10.1002/chem.201700169

Dar, A. C., and Shokat, K. M. (2011). The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling. *Annu. Rev. Biochem.* 80, 769–795. doi: 10.1146/annurev-biochem-090308-173656

Darmon, D., and Rapp, P. E. (2017). Specific transfer entropy and other state-dependent transfer entropies for continuous-state input-output systems. *Phys. Rev. E* 96:022121. doi: 10.1103/PhysRevE.96.022121

Decherchi, S., Berteotti, A., Bottegoni, G., Rocchia, W., and Cavalli, A. (2015). The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. *Nat. Commun.* 6:6155. doi: 10.1038/ncomms7155

Degiacomi, M. T. (2019). Coupling molecular dynamics and deep learning to mine protein conformational space. *Structure* 27, 1034–1040.e1033. doi: 10.1016/j.str.2019.03.018

del Sol, A., Fujihashi, H., Amoros, D., and Nussinov, R. (2006). Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.* 2:2006.0019. doi: 10.1038/msb4100063

del Sol, A., and O'Meara, P. (2005). Small-world network approach to identify key residues in protein-protein interaction. *Proteins* 58, 672–682. doi: 10.1002/prot.20348

del Sol, A., Tsai, C. J., Ma, B., and Nussinov, R. (2009). The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure* 17, 1042–1050. doi: 10.1016/j.str.2009.06.008

Delvenne, J. C., Lambiotte, R., and Rocha, L. E. (2015). Diffusion on networked systems is a question of time or structure. *Nat. Commun.* 6:7366. doi: 10.1038/ncomms8366

Di Paola, L., and Giuliani, A. (2015). Protein contact network topology: a natural language for allostery. *Curr. Opin. Struct. Biol.* 31, 43–48. doi: 10.1016/j.sbi.2015.03.001

Dimitrov, T., Kreisbeck, C., Becker, J. S., Aspuru-Guzik, A., and Saikin, S. K. (2019). Autonomous molecular design: then and now. *ACS Appl. Mater. Interfaces* 11, 24825–24836. doi: 10.1021/acsami.9b01226

Dixit, A., and Verkhivker, G. M. (2009). Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations. *PLoS Comput. Biol.* 5:e1000487. doi: 10.1371/journal.pcbi.1000487

Dixit, A., and Verkhivker, G. M. (2011a). Computational modeling of allosteric communication reveals organizing principles of mutation-induced signaling in ABL and EGFR kinases. *PLoS Comput. Biol.* 7:e1002179. doi: 10.1371/journal.pcbi.1002179

Dixit, A., and Verkhivker, G. M. (2011b). The energy landscape analysis of cancer mutations in protein kinases. *PLoS ONE* 6:e26071. doi: 10.1371/journal.pone.0026071

Dixit, A., and Verkhivker, G. M. (2012). Probing molecular mechanisms of the Hsp90 chaperone: biophysical modeling identifies key regulators of functional dynamics. *PLoS ONE* 7:e37605. doi: 10.1371/journal.pone.0037605

Dixit, A., and Verkhivker, G. M. (2014). Structure-functional prediction and analysis of cancer mutation effects in protein kinases. *Comput. Math. Methods Med.* 2014:653487. doi: 10.1155/2014/653487

Dokholyan, N. V. (2016). Controlling allosteric networks in proteins. *Chem. Rev.* 116, 6463–6487. doi: 10.1021/acs.chemrev.5b00544

Donnelly, A., and Blagg, B. S. (2008). Novobiocin and additional inhibitors of the Hsp90 C-terminal nucleotide-binding pocket. *Curr. Med. Chem.* 15, 2702–2717. doi: 10.2174/092986708786242895

Edler, D., Guedes, T., Zizka, A., Rosvall, M., and Antonelli, A. (2017). Infomap bioregions: interactive mapping of biogeographical regions from species distributions. *Syst. Biol.* 66, 197–204. doi: 10.1093/sysbio/syw087

Fajer, M., Meng, Y., and Roux, B. (2017). The activation of c-Src tyrosine kinase: conformational transition pathway and free energy landscape. *J. Phys. Chem. B* 121, 3352–3363. doi: 10.1021/acs.jpcb.6b08409

Fang, Z., Grutter, C., and Rauh, D. (2013). Strategies for the selective regulation of kinases with allosteric modulators: exploiting exclusive structural features. *ACS Chem. Biol.* 8, 58–70. doi: 10.1021/cb300663j

Farabella, I., Pham, T., Henderson, N. S., Geibel, S., Phan, G., Thanassi, D. G., et al. (2014). Allosteric signalling in the outer membrane translocation domain of PapC usher. *Elife* 3:e0532. doi: 10.7554/eLife.03532.020

Ferraro, M., D'Annessa, I., Moroni, E., Morra, G., Paladino, A., Rinaldi, S., et al. (2019). Allosteric modulators of HSP90 and HSP70: dynamics meets function through structure-based drug design. *J. Med. Chem.* 62, 60–87. doi: 10.1021/acs.jmedchem.8b00825

Frederick, K. K., Marlow, M. S., Valentine, K. G., and Wand, A. J. (2007). Conformational entropy in molecular recognition by proteins. *Nature* 448, 325–329. doi: 10.1038/nature05959

Garg, G., Forsberg, L. K., Zhao, H., and Blagg, B. S. J. (2017a). Development of phenyl cyclohexylcarboxamides as a novel class of Hsp90 C-terminal inhibitors. *Chemistry* 23, 16574–16585. doi: 10.1002/chem.201703206

Garg, G., Khandelwal, A., and Blagg, B. S. (2016). Anticancer inhibitors of Hsp90 function: beyond the usual suspects. *Adv. Cancer Res.* 129, 51–88. doi: 10.1016/bs.acr.2015.12.001

Garg, G., Zhao, H., and Blagg, B. S. (2017b). Design, synthesis and biological evaluation of alkylamino biphenylamides as Hsp90 C-terminal inhibitors. *Bioorg. Med. Chem.* 25, 451–457. doi: 10.1016/j.bmc.2016.11.030

Gasper, P. M., Fuglestad, B., Komives, E. A., Markwick, P. R., and McCammon, J. A. (2012). Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21216–21222. doi: 10.1073/pnas.1218414109

General, I. J., Liu, Y., Blackburn, M. E., Mao, W., Gierasch, L. M., and Bahar, I. (2014). ATPase subdomain IA is a mediator of interdomain allostery in Hsp70 molecular chaperones. *PLoS Comput. Biol.* 10:e1003624. doi: 10.1371/journal.pcbi.1003624

Gestwicki, J. E., and Shao, H. (2019). Inhibitors and chemical probes for molecular chaperone networks. *J. Biol. Chem.* 294, 2151–2161. doi: 10.1074/jbc.TM118.002813

Ghosh, A., Sakaguchi, R., Liu, C., Vishveshwara, S., and Hou, Y. M. (2011). Allosteric communication in cysteinyl tRNA synthetase: a network of direct and indirect readout. *J. Biol. Chem.* 286, 37721–37731. doi: 10.1074/jbc.M111.246702

Ghosh, A., and Vishveshwara, S. (2007). A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15711–15716. doi: 10.1073/pnas.0704459104

Ghosh, A., and Vishveshwara, S. (2008). Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes. *Biochemistry* 47, 11398–11407. doi: 10.1021/bi8007559

Goh, G. B., Hodas, N. O., and Vishnu, A. (2017). Deep learning for computational chemistry. *J. Comput. Chem.* 38, 1291–1307. doi: 10.1002/jcc.24764

Gomez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernandez-Lobato, J. M., Sanchez-Lengeling, B., Sheberla, D., et al. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* 4, 268–276. doi: 10.1021/acscentsci.7b00572

Goncearenco, A., Mitternacht, S., Yong, T., Eisenhaber, B., Eisenhaber, F., and Berezovsky, I. N. (2013). SPACER: server for predicting allosteric communication and effects of regulation. *Nucleic Acids Res.* 41, W266–W272. doi: 10.1093/nar/gkt460

Granata, D., Camilloni, C., Vendruscolo, M., and Laio, A. (2013). Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6817–6822. doi: 10.1073/pnas.1218350110

Grebner, C., Matter, H., Plowright, A. T., and Hessler, G. (2020). Automated *de novo* design in medicinal chemistry: which types of chemistry does a generative neural network learn? *J. Med. Chem.* doi: 10.1021/acs.jmedchem.9b02044. [Epub ahead of print].

Grutsch, S., Bruschweiler, S., and Tollinger, M. (2016). NMR methods to study dynamic allostery. *PLoS Comput. Biol.* 12:e1004620. doi: 10.1371/journal.pcbi.1004620

Guimaraes, G. L., Sanchez-Lengeling, B., Outeiral, C., Farias, P.L. C., and Aspuru-Guzik, A. (2017). Objective-reinforced generative adversarial networks (organ) for sequence generation models. *arXiv [Preprint]*. arXiv:1705.10843.

Gunasekaran, K., Ma, B., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? *Proteins* 57, 433–443. doi: 10.1002/prot.20232

Guo, J., and Zhou, H. X. (2016). Protein allostery and conformational dynamics. *Chem. Rev.* 116, 6503–6515. doi: 10.1021/acs.chemrev.5b00590

Gupta, A., Muller, A. T., Huisman, B. J. H., Fuchs, J. A., Schneider, P., and Schneider, G. (2018). Generative recurrent networks for *de novo* drug design. *Mol. Inform.* 37:1700111. doi: 10.1002/minf.201700111

Hacisuleyman, A., and Erman, B. (2017a). Causality, transfer entropy, and allosteric communication landscapes in proteins with harmonic interactions. *Proteins* 85, 1056–1064. doi: 10.1002/prot.25272

Hacisuleyman, A., and Erman, B. (2017b). Entropy transfer between residue pairs and allostery in proteins: quantifying allosteric communication in ubiquitin. *PLoS Comput. Biol.* 13:e1005319. doi: 10.1371/journal.pcbi.1005319

Halabi, N., Rivoire, O., Leibler, S., and Ranganathan, R. (2009). Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138, 774–786. doi: 10.1016/j.cell.2009.07.038

Hall, J. A., Seedarala, S., Zhao, H., Garg, G., Ghosh, S., and Blagg, B. S. (2016). Novobiocin analogues that inhibit the MAPK pathway. *J. Med. Chem.* 59, 925–933. doi: 10.1021/acs.jmedchem.5b01354

Han, B., Liu, Y., Ginzinger, S. W., and Wishart, D. S. (2011). SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* 50, 43–57. doi: 10.1007/s10858-011-9478-4

Han, R., Chen, K., and Tan, C. (2020). Curiosity-driven recommendation strategy for adaptive learning via deep reinforcement learning. *Br. J. Math. Stat. Psychol.* doi: 10.1111/bmsp.12199. [Epub ahead of print].

Hansia, P., Ghosh, A., and Vishveshwara, S. (2009). Ligand dependent intra and inter subunit communication in human tryptophanyl tRNA synthetase as deduced from the dynamics of structure networks. *Mol. Biosyst.* 5, 1860–1872. doi: 10.1039/b903807h

Harrigan, M. P., Sultan, M. M., Hernandez, C. X., Husic, B. E., Eastman, P., Schwantes, C. R., et al. (2017). MSMBuilder: statistical models for biomolecular dynamics. *Biophys. J.* 112, 10–15. doi: 10.1016/j.bpj.2016.10.042

Hart, K. M., Ho, C. M., Dutta, S., Gross, M. L., and Bowman, G. R. (2016). Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat. Commun.* 7:12965. doi: 10.1038/ncomms12965

Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron* 95, 245–258. doi: 10.1016/j.neuron.2017.06.011

Hayatshahi, H. S., Ahuactzin, E., Tao, P., Wang, S., and Liu, J. (2019). Probing protein allostery as a residue-specific concept via residue response maps. *J. Chem. Inf. Model.* 59, 4691–4705. doi: 10.1021/acs.jcim.9b00447

Hernandez, C. X., Wayment-Steele, H. K., Sultan, M. M., Husic, B. E., and Pande, V. S. (2018). Variational encoding of complex dynamics. *Phys. Rev. E* 97:062412. doi: 10.1103/PhysRevE.97.062412

Hertig, S., Latorraca, N. R., and Dror, R. O. (2016). Revealing atomic-level mechanisms of protein allostery with molecular dynamics simulations. *PLoS Comput. Biol.* 12:e1004746. doi: 10.1371/journal.pcbi.1004746

Hey, T., Butler, K., Jackson, S., and Thiyagalingam, J. (2020). Machine learning and big scientific data. *Philos. Trans. A Math. Phys. Eng. Sci.* 378:20190054. doi: 10.1098/rsta.2019.0054

Hilser, V. J., Wrabl, J. O., and Motlagh, H. N. (2012). Structural and energetic basis of allostery. *Annu. Rev. Biophys.* 41, 585–609. doi: 10.1146/annurev-biophys-050511-102319

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647

Hlavácková-Schindler, K., Paluš, M., Vejmelka, M., and Bhattacharya, J. (2007). Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.* 441, 1–46. doi: 10.1016/j.physrep.2006.12.004

Holzinger, A., Dehmer, M., and Jurisica, I. (2014). Knowledge discovery and interactive data mining in bioinformatics–state-of-the-art, future challenges and research directions. *BMC Bioinformatics* 15(Suppl. 6):I1. doi: 10.1186/1471-2105-15-S6-I1

Hu, J., Stites, E. C., Yu, H., Germino, E. A., Meharena, H. S., Stork, P. J., et al. (2013). Allosteric activation of functionally asymmetric RAF kinase dimers. *Cell* 154, 1036–1046. doi: 10.1016/j.cell.2013.07.046

Hu, S., Ferraro, M., Thomas, A. P., Chung, J. M., Yoon, N. G., Seol, J. H., et al. (2020). Dual binding to orthosteric and allosteric sites enhances the anticancer activity of a TRAP1-targeting drug. *J. Med. Chem.* 63, 2930–2940. doi: 10.1021/acs.jmedchem.9b01420

Huang, C., and Kalodimos, C. G. (2017). Structures of large protein complexes determined by nuclear magnetic resonance spectroscopy. *Annu. Rev. Biophys.* 46, 317–336. doi: 10.1146/annurev-biophys-070816-033701

Husic, B. E., and Pande, V. S. (2018). Markov state models: from an art to a science. *J. Am. Chem. Soc.* 140, 2386–2396. doi: 10.1021/jacs.7b12191

Ito, S. (2016). Backward transfer entropy: informational measure for detecting hidden markov models and its interpretations in thermodynamics, gambling and causality. *Sci. Rep.* 6:36831. doi: 10.1038/srep36831

Jaderberg, M., Czarnecki, W. M., Dunning, I., Marris, L., Lever, G., Castaneda, A. G., et al. (2019). Human-level performance in 3D multiplayer games with population-based reinforcement learning. *Science* 364, 859–865. doi: 10.1126/science.aau6249

James, K. A., and Verkhivker, G. M. (2014). Structure-based network analysis of activation mechanisms in the ErbB family of receptor tyrosine kinases: the regulatory spine residues are global mediators of structural stability and allosteric interactions. *PLoS ONE* 9:e113488. doi: 10.1371/journal.pone.0113488

Jarymowycz, V. A., and Stone, M. J. (2006). Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* 106, 1624–1671. doi: 10.1021/cr040421p

Jiang, Y., and Kalodimos, C. G. (2017). NMR studies of large proteins. *J. Mol. Biol.* 429, 2667–2676. doi: 10.1016/j.jmb.2017.07.007

Jura, N., Zhang, X., Endres, N. F., Seeliger, M. A., Schindler, T., and Kuriyan, J. (2011). Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. *Mol. Cell.* 42, 9–22. doi: 10.1016/j.molcel.2011.03.004

Kadurin, A., Aliper, A., Kazennov, A., Mamoshina, P., Vanhaelen, Q., Khrabrov, K., et al. (2017a). The cornucopia of meaningful leads: applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* 8, 10883–10890. doi: 10.18632/oncotarget.14073

Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., and Zhavoronkov, A. (2017b). druGAN: an advanced generative adversarial autoencoder model for *de novo* generation of new molecules with desired molecular properties *in silico*. *Mol. Pharm.* 14, 3098–3104. doi: 10.1021/acs.molpharmaceut.7b00346

Kalbitzer, H. R., Rosnizeck, I. C., Munte, C. E., Narayanan, S. P., Kropf, V., and Spoerner, M. (2013). Intrinsic allosteric inhibition of signaling proteins by targeting rare interaction states detected by high-pressure NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* 52, 14242–14246. doi: 10.1002/anie.201305741

Kalescky, R., Liu, J., and Tao, P. (2015). Identifying key residues for protein allostery through rigid residue scan. *J. Phys. Chem. A* 119, 1689–1700. doi: 10.1021/jp5083455

Kalescky, R., Zhou, H., Liu, J., and Tao, P. (2016). Rigid residue scan simulations systematically reveal residue entropic roles in protein allostery. *PLoS Comput. Biol.* 12:e1004893. doi: 10.1371/journal.pcbi.1004893

Kalodimos, C. G. (2011). NMR reveals novel mechanisms of protein activity regulation. *Protein Sci.* 20, 773–782. doi: 10.1002/pro.614

Kalodimos, C. G. (2012). Protein function and allostery: a dynamic relationship. *Ann. N Y Acad. Sci.* 1260, 81–86. doi: 10.1111/j.1749-6632.2011.06319.x

Kawamoto, T., and Rosvall, M. (2015). Estimating the resolution limit of the map equation in community detection. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* 91:012809. doi: 10.1103/PhysRevE.91.012809

Kay, L. E. (2011). NMR studies of protein structure and dynamics - a look backwards and forwards. *J. Magn. Reson.* 213, 492–494. doi: 10.1016/j.jmr.2011.08.010

Kay, L. E. (2016). New views of functionally dynamic proteins by solution NMR spectroscopy. *J. Mol. Biol.* 428(Pt A), 323–331. doi: 10.1016/j.jmb.2015.11.028

Kaya, C., Armutlulu, A., Ekesan, S., and Haliloglu, T. (2013). MCPath: monte carlo path generation approach to predict likely allosteric pathways and functional residues. *Nucleic Acids Res.* 41, W249–W255. doi: 10.1093/nar/gkt284

Kern, D., and Zuiderweg, E. R. (2003). The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* 13, 748–757. doi: 10.1016/j.sbi.2003.10.008

Khandelwal, A., Crowley, V. M., and Blagg, B. S. J. (2016). Natural product inspired N-terminal Hsp90 inhibitors: from bench to bedside? *Med. Res. Rev.* 36, 92–118. doi: 10.1002/med.21351

Kim, J., Ahuja, L. G., Chao, F. A., Xia, Y., McClendon, C. L., Kornev, A. P., et al. (2017). A dynamic hydrophobic core orchestrates allostery in protein kinases. *Sci. Adv.* 3:e1600663. doi: 10.1126/sciadv.1600663

Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H. J., et al. (2019). Deep learning in medical imaging. *Neurospine* 16, 657–668. doi: 10.14245/ns.1938396.198

Kimanius, D., Pettersson, I., Schluckebier, G., Lindahl, E., and Andersson, M. (2015). SAXS-guided metadynamics. *J. Chem. Theory Comput.* 11, 3491–3498. doi: 10.1021/acs.jctc.5b00299

Kohlhoff, K. J., Robustelli, P., Cavalli, A., Salvatella, X., and Vendruscolo, M. (2009). Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.* 131, 13894–13895. doi: 10.1021/ja903772t

Kornev, A. P., and Taylor, S. S. (2015). Dynamics-driven allostery in protein kinases. *Trends Biochem. Sci.* 40, 628–647. doi: 10.1016/j.tibs.2015.09.002

Kornev, A. P., Taylor, S. S., and Ten Eyck, L. F. (2008). A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14377–14382. doi: 10.1073/pnas.0807988105

Korotcov, A., Tkachenko, V., Russo, D. P., and Ekins, S. (2017). Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. *Mol. Pharm.* 14, 4462–4475. doi: 10.1021/acs.molpharmaceut.7b00578

Korzhnev, D. M., and Kay, L. E. (2008). Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* 41, 442–451. doi: 10.1021/ar700189y

Koshland, D. E. Jr. (1998). Conformational changes: how small is big enough? *Nat. Med.* 4, 1112–1114. doi: 10.1038/2605

Krivdin, L. B. (2017). Calculation of (15)N NMR chemical shifts: recent advances and perspectives. *Prog. Nucl. Magn. Reson. Spectrosc.* 102–103, 98–119. doi: 10.1016/j.pnmrs.2017.08.001

Kumar MV, V., Ebna Noor, R., Davis, R. E., Zhang, Z., Sipavicius, E., Keramisanou, D., et al. (2018). Molecular insights into the interaction of Hsp90 with allosteric inhibitors targeting the C-terminal domain. *MedChemComm* 9, 1323–1331. doi: 10.1039/C8MD00151K

Kuzmanic, A., Bowman, G. R., Juarez-Jimenez, J., Michel, J., and Gervasio, F. L. (2020). Investigating cryptic binding sites by molecular dynamics simulations. *Acc. Chem. Res.* 53, 654–661. doi: 10.1021/acs.accounts.9b00613

Kuzmanic, A., Sutto, L., Saladino, G., Nebreda, A. R., Gervasio, F. L., and Orozco, M. (2017). Changes in the free-energy landscape of p38alpha MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations. *Elife* 6:e22175. doi: 10.7554/eLife.22175.024

La Sala, G., Decherchi, S., De Vivo, M., and Rocchia, W. (2017). Allosteric communication networks in proteins revealed through pocket crosstalk analysis. *ACS Cent. Sci.* 3, 949–960. doi: 10.1021/acscentsci.7b00211

Lambiotte, R., Rosvall, M., and Scholtes, I. (2019). From networks to optimal higher-order models of complex systems. *Nat. Phys.* 15, 313–320. doi: 10.1038/s41567-019-0459-y

Lambiotte, R., Sinatra, R., Delvenne, J. C., Evans, T. S., Barahona, M., and Latora, V. (2011). Flow graphs: interweaving dynamics and structure. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* 84(Pt. 2):017102. doi: 10.1103/PhysRevE.84.017102

Lancichinetti, A., and Fortunato, S. (2009). Community detection algorithms: a comparative analysis. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* 80(Pt. 2):056117. doi: 10.1103/PhysRevE.80.056117

Lawless, N., Blacklock, K., Berrigan, E., and Verkhivker, G. (2013). Structural bioinformatics and protein docking analysis of the molecular chaperone-kinase interactions: towards allosteric inhibition of protein kinases by targeting the hsp90-cdc37 chaperone machinery. *Pharmaceuticals* 6, 1407–1428. doi: 10.3390/ph6111407

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Levinson, N. M. (2018). The multifaceted allosteric regulation of aurora kinase A. *Biochem. J.* 475, 2025–2042. doi: 10.1042/BCJ20170771

Levinson, N. M., and Boxer, S. G. (2012). Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. *PLoS ONE* 7:e29828. doi: 10.1371/journal.pone.0029828

Levinson, N. M., and Boxer, S. G. (2014). A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. *Nat. Chem. Biol.* 10, 127–132. doi: 10.1038/nchembio.1404

Levinson, N. M., Kuchment, O., Shen, K., Young, M. A., Koldobskiy, M., Karplus, M., et al. (2006). A Src-like inactive conformation in the abl tyrosine kinase domain. *PLoS Biol* 4:e144. doi: 10.1371/journal.pbio.0040144

Li, C., Ma, N., Wang, Y., and Chen, G. (2014). Molecular dynamics simulation studies on the positive cooperativity of the Kemptide substrate with protein kinase A induced by the ATP ligand. *J. Phys. Chem. B* 118, 1273–1287. doi: 10.1021/jp411111g

Li, Q., Luo, R., and Chen, H. F. (2019). Dynamical important residue network (DIRN): network inference via conformational change. *Bioinformatics* 35, 4664–4670. doi: 10.1093/bioinformatics/btz298

Li, Z., Kermode, J. R., and De Vita, A. (2015). Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* 114:096405. doi: 10.1103/PhysRevLett.114.096405

Liang, Z., Verkhivker, G. M., and Hu, G. (2019). Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *Brief. Bioinform.* 21, 815–835. doi: 10.1093/bib/bbz029

Limongelli, V., Bonomi, M., and Parrinello, M. (2013). Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6358–6363. doi: 10.1073/pnas.1303186110

Lin, Y. L., Meng, Y., Huang, L., and Roux, B. (2014). Computational study of gleevec and G6G reveals molecular determinants of kinase inhibitor selectivity. *J. Am. Chem. Soc.* 136, 14753–14762. doi: 10.1021/ja504146x

Lin, Y. L., Meng, Y., Jiang, W., and Roux, B. (2013). Explaining why gleevec is a specific and potent inhibitor of Abl kinase. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1664–1669. doi: 10.1073/pnas.1214330110

Lin, Y. L., and Roux, B. (2013). Computational analysis of the binding specificity of gleevec to Abl, c-Kit, Lck, and c-Src tyrosine kinases. *J. Am. Chem. Soc.* 135, 14741–14753. doi: 10.1021/ja405939x

Lisi, G. P., and Loria, J. P. (2016). Solution NMR spectroscopy for the study of enzyme allostery. *Chem. Rev.* 116, 6323–6369. doi: 10.1021/acs.chemrev.5b00541

Lisi, G. P., and Loria, J. P. (2017). Allostery in enzyme catalysis. *Curr. Opin. Struct. Biol.* 47, 123–130. doi: 10.1016/j.sbi.2017.08.002

Liu, J., and Nussinov, R. (2016). Allostery: an overview of its history, concepts, methods, and applications. *PLoS Comput. Biol.* 12:e1004966. doi: 10.1371/journal.pcbi.1004966

Lockless, S. W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286, 295–299. doi: 10.1126/science.286.5438.295

Long, D., Bouvignies, G., and Kay, L. E. (2014). Measuring hydrogen exchange rates in invisible protein excited states. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8820–8825. doi: 10.1073/pnas.1405011111

Long, D., and Bruschweiler, R. (2011). Atomistic kinetic model for population shift and allostery in biomolecules. *J. Am. Chem. Soc.* 133, 18999–19005. doi: 10.1021/ja208813t

Ma, B., Tsai, C. J., Haliloglu, T., and Nussinov, R. (2011). Dynamic allostery: linkers are not merely flexible. *Structure* 19, 907–917. doi: 10.1016/j.str.2011.06.002

Malmstrom, R. D., Kornev, A. P., Taylor, S. S., and Amaro, R. E. (2015). Allostery through the computational microscope: cAMP activation of a canonical signalling domain. *Nat. Commun.* 6:7588. doi: 10.1038/ncomms8588

Marcu, M. G., Chadli, A., Bouhouche, I., Catelli, M., and Neckers, L. M. (2000a). The heat shock protein 90 antagonist novobiocin interacts with a previously unrecognized ATP-binding domain in the carboxyl terminus of the chaperone. *J. Biol. Chem.* 275, 37181–37186. doi: 10.1074/jbc.M003701200

Marcu, M. G., Schulte, T. W., and Neckers, L. (2000b). Novobiocin and related coumarins and depletion of heat shock protein 90-dependent signaling proteins. *J. Natl. Cancer Inst.* 92, 242–248. doi: 10.1093/jnci/92.3.242

Mardt, A., Pasquali, L., Wu, H., and Noe, F. (2018). VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* 9:5. doi: 10.1038/s41467-017-02388-1

Masgras, I., Sanchez-Martin, C., Colombo, G., and Rasola, A. (2017). The chaperone TRAP1 as a modulator of the mitochondrial adaptations in cancer cells. *Front. Oncol.* 7:58. doi: 10.3389/fonc.2017.00058

Mater, A. C., and Coote, M. L. (2019). Deep learning in chemistry. *J. Chem. Inf. Model.* 59, 2545–2559. doi: 10.1021/acs.jcim.9b00266

Matts, R. L., Brandt, G. E. L., Lu, Y. M., Dixit, A., Mollapour, M., Wang, S. Q., et al. (2011a). A systematic protocol for the characterization of Hsp90 modulators. *Bioorg. Med. Chem.* 19, 684–692. doi: 10.1016/j.bmc.2010.10.029

Matts, R. L., Dixit, A., Peterson, L. B., Sun, L., Voruganti, S., Kalyanaraman, P., et al. (2011b). Elucidation of the Hsp90 C-terminal inhibitor binding site. *ACS Chem. Biol.* 6, 800–807. doi: 10.1021/cb200052x

Maziarka, Ł., Pocha, A., Kaczmarczyk, J., Rataj, K., Danel, T., and Warchoł, M. (2020). Mol-CycleGAN: a generative model for molecular optimization. *J. Cheminform.* 12:2. doi: 10.1186/s13321-019-0404-1

McClendon, C. L., Friedland, G., Mobley, D. L., Amirkhani, H., and Jacobson, M. P. (2009). Quantifying correlations between allosteric sites in thermodynamic ensembles. *J. Chem. Theory Comput.* 5, 2486–2502. doi: 10.1021/ct9001812

McGibbon, R. T., Schwantes, C. R., and Pande, V. S. (2014). Statistical model selection for markov models of biomolecular dynamics. *J. Phys. Chem. B* 118, 6475–6481. doi: 10.1021/jp411822r

McLaughlin, R. N. Jr., Poelwijk, F. J., Raman, A., Gosal, W. S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. *Nature* 491, 138–142. doi: 10.1038/nature11500

McLeish, T. C., Rodgers, T. L., and Wilson, M. R. (2013). Allostery without conformation change: modelling protein dynamics at multiple scales. *Phys. Biol.* 10:056004. doi: 10.1088/1478-3975/10/5/056004

Meharena, H. S., Chang, P., Keshwani, M. M., Oruganty, K., Nene, A. K., Kannan, N., et al. (2013). Deciphering the structural basis of eukaryotic protein kinase regulation. *PLoS Biol.* 11:e1001680. doi: 10.1371/journal.pbio.1001680

Meng, Y., Pond, M. P., and Roux, B. (2017). Tyrosine kinase activation and conformational flexibility: lessons from src-family tyrosine kinases. *Acc. Chem. Res.* 50, 1193–1201. doi: 10.1021/acs.accounts.7b00012

Meng, Y., and Roux, B. (2014). Locking the active conformation of c-Src kinase through the phosphorylation of the activation loop. *J. Mol. Biol.* 426, 423–435. doi: 10.1016/j.jmb.2013.10.001

Ming, D., and Wall, M. E. (2005). Quantifying allosteric effects in proteins. *Proteins* 59, 697–707. doi: 10.1002/prot.20440

Ming, D., and Wall, M. E. (2006). Interactions in native binding sites cause a large change in protein dynamics. *J. Mol. Biol.* 358, 213–223. doi: 10.1016/j.jmb.2006.01.097

Mittermaier, A., and Kay, L. E. (2006). New tools provide new insights in NMR studies of protein dynamics. *Science* 312, 224–228. doi: 10.1126/science.1124964

Mittermaier, A. K., and Kay, L. E. (2009). Observing biological dynamics at atomic resolution using NMR. *Trends Biochem. Sci.* 34, 601–611. doi: 10.1016/j.tibs.2009.07.004

Mitternacht, S., and Berezovsky, I. N. (2011). Binding leverage as a molecular basis for allosteric regulation. *PLoS Comput. Biol.* 7:e1002148. doi: 10.1371/journal.pcbi.1002148

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529–533. doi: 10.1038/nature14236

Monod, J., Wyman, J., and Changeux, J. P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12, 88–118. doi: 10.1016/S0022-2836(65)80285-6

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., et al. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1293–1301. doi: 10.1073/pnas.1111471108

Morra, G., Neves, M. A. C., Plescia, C. J., Tsustsumi, S., Neckers, L., Verkhivker, G., et al. (2010). Dynamics-based discovery of allosteric inhibitors: selection of new ligands for the C-terminal domain of Hsp90. *J. Chem. Theory Comput.* 6, 2978–2989. doi: 10.1021/ct100334n

Morra, G., Potestio, R., Micheletti, C., and Colombo, G. (2012). Corresponding functional dynamics across the Hsp90 chaperone family: insights from a multiscale analysis of MD simulations. *PLoS Comput. Biol.* 8:e1002433. doi: 10.1371/journal.pcbi.1002433

Morra, G., Verkhivker, G., and Colombo, G. (2009). Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer. *PLoS Comput. Biol.* 5:e1000323. doi: 10.1371/journal.pcbi.1000323

Motlagh, H. N., Wrabl, J. O., Li, J., and Hilser, V. J. (2014). The ensemble nature of allostery. *Nature* 508, 331–339. doi: 10.1038/nature13001

Munte, C. E., Beck Erlach, M., Kremer, W., Koehler, J., and Kalbitzer, H. R. (2013). Distinct conformational states of the alzheimer beta-amyloid peptide can be detected by high-pressure NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* 52, 8943–8947. doi: 10.1002/anie.201301537

Narayanan, C., Bafna, K., Roux, L. D., Agarwal, P. K., and Doucet, N. (2017). Applications of NMR and computational methodologies to study protein dynamics. *Arch. Biochem. Biophys.* 628, 71–80. doi: 10.1016/j.abb.2017.05.002

Nerli, S., McShan, A. C., and Sgourakis, N. G. (2018). Chemical shift-based methods in NMR structure determination. *Prog. Nucl Magn. Reson. Spectrosc.* 106–107, 1–25. doi: 10.1016/j.pnmrs.2018.03.002

Nesmelova, I. V., Ermakova, E., Daragan, V. A., Pang, M., Menendez, M., Lagartera, L., et al. (2010). Lactose binding to galectin-1 modulates structural dynamics, increases conformational entropy, and occurs with apparent negative cooperativity. *J. Mol. Biol.* 397, 1209–1230. doi: 10.1016/j.jmb.2010.02.033

Nishi, H., Hashimoto, K., and Panchenko, A. R. (2011). Phosphorylation in protein-protein binding: effect on stability and function. *Structure* 19, 1807–1815. doi: 10.1016/j.str.2011.09.021

Noe, F., and Clementi, C. (2017). Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. *Curr. Opin. Struct. Biol.* 43, 141–147. doi: 10.1016/j.sbi.2017.02.006

Nussinov, R. (2012). How do dynamic cellular signals travel long distances? *Mol. Biosyst.* 8, 22–26. doi: 10.1039/C1MB05205E

Nussinov, R., and Tsai, C. J. (2015). Allostery without a conformational change? Revisiting the paradigm. *Curr. Opin. Struct. Biol.* 30, 17–24. doi: 10.1016/j.sbi.2014.11.005

Oleinikovas, V., Saladino, G., Cossins, B. P., and Gervasio, F. L. (2016). Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *J. Am. Chem. Soc.* 138, 14257–14263. doi: 10.1021/jacs.6b05425

Olivecrona, M., Blaschke, T., Engkvist, O., and Chen, H. (2017). Molecular *de-novo* design through deep reinforcement learning. *J. Cheminform.* 9:48. doi: 10.1186/s13321-017-0235-x

Olson, R. S., La Cava, W., Orzechowski, P., Urbanowicz, R. J., and Moore, J. H. (2017). PMLB: a large benchmark suite for machine learning evaluation and comparison. *BioData Min.* 10:36. doi: 10.1186/s13040-017-0154-4

Olsson, S., Wu, H., Paul, F., Clementi, C., and Noe, F. (2017). Combining experimental and simulation data of molecular processes via augmented markov models. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8265–8270. doi: 10.1073/pnas.1704803114

Paladino, A., Morra, G., and Colombo, G. (2015). Structural stability and flexibility direct the selection of activating mutations in epidermal growth factor receptor kinase. *J. Chem. Inf. Model.* 55, 1377–1387. doi: 10.1021/acs.jcim.5b00270

Palazzesi, F., Barducci, A., Tollinger, M., and Parrinello, M. (2013). The allosteric communication pathways in KIX domain of CBP. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14237–14242. doi: 10.1073/pnas.1313548110

Palazzesi, F., Valsson, O., and Parrinello, M. (2017). Conformational entropy as collective variable for proteins. *J. Phys. Chem. Lett.* 8, 4752–4756. doi: 10.1021/acs.jpclett.7b01770

Palermo, G., Ricci, C. G., Fernando, A., Basak, R., Jinek, M., Rivalta, I., et al. (2017). Protospacer adjacent motif-induced allostery activates CRISPR-Cas9. *J. Am. Chem. Soc.* 139, 16028–16031. doi: 10.1021/jacs.7b05313

Pande, V. S. (2014). Understanding protein folding using markov state models. *Adv. Exp. Med. Biol.* 797, 101–106. doi: 10.1007/978-94-007-7606-7_8

Panjkovich, A., and Daura, X. (2012). Exploiting protein flexibility to predict the location of allosteric sites. *BMC Bioinformatics* 13:273. doi: 10.1186/1471-2105-13-273

Panjkovich, A., and Daura, X. (2014). PARS: a web server for the prediction of protein allosteric and regulatory sites. *Bioinformatics* 30, 1314–1315. doi: 10.1093/bioinformatics/btu002

Perez-Hernandez, G., and Noe, F. (2016). Hierarchical time-lagged independent component analysis: computing slow modes and reaction coordinates for large molecular systems. *J. Chem. Theory Comput.* 12, 6118–6129. doi: 10.1021/acs.jctc.6b00738

Polykovskiy, D., Zhebrak, A., Vetrov, D., Ivanenkov, Y., Aladinskiy, V., Mamoshina, P., et al. (2018). Entangled conditional adversarial autoencoder for *de novo* drug discovery. *Mol. Pharm.* 15, 4398–4405. doi: 10.1021/acs.molpharmaceut.8b00839

Popova, M., Isayev, O., and Tropsha, A. (2018). Deep reinforcement learning for *de novo* drug design. *Sci. Adv.* 4:eaap7885. doi: 10.1126/sciadv.aap7885

Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006). Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* 13, 831–838. doi: 10.1038/nsmb1132

Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., et al. (2011). Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* 134:174105. doi: 10.1063/1.3565032

Putin, E., Asadulaev, A., Ivanenkov, Y., Aladinskiy, V., Sanchez-Lengeling, B., Aspuru-Guzik, A., et al. (2018a). Reinforced adversarial neural computer for *de novo* molecular design. *J. Chem. Inf. Model.* 58, 1194–1204. doi: 10.1021/acs.jcim.7b00690

Putin, E., Asadulaev, A., Vanhaelen, Q., Ivanenkov, Y., Aladinskaya, A. V., Aliper, A., et al. (2018b). Adversarial threshold neural computer for molecular *de novo* design. *Mol. Pharm.* 15, 4386–4397. doi: 10.1021/acs.molpharmaceut.7b01137

Racz, A., Bajusz, D., and Heberger, K. (2019). Multi-level comparison of machine learning classifiers and their performance metrics. *Molecules* 24:2811. doi: 10.3390/molecules24152811

Ricci, C. G., Silveira, R. L., Rivalta, I., Batista, V. S., and Skaf, M. S. (2016). Allosteric pathways in the PPARgamma-RXRalpha nuclear receptor complex. *Sci. Rep.* 6:19940. doi: 10.1038/srep19940

Rinaldi, S., Assimon, V. A., Young, Z. T., Morra, G., Shao, H., Taylor, I. R., et al. (2018). A local allosteric network in heat shock protein 70 (Hsp70) links inhibitor binding to enzyme activity and distal protein-protein interactions. *ACS Chem. Biol.* 13, 3142–3152. doi: 10.1021/acschembio.8b00712

Rivalta, I., Sultan, M. M., Lee, N. S., Manley, G. A., Loria, J. P., and Batista, V. S. (2012). Allosteric pathways in imidazole glycerol phosphate synthase. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1428–1436. doi: 10.1073/pnas.1120536109

Robustelli, P., Kohlhoff, K., Cavalli, A., and Vendruscolo, M. (2010). Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* 18, 923–933. doi: 10.1016/j.str.2010.04.016

Robustelli, P., Stafford, K. A., and Palmer, A. G. 3rd. (2012). Interpreting protein structural dynamics from NMR chemical shifts. *J. Am. Chem. Soc.* 134, 6365–6374. doi: 10.1021/ja300265w

Rosenzweig, R., and Kay, L. E. (2014). Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu. Rev. Biochem.* 83, 291–315. doi: 10.1146/annurev-biochem-060713-035829

Rosvall, M., and Bergstrom, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7327–7331. doi: 10.1073/pnas.0611034104

Rosvall, M., and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1118–1123. doi: 10.1073/pnas.0706851105

Rosvall, M., and Bergstrom, C. T. (2010). Mapping change in large networks. *PLoS ONE* 5:e8694. doi: 10.1371/journal.pone.0008694

Rosvall, M., and Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS ONE* 6:e18209. doi: 10.1371/journal.pone.0018209

Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D., and Lambiotte, R. (2014). Memory in network flows and its effects on spreading dynamics and community detection. *Nat. Commun.* 5:4630. doi: 10.1038/ncomms5630

Ruff, E. F., Muretta, J. M., Thompson, A. R., Lake, E. W., Cyphers, S., Albanese, S. K., et al. (2018). A dynamic mechanism for allosteric activation of aurora kinase A by activation loop phosphorylation. *Elife* 7:e32766. doi: 10.7554/eLife.32766.019

Rupp, M., Tkatchenko, A., Muller, K. R., and von Lilienfeld, O. A. (2012). Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* 108:058301. doi: 10.1103/PhysRevLett.108.058301

Saleh, T., Rossi, P., and Kalodimos, C. G. (2017). Atomic view of the energy landscape in the allosteric regulation of Abl kinase. *Nat. Struct. Mol. Biol.* 24, 893–901. doi: 10.1038/nsmb.3470

Salnikov, V., Schaub, M. T., and Lambiotte, R. (2016). Using higher-order markov models to reveal flow-based communities in networks. *Sci. Rep.* 6:23194. doi: 10.1038/srep23194

Sanchez-Martin, C., Moroni, E., Ferraro, M., Laquatra, C., Cannino, G., Masgras, I., et al. (2020). Rational design of allosteric and selective inhibitors of the molecular chaperone TRAP1. *Cell. Rep.* 31:107531. doi: 10.1016/j.celrep.2020.107531

Sattin, S., Tao, J., Vettoretti, G., Moroni, E., Pennati, M., Lopergolo, A., et al. (2015). Activation of Hsp90 enzymatic activity and conformational dynamics through rationally designed allosteric ligands. *Chemistry* 21, 13598–13608. doi: 10.1002/chem.201502211

Schaub, M. T., Lambiotte, R., and Barahona, M. (2012). Encoding dynamics for multiscale community detection: markov time sweeping for the map equation. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* 86(Pt. 2):026112. doi: 10.1103/PhysRevE.86.026112

Schoepfer, J., Jahnke, W., Berellini, G., Buonamici, S., Cotesta, S., Cowan-Jacob, S. W., et al. (2018). Discovery of asciminib (ABL001), an allosteric inhibitor of the tyrosine kinase activity of BCR-ABL1. *J. Med. Chem.* 61, 8120–8135. doi: 10.1021/acs.jmedchem.8b01040

Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464. doi: 10.1103/PhysRevLett.85.461

Schulze, J. O., Saladino, G., Busschots, K., Neimanis, S., Suss, E., Odadzic, D., et al. (2016). Bidirectional allosteric communication between the ATP-binding site and the regulatory PIF pocket in PDK1 protein kinase. *Cell. Chem. Biol.* 23, 1193–1205. doi: 10.1016/j.chembiol.2016.06.017

Sekhar, A., and Kay, L. E. (2013). NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12867–12874. doi: 10.1073/pnas.1305688110

Sekhar, A., and Kay, L. E. (2019). An NMR view of protein dynamics in health and disease. *Annu. Rev. Biophys.* 48, 297–319. doi: 10.1146/annurev-biophys-052118-115647

Selvaratnam, R., Chowdhury, S., VanSchouwen, B., and Melacini, G. (2011). Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6133–6138. doi: 10.1073/pnas.1017311108

Selvaratnam, R., VanSchouwen, B., Fogolari, F., Mazhab-Jafari, M. T., Das, R., and Melacini, G. (2012). The projection analysis of NMR chemical shifts reveals extended EPAC autoinhibition determinants. *Biophys. J.* 102, 630–639. doi: 10.1016/j.bpj.2011.12.030

Sengupta, U., and Strodel, B. (2018). Markov models for the elucidation of allosteric regulation. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 373:20170178. doi: 10.1098/rstb.2017.0178

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2019). Protein structure prediction using multiple deep neural networks in the 13th critical assessment of protein structure prediction (CASP13). *Proteins* 87, 1141–1148. doi: 10.1002/prot.25834

Sethi, A., Eargle, J., Black, A. A., and Luthey-Schulten, Z. (2009). Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6620–6625. doi: 10.1073/pnas.0810961106

Sethi, A., Tian, J., Derdeyn, C. A., Korber, B., and Gnanakaran, S. (2013). A mechanistic understanding of allosteric immune escape pathways in the HIV-1 envelope glycoprotein. *PLoS Comput. Biol.* 9:e1003046. doi: 10.1371/journal.pcbi.1003046

Shamsi, Z., Cheng, K. J., and Shukla, D. (2018). Reinforcement learning based adaptive sampling: REAPing rewards by exploring protein conformational landscapes. *J. Phys. Chem. B* 122, 8386–8395. doi: 10.1021/acs.jpcb.8b06521

Shan, Y., Arkhipov, A., Kim, E. T., Pan, A. C., and Shaw, D. E. (2013). Transitions to catalytically inactive conformations in EGFR kinase. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7270–7275. doi: 10.1073/pnas.1220843110

Shan, Y., Eastwood, M. P., Zhang, X., Kim, E. T., Arkhipov, A., Dror, R. O., et al. (2012). Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell* 149, 860–870. doi: 10.1016/j.cell.2012.02.063

Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., Seeliger, M. A., and Shaw, D. E. (2011). How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* 133, 9181–9183. doi: 10.1021/ja202726y

Shaw, A. S., Kornev, A. P., Hu, J., Ahuja, L. G., and Taylor, S. S. (2014). Kinases and pseudokinases: lessons from RAF. *Mol. Cell. Biol.* 34, 1538–1546. doi: 10.1128/MCB.00057-14

Shcherbinin, D., Veselovsky, A., Rubtsova, M., Grigorenko, V., and Egorov, A. (2019). The impact of long-distance mutations on the omega-loop conformation in TEM type beta-lactamases. *J. Biomol. Struct. Dyn.* 38, 2369–2376. doi: 10.1080/07391102.2019.1634642

Shen, Y., and Bax, A. (2010). SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* 48, 13–22. doi: 10.1007/s10858-010-9433-9

Shi, L., and Kay, L. E. (2014). Tracing an allosteric pathway regulating the activity of the HslV protease. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2140–2145. doi: 10.1073/pnas.1318476111

Shukla, D., Hernandez, C. X., Weber, J. K., and Pande, V. S. (2015). Markov state models provide insights into dynamic modulation of protein function. *Acc. Chem. Res.* 48, 414–422. doi: 10.1021/ar5002999

Shukla, D., Meng, Y., Roux, B., and Pande, V. S. (2014). Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* 5:3397. doi: 10.1038/ncomms4397

Shukla, S., Shamsi, Z., Moffett, A. S., Selvam, B., and Shukla, D. (2017). Application of hidden markov models in biomolecular simulations. *Methods Mol. Biol.* 1552, 29–41. doi: 10.1007/978-1-4939-6753-7_3

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of go without human knowledge. *Nature* 550, 354–359. doi: 10.1038/nature24270

Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M., and Marino Buslje, C. (2013). MISTIC: mutual information server to infer coevolution. *Nucleic Acids Res.* 41, W8–14. doi: 10.1093/nar/gkt427

Smock, R. G., and Gierasch, L. M. (2009). Sending signals dynamically. *Science* 324, 198–203. doi: 10.1126/science.1169377

Socolich, M., Lockless, S. W., Russ, W. P., Lee, H., Gardner, K. H., and Ranganathan, R. (2005). Evolutionary information for specifying a protein fold. *Nature* 437, 512–518. doi: 10.1038/nature03991

Sorin, V., Barash, Y., Konen, E., and Klang, E. (2020). Creating artificial images for radiology applications using generative adversarial networks (GANs) – a systematic review. *Acad. Radiol.* doi: 10.1016/j.acra.2019.12.024. [Epub ahead of print].

Sprangers, R., Velyvis, A., and Kay, L. E. (2007). Solution NMR of supramolecular complexes: providing new insights into function. *Nat. Methods* 4, 697–703. doi: 10.1038/nmeth1080

Stetz, G., Tse, A., and Verkhivker, G. M. (2017). Ensemble-based modeling and rigidity decomposition of allosteric interaction networks and communication pathways in cyclin-dependent kinases: differentiating kinase clients of the Hsp90-Cdc37 chaperone. *PLoS ONE* 12:e0186089. doi: 10.1371/journal.pone.0186089

Stetz, G., and Verkhivker, G. M. (2015). Dancing through life: molecular dynamics simulations and network-centric modeling of allosteric mechanisms in Hsp70 and Hsp110 chaperone proteins. *PLoS ONE* 10:e0143752. doi: 10.1371/journal.pone.0143752

Stetz, G., and Verkhivker, G. M. (2016). Probing allosteric inhibition mechanisms of the Hsp70 chaperone proteins using molecular dynamics simulations and analysis of the residue interaction networks. *J. Chem. Inf. Model.* 56, 1490–1517. doi: 10.1021/acs.jcim.5b00755

Stetz, G., and Verkhivker, G. M. (2017). Computational analysis of residue interaction networks and coevolutionary relationships in the Hsp70 chaperones: a community-hopping model of allosteric regulation and communication. *PLoS Comput. Biol.* 13:e1005299. doi: 10.1371/journal.pcbi.1005299

Stetz, G., and Verkhivker, G. M. (2018). Functional role and hierarchy of the intermolecular interactions in binding of protein kinase clients to the Hsp90-Cdc37 chaperone: structure-based network modeling of allosteric regulation. *J. Chem. Inf. Model.* 58, 405–421. doi: 10.1021/acs.jcim.7b00638

Stevens, S. Y., Sanker, S., Kent, C., and Zuiderweg, E. R. (2001). Delineation of the allosteric mechanism of a cytidylyltransferase exhibiting negative cooperativity. *Nat. Struct. Biol.* 8, 947–952. doi: 10.1038/nsb1101-947

Stock, G., and Hamm, P. (2018). A non-equilibrium approach to allosteric communication. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 373:20170187. doi: 10.1098/rstb.2017.0187

Stolzenberg, S., Michino, M., LeVine, M. V., Weinstein, H., and Shi, L. (2016). Computational approaches to detect allosteric pathways in transmembrane molecular machines. *Biochim. Biophys. Acta* 1858(Pt. B), 1652–1662. doi: 10.1016/j.bbamem.2016.01.010

Suel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* 10, 59–69. doi: 10.1038/nsb881

Sutto, L., and Gervasio, F. L. (2013). Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10616–10621. doi: 10.1073/pnas.1221953110

Sutton, R. S., and Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* 88, 135–170. doi: 10.1037/0033-295X.88.2.135

Swain, J. F., and Gierasch, L. M. (2006). The changing landscape of protein allostery. *Curr. Opin. Struct. Biol.* 16, 102–108. doi: 10.1016/j.sbi.2006.01.003

Taylor, S. S., Ilouz, R., Zhang, P., and Kornev, A. P. (2012a). Assembly of allosteric macromolecular switches: lessons from PKA. *Nat. Rev. Mol. Cell. Biol.* 13, 646–658. doi: 10.1038/nrm3432

Taylor, S. S., Keshwani, M. M., Steichen, J. M., and Kornev, A. P. (2012b). Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 367, 2517–2528. doi: 10.1098/rstb.2012.0054

Taylor, S. S., and Kornev, A. P. (2011). Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* 36, 65–77. doi: 10.1016/j.tibs.2010.09.006

Tehver, R., Chen, J., and Thirumalai, D. (2009). Allostery wiring diagrams in the transitions that drive the GroEL reaction cycle. *J. Mol. Biol.* 387, 390–406. doi: 10.1016/j.jmb.2008.12.032

Toledano, D. T., Fernandez-Gallego, M. P., and Lozano-Diez, A. (2018). Multi-resolution speech analysis for automatic speech recognition using deep neural networks: experiments on TIMIT. *PLoS ONE* 13:e0205355. doi: 10.1371/journal.pone.0205355

Tsai, C. J., del Sol, A., and Nussinov, R. (2008). Allostery: absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.* 378, 1–11. doi: 10.1016/j.jmb.2008.02.034

Tsai, C. J., Del Sol, A., and Nussinov, R. (2009). Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol. Biosyst.* 5, 207–216. doi: 10.1039/b819720b

Tsai, C. J., and Nussinov, R. (2014). A unified view of "how allostery works". *PLoS Comput. Biol.* 10:e1003394. doi: 10.1371/journal.pcbi.1003394

Tse, A., and Verkhivker, G. M. (2015a). Molecular determinants underlying binding specificities of the ABL kinase inhibitors: combining alanine scanning of binding hot spots with network analysis of residue interactions and coevolution. *PLoS ONE* 10:e0130203. doi: 10.1371/journal.pone.0130203

Tse, A., and Verkhivker, G. M. (2015b). Molecular dynamics simulations and structural network analysis of c-Abl and c-Src kinase core proteins: capturing allosteric mechanisms and communication pathways from residue centrality. *J. Chem. Inf. Model.* 55, 1645–1662. doi: 10.1021/acs.jcim.5b00240

Tse, A., and Verkhivker, G. M. (2015c). Small-world networks of residue interactions in the Abl kinase complexes with cancer drugs: topology of allosteric communication pathways can determine drug resistance effects. *Mol. Biosyst.* 11, 2082–2095. doi: 10.1039/C5MB00246J

Tsuchiya, Y., Taneishi, K., and Yonezawa, Y. (2019). Autoencoder-based detection of dynamic allostery triggered by ligand binding based on molecular dynamics. *J. Chem. Inf. Model.* 59, 4043–4051. doi: 10.1021/acs.jcim.9b00426

Tzeng, S. R., and Kalodimos, C. G. (2009). Dynamic activation of an allosteric regulatory protein. *Nature* 462, 368–372. doi: 10.1038/nature08560

Tzeng, S. R., and Kalodimos, C. G. (2011). Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* 21, 62–67. doi: 10.1016/j.sbi.2010.10.007

Vallurupalli, P., Bouvignies, G., and Kay, L. E. (2012). Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* 134, 8148–8161. doi: 10.1021/ja3001419

Vanwart, A. T., Eargle, J., Luthey-Schulten, Z., and Amaro, R. E. (2012). Exploring residue component contributions to dynamical network models of allostery. *J. Chem. Theory Comput.* 8, 2949–2961. doi: 10.1021/ct300377a

Vendruscolo, M., Dokholyan, N. V., Paci, E., and Karplus, M. (2002). Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* 65(Pt. 1):061910. doi: 10.1103/PhysRevE.65.061910

Verkhivker, G., Blacklock, K., and Buchner, J. (2016). Dissecting allosteric regulatory mechanisms of the Hsp90 chaperone interactions with the protein kinase clients: Integrating structural bioinformatics with multiscale atomistic simulations and biophysical experiments. *Abst. Pap. Am. Chem. Soc.* 251:2.

Verkhivker, G. M. (2014). Computational studies of allosteric regulation in the hsp90 molecular chaperone: from functional dynamics and protein structure networks to allosteric communications and targeted anti-cancer modulators. *Israel J. Chem.* 54, 1052–1064. doi: 10.1002/ijch.201300143

Verkhivker, G. M. (2018a). Computational modeling of the Hsp90 interactions with cochaperones and small-molecule inhibitors. *Methods Mol. Biol.* 1709, 253–273. doi: 10.1007/978-1-4939-7477-1_19

Verkhivker, G. M. (2018b). Dynamics-based community analysis and perturbation response scanning of allosteric interaction networks in the TRAP1 chaperone structures dissect molecular linkage between conformational asymmetry and sequential ATP hydrolysis. *Biochim. Biophys. Acta Proteins Proteom.* 1866, 899–912. doi: 10.1016/j.bbapap.2018.04.008

Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Freer, S. T., and Rose, P. W. (2002). Complexity and simplicity of ligand-macromolecule interactions: the energy landscape perspective. *Curr. Opin. Struct. Biol.* 12, 197–203. doi: 10.1016/S0959-440X(02)00310-X

Verkhivker, G. M., Dixit, A., Morra, G., and Colombo, G. (2009). Structural and computational biology of the molecular chaperone Hsp90: from understanding molecular mechanisms to computer-based inhibitor design. *Curr. Top. Med. Chem.* 9, 1369–1385. doi: 10.2174/156802609789895700

Vijayabaskar, M. S., and Vishveshwara, S. (2010). Interaction energy based protein structure networks. *Biophys. J.* 99, 3704–3715. doi: 10.1016/j.bpj.2010.08.079

Wang, F., Shen, L., Zhou, H., Wang, S., Wang, X., and Tao, P. (2019). Machine learning classification model for functional binding modes of TEM-1 beta-lactamase. *Front. Mol. Biosci.* 6:47. doi: 10.3389/fmolb.2019.00047

Williamson, M. P., and Kitahara, R. (2019). Characterization of low-lying excited states of proteins by high-pressure NMR. *Biochim. Biophys. Acta Proteins Proteom.* 1867, 350–358. doi: 10.1016/j.bbapap.2018.10.014

Wodak, S. J., Paci, E., Dokholyan, N. V., Berezovsky, I. N., Horovitz, A., Li, J., et al. (2019). Allostery in its many disguises: from theory to applications. *Structure* 27, 566–578. doi: 10.1016/j.str.2019.01.003

Wrabl, J. O., Gu, J., Liu, T., Schrank, T. P., Whitten, S. T., and Hilser, V. J. (2011). The role of protein conformational fluctuations in allostery, function, and evolution. *Biophys. Chem.* 159, 129–141. doi: 10.1016/j.bpc.2011.05.020

Wu, H., Paul, F., Wehmeyer, C., and Noe, F. (2016). Multiensemble markov models of molecular thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3221–E3230. doi: 10.1073/pnas.1525092113

Wu, J., Yilmaz, E., Zhang, M., Li, H., and Tan, K. C. (2020). Deep spiking neural networks for large vocabulary automatic speech recognition. *Front. Neurosci.* 14:199. doi: 10.3389/fnins.2020.00199

Wylie, A. A., Schoepfer, J., Jahnke, W., Cowan-Jacob, S. W., Loo, A., Furet, P., et al. (2017). The allosteric inhibitor ABL001 enables dual targeting of BCR-ABL1. *Nature* 543, 733–737. doi: 10.1038/nature21702

Xia, J., Deng, N. J., and Levy, R. M. (2013). NMR relaxation in proteins with fast internal motions and slow conformational exchange: model-free framework and markov state simulations. *J. Phys. Chem. B* 117, 6625–6634. doi: 10.1021/jp400797y

Yang, S., Banavali, N. K., and Roux, B. (2009). Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3776–3781. doi: 10.1073/pnas.0808261106

Yang, S., and Roux, B. (2008). Src kinase conformational activation: thermodynamics, pathways, and mechanisms. *PLoS Comput. Biol.* 4:e1000047. doi: 10.1371/journal.pcbi.1000047

Yang, Y. I., Niu, H., and Parrinello, M. (2018). Combining metadynamics and integrated tempering sampling. *J. Phys. Chem. Lett.* 9, 6426–6430. doi: 10.1021/acs.jpclett.8b03005

Yu, L., Zhang, W., Wang, J., and Yu, Y. (2017). SeqGAN: sequence generative adversarial nets with policy gradient. *arXiv [Preprint]*. arXiv:1609.05473.

Yuwen, T., Sekhar, A., and Kay, L. E. (2017). Separating dipolar and chemical exchange magnetization transfer processes in (1) H-CEST. *Angew. Chem. Int. Ed. Engl.* 56, 6122–6125. doi: 10.1002/anie.201610759

Zaiter, S. S., Huo, Y., Tiew, F. Y., Gestwicki, J. E., and McAlpine, S. R. (2019). Designing *de novo* small molecules that control heat shock protein 70 (Hsp70) and heat shock organizing protein (HOP) within the chaperone protein-folding machinery. *J. Med. Chem.* 62, 742–761. doi: 10.1021/acs.jmedchem.8b01436

Zhang, J., Adrian, F. J., Jahnke, W., Cowan-Jacob, S. W., Li, A. G., Iacob, R. E., et al. (2010). Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* 463, 501–506. doi: 10.1038/nature08675

Zhang, X., Gureasko, J., Shen, K., Cole, P. A., and Kuriyan, J. (2006). An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* 125, 1137–1149. doi: 10.1016/j.cell.2006.05.013

Zhong, G., Gao, W., Liu, Y., Yang, Y., Wang, D. H., and Huang, K. (2020). Generative adversarial networks with decoder-encoder output noises. *Neural Netw.* 127, 19–28. doi: 10.1016/j.neunet.2020.04.005

Zhou, H., Dong, Z., and Tao, P. (2018a). Recognition of protein allosteric states and residues: machine learning approaches. *J. Comput. Chem.* 39, 1481–1490. doi: 10.1002/jcc.25218

Zhou, H., Dong, Z., Verkhivker, G., Zoltowski, B. D., and Tao, P. (2019a). Allosteric mechanism of the circadian protein vivid resolved through markov state model and machine learning analysis. *PLoS Comput. Biol.* 15:e1006801. doi: 10.1371/journal.pcbi.1006801

Zhou, H., and Tao, P. (2018). Dynamics sampling in transition pathway space. *J. Chem. Theory Comput.* 14, 14–29. doi: 10.1021/acs.jctc.7b00606

Zhou, H., and Tao, P. (2019). REDAN: Relative entropy-based dynamical allosteric network model. *Mol. Phys.* 117, 1334–1343. doi: 10.1080/00268976.2018.1543904

Zhou, H., Wang, F., Bennett, D. I. G., and Tao, P. (2019b). Directed kinetic transition network model. *J. Chem. Phys.* 151:144112. doi: 10.1063/1.5110896

Zhou, H., Wang, F., and Tao, P. (2018b). t-Distributed stochastic neighbor embedding method with the least information loss for macromolecular simulations. *J. Chem. Theory Comput.* 14, 5499–5510. doi: 10.1021/acs.jctc.8b00652

Zhou, H., Zoltowski, B. D., and Tao, P. (2017). Revealing hidden conformational space of LOV protein VIVID through rigid residue scan simulations. *Sci. Rep.* 7:46626. doi: 10.1038/srep46626

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)* (Venice: IEEE), 2242–2251. doi: 10.1109/ICCV.2017.244

Zhuravlev, P. I., and Papoian, G. A. (2010). Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. *Q Rev. Biophys.* 43, 295–332. doi: 10.1017/S0033583510000119

Zimmerman, M. I., and Bowman, G. R. (2015). FAST conformational searches by balancing exploration/exploitation trade-offs. *J. Chem. Theory Comput.* 11, 5747–5757. doi: 10.1021/acs.jctc.5b00737

Zimmerman, M. I., Porter, J. R., Sun, X., Silva, R. R., and Bowman, G. R. (2018). Choice of adaptive sampling strategy impacts state discovery, transition probabilities, and the apparent mechanism of conformational changes. *J. Chem. Theory Comput.* 14, 5459–5475. doi: 10.1021/acs.jctc.8b00500

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 10 July 2020
doi: 10.3389/fmolb.2020.00145

Nucleotide-Specific Autoinhibition of Full-Length K-Ras4B Identified by Extensive Conformational Sampling

*Balint Dudas[1,2], Franci Merzel[3], Hyunbum Jang[4], Ruth Nussinov[4,5], David Perahia[6] and Erika Balog[1]**

[1] Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, Hungary, [2] Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary, [3] Theory Department, National Institute of Chemistry, Ljubljana, Slovenia, [4] Computational Structural Biology Section, Basic Science Program, Frederick National Laboratory for Cancer Research, Frederick, MD, United States, [5] Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel, [6] Laboratoire de Biologie et de Pharmacologie Appliquée, Ecole Normale Supérieure Paris-Saclay, Gif-sur-Yvette, France

K-Ras is one of the most frequently mutated oncogenes in human tumor cells. It consists of a well-conserved globular catalytic domain and a flexible tail-like hypervariable region (HVR) at its C-terminal end. It plays a key role in signaling networks in proliferation, differentiation, and survival, undergoing a conformational switch between the active and inactive states. It is regulated through the GDP-GTP cycle of the inactive GDP-bound and active GTP-bound states. Here, without imposing any prior constraints, we mapped the interaction pattern between the catalytic domain and the HVR using Molecular Dynamics with excited Normal Modes (MDeNM) starting from an initially extended HVR conformation for both states. Our sampling captured similar interaction patterns in both GDP- and GTP-bound states with shifted populations depending on the bound nucleotide. In the GDP-bound state, the conformations where the HVR interacts with the effector lobe are more populated than in the GTP-bound state, forming a buried thus autoinhibited catalytic site; in the GTP-bound state conformations where the HVR interacts with the allosteric lobe are more populated, overlapping the α3/α4 dimerization interface. The interaction of the GTP with Switch I and Switch II is stronger than that of the GDP in line with a decrease in the fluctuation upon GTP binding.

Keywords: K-Ras4B, KRas4B, autoinhibition, molecular dynamics, normal modes, conformational search, MDeNM

OPEN ACCESS

Edited by:
Pemra Doruker, University of Pittsburgh, United States

Reviewed by:
Turkan Haliloglu, Bogaziçi University, Turkey
*Igor N. Berezovsky, Bioinformatics Institute (A*STAR), Singapore*

***Correspondence:**
Erika Balog
balog.erika@med.semmelweis-univ.hu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *19 April 2020*
Accepted: *11 June 2020*
Published: *10 July 2020*

Citation:
Dudas B, Merzel F, Jang H, Nussinov R, Perahia D and Balog E (2020) Nucleotide-Specific Autoinhibition of Full-Length K-Ras4B Identified by Extensive Conformational Sampling. Front. Mol. Biosci. 7:145. doi: 10.3389/fmolb.2020.00145

INTRODUCTION

Members of the Ras (Rat sarcoma) family of small GTPases are conformational switches involved in signal transduction originating from receptor-mediated extracellular signals to the nucleus, controlling cellular proliferation, differentiation, and survival (Hernandez-Alcoceba et al., 2000; Cherfils and Zeghouf, 2013; van Hattum and Waldmann, 2014).

Ras signaling is determined by the GTPase cycle: inactive GDP-bound and active GTP-bound states, which change the conformation of Ras and its affinity to bind to downstream effectors—such as Raf kinase (Pacold et al., 2000; Fetics et al., 2015) and phosphatidylinositol 3-kinase (PI3K) (Pacold et al., 2000). The intrinsically very slow GDP/GTP exchange and GTP hydrolysis rates are increased by two types of regulatory proteins. Guanine nucleotide exchange factors (GEFs) catalyze the release of GDP which is followed by GTP binding, and the GTPase activating

proteins (GAPs) catalyze GTP hydrolysis (Bos et al., 2007), resulting in GDP-bound conformations. Oncogenic mutations lock Ras in its active, GTP-bound conformation, being always available to downstream effectors, which leads to uncontrolled cell growth and cancer (Prior et al., 2012). Mutant GDP-bound proteins may also shift their conformations to a GTP-bound-like state.

Three main isoforms of human Ras, N-, H-, and K-Ras, the latter having two splice variants, K-Ras4A and K-Ras4B, are known. K-Ras4B has been observed at higher levels (Koera et al., 1997; Plowman et al., 2003) and plays an essential role in cell growth and development. All Ras isoforms share a highly conserved catalytic domain (sequence identity >89%) and a flexible, C-terminal hypervariable region (HVR, sequence homology < 15%) (Castellano and Santos, 2011). For structural details see **Figure 1**.

To anchor in the membrane and signal, K-Ras4B undergoes post-translational modifications (PTM), including methylation and farnesylation of its CAAX C-terminal box at Cys-185. This farnesyl group together with the lysine-rich HVR direct the interaction of K-Ras4B with the negatively charged lipids of the inner face of the plasma membrane (Brunsveld et al., 2009; Jang et al., 2015, 2016a; Prakash et al., 2016; Prakash and Gorfe, 2017). Emerging nuclear magnetic resonance (NMR) spectroscopy data and computational studies show that in solution, the catalytic domain and the HVR can interact (Abraham et al., 2009; Lu et al., 2015). NMR measurements showed that the HVR of H-Ras dynamically interacts with the catalytic domain, which exhibits increased flexibility in the truncation of the HVR (Thapar et al., 2004). Based on NMR interaction patterns, recent computational studies provided several possible models of the complete K-Ras4B in solution (Chavan et al., 2015; Lu et al., 2015; Jang et al., 2016a). The molecular dynamics (MD) simulations showed that full-length GDP-bound K-Ras4B could promote an autoinhibited state through HVR-catalytic domain interactions, while looser interactions have been detected for the GTP-bound state, which could release autoinhibition. Autoinhibition is typically a transient state, which explains the difficulties in obtaining its structure experimentally, e.g., by NMR or crystallography. Autoinhibition protects against spurious activation and proteolysis (reviewed in Nussinov et al., 2018). Even if the interaction of the autoinhibiting segment is weak, its large population at the active/functional site effectively shields it, which explains why oncogenic drivers often aim to release it (Nussinov et al., 2020).

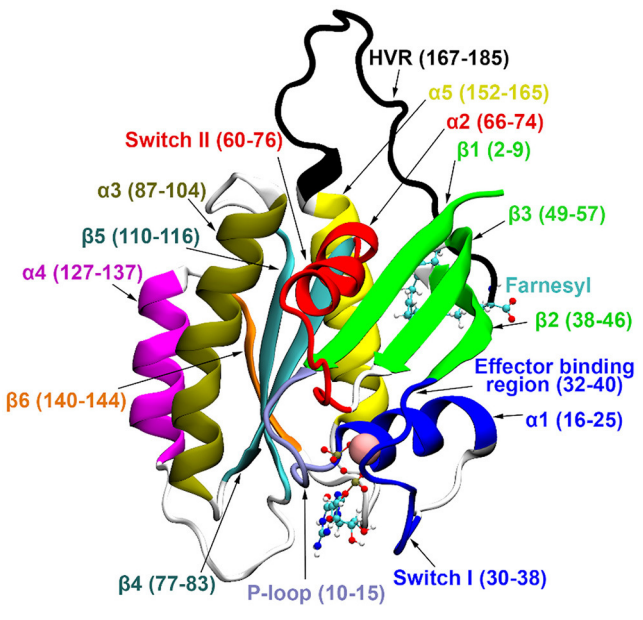


FIGURE 1 | K-Ras4B secondary structure elements. GDP and farnesyl are represented in CPK representation, Mg^{2+} with vdW sphere.

Here, we perform a conformational space mapping of the full-length GDP- and GTP-bound K-Ras4B in solution without imposing any prior knowledge or constraints on the catalytic domain-HVR interactions. We start from an extended HVR conformation and use a computational method that combines MD and the normal mode approach, called Molecular Dynamics with excited Normal Modes (MDeNM) (Costa et al., 2015). In agreement with previous studies, we found that GDP-bound full-length K-Ras4B favors an autoinhibited conformation, while in solution the GTP-bound protein favors a conformation where the autoinhibition is lifted leading to an active state. The autoinhibition in the GDP-bound state is realized by the interaction of HVR with Switch I and Switch II at the effector lobe, thereby blocking the effector binding site. In turn, the conformations populated in the GTP-bound state exhibit an HVR interaction with α4 at the allosteric lobe, thus blocking the α3/α4 dimerization site. In the GDP-bound state, these conformations hardly exist. In particular, our results show that compared to the GDP-bound form, the interaction of the GTP nucleotide is stronger with both Switch I and Switch II in agreement with the decrease in fluctuation upon GTP binding, and the measured binding affinity, Kd values (John et al., 1993).

MATERIALS AND METHODS

Molecular Dynamics with excited Normal Modes (MDeNM) (Costa et al., 2015) simulations were carried out on full-length (residues 1-185) farnesylated and methylated (FME) K-Ras4B in its inactive GDP-bound and active GTP-bound forms. The starting coordinates of the catalytic domain were taken from crystal structures with PDB ID 4OBE (Hunter et al., 2014) and 3GFT, respectively. For the active state, the GTP analog GppNHp was modified to GTP, and His61 was mutated back to the native Gln61.

MDeNM simulations and analysis were performed with CHARMM (Brooks et al., 2009) using CHARMM all-atom additive force field C36 (Best et al., 2012) while conventional MD simulations were carried out with NAMD (Phillips et al., 2005) using the same CHARMM force field mentioned above. The GDP/GTP parameters were combined from the ADP and guanine parameters existing in CHARMM, while the parameters for farnesylated Cys were taken from our previous studies (Jang et al., 2016b).

Our main objective in carrying out the MDeNM simulations is a deep exploration of the conformational preferences of the two Ras forms. The structures of the catalytic domain deposited in Protein Data Bank differ in length (e.g., in 4EPT the structure is determined up to amino acid 166; in 3GFT up to 167, in 4OBE to 169), indicating that the C-terminal amino acids have higher fluctuation. Consequently, for the GDP bound state we created three initial structures with different HVR orientations. In these structures, the HVR was built starting at residues 167, 168, and 169, respectively, from the internal coordinate table of CHARMM, yielding a linear conformation distant from the catalytic domain. In order to relax the HVR, the first 100 steps of steepest descent were followed by 1,000 steps of adopted basis Newton-Raphson energy minimization. Then the HVR of these structures was heated to 300 K in 10 ps followed by a 90 ps equilibration while the catalytic domain was kept fixed. This procedure leads to HVR conformations that face different sides of the catalytic domain, as can be seen in **Supplementary Figure 1**. In order to have the same initial HVR conformations to study the nucleotide dependent catalytic domain-HVR interactions in an unbiased manner, the same three orientations of the HVR were applied to the GTP-bound state using 3GFT for the structure of the catalytic domain as described above.

The six obtained structures were solvated using CHARMM-GUI (Jo et al., 2008; Lee et al., 2016). In all cases, a rectangular box containing TIP3 water molecules extending 15 Å in all directions from the surface of the protein was generated with a concentration of 0.10 M NaCl. For the energy calculations, the dielectric constant was set to 1. The Particle Mesh Ewald (PME) method was used to calculate the electrostatic interactions with a grid spacing of 1 Å or less having the order of 6; the real space summation was truncated at 12.0 Å, and the width of the Gaussian distribution was set to 0.34 $Å^{-1}$. Van der Waals (vdW) interactions were reduced to zero by "switch" truncation operating between 10.0 and 12.0 Å.

Solvated systems were energy minimized with progressively decreasing harmonic restraints: first, the steepest descent was used with the harmonic force constant decreased every 500 steps of 10, 1, and 0.1 kcal/mol/$Å^2$, followed by 200 steps of conjugate gradient with a 0.1 kcal/mol/$Å^2$ force constant. Unrestrained minimization was then applied for 100 steps with the steepest descent, 200 steps with the conjugate gradient, and 1,000 steps with the adopted basis Newton-Raphson method. The energy-minimized structures were heated and equilibrated at 300 K for 200 ps in an NVT ensemble, followed by a 5 ns NPT run at a pressure of 1 atm. The Langevin dynamics was used with the damping coefficient of 1 ps^{-1}, a piston oscillation period of 50 fs, and a piston oscillation decay time of 25 fs. The integration time step was set to 2 fs.

MDeNM Simulations

In order to map the conformational space more efficiently than with classical MD simulations, the Molecular Dynamics with excited Normal Modes (MDeNM) (Costa et al., 2015) method was used. The normal modes necessary for the MDeNM simulations were calculated by considering the final structures resulting from the 5 ns MD run for both GDP- and GTP-bound FME full-length K-Ras4B. The energy of the two structures was minimized using the steepest descent method, the harmonic force constant decreasing every 1,000 steps, adopting the values 10, 1, 0.1, and 0 kcal/mol/$Å^2$, followed by 50,000 steps of adopted basis Newton-Raphson method. After the energy minimization, the normal modes were calculated using the VIBRAN module of CHARMM. For the MDeNM calculations, based on their RMSF contribution, the 10 lowest frequency normal modes were taken.

In the second step, the final structure of the 5 ns MD run for all 3 models of both GDP- and GTP-bound states (in total six systems) was considered as initial structures for MDeNM simulations: randomized linear combinations of the first 10 lowest frequency normal modes were generated, giving excitation directions. These excitation directions were then used to kinetically excite the systems during MD simulations yielding to different replicas. The excitations were carried out during the MD simulations with successive kinetic energy injection along the direction of the combined mode in the form of velocity increment equivalent to an overall 10 K temperature increment of the system. Excitations were performed in the same direction at every 2,500 steps of the MD simulation, giving a relaxation time of 5 ps for the system in each excitation cycle. The dissipation of the energy inserted in each cycle was checked by to ensure that there is no appreciable accumulation of kinetic energy along the excitation direction. Each period of excitation-relaxation yields a given conformation. The other parameters of the MD remained the same as those described above.

In total 264 MDeNM replica simulations were carried out corresponding to different NM combination directions for the three models in GDP- and GTP-bound states. In order to ensure an exhaustive search of the conformational space, the newly generated replicas were compared to the previously accepted ones and were only kept if the root-mean-square deviation (RMSD) value—between the structures displaced by 1 Å along the mode combinations—was greater than 1.65 Å. 32 excitations per replica were generated, resulting in 8,448 structures for each of the systems.

MD Simulations

In order to compare the conformational space mapped by MDeNM to the conformations accessible by conventional MD simulation, three parallel 200 ns long MD simulations were performed for both the GDP- and GTP-bound states, having the same starting structures as the MDeNM simulations (i.e., the final structure of the 5 ns equilibration run). The parameters for the 200 ns run were identical to those of the 5 ns equilibration.

Interaction Energy Calculations

The binding preferences of the HVR with the allosteric or effector lobe regions in the GDP- or GTP-bound K-Ras4B were analyzed by comparing distributions of HVR interaction energies, ΔE_{HVR}. The interaction energy of a given structure was evaluated as a difference between the energies of bound and a reference unbound state in which the HVR is completely detached, as

given by

$$\Delta E_{\mathrm{HVR(a/e)}} = E_{\mathrm{bound(a/e)}} - \langle E_{\mathrm{non-bound}} \rangle \quad (1) \qquad (1)$$

where $E_{bound(a/e)}$ is the HVR interaction energy of a structure within either the allosteric (a) or effector (e) lobe and $<E_{non-bound}>$ the reference energy of its unbound state. The interaction energy is evaluated as a sum of pairwise electrostatic and vdW potential energy contributions between the HVR and its environment including the protein and the solvent by using CHARMM36 force field. The interaction energy of the reference state, in which the HVR is exposed exclusively to the solvent composed of explicit water molecules and ions, is obtained as an average over a sufficiently large number of conformations (15,500) to ensure a converging value. The simulations were carried out using the Hungarian KIFU supercomputing facility.

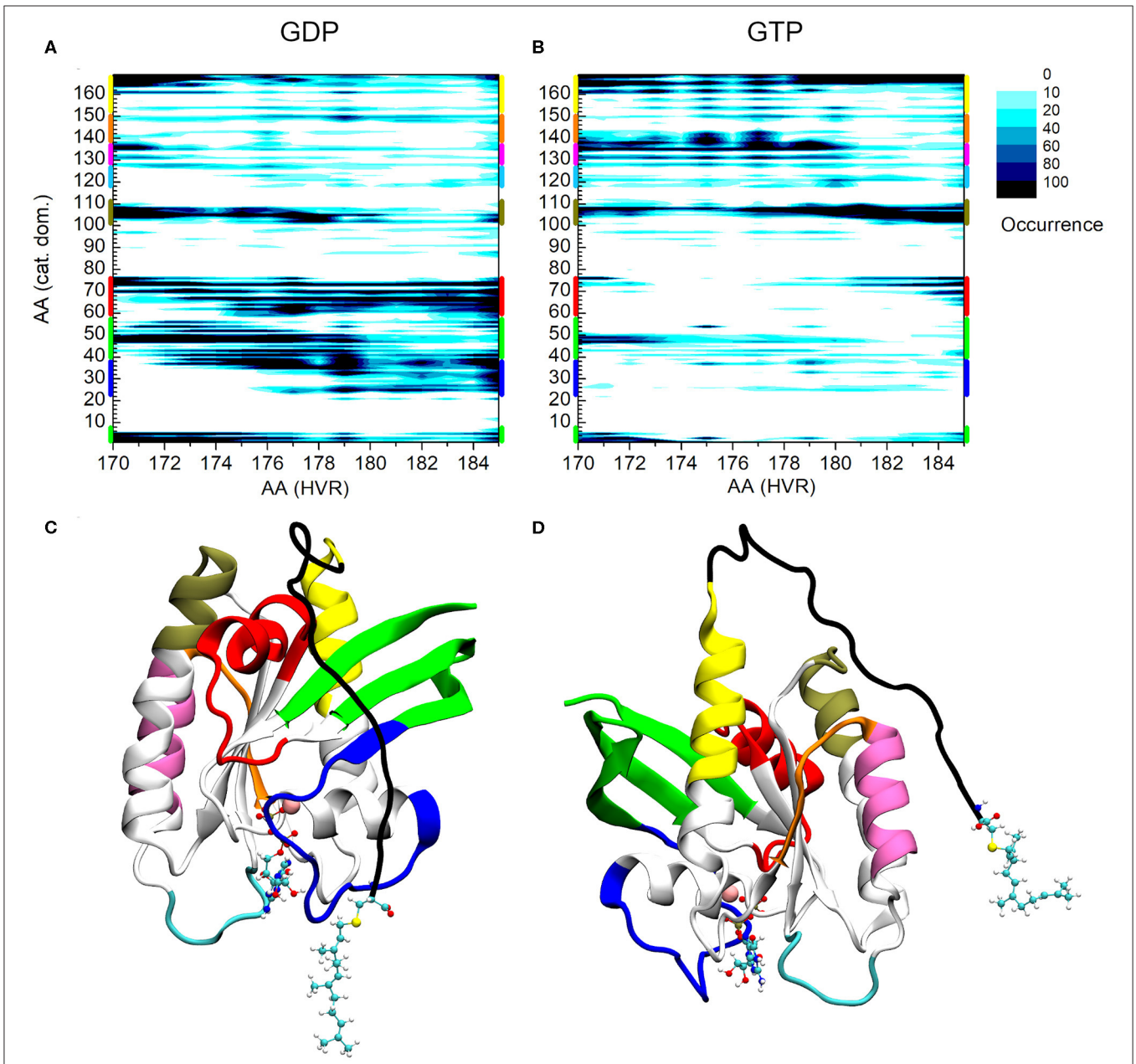


FIGURE 2 | Distance-based interaction map between the catalytic domain and the HVR of K-Ras4B. **(A)** GDP-bound, **(B)** GTP-bound state. **(C)** 3D structure of K-Ras4B with the effector lobe and **(D):** the allosteric lobe in front. The nucleotide and the farnesyl group are represented as CPK, Mg^{2+} as VDW sphere. Structural regions of the catalytic domain color-coded along the y-axis of **(A)** and **(B)** are mapped with identical coloring onto the C and D part of the figure.

RESULTS AND DISCUSSION

Interactions of the HVR With the Catalytic Domain

In order to identify the interacting residues of the catalytic domain and the HVR, a distance-based criterion was applied to the conformations generated by MDeNM: if the distance between two heavy atoms of a residue within the catalytic domain and a residue in the HVR was less than 5.5 Å, these two residues were considered interacting (Bowerman and Wereszczynski, 2016). Based on this criterion, our analysis shows that the interacting catalytic domain-HVR residues detected by MDeNM contain all the interacting residues identified by NMR measurements (Thapar et al., 2004; Chavan et al., 2015; Lu et al., 2015; Jang et al., 2016a).

Figure 2 shows the interaction map based on the calculations described above for both the GDP- and GTP-bound structures. The number of interactions within a given population goes from white (no structure with such an interaction) to black (maximum number of structures with the given interaction), as indicated in the legend. On the x-axis HVR residue numbers, on the y-axis those of the catalytic domain are presented. **Figure 2A** shows the pairs of interacting residues for the GDP-bound state, **Figure 2B** those for the GTP-bound state. In order to identify the structural elements of the catalytic domain which interact with the HVR, different regions are designated by bars of different colors on the vertical edges of the graphs. The same color coding is kept for **Figures 2C,D**, where the three-dimensional structure of the protein is represented from different views.

A striking difference can be observed between the GDP- and GTP-bound states. In the GDP-bound form, the HVR shows extensive interactions with Switch I (blue) (including the effector binding site), β1, β2, and β3 (green), and Switch II (red) (**Figure 2A**), while in the GTP-bound form, these interactions are either sparsely populated or non-existent (**Figure 2B**). This is in agreement with previous observations (Chavan et al., 2015; Lu et al., 2015; Jang et al., 2016a), demonstrating that in the GDP-bound state, the HVR hinder approach to the effector binding

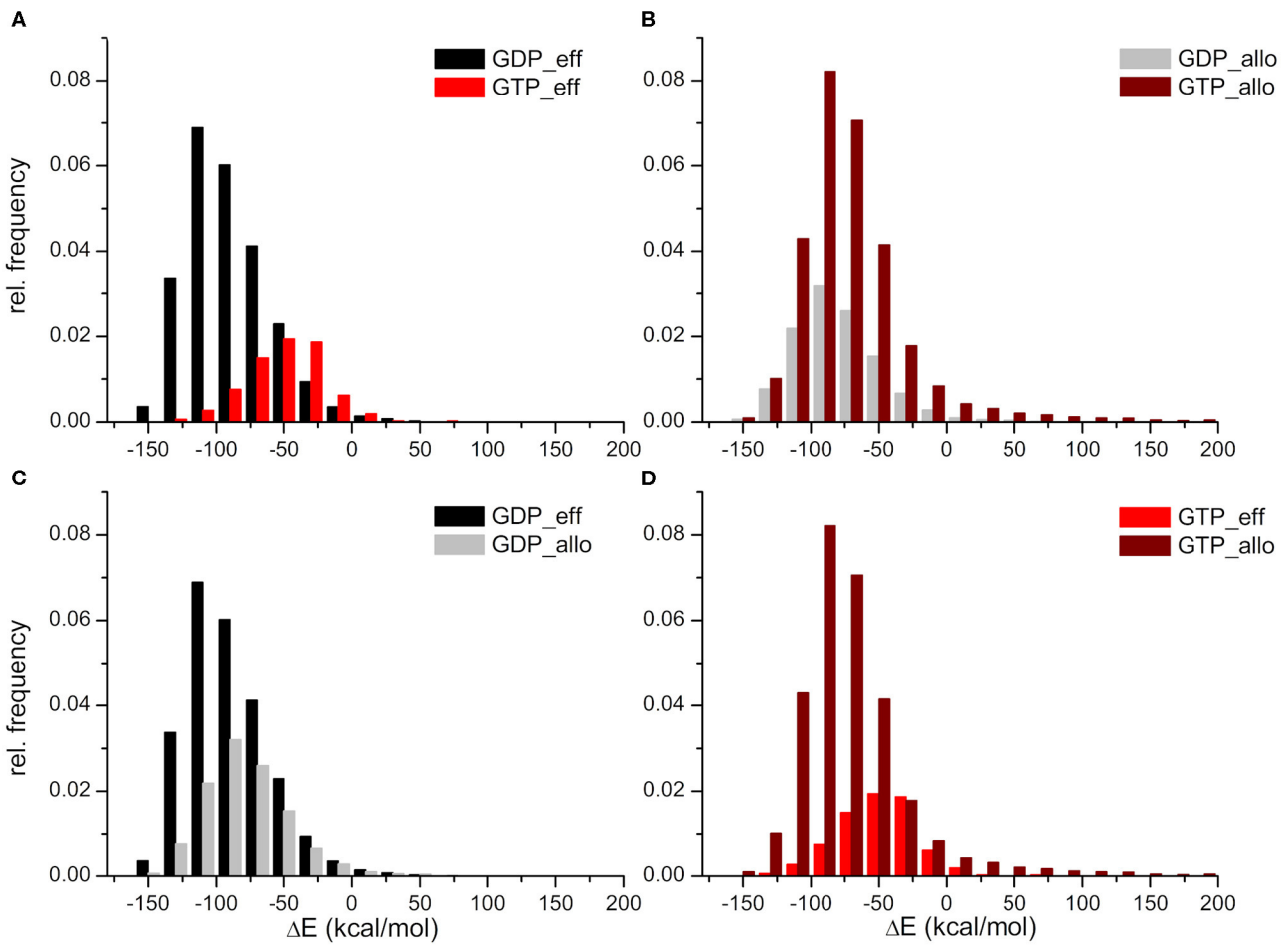


FIGURE 3 | Distributions of interaction energies between the HVR and the catalytic domain of K-Ras4B, within the GDP- and GTP-bound population. HVR interactions **(A)** with the effector lobe in the GDP-bound (black) and GTP-bound (red) state; **(B)** with the allosteric lobe in the GDP-bound (gray) and GTP-bound (wine) state; **(C)** with the effector (black) and the allosteric (gray) lobe in the GDP-bound state; and finally **(D)** with the effector (red) and the allosteric (wine) lobe in the GTP-bound state.

site, by overlapping with Switch I and Switch II, brings the system to autoinhibition (**Figure 2C**).

For the GTP-bound state, there are interactions between the very first residues (170-173) of the HVR and the N-terminal part of the catalytic domain (residues 1-6) and the loop, L3, (residues 47-50)—which connects β1 and β2 strands (noted in green) (**Figure 2B**). However, the interactions between the second part of the HVR and Switch I (blue) and Switch II (red) are sparsely populated with only the last residues at the farnesylated C-terminal end of the HVR interacting with this region. This indicates that in the GTP-bound state, the C-terminal end of the HVR shifts toward the allosteric lobe or is detached from the catalytic domain (**Figure 2D**), but the interaction with the effector lobe is very weakly populated.

The olive-colored region shows the HVR interaction with the C-terminal end of α3 and the loop between α3-β5, while the yellow-colored region shows the HVR interaction with the C-terminal end of α5 (**Figures 2C,D**). The main difference between the GDP- and GTP-bound states is that in the GDP-bound form, the N-terminal region of the HVR interacts with α3/α5, and the C-terminal region interacts with the Switch II region at the effector lobe. However, in the GTP-bound form, the more populated interactions extend to the C-terminal region of the HVR that interacts with the catalytic domain residues located in spatial proximity of the N-terminal region of HVR, such HVR forming a loop by itself.

Unlike the interacting regions in the effector lobe, HVR interactions at the regions designated by purple and orange of the allosteric lobe are almost non-existent in the GDP-bound state, while they are strongly present in the GTP-bound form. We observed that HVR residues, Asp173, Lys175, and Lys177 interact with residues 135-142 of α4 (noted purple), loop L9 and β6 (noted orange) in the catalytic domain. Interestingly, this region is part of the allosteric Ca^{2+}-acetate binding site described by (Buhrman et al., 2010).

A possible dimerization interface of K-Ras4B was observed at the α3/α4 region of the allosteric lobe (Muratcioglu et al., 2015). In this dimerization mode, the effector binding sites are exposed, allowing the recruitment of Raf and its dimerization, which is a prerequisite for its activation (Inouye et al., 2000). Our results indicate that the HVR-α4 interaction, which is almost non-existent in the GDP-bound form, but highly populated in the GTP-bound state, disfavors dimerization at this helical interface in the GTP-bound form in solution. On the other hand, in the GDP-bound state with the HVR associating with the effector binding region, it blocks the β-sheet dimer formation interface (Muratcioglu et al., 2015). These results may explain why GTP-bound K-Ras4B is predominantly monomeric in solution, even at high protein concentrations. *In vivo*, in the presence of the membrane, the HVR largely associates with the membrane in the GTP-bound state, so this GTP-bound inhibitory scenario is unlikely to play a significant role. Nonetheless, it still provides some insight into the inherent tendencies of the HVR behavior.

To understand the differences in the sparsity of the distance map between GDP- and GTP-bound states, we calculated the interaction energy between the HVR and the catalytic domain, ΔE_{HVR}. The conformations generated by the simulation protocol (above) were divided into three groups: (i) conformations with the HVR detached from the catalytic domain giving rise to the reference state, (ii) conformations with the HVR interacting with the effector lobe of the catalytic domain, and (iii) conformations with the HVR interacting with the allosteric lobe of the catalytic domain. Interaction energies of the HVR were then calculated in the presence of explicit solvent for all conformations in the three groups according to Equation (1).

The HVR interaction energy distributions illustrate a higher relative frequency of structures when the HVR interacts with the effector lobe within the GDP- than the GTP-bound population with respect to their whole structural population (**Figure 3A**). In contrast, higher relative frequency can be seen of the HVR interaction with the allosteric lobe in the GTP-bound state (**Figure 3B**). This suggests that the HVR of GDP-bound K-Ras4B favors the interaction with the effector lobe (**Figure 3C**), while that of GTP-bound K-Ras4B tends to reside in the allosteric lobe (**Figure 3D**), resulting in the differences in the sparsity between the lower parts of the distance based interaction maps (**Figures 2A,B**). **Figure 3A** also shows that the GDP-bound interaction energy distribution falls deeper, indicating stronger interactions of the HVR with the effector lobe within the GDP-bound than the GTP-bound population. While (as **Figure 3B** shows) in the HVR interaction with the allosteric lobe, we observe that for both GDP- and GTP-bound states the peaks of the distributions are located close to each other, indicating no difference in interaction energy of the HVR with the allosteric lobe irrespectively of the bound nucleotide.

In summary, whereas in the GDP-bound state interactions with the effector lobe are more populated and are shifted toward the low energy values, in the GTP-bound state, it is the opposite: structures interacting with the allosteric lobe are much more numerous than those interacting with the effector lobe, and the interaction energy of HVR with the allosteric lobe falls deeper than with the effector lobe.

Interactions of the Nucleotides With the Switch Regions

In order to characterize the interaction of the nucleotides with the catalytic site, the root-mean-square fluctuation (RMSF) of the C_α atoms on the MDeNM generated conformations for both GDP- and GTP-bound forms was calculated (**Figure 4**). This agrees with the previous findings: in the calculations, we observed that the GDP-bound form exhibits a higher RMSF at both Switch I and Switch II regions. As expected, the HVR (residues 167-185) has a considerably larger fluctuation in both systems compared to the catalytic domain.

To elucidate the behavior of the switch regions from an energetic point of view, the interaction energies between the nucleotides (with the coordinating Mg^{2+}: Mg^{2+}-GDP/Mg^{2+}-GTP) and the Switch I/Switch II regions were calculated for the MDeNM conformations. In the interaction with Switch I (**Figure 5A**), the interaction energy of Mg^{2+}-GTP is considerably

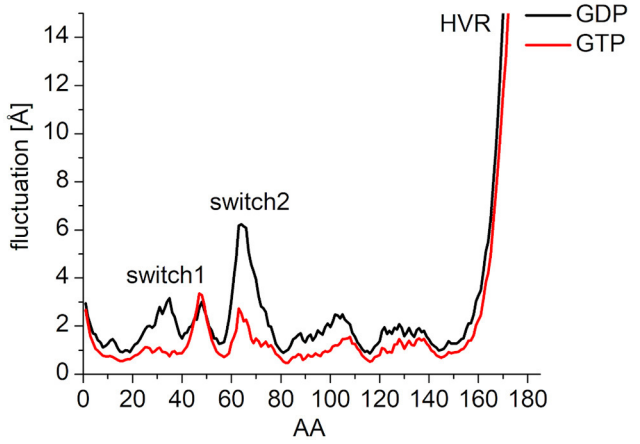


FIGURE 4 | RMS fluctuation of C_α atoms on MDeNM generated conformations of GDP- (black) and GTP-bound (red) K-Ras4B. For the better visualization of the catalytic domain regions the y-axis is enlarged up to and cropped at 15 Å.

more favorable than that of Mg^{2+}-GDP, suggesting the well-known stiffening conformation of Switch I in the active form, which is more available to downstream effectors (Cherfils and Zeghouf, 2013; Lu et al., 2016). In the interaction with Switch II (**Figure 5B**), the Mg^{2+}-GTP-bound population shows two peaks in its distribution, corresponding to two different conformational states, both falling deeper than that of the Mg^{2+}-GDP, which is in agreement with the decrease of RMSF in the GTP-bound form and explains the stronger binding of GTP vs. GDP (John et al., 1993).

As seen in **Figure 5B**, the two conformational states of Mg^{2+}-GTP/Switch II are likely to be identical to state 1 (higher energy conformation) and to state 2 (lower energy conformation) (Spoerner et al., 2010), which are in chemical equilibrium in solution. State 1 was identified as having a low affinity for effectors and low intrinsic hydrolysis, and state 2 as an "effector-binding state" showing high affinity for effectors and being stabilized by them. **Figure 6** shows the environment of a conformation from state 1 and from state 2 with superimposed nucleotides. The conformations of state 1 show weaker interactions between the nucleotide (Mg^{2+}-GTP) and the Switch II, lacking the H-bond formed between the donor N atom of Gly60 and the oxygen acceptor atom of γ-phosphate of GTP. This H-bond only exists in the state 2 conformations. Since this strong H-bond is present in the initial structure, showing a distance of 2.80 Å between GTP:O1G (acceptor) - Gly60:N (donor), the existence of the state 1 can be interpreted as the result of the moderate excitation kinetic energies introduced by MDeNM being capable of breaking the H-bond and contributing to a more fluctuating Switch II, resembling the inactive, GDP-bound K-Ras4B. This interaction is completely absent due to the lack of γ-phosphate in the GDP-bound state.

As stated previously, the HVR has a considerably larger fluctuation in both systems compared to the catalytic domain. To elucidate the internal flexibility of the HVR, we analyzed the rotational angle space of its residues. The dihedral angles

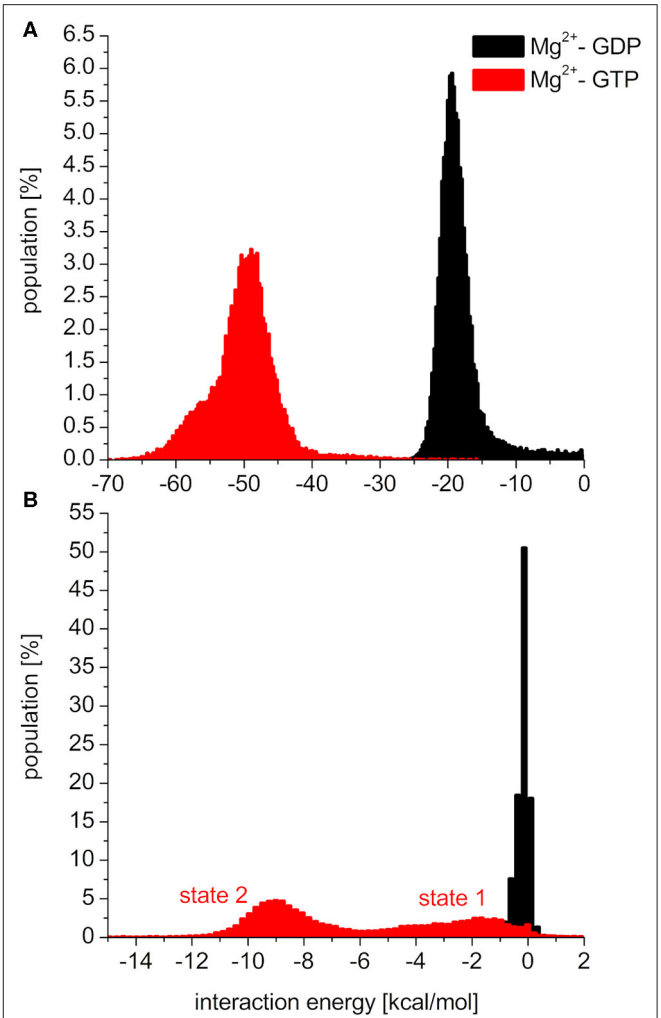


FIGURE 5 | Distributions of interaction energies between the nucleotide (Mg^{2+}-GDP: black, Mg^{2+}-GTP: red) and **(A)**: switch1 region, **(B)**: switch2 region of K-Ras4B.

defined by consecutive Cα atom quadruplets were calculated within the HVR for the GDP- and GTP-bound population (**Supplementary Figure 2**). The base of the HVR (N-terminal part of it) shows a bent configuration in both states with a richer rotational distribution in the case of the GDP-bound state. Both states exhibit a distributional peak corresponding to a rather elongated configuration for the parts following the bend of the base of the HVR. The difference between the activation states is that the HVR of the GTP-bound population has a slightly more restricted rotational profile; still, in some regions, two population peaks are observed one corresponding to an elongated structure and the other to a sharp bending located at the residue levels of 169, 172, and 175.

Comparison of MDeNM and MD

In order to compare the conformational space mapped by MDeNM to conformations accessible by classical MD simulation,

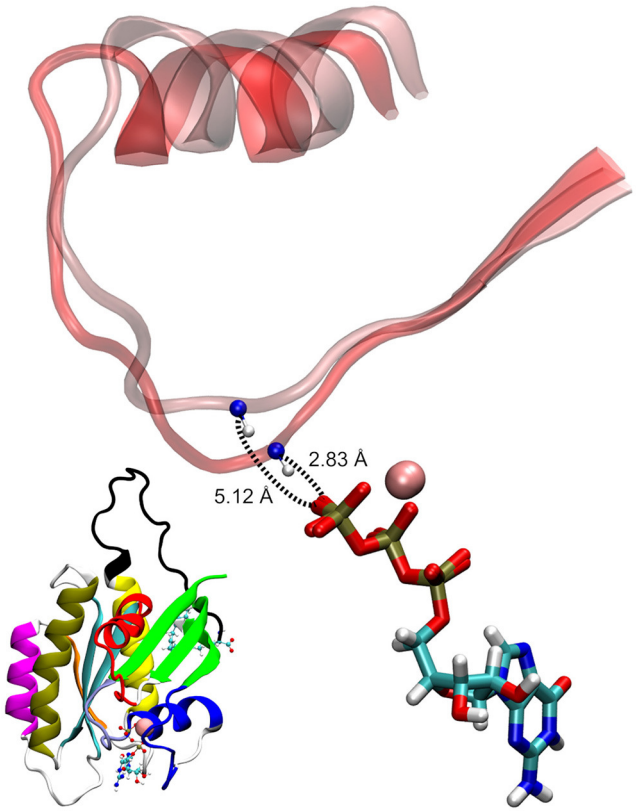


FIGURE 6 | Examples of GTP-bound K-Ras4B in state 1 and state 2, superimposed at their nucleotides. The shorter donor-acceptor distance corresponds to state 2.

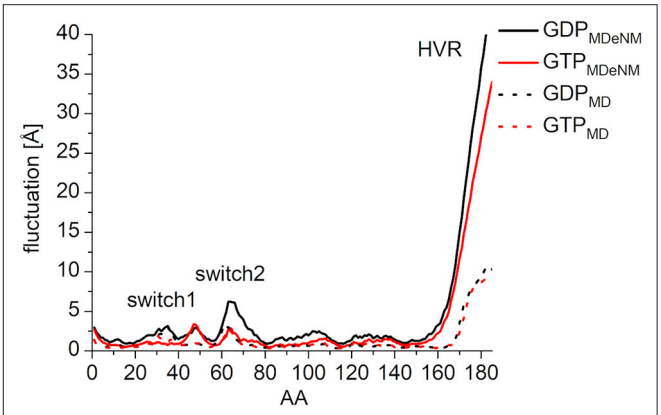


FIGURE 7 | RMS fluctuation of C_α atoms among the MDeNM (solid) and MD (dotted) generated conformations for GDP- (black) and GTP-bound (red) K-Ras4B.

three parallel MD simulations of 200 ns were performed for both the GDP- and GTP-bound forms, having the same starting structures as the MDeNM simulations. In general, the RMSF follows the same pattern both on MDeNM (solid line) and MD (dotted line) generated structures (**Figure 7**). However, two major differences can be noted: (i) the HVR shows approximately four times higher values for MDeNM compared to the MD conformations, demonstrating that MDeNM maps a considerably larger HVR conformational space; (ii) the GDP-bound Switch II shows a 2-fold RMSF in the case of MDeNM compared to the classical MD. This difference is not visible in the GTP-bound form, due to the presence of γ-phosphate.

To compare the two methods, we present the distance-based interaction maps between the catalytic domain and HVR among the conformations of both MDeNM (**Figures 8A,B**) and MD (**Figures 8C,D**) simulations for the two nucleotide-binding states. The striking difference between the results of the two methods is that the HVR-catalytic domain contacts are more dispersed in the case of MDeNM than for classical MD, the latter being more localized. For the GDP-bound form (**Figures 8A,C**), both methods show contacts between the HVR and the C-terminal of the catalytic domain (yellow); the C-terminal end of α3 (olive); α2 (red); β1, β2, β3 (green); the C-terminal end of α4 (magenta) and β6 (orange), which simply form the "northern hemisphere" of the catalytic domain, with the north-pole being the N-terminal region of HVR. The major difference between MDeNM and MD manifested at the active site of the protein is the degree of autoinhibition, which is more populated by MDeNM and the extent of the HVR-Switch I/Switch II (blue and red respectively) contact and its versatility being greater again by MDeNM. The GTP-bound form also shows contact with the "northern hemisphere" (yellow, olive, red, green, magenta, orange) indicating a broader conformational mapping with the MDeNM method than with the classical MD. Two regions are more populated in the case of MD than MDeNM: the interaction of the farnesylated C-terminal end with α4/β6 (magenta/orange), and with Switch I (blue). In both cases, visual examination of the MD trajectory revealed that after the HVR found an energetically favorable position it spends considerably more time around the given local minima before moving further, thus exploring the conformational space in a more limited manner.

Another measure to compare the ensembles generated by MDeNM and MD for both states of the enzyme is the overall distribution of the RMSD values with respect to the starting structure (**Supplementary Figure 3**). If the RMSD of the catalytic domain (without HVR) is calculated, the distribution of the RMSD for the GDP- bound state is 1.9 ± 1.7 Å on the MDeNM, and 1.5 ± 1.2 Å on the MD generated ensemble, while for the GTP-bound state it is 1.7 ± 1.5 Å for MDeNM and 1.2 ± 0.2 Å for MD. On one hand, these values demonstrate that the GDP-bound catalytic domain is more flexible than the GTP-bound one; on the other hand, they also show that MDeNM explores a larger conformational space (both the average RMSD values and the deviations are higher).

If for the RMSD calculations the full-length protein is considered, for the GDP-bound state the values are 12.1 ± 4.0 Å for MDeNM and 11.6 ± 1.9 Å for MD, while for the GTP-bound state 12.4 ± 4.0 Å on the MDeNM and 10.8 ± 2.1 Å on the MD generated structures. The tendency

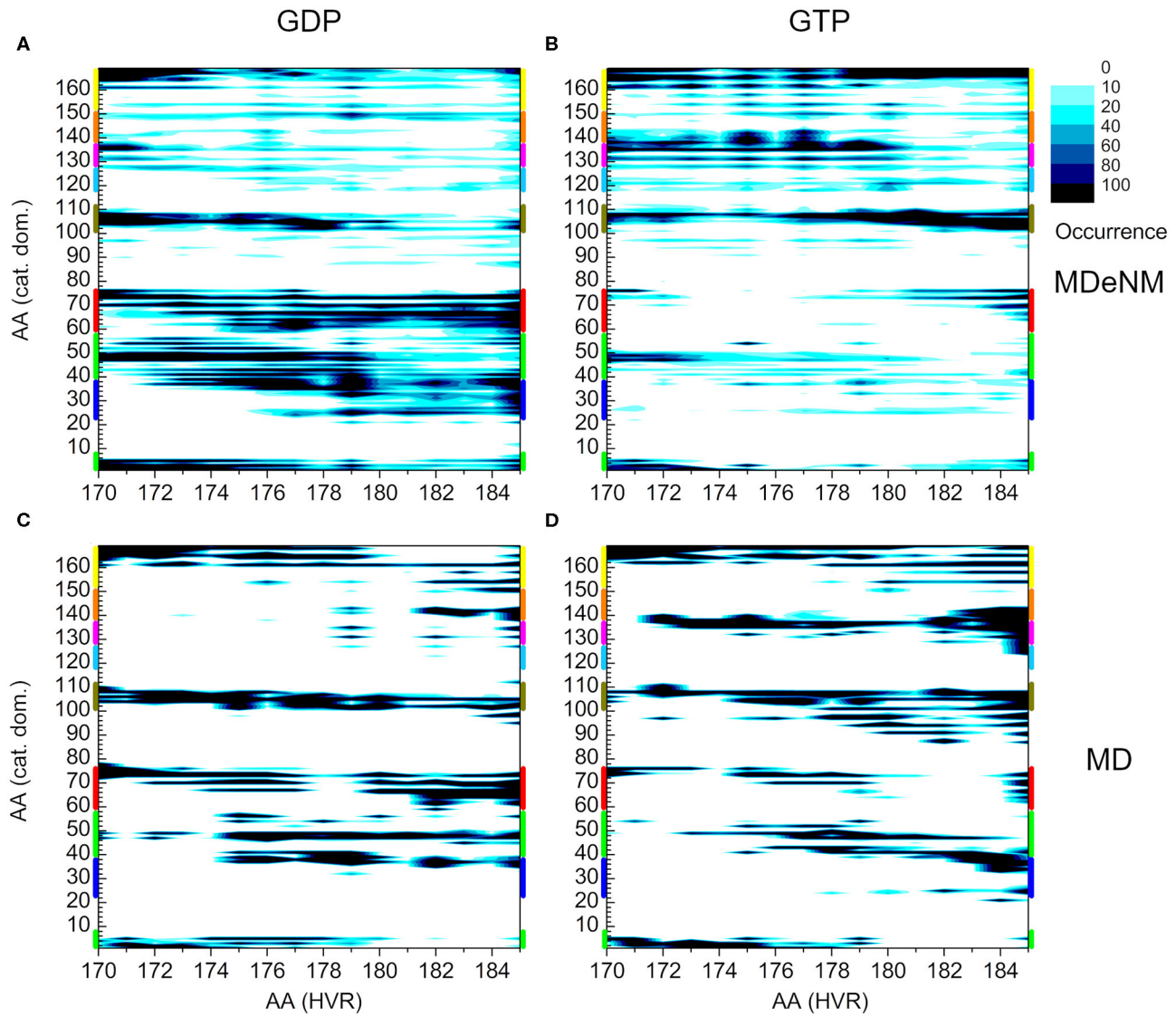


FIGURE 8 | Distance-based interaction map between the catalytic domain and the HVR of K-Ras4B. **(A)** and **(B)** correspond to conformations of MDeNM for the GDP- and GTP-bound population, respectively. **(C)** and **(D)** correspond to conformations of MD for the GDP- and GTP-bound state, respectively. The color-coding of the structural regions within the catalytic domain along the y-axis is identical to **Figure 2**.

in the full-length case is similar to the one shown for the catalytic domain by itself, magnified by the fluctuation of the HVR.

CONCLUSION

As MDeNM revealed and was confirmed by MD simulations, almost all catalytic domain-HVR interactions exist in both GDP- and GTP-bound K-Ras4B. In the GDP-bound form, the population of the conformers is shifted toward a state where the farnesylated C-terminal HVR interacts with the effector lobe of the catalytic domain, blocking both effector binding and dimerization through β-sheet formation. In contrast, in the GTP-bound form, these conformations either do not exist or are poorly populated. In addition, in the GTP-bound state, the population of the conformers is shifted in such a manner that becomes highly populated when the HVR interacts with the allosteric lobe, exposing the effector binding sites. The GTP-bound form provides the HVR conformation overlapping with the α-helix dimer interface at the allosteric lobe, but these conformations are much less populated in the GDP-bound state.

Thus, MDeNM proves capable of an extensive sampling of the interaction of the HVR with the catalytic domain, capturing ensembles that were shown earlier for the GDP-bound K-Ras4B states and revealing new distributions for the GTP-bound states where HVR binding overlaps the α3/α4

dimerization surface. Even though in the presence of the membrane the extent of these interactions is unclear, MDeNM captures the HVR tendencies, which may nonetheless be partially populated.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

This work was designed by EB, DP, and RN. The simulations were carried out by BD, and were analyzed by BD and FM. The manuscript was written by BD, FM, HJ, RN, DP, and EB. All authors contributed to the article and approved the submitted version.

FUNDING

This project has been funded in part within the frame of the Erasmus+ traineeship program (15/KA1HE/137/SMP-326) financed by the European Union (BD), with federal funds from the National Cancer Institute, National Institutes of Health (HHSN26120080001E) and by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research (HJ and RN). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00145/full#supplementary-material

REFERENCES

Abraham, S. J., Nolet, R. P., Calvert, R. J., Anderson, L. M., and Gaponenko, V. (2009). The hypervariable region of K-Ras4B is responsible for its specific interactions with calmodulin. *Biochemistry* 48, 7575–7583. doi: 10.1021/bi900769j

Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., et al. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* 8, 3257–3273. doi: 10.1021/ct300400x

Bos, J. L., Rehmann, H., and Wittinghofer, A. (2007). GEFs and GAPs: critical elements in the control of small G proteins. *Cell* 129, 865–877. doi: 10.1016/j.cell.2007.05.018

Bowerman, S., and Wereszczynski, J. (2016). Detecting allosteric networks using molecular dynamics simulation. *Methods Enzymol.* 578, 429–447. doi: 10.1016/bs.mie.2016.05.027

Brooks, B. R., Brooks, C. L. III., Mackerell, A. D. Jr., Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614. doi: 10.1002/jcc.21287

Brunsveld, L., Waldmann, H., and Huster, D. (2009). Membrane binding of lipidated Ras peptides and proteins–the structural point of view. *Biochim. Biophys. Acta* 1788, 273–288. doi: 10.1016/j.bbamem.2008.08.006

Buhrman, G., Holzapfel, G., Fetics, S., and Mattos, C. (2010). Allosteric modulation of Ras positions Q61 for a direct role in catalysis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4931–4936. doi: 10.1073/pnas.0912226107

Castellano, E., and Santos, E. (2011). Functional specificity of ras isoforms: so similar but so different. *Genes Cancer* 2, 216–231. doi: 10.1177/1947601911408081

Chavan, T. S., Jang, H., Khavrutskii, L., Abraham, S. J., Banerjee, A., Freed, B. C., et al. (2015). High-affinity interaction of the K-Ras4B hypervariable region with the Ras active site. *Biophys. J.* 109, 2602–2613. doi: 10.1016/j.bpj.2015.09.034

Cherfils, J., and Zeghouf, M. (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol. Rev.* 93, 269–309. doi: 10.1152/physrev.00003.2012

Costa, M. G. S., Batista, P. R., Bisch, P. M., and Perahia, D. (2015). Exploring free energy landscapes of large conformational changes: molecular dynamics with excited normal modes. *J. Chem. Theory Comput.* 11, 2755–2767. doi: 10.1021/acs.jctc.5b00003

Fetics, S. K., Guterres, H., Kearney, B. M., Buhrman, G., Ma, B., Nussinov, R., et al. (2015). Allosteric effects of the oncogenic RasQ61L mutant on Raf-RBD. *Structure* 23, 505–516. doi: 10.1016/j.str.2014.12.017

Hernandez-Alcoceba, R., del Peso, L., and Lacal, J. C. (2000). The Ras family of GTPases in cancer cell invasion. *Cell. Mol. Life Sci.* 57, 65–76. doi: 10.1007/s000180050499

Hunter, J. C., Gurbani, D., Ficarro, S. B., Carrasco, M. A., Lim, S. M., Choi, H. G., et al. (2014). *In situ* selectivity profiling and crystal structure of SML-8-73-1, an active site inhibitor of oncogenic K-Ras G12C. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8895–8900. doi: 10.1073/pnas.1404639111

Inouye, K., Mizutani, S., Koide, H., and Kaziro, Y. (2000). Formation of the Ras dimer is essential for Raf-1 activation. *J. Biol. Chem.* 275, 3737–3740. doi: 10.1074/jbc.275.6.3737

Jang, H., Abraham, S. J., Chavan, T. S., Hitchinson, B., Khavrutskii, L., Tarasova, N. I., et al. (2015). Mechanisms of membrane binding of small GTPase K-Ras4B farnesylated hypervariable region. *J. Biol. Chem.* 290, 9465–9477. doi: 10.1074/jbc.M114.620724

Jang, H., Banerjee, A., Chavan, T. S., Lu, S., Zhang, J., Gaponenko, V., et al. (2016a). The higher level of complexity of K-Ras4B activation at the membrane. *FASEB J.* 30, 1643–1655. doi: 10.1096/fj.15-279091

Jang, H., Muratcioglu, S., Gursoy, A., Keskin, O., and Nussinov, R. (2016b). Membrane-associated Ras dimers are isoform-specific: K-Ras dimers differ from H-Ras dimers. *Biochem. J.* 473, 1719–1732. doi: 10.1042/BCJ20160031

Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29, 1859–1865. doi: 10.1002/jcc.20945

John, J., Rensland, H., Schlichting, I., Vetter, I., Borasio, G. D., Goody, R. S., et al. (1993). Kinetic and structural analysis of the Mg(2+)-binding site of the guanine nucleotide-binding protein p21H-ras. *J. Biol. Chem.* 268, 923–929.

Koera, K., Nakamura, K., Nakao, K., Miyoshi, J., Toyoshima, K., Hatta, T., et al. (1997). K-ras is essential for the development of the mouse embryo. *Oncogene* 15, 1151–1159. doi: 10.1038/sj.onc.1201284

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., et al. (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* 12, 405–413. doi: 10.1021/acs.jctc.5b00935

Lu, S., Banerjee, A., Jang, H., Zhang, J., Gaponenko, V., and Nussinov, R. (2015). GTP binding and oncogenic mutations may attenuate hypervariable region (HVR)-catalytic domain interactions in small GTPase K-Ras4B, exposing the effector binding site. *J. Biol. Chem.* 290, 28887–28900. doi: 10.1074/jbc.M115.664755

Lu, S., Jang, H., Muratcioglu, S., Gursoy, A., Keskin, O., Nussinov, R., et al. (2016). Ras conformational ensembles, allostery, and signaling. *Chem. Rev.* 116, 6607–6665. doi: 10.1021/acs.chemrev.5b00542

Muratcioglu, S., Chavan, T. S., Freed, B. C., Jang, H., Khavrutskii, L., Freed, R. N., et al. (2015). GTP-dependent K-Ras dimerization. *Structure* 23, 1325–1335. doi: 10.1016/j.str.2015.04.019

Nussinov, R., Tsai, C. J., and Jang, H. (2020). Autoinhibition can identify rare driver mutations and advise pharmacology. *FASEB J.* 34, 16–29. doi: 10.1096/fj.201901341R

Nussinov, R., Zhang, M., Tsai, C. J., Liao, T. J., Fushman, D., and Jang, H. (2018). Autoinhibition in Ras effectors Raf, PI3Kalpha, and RASSF5: a comprehensive review underscoring the challenges in pharmacological intervention. *Biophys. Rev.* 10, 1263–1282. doi: 10.1007/s12551-018-0461-0

Pacold, M. E., Suire, S., Perisic, O., Lara-Gonzalez, S., Davis, C. T., Walker, E. H., et al. (2000). Crystal structure and functional analysis of Ras binding to its effector phosphoinositide 3-kinase gamma. *Cell* 103, 931–943. doi: 10.1016/s0092-8674(00)00196-3

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289

Plowman, S. J., Williamson, D. J., O'Sullivan, M. J., Doig, J., Ritchie, A. M., Harrison, D. J., et al. (2003). While K-ras is essential for mouse development, expression of the K-ras 4A splice variant is dispensable. *Mol. Cell. Biol.* 23, 9245–9250. doi: 10.1128/mcb.23.24.9245-9250.2003

Prakash, P., and Gorfe, A. A. (2017). Membrane orientation dynamics of lipid-modified small GTPases. *Small GTPases* 8, 129–138. doi: 10.1080/21541248.2016.1211067

Prakash, P., Zhou, Y., Liang, H., Hancock, J. F., and Gorfe, A. A. (2016). Oncogenic K-Ras binds to an anionic membrane in two distinct orientations: a molecular dynamics analysis. *Biophys. J.* 110, 1125–1138. doi: 10.1016/j.bpj.2016.01.019

Prior, I. A., Lewis, P. D., and Mattos, C. (2012). A comprehensive survey of Ras mutations in cancer. *Cancer Res.* 72, 2457–2467. doi: 10.1158/0008-5472.CAN-11-2612

Spoerner, M., Hozsa, C., Poetzl, J. A., Reiss, K., Ganser, P., Geyer, M., et al. (2010). Conformational states of human rat sarcoma (Ras) protein complexed with its natural ligand GTP and their role for effector interaction and GTP hydrolysis. *J. Biol. Chem.* 285, 39768–39778. doi: 10.1074/jbc.M110.145235

Thapar, R., Williams, J. G., and Campbell, S. L. (2004). NMR characterization of full-length farnesylated and non-farnesylated H-Ras and its implications for Raf activation. *J. Mol. Biol.* 343, 1391–1408. doi: 10.1016/j.jmb.2004.08.106

van Hattum, H., and Waldmann, H. (2014). Chemical biology tools for regulating RAS signaling complexity in space and time. *Chem. Biol.* 21, 1185–1195. doi: 10.1016/j.chembiol.2014.08.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past collaboration with one of the authors DP.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 25 August 2020
doi: 10.3389/fmolb.2020.00210

A Coarse-Grained Methodology Identifies Intrinsic Mechanisms That Dissociate Interacting Protein Pairs

*Haleh Abdizadeh[1], Farzaneh Jalalypour[2], Ali Rana Atilgan[2] and Canan Atilgan[2]**

[1] Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, Netherlands, [2] Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey

We address the problem of triggering dissociation events between proteins that have formed a complex. We have collected a set of 25 non-redundant, functionally diverse protein complexes having high-resolution three-dimensional structures in both the unbound and bound forms. We unify elastic network models with perturbation response scanning (PRS) methodology as an efficient approach for predicting residues that have the propensity to trigger dissociation of an interacting protein pair, using the three-dimensional structures of the bound and unbound proteins as input. PRS reveals that while for a group of protein pairs, residues involved in the conformational shifts are confined to regions with large motions, there are others where they originate from parts of the protein unaffected structurally by binding. Strikingly, only a few of the complexes have interface residues responsible for dissociation. We find two main modes of response: In one mode, remote control of dissociation in which disruption of the electrostatic potential distribution along protein surfaces play the major role; in the alternative mode, mechanical control of dissociation by remote residues prevail. In the former, dissociation is triggered by changes in the local environment of the protein, e.g., pH or ionic strength, while in the latter, specific perturbations arriving at the controlling residues, e.g., via binding to a third interacting partner is required for decomplexation. We resolve the observations by relying on an electromechanical coupling model which reduces to the usual elastic network result in the limit of the lack of coupling. We validate the approach by illustrating the biological significance of top residues selected by PRS on select cases where we show that the residues whose perturbation leads to the observed conformational changes correspond to either functionally important or highly conserved residues in the complex.

Keywords: perturbation response scanning, elastic network model, protein complexes, structural motifs, electrostatic potential distribution, protein–protein dissociation, allostery, cooperative conformational change

OPEN ACCESS

Edited by:
Pemra Doruker, University of Pittsburgh, United States

Reviewed by:
Ruth Nussinov, National Cancer Institute (NCI), United States
Pavel Srb, Academy of Sciences of the Czech Republic (ASCR), Czechia

***Correspondence:**
Canan Atilgan
canan@sabanciuniv.edu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *24 June 2020*
Accepted: *03 August 2020*
Published: *25 August 2020*

Citation:
Abdizadeh H, Jalalypour F, Atilgan AR and Atilgan C (2020) A Coarse-Grained Methodology Identifies Intrinsic Mechanisms That Dissociate Interacting Protein Pairs. Front. Mol. Biosci. 7:210. doi: 10.3389/fmolb.2020.00210

INTRODUCTION

Chemical and physical processes within assemblies of proteins in the cellular environment are events often encompassing multiple time and length scales. Therefore, different modeling tools are commonly used to describe the network of interactions featuring the protein dynamics (Adcock and McCammon, 2006; Tozzini, 2010). In this vein, coarse-grained models (CG), with several

atoms of the protein grouped into one bead and in the absence of atomic details of the solvent molecules have been developed to supplement the extremely expensive atomistic modeling of large scale motions of biomolecular aggregates (Orozco, 2014; Atilgan, 2018). CG models have proved to aid sampling efficiency, predict allosteric regulations (Ming and Wall, 2005) and describe conformational transition pathways (Kim et al., 2002; Miyashita et al., 2003). One useful measure of large-scale protein mechanics in the context of CG models is the elastic network model (ENM) (Tirion, 1996; Bahar et al., 1997; Hinsen, 1998). ENMs are based on the assumption that the potential energy of the system is approximated harmonically about a single minimum energy conformation. Methodologically, in ENM, the atomic details of the biomolecule structure are reduced to a network of nodes (typically one site per residue) connected by harmonic springs. Since all springs are in a relaxed state in the network, no energy minimization is required, in comparison to normal mode analysis in which an expensive initial energy minimization is required prior to calculating the Hessian matrix. For large biomolecules and multi-protein complexes, ENM models with a resolution lower than standard have been used (Durand et al., 1994; Doruker et al., 2002; Chennubhotla et al., 2005; Ahmed and Gohlke, 2006; Kurkcuoglu et al., 2009; Ross C. et al., 2018). The gross representation of large assemblies has proven to predict dynamics of the rigid and flexible parts of the proteins (Ross C.J. et al., 2018). Anisotropic network model (ANM) and Gaussian network model (GNM) are amongst the most widely used ENM-based methods (Bahar et al., 1997; Atilgan et al., 2001). While GNM is applied to produce the amplitudes of isotropic thermal motions, ANM describes both amplitudes and directionality of anisotropic motions. Increased amount of data for proteins of different forms (free, liganded, or complexed), elucidates the correlation between protein function observed in experiments and the global motions predicted by ANM/GNM analyses. Numerous studies have employed ENM-based models to explore various aspects of protein structural dynamics. These include identifying and visualizing collective motions (Kong et al., 2003), predicting modes of motion underlying function (Baysal and Atilgan, 2001b; Keskin et al., 2002; Zheng and Doniach, 2003; Whitford et al., 2007), and explaining details of conformational changes of various types and amplitudes (Tama and Sanejouand, 2001; Krebs et al., 2002; Zheng et al., 2007). ENMs may be applied to the refinement of medium to low-resolution structures of electron microscopy density maps of large macromolecular complexes or molecular envelopes derived from small-angle x-ray scattering (SAXS) data (Delarue and Dumas, 2004; Hinsen et al., 2005). Within the concept of ENM, researchers have developed approaches to generate feasible pathways for conformational transitions between two end conformers (Kim et al., 2002; Orellana et al., 2019), removing the need for expensive molecular dynamics (MD) simulations and all-atom empirical force fields to set up intermediate conformations. ENMs are also applied to determine the main evolutionary transformations of structural changes among homologous proteins (Leo-Macias et al., 2005; Han et al., 2008). In such an approach, for a given set of proteins, evolutionary direction is argued to be a combination of a small subspace projected by a few low frequency modes imposed by inter-residue contact topology.

We have extended ENM to analyze allosterically significant residues and function-related motions of proteins via a technique named perturbation-response scanning (PRS) (Atilgan and Atilgan, 2009; Atilgan et al., 2010a). The methodology inserts fictitious forces on selected atoms and predicts the response within the realm of linear response theory (LRT). *In vivo*, the perturbation may arrive in the form of changing environmental conditions such as pH or ionic strength (Abdizadeh et al., 2015a; Sensoy et al., 2017), or it may act directly on the chain as in pulling (Carrion-Vazquez et al., 2003; Dietz et al., 2006) or other single-molecule experiments, as well as through mutations or ligand binding. PRS serves as a tool to gain insight into molecular origins of mechanical feedbacks of bimolecular structures through recording response to each inserted force on each residue of a protein (Atilgan and Atilgan, 2009; Atilgan et al., 2010a, 2011; Abdizadeh et al., 2015b; Abdizadeh and Atilgan, 2016). It is further capable of recognizing how directionality of the inserted force may coordinate the response of the protein in a functional motion (Jalalypour et al., 2020). PRS requires two distinct conformations of a protein, determined, e.g., by x-ray crystallography, as input; and relies on LRT to relate virtual external forces acting on a protein to the perturbed positions of the residues (Ikeguchi et al., 2005). In PRS, one scans a protein structure residue by residue through applying forces in many directions and records the subset of responses that encourage conversion to another known conformation of the protein. Thus, one can map the regions on the protein surface whose perturbation might lead to the expected conformational change. Besides mapping active site residues that are prime regions for invoking conformational transitions, this approach also has the potential of pointing out allosteric locations and drug target regions. For example, previously, we have studied the proteins calmodulin (Atilgan et al., 2011) and ferric binding protein (Atilgan and Atilgan, 2009) via PRS. By mutating those residues that were implicated in allosteric communication, we later verified through classical MD simulations that they affect the conformation distributions (Aykut et al., 2013; Guven et al., 2014). In a later study, we have performed PRS on subtilisin in complex with its inhibitor to pinpoint hot residues involved in catalytic mechanism and stability of the enzyme (Abdizadeh et al., 2015b). PRS has also been used in the conformation generation step of a flexible docking scheme for exploring protein-ligand interactions (Bolia et al., 2014). In a similar methodology, Mottonen et al. (2010) and Jacobs et al. (2012) have used a method based on distance constraint model to impose constraints on the torsional degrees of freedom of the protein to mimic a hypothetical ligand-binding situation.

In this manuscript, we utilize these CG approaches to address a major challenge for structural biology in providing a mechanistic view of the behavior of molecular complexes and their conformational changes. Protein-ligand and protein–protein interactions (PPI) govern most of the cellular processes (De Las Rivas and Fontanillo, 2010). Many studies investigate the protein-ligand complexes and look for functional regions, binding sites or druggable cavities (Lichtarge and Sowa, 2002;

Xie et al., 2009; Siragusa et al., 2015). On the other hand, PPI allow a protein to perform its biological function by interacting with another partner protein (Ozdemir et al., 2019). Therefore, the interface is usually considered as a candidate to be targeted by a potential drug such as orthosteric or allosteric PPI modulators (Xie et al., 2009; Ni et al., 2019). Studying PPIs and protein-interaction networks may provide insights into new opportunities in the medical, biotechnological, and pharmaceutical fields. Hence, several approaches have been proposed to study PPIs (Ozdemir et al., 2019). A number of bioinformatics techniques have been developed to predict PPI networks based on genomic-context, sequence homology and structural similarity (Shi et al., 2005). Most systematic studies involving protein–protein complexes focuses on the interaction interface to determine compatibility of the structures or attempts to study individual PPI and predicts residues, called hotspots, effective in recognition of partner proteins (Liu Q. et al., 2018; Qiao et al., 2018). Alanine scanning mutagenesis is the major experimental method to identify hotspots (Kenneth Morrow and Zhang, 2012). In one study, non-covalent interactions (hydrophobic, van der Waals, and hydrogen bonding) are found to account as the major forces operating at the PPI interfaces (Gao et al., 2004). Available computational techniques for PPI hotspot prediction are roughly divided to two groups whereby most use the complex structure and a few utilize unbound structures (Ozdemir et al., 2019). Generally, hotspots resulting from the computations are compared to those from alanine scanning mutagenesis experiments (Bradshaw et al., 2011; Ibarra et al., 2019). In addition, machine learning-based methods have been developed to predict hotspots, considering the amino-acid features and conservation information (Liu S. et al., 2018). Most recently, by ignoring the internal structures of the molecules and scanning the protein surface for the so-called "interaction fingerprints," geometric deep learning algorithms have been developed for predicting protein-protein complexes (Gainza et al., 2020). Although these approaches attempt to define a general interaction pattern based on parameters such as structure, hydrophobicity or polarity, there is no general rule to be used in PPI prediction due to their diversity (Ni et al., 2019).

In this work we address a reverse problem: How is it possible to trigger dissociation events between proteins which have already formed a complex? We study 25 sets of protein complexes utilizing PRS with the ENM harmonic potential to determine regions responsible for rendering known complexes incompatible. Elastic network construction helps one to probe conformational changes due to altered physical and chemical environment (Atilgan et al., 2012). Accordingly, the information needed for assessing protein–protein interactions can be derived from knowledge of inter-residue contact topology, buried in the Hessian matrix deduced from ENM (Bahar et al., 2010). Rather than focusing on the interface of the interacting subunits, we relate the physical effects of the internal dynamics of protein complexes to the motions involved in their dissociation. We show that PRS maps residues that may alone initiate the structural change between the bound and unbound forms during dissociation processes of the protein complexes.

MODELS AND METHODS

The conformational change was analyzed via PRS between two different conformers of a protein, one in its complexed form with another protein, and the other in its unbound form. The propensity to convert between conformations was examined for these two states of the protein by employing fictitious forces. The detailed theory of PRS has been laid-out in previous studies (Atilgan and Atilgan, 2009; Atilgan et al., 2010a). In brief, the unbound state of a protein may be described by a perturbation of the Hamiltonian of the bound state, or vice versa. Under LRT, the shift in the coordinates due to unbinding is approximated by Yilmaz and Atilgan (2000) and Atilgan and Atilgan (2009):

$$\Delta\mathbf{R}_1 = \langle\mathbf{R}\rangle_1 - \langle\mathbf{R}\rangle_0 \simeq \frac{1}{3k_BT}\langle\Delta\mathbf{R}\Delta\mathbf{R}^{\mathrm{T}}\rangle_0\Delta\mathbf{F} = \frac{1}{3k_BT}\mathbf{H}^{-1}\Delta\mathbf{F} \tag{1}$$

where the subscripts 1 and 0 denote perturbed and unperturbed configurations of the protein, k_B is the Boltzmann constant and T is temperature. $\Delta\mathbf{F}$ vector contains the components of the externally inserted force vectors on the selected residues; e.g., for the perturbation of a single residue i, $(\Delta\mathbf{F})^T = \left\{000\ldots\Delta F_x^i\Delta F_y^i\Delta F_z^i\ldots000\right\}_{1\times3N}$. $\mathbf{H}^{-1}$ is the variance-covariance matrix which may be obtained from either the atomic coordinate trajectories of MD simulations of suitable length (Atilgan et al., 2012), or by imposing the approximation of harmonic springs between pairs of interacting atoms. In this work, we generate the $\mathbf{H}^{-1}$ matrix from a coarse-grained approach, constructing a network of N nodes on the C_α atoms of the protein complexes whose coordinates are directly used from their protein data bank (PDB) entries. Any given pair of nodes interacts in accord with a conventional harmonic potential, if the nodes are within a cut-off distance, r_c, of each other. This leads to a total of M interactions. Within the scope of an elastic network of residues that are connected to their neighbors by linear-elastic springs, one gets the $3N \times M$ direction cosine matrix $\mathbf{B}$ (Yilmaz and Atilgan, 2000). $\mathbf{BB}^T$ is exactly the Hessian if harmonic interactions with uniform force constants for all M bonds in the network are assumed. $(\mathbf{BB}^T)^{-1}$ is the covariance matrix $\mathbf{H}^{-1}$ for a given configuration, which is also an $N \times N$ supermatrix whose ij^{th} element is the 3×3 matrix of correlations between the x-, y-, and z-components of the fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ of residues i and j.

$\mathbf{H}$ of the system has at least six zero eigen-values corresponding to the purely translational and rotational motions. The eigen-value distribution of the Hessian of proteins is such that the low-frequency region is more crowded than expected of polymers or other condensed matter (Ben-Avraham, 1993). Thus, the choice of the cutoff distance, r_c, for the construction of the Hessian is critical for extracting protein-like properties from the systems studied (Atilgan et al., 2010b). For all the proteins studied in this work, we coarse-grain the crystal structure so that each residue is represented by the coordinates of its C_α atom. To account for the flexibility of proteins, we repeat the PRS analysis for a variety of cut-off distances in the range of 10.0–14.0 Å in increments of 1 Å; the lower limit of 10 Å agrees with the definition of first coordination shell of residues in proteins

(∼ 7 Å). For each network structure, we ensure that the system has six zero eigen-values corresponding to the translational and rotational degrees of freedom of the protein. The smallest common r_c at which we obtain six zero eigenvalues for all the proteins tested is 9 Å.

PRS technique relies on repeating the above LRT calculation (Equation 1) by scanning the residues of the protein one-by-one and focusing further on those perturbations that overlap with the conformational change, $\Delta\mathbf{R}_1 = \langle\mathbf{R}\rangle_1 - \langle\mathbf{R}\rangle_0$. There is no *a priori* assumption on how a force might be generated at a particular point. Conversely, after finding the force/residue pair that best leads to the conformational change of interest, we relate this finding to possible causes. In this study, PRS is applied by scanning each residue in 500 random directions.

To assess the quality of the predicted displacements of all residues resulting from a force applied on selected residue *i*, we use the correlation coefficient between the predicted and experimental displacements, averaged over all the affected residues, *k*:

$$C_i = \frac{\sum_{k=1}^{N}\left[(\Delta R_{\mathrm{k}})^{\mathrm{i}} - \left(\overline{\Delta R}\right)^{\mathrm{i}}\right]\left(\Delta S_{\mathrm{k}} - \overline{\Delta S}\right)}{(N-1)\,\sigma_R\sigma_S} \qquad (2)$$

In equation 2, the overbar indicates the average, $\Delta\mathbf{S}_k$ are the displacements between the initial and the target forms obtained from the PDB structures, σ_S and σ_R are the respective standard deviations for experimental and calculated displacements. A value close to 1 implies good agreement with the experimental conformational change, while a value of zero indicates lack of correlation between experimental and theoretical findings. Several approaches were taken to select the residues that are effective in directing the protein toward alternative conformations depending on the distribution of the maximum of the C_i values, C_i^{max}, obtained from the 500 perturbations and calculated through equation 2. We first list C_i^{max} in ascending order: (1) If there is a sharp decrease in the Pearson correlation values, we list the top residues until that gap. (2) If there is a smooth decrease in the Pearson correlation values, we list the residues that are common among top 10 residues of all cut-off values. We also check the location of the residues that do not survive these selection criteria. We have found that the variable residues observed among top residues in different cut-off values are spatial neighbors of the listed ones.

Protein Complex Selection

We analyzed a set of 25 protein complexes in their bound and unbound forms (**Table 1**, column 2). We collected protein pair structures from those reported in Benchmark 0.0 of ZDOCK (Chen and Weng, 2002). The complexes are non-redundant and they have X-ray structure solved at better than 2.90 Å resolution. They include a wide variety of function and affinities; they belong to different biological families. The constituent unbound forms of all the 25 complexes are available in the PDB with solution NMR or X-ray structures solved at better than 3.50 Å resolution. For the proteins resolved by solution NMR, we always use the first model for the PRS calculations. More specifically, we have chosen the protein complexes with no less than two missing residues along the protein chain, either in the bound or unbound form. Furthermore, if the number of residues in the bound and unbound components differ, we only analyze the common parts of the bound and unbound proteins.

Optimization of the Cut-Off Distance

For each pair of experimental structures, the unbound form is superimposed on the bound form, followed by the residue displacement vectors computation, $\Delta\mathbf{S}$. In this study, we perturb the bound form of each protein by applying a random force to the C_α atom of each residue in the complex. We select residues whose perturbation leads to the $\Delta\mathbf{R}$ vector (equation 1) that best resembles the dissociated proteins using equation 2 as the measure. For a given protein, we select the cut-off distance, r_{opt} that yields the closest agreement with the displacement vectors from experiments for at least one residue (**Table 1**, column 8). We verify that the residue indices that provide the best Pearson correlation value are always present within the range perceived as the highest correlation value for all cut-off distances studied. We note that the correlation values reported in **Table 2** (column 5 and 6) are not affected for the range of values $r_{opt} \pm 1$ Å for all of the 25 protein complexes. Strikingly, although many of these proteins display large Pearson correlation values, the numbers of residues yielding the highest values differ among proteins. For some proteins, there is not much specificity on the residue to be perturbed to reproduce the conformational change. For others, by perturbing a very specific location, the complete conformational change is obtained. The former category is exemplified by fibrinogen-binding protein (pdb code: 3D5S; Haspel et al., 2008) while serine protease and its proteinaceous inhibitor (pdb code: 1D6R; Koepke et al., 2000) is an example of a protein complex that we need to perturb in a specific location to mimic the dissociated conformation.

RESULTS AND DISCUSSION

The structural difference between the bound and the unbound forms, based on the C_α superposition of the binding partners, show that while for 22 of the cases the interface RMSD is less than 1.5 Å; for two cases interface RMSD is 1.5–2.2 Å and two cases have interface RMSD greater than 2.2 Å (Humphrey et al., 1996). For the pair of initial and target forms of the proteins present in bound and unbound form, we perform STAMP structural alignment, implemented in VMD 1.9.1 MultiSeq plugin (Humphrey et al., 1996). We record the RMSD between the structures of the target forms with the initial structure which vary between 0.4 and 4.3Å (**Table 1**, column 6). Although large motions of side chains and surface loops is always present as a local conformational change, we do not detect any discernible shape change on a global scale in proteins in their bound form compared to the respective unbound constituents. For all cases, the main chains essentially have the same conformation in bound and unbound forms and the binding partners interact as rigid bodies. However, the superposition of mobile segments of the proteins in the complex and corresponding unbound forms produces higher RMSDs (**Table 1**, column 7) and can be as large

TABLE 1 | General features of protein complexes studied.

Type protein complex		PDB codes (Bound: chain/Unbound: chain) (X-ray resolution, Å)	Amino acid length	*N*	Global RMSD (Å)	Local RMSD Regions of motion (residue:Å)	Cutoff, r_{opt} (Å)
Type I	**Alpha amylase: Alpha amylase inhibitor**	1CLV:A/1JAE:A (2.00/1.65)	2–471	470	0.5	81–90: 1.8 290–300: 2.5 440–450:1.7	12
		1CLV:I/1HTX:A (2.00/NMR)	501–532	32	0.9	511–520:1.0 520–525:1.3	
	MAP kinase-activated protein kinase 2:Mitogen-activated protein kinase 14	2OZA:A/1KWP:A (2.70/2.80)	51–215	165	2.0	65–70:6.2 70–75:8.0 75–80:3.0	14
		2OZA:B/1P38:A (2.70/2.10)	16–169	154	1.7	16–21:1.8 25–30:2.3 30–35:5.3 35–40:2.6 55–58:2.4 97–100:2.2 115–120:2.9 120–125:2.9	
	Trypsin:Trypsin inhibitor	1AVX:A/1QQU:A (1.90/1.63)	16–245	230	0.5	95–100:0.85 165–170:0.85 215–220:0.90 240–245:0.75	13
		1AVX:B/1BA7:A (1.90/2.50)	501–623	123	0.5	545–550:0.6 595–600:0.8	
	Ras-related protein Ral-A:Mono-ADP-ribosyltransferase C3	2A9K:A/1U90:A (1.73/2.00)	13–178	166	0.6	47–51:0.8 56–62:0.8 70–75:2.1	11
		2A9K:B/2C8B:X (1.73/1.70)	45–245	201	1.2	70–82:1.2 140–150:1.5 180–185:2.9 209–216:3.3	
	Ribonuclease SA:Barstar	1AY7:A/1RGH:A (1.70/1.20)	1–96	96	0.5	28–30:0.7 39–41:0.7 61–65:0.9	13
		1AY7:B/1A19:A (1.70/2.76)	1–89	89	0.6	5–15:0.95 55–66:0.95	
	Carboxypeptidase A:Metallocarboxypeptidase inhibitor	4CPA:A/1YME:A (2.50/1.53)	1–307	307	0.5	132–136:1.8 245–249:1.2	13
		4CPA:I/1H20:A (2.50/NMR)	4–37	34	1.1	10–16:1.6 17–21:2.1	
	Ribonuclease A:Ribonuclease inhibitor	1DFJ:E/9RSA:A (2.50/1.80)	1–124	124	0.7	15–17:0.8 90–95:1.2 110–114:1.0	12
		1DFJ:I/2BNH:A (2.50/2.30)	1–456	456	1.5	1–30:2.9 40–55:2.6 414–423:2.7 443–451:3.1	
	Superoxide dismutase:copper chaperone for superoxide dismutase	1JK9:A/2JCW:A (2.90/1.70)	1–153	153	0.8	51–62:2.5	13
		1JK9:B/1QUP:A (2.90/1.80)	3–222	220	4.3	3–10:5.4 30–40:5.3 50–70:6.9 160–170:5.7	
	Carboxypeptidase A1:Metallocarboxypeptidase inhibitor	2ABZ:A/1M4L:A (2.16/1.25)	5–305	301	0.4	245–250:1.3	13

(Continued)

TABLE 1 | Continued

Type protein complex		PDB codes (Bound: chain/Unbound: chain) (X-ray resolution, Å)	Amino acid length	*N*	Global RMSD (Å)	Local RMSD Regions of motion (residue:Å)	Cutoff, r_{opt} (Å)
		2ABZ:C/1DTV:A (2.16/NMR)	5–65	61	1.3	12–15:2.1 52–56:1.5	
	Ferredoxin-NADP reductase:Ferredoxin I	1EWY:A/1GJR:A (2.38/2.10)	9–303	295	1.1	50–54:1.4 67–74:3.3 102–112:2.0 221–236:2.3 262–270:1.4 280–287:1.2	12
		1EWY:C/1CZP:A (2.38/1.17)	1–98	98	0.8	10–15:1.5 50–67:1.5	
	Cysteine protease:cysteine protease Inhibitor	1PXV:A/1X9Y:A (1.80/2.50)	223–392	170	2.5	330–339:9.1 375–382:3.9	11
		1PXV:C/1NYC:A (1.80/1.40)	0–109	110	0.9	0–5:1.6 91–96:1.3 104–109:2.7	
Type II	**Chemotaxis protein CHEY:Chemotaxis protein CHEA**	1FFW:A/3CHY:A (2.70/1.66)	2–129	128	0.5	90–92:1.1 110–112:0.9 124–130:1.3	12
		1FFW:B/1FWP:A (2.70/NMR)	160–226	67	1.8	165–170:2.6 200–205:2.6 210–215:2.3	
	Cell division protein FTSZ:Hypothetical protein PA3008	1OFU:A/2VAW:A (2.10/2.90)	11–316	306	0.5	70–72:0.9 122–125:0.9 202–208:2.7 231–235:1.2 268–273:1.8 288–293:0.9 299–306:2.0	12
		1OFU:X/1OFT:A (2.10/2.90)	45–160	116	0.8	70–72:1.1 87–90:1.0	
	Complement C3:Fibrinogen-binding protein	3D5S:A/1C3D:A (2.30/1.80)	8–298	291	0.4	44–51:1.7 165–171:1.0	12
		3D5S:C/2GOM:A (2.30/1.25)	15–75	61	0.4	15–19:0.7 70–75:0.9	
	Bovine hymotrypsinogen A:human pancreatic secretory trypsin inhibitor (Kazal-type)	1CGI:E/2CGA:A (2.30/1.80)	1–245	245	1.5	31–39:1.2 140–154:4.5 183–195:3.5 216–223:1.6	13
		1CGI:I/1HPT:A (2.30/2.30)	(1–56)	56	1.8	1–5:4.4 10–14:3.2 17–20:1.5	
	Elastase:Elafin	1FLE:E/1QNJ:A (1.90/1.10)	16–245	229	0.9	59–62:1.4 96–106:1.6 166–176:1.7 186–196:1.3 216–226:1.7	13 Å
		1FLE:I/2REL:A (1.90/NMR)	11–57	47	2.8	11–43:3.5	
Type III	**Interleukin-6 receptor beta chain Leukemia inhibitory factor**	1PVH:A/1BQU:B (2.50/2.00)	101–301	201	0.9	130–135:1.7 210–215:1.3 240–245:1.3 295–301:1.8	12
		1PVH:B/1EMR:A (2.50/3.50)	22–180	159	0.9	135–140:1.3 140–150:2.0 150–160:1.6	

(Continued)

TABLE 1 | Continued

Type protein complex		PDB codes (Bound: chain/Unbound: chain) (X-ray resolution, Å)	Amino acid length	*N*	Global RMSD (Å)	Local RMSD Regions of motion (residue:Å)	Cutoff, r_{opt} (Å)
	Alkaline metalloproteinase:Proteinase inhibitor	1JIW:P/1AKL:A (1.74/2.00)	1–470	470	1.2	20–22:1.5 185–195:7.3	10
		1JIW:I/2RN4:A (1.74/NMR)	8–105	98	1.4	18–25:3.4 50–55:1.2 95–98:2.2	
	Heat shock protein HSP82:HSP90 Co-chaperone CDC37	1US7:A/1AH6:A (2.30/1.80)	2–207	206	0.8	54–57:1.3 91–99:3.5 100–117:1.7 198–203:1.3	10
		1US7:B/2W0G:A (2.30/1.88)	148–276	129	1.0	224–228:1.4 235–240:1.3 240–255:2.7 270–276:1.9	
	Trypsinogen:Bowman-Birk proteinase inhibitor precursor	1D6R:A/2TGT:A (2.30/1.70)	19–245	217	0.6	96–99:0.7 144–146:0.8 188–194:2.6 217–220:1.6	11
		1D6R:I/1K9B:A (2.30/2.80)	7–63	57	1.0	30–35:1.3 42–46:1.4 57–59:1.1	
Type IV	**Falcipain 2:Chagasin**	2OUL:A/2GHU:A (2.20/3.10)	15 to 224	240	0.6	1–5:2.1 115–120:1.0 153–158:1.2 185–195:2.6	12
		2OUL:B/2H7W:A (2.20/1.70)	4–110	107	0.6	60–65:1.7	
	ATP-dependent Clp protease ATP-binding subunit clpA: ATP-dependent Clp protease adaptor protein clpS	1R6Q:A/1R6B:X (2.35/2.25)	1–141	141	1.0	15–20:1.2 68–76:2.8 91–102:1.6	13
		1R6Q:C/3O1F:A (2.35/1.40)	26–106	81	0.4	37–40:0.6	
	NAD(P) transhydrogenase subunit alpha part 1: NAD(P) transhydrogenase subunit beta	2OOR:A/1L7D:A (2.32/1.81)	1–220	220	0.6	45–50:0.9 80–85:1.3 164–170:1.6 204–207:1.2 214–220:1.1	12
		2OOR:C/1E3T:A (2.32/NMR)	30–201	172	2.2	30–50:2.8 140–150:2.9 170–190:2.4	
	Alpha-Chymotrypsin:Protease inhibitor LCMI II	1GL1:A/1MTN:F (2.10/2.80)	16–146	131	0.5	71–82:1.3 143–146:1.0	12
		1GL1:I/1PMC:A (2.10/NMR)	2–33	32	1.6	7–23:2.1	
	Alpha amylase:Tendamistat	1BVN:P/1PIG:A (2.50/2.20)	2–496	495	0.7	51–55:0.8 108–112:2.3 140–150:1.5 238–243:1.9 303–308:2.7 347–351:1.4	13
		1BVN:T/1HOE:A (2.50/2.00)	804–874	71	0.5	827–830:0.6 838–840:0.7 860–862:0.7	

as 9.1 Å (see the case of cysteine protease and its inhibitor; Filipek et al., 2003), the largest deviations being mostly confined to flexible unstructured stretches, i.e., turns and bends.

Recording the residues identified by PRS (**Table 2**; those with C_i values exceeding the value listed in column 6), displays those encouraging the conformational change from the bound form

TABLE 2 | PRS results and classification of the protein complexes.

Type protein complex		PDB codes (Bound:chain/ Unbound:chain) (X-ray resolution, Å)	Correlation Bound/Unbound	C_i^{max}	$>C_i$	Interface residues	First shell residues	Remote site residues
Type I	**Alpha amylase:Alpha amylase inhibitor**	1CLV:A/1JAE:A (2.00/1.65)	0.92	0.87	0.87	G292:A	N/A	N/A
		1CLV:I/1HTX:A (2.00/NMR)	0.75	0.76	0.66	W57:A, N138:A, V151:A, G152:A	N/A	N/A
	MAP kinase-activated protein kinase 2:Mitogen-activated protein kinase 14	2OZA:A/1KWP:A (2.70/2.80)	0.97	0.68	0.65	L70:A	G71:A	N/A
		2OZA:B/1P38:A (2.70/2.10)	0.96	0.80	0.80	N/A	N/A	G171:A, Q175:A, Y176:A
	Trypsin:Trypsin inhibitor	1AVX:A/1QQU:A (1.90/1.63)	0.9	0.62	0.59	N/A	N/A	L520:B, K552:B
		1AVX:B/1BA7:A (1.90/2.50)	0.87	0.42	0.40	N/A	N/A	S579:B, F580:B, A581:B, D598:B, K611:B
	Ras-related protein Ral-A:Mono-ADP-ribosyltransferase C3	2A9K:A/1U90:A (1.73/2.00)	0.78	0.8	0.79	N/A	N/A	Y66:B, G67:B, L68:B, S69:B, D112:B
		2A9K:B/2C8B:X (1.73/1.70)	0.73	0.68	0.62	N/A	N/A	S181:B, F209:B, A210:B, G211:B
	Ribonuclease SA:Barstar	1AY7:A/1RGH:A (1.70/1.20)	0.93	0.72	0.69	N/A	N/A	L41:B, T42:B, G43:B, W44:B
		1AY7:B/1A19:A (1.70/2.76)	0.75	0.52	0.50	N/A	N/A	E8:B, E57:B, Q58:B
	Carboxypeptidase A:Metallocarboxypeptidase inhibitor	4CPA:A/1YME:A (2.50/1.53)	0.97	0.70	0.70	N/A	N/A	S134:A
		4CPA:I/1H20:A (2.50/NMR)	0.50	0.70	0.66	N/A	N/A	K177:A, S199:A, I274:A
	Ribonuclease A:Ribonuclease inhibitor	1DFJ:E/9RSA:A (2.50/1.80)	0.76	0.59	0.58	N/A	N/A	G186:I, D213:I, P450:I, G451:I
		1DFJ:I/2BNH:A (2.50/2.30)	0.80	0.91	0.89	N/A	N/A	L22:I, A46:I, L47:I, R48:I, A49:I
	Superoxide dismutase:copper chaperone for superoxide dismutase	1JK9:A/2JCW:A (2.90/1.70)	0.70	0.78	0.76	N/A	I69:B	D67:B, A68:B
		1JK9:B/1QUP:A (2.90/1.80)	0.70	0.60	0.57	N/A	N/A	C27:B, P54:B, S55:B
	Carboxypeptidase A1:Metallocarboxypeptidase inhibitor	2ABZ:A/1M4L:A (2.16/1.25)	0.96	0.75	0.73	Q16:C, V17:C	C18:C	N/A
		2ABZ:C/1DTV:A (2.16/NMR)	0.56	0.75	0.75	N/A	E31:C	N/A
	Ferredoxin-NADP reductase:Ferredoxin I	1EWY:A/1GJR:A (2.38/2.10)	1.00	0.70	0.67	N/A	N/A	V67:A, D68:A, K69:A
		1EWY:C/1CZP:A (2.38/1.17)	0.80	0.74	0.72	N/A	I62:A	T164:A, F183:A

(Continued)

TABLE 2 | Continued

Type protein complex		PDB codes (Bound:chain/ Unbound:chain) (X-ray resolution, Å)	Correlation Bound/Unbound	C_i^{max}	$>C_i$	Interface residues	First shell residues	Remote site residues
	Cysteine protease:cysteine protease Inhibitor	1PXV:A/1X9Y:A (1.80/2.50)	0.55	0.9	0.85	N/A	H54:C	V10:C, Y11:C, H44:C
		1PXV:C/1NYC:A (1.80/1.40)	0.72	0.80	0.80	N/A	N/A	V109:C
Type II	**Chemotaxis protein CHEY:Chemotaxis protein CHEA**	1FFW:A/3CHY:A (2.70/1.66)	0.83	0.75	0.74	N/A	N/A	K190:B, G191:B, L195:B, A197:B
		1FFW:B/1FWP:A (2.70/NMR)	0.77	0.61	0.54	N/A	N/A	G52:A, V54:A, D57:A, N59:A
	Cell division protein FTSZ:Hypothetical protein PA3008	1OFU:A/2VAW:A (2.10/2.90)	0.92	0.83	0.82	H89:X, R93:X	L87:X, T88:X	N/A
		1OFU:X/1OFT:A (2.10/2.90)	0.90	0.58	0.56	I207:A	D210:A, L271:A, S272:A	N/A
	Complement C3:Fibrinogen-binding protein	3D5S:A/1C3D:A (2.30/1.80)	0.92	0.70	0.69	N/A	N67:C, K70:C, Q71:C	N/A
		3D5S:C/2GOM:A (2.30/1.25)	0.78	0.80	0.78	N/A	N/A	R10:A, L11:A, K12:A, H13:A, L14:A, I15:A, V16:A, T17:A
	Bovine hymotrypsinogen A:human pancreatic secretory trypsin inhibitor (Kazal-type)	1CGI:E/2CGA:A (2.30/1.80)	0.60	0.70	0.68	T30:I, Y31:I, P32:I	N/A	N/A
		1CGI:I/1HPT:A (2.30/2.30)	0.48	0.83	0.79	G197:E	A179:E	N/A
	Elastase:Elafin	1FLE:E/1QNJ:A (1.90/1.10)	0.96	0.75	0.65	N/A	N/A	T11:I, K12:I, P13:I, L33:I, K34:I
		1FLE:I/2REL:A (1.90/NMR)	-0.30	0.71	0.66	N/A	N/A	L123:E, A208:E, V209:E
Type III	**Interleukin-6 receptor beta chain Leukemia inhibitory factor**	1PVH:A/1BQU:B (2.50/2.00)	0.89	0.85	0.83	N/A	N/A	R276:A, I277:A, E294:A, A295:A, S296:A, G297:A
		1PVH:B/1EMR:A (2.50/3.50)	0.84	0.58	0.54	N/A	N/A	G147:B, P148:B, D149:B, T150:B
	Alkaline metalloproteinase: Proteinase inhibitor	1JIW:P/1AKL:A (1.74/2.00)	0.98	0.85	0.78	N191:P, A192:P	G193:P	N/A
		1JIW:I/2RN4:A (1.74/NMR)	0.80	0.79	0.73	N/A	N/A	E21:I, A22:I
	Heat shock protein HSP82:HSP90 Co-chaperone CDC37	1US7:A/1AH6:A (2.30/1.80)	0.80	0.79	0.79	N/A	A97:A	N/A
		1US7:B/2W0G:A (2.30/1.88)	0.92	0.79	0.79	N/A	N/A	A244:B
	Trypsinogen:Bowman-Birk proteinase inhibitor precursor	1D6R:A/2TGT:A (2.30/1.70)	0.96	0.78	0.78	N/A	N/A	N/A

(Continued)

TABLE 2 | Continued

Type protein complex		PDB codes (Bound:chain/ Unbound:chain) (X-ray resolution, Å)	Correlation Bound/Unbound	C_i^{max}	$>C_i$	Interface residues	First shell residues	Remote site residues
		1D6R:I/1K9B:A (2.30/2.80)	0.50	0.70	0.70	N/A	N/A	H33:I, S34:I
Type IV	**Falcipain 2:Chagasin**	2OUL:A/2GHU:A (2.20/3.10)	1.00	0.87	0.83	N/A	D148:A, Y37:B, G41:B	G40:B
		2OUL:B/2H7W:A (2.20/1.70)	0.25	0.78	0.76	N/A	I68:A	S113:A, V114:A, D148:A, F219:A
	ATP-dependent Clp protease ATP-binding subunit clpA:ATP-dependent Clp protease adaptor protein clpS	1R6Q:A/1R6B:X (2.35/2.25)	0.85	0.82	0.77	N/A	E73:A, K49:C	D45:C, L61:C
		1R6Q:C/3O1F:A (2.35/1.40)	0.76	0.76	0.74	S118:A	N/A	E7:A, Y122:A
	NAD(P) transhydrogenase subunit alpha part 1:NAD(P) transhydrogenase subunit beta	2OOR:A/1L7D:A (2.32/1.81)	0.23	0.68	0.68	N/A	N103:C, P105:C	L214:A, T220:A
		2OOR:C/1E3T:A (2.32/NMR)	0.52	0.50	0.49	A166:A	N/A	M167:C
	Alpha-Chymotrypsin:Protease inhibitor LCMI II	1GL1:A/1MTN:F (2.10/2.80)	1.00	0.76	0.52	N/A	N/A	S76:A, S77:A
		1GL1:I/1PMC:A (2.10/NMR)	−0.40	0.87	0.76	C58:A, C14:I	K13:I	N/A
	Alpha amylase:Tendamistat	1BVN:P/1PIG:A (2.50/2.20)	0.89	0.64	0.64	A823:T	C811:T	N/A
		1BVN:T/1HOE:A (2.50/2.00)	0.68	0.57	0.53	N/A	N/A	Q5:P, T6:P, Q7:P, S8:P, R10:P, V804:T

to unbound form do not necessarily reside on these flexible structures. We have collected 161 effector residues from PRS calculations (**Table 2**, columns 7–9). While 63 of them are located on flexible loops with large motions and high RMSD values, we identify 30 residues residing on α-helices and 35 on β-strands. The remaining are on loops that do not display any large structural change upon binding.

The average motions of the proteins, quantified by the root mean square fluctuations (RMSFs), are usually expected to dampen upon binding, especially at the binding interface residues, even when the protein conformation is unaltered (Baysal and Atilgan, 2001a). RMSFs of each protein complex constituents in their bound and unbound form are derived from auto-correlation of the residues in each protein pair. By treating the $\mathbf{H}^{-1}$ matrix as an $N \times N$ supermatrix, whose ij^{th} element is the 3 × 3 second moment matrix of correlations between the x-, y-, and z-components of the fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ of residues i and j (Baysal and Atilgan, 2001b) are calculated, whose diagonal elements predict the RMSFs (Atilgan et al., 2001). The cut-off distances, r_{opt}, optimized for building the Hessian matrix of each protein complex have the same values as r_c chosen for PRS analysis. We report the correlation values between proteins in their bound and unbound form for all the protein pairs; the similarity between the RMSF profiles of a protein in its bound and unbound form is expressed as a Pearson correlation and is listed in column 4 of **Table 2**. We observe that there is a significant change in RMSF of binding region residues in one of the constituents in each pair, so far as to have a negative correlation in some cases; e.g., fluctuation patterns in some regions of the protein is reversed upon complexation. This means that the local fluctuations of the interface area vary in at least one protein upon complex formation, while local fluctuations of their binding partners display the same pattern as the unbound form. However, in an exceptional case of transhydrogenase complex (pdb code: 2OOR; Bhakta et al., 2007), we observe low correlations, 0.23 and 0.52, between the binding proteins and their unbound form. For this protein complex, the RMSF curves of both the constituents display a significant transformation of fluctuation pattern upon protein binding.

Features of the Amino Acids Involved in Dissociation Event

From the 8828 residues, 161 of them are selected by PRS. These residues, whose perturbation encourages the unbound over

TABLE 3 | Types of amino acids selected by PRS compared to all residues in the study.

	Non-polar				Charged									Polar amino						
Residue type	**A**	**V**	**L**	**I**	**M**	**F**	**G**	**W**	**P**	**D**	**E**	**R**	**K**	**N**	**Q**	**T**	**Y**	**S**	**H**	**C**
PRS residues (%)	10	7	9	4	1	3	13	1	3	6	4	3	6	4	4	5	4	9	2	3
All residues (%)	8	7	9	5	2	4	8	2	4	6	6	4	6	5	4	6	4	7	2	3

TABLE 4 | Secondary structure attributes of amino acids selected by PRS compared to all residues in the study.

	α-helix	**π-helix**	**3-10 helix**	**β-strand**	**Isolated bridge**	**Turn**	**Coil**	**Total**
PRS residues (%)	19	0.0	2	20	1	29	29	100.0
All residues (%)	26	0.0	4	24	2	26	18	100.0

the bound form, are distributed among all amino acid types. The percentage of each amino acid type in our analysis pool and their contributions in PRS analysis are listed in **Table 3**. PRS does not display any preference over amino acid types and any contribution to PRS selection is corroborated to the population of the amino acid type in the total analysis pool. For example, methionines and tryptophans, each with 1% frequency are the least detected residues by PRS. They are also less frequently seen in the analysis pool (2% of the population). The only residue type that is observed significantly more that in the average pool is glycine which constitutes 13% of all PRS selected residues, compared to its 8% abundance in the residue data set.

In **Table 4**, we report the local secondary structure attributes of residues detected by PRS compared to all residues. The secondary structures are assigned by the "Timeline" plugin of VMD and are calculated based on the STRIDE algorithm (Humphrey et al., 1996). Among all the protein complexes, we do not find any π-helix type of structure. Residues in the total analysis pool are mostly populated by turns, α-helices and β-strands with 26, 26, and 24% distribution, respectively. However, we find that most preferred regions by PRS defined residues are on coils and turns, each with 29% of all PRS defined residues, although they populate only 44% of the analysis pool. In particular, the enhancement of coil residues in the PRS selection is statistically significant, as these are represented by 29% in the PRS sub-pool, up from 18% of all residues in the original pool of residues.

We divide the protein structure into three zones; interface, first coordination and remote, so as to categorize the location of the PRS selected residues. The interface of the two proteins present in the complex is defined by the C_α atoms of the residues from the two sides of the pair residing within 7 Å cut-off distance of each other. We define first coordination shell residues as those located within 7 Å cut-off distance from any interface residue. All remaining residues are classified as remote, defined as those residing beyond the first coordination shell of the interface. We observe that except for the case of alpha amylase and its inhibitor (pdb code:1CLV; Pereira et al., 1999), PRS selects for remote residues (**Table 2**). In fact, in 9 cases PRS selected only residues away from the interface. The remaining 16 protein complexes display residues from different parts of the protein in their PRS analysis, including, but not limited to the interface. In fact, these residues are overwhelmingly located on or near the outer surface, as indicated by their depth values from the surface as calculated by the DEPTH server (Tan et al., 2013) and listed in **Supplementary Table S1**. In fact, those few that are deeply located (depth greater than 5 Å; shown in bold) are part of a network of interactions whose one end is located on the surface. Thus, the interface of a protein complex is not the controlling region for dissociation of the two proteins.

Remarkably, residues signaling the dissociation of each protein in a given complex are not located on the same protein in all cases. Accordingly, effective sites involved in the dissociation of various protein complexes found in PRS analysis are categorized into four groups based on their responses to the perturbations on the protein. Proteins in which dissociation is signaled through remote residues of the complex are labeled as Type I (11 cases). In this group, PRS top rated residues are all confined to one of the binding proteins. Thus, residues on this protein also control the conformational changes of the binding partner. Type II are the proteins in which residues confined to one of the proteins control the other binding protein and vice versa. Type III are the proteins in which each constituent controls its own dissociation event; therefore, essential residues are clustered on the "self" protein. Finally, in Type IV both partners are essential for the dissociation to occur, as residues signaling the dissociation are scattered on both binding partners. **Tables 1**, **2** are organized according to these four distinct groups (I–IV).

Long Range Control of Dissociation Is Coupled to Electrostatic Effects

In a subset of the cases, perturbation of specific sites on only one of the constituents in the protein complex modulates dissociation. We label these as Type I group of protein complexes. The functional amino acids defined by PRS which are involved in disintegrating the contact network displays no specific perturbation location in Type I; they may be located on the interface, first shell or remote locations of the protein tertiary structure. Thus, the local perturbations which lead to global conformational shifts between bound and unbound states are not bundled in a specific region.

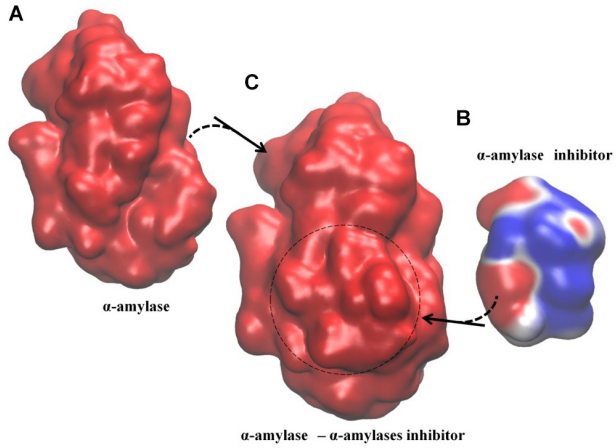


FIGURE 1 | Example of Type I electrostatic isocontour shifts upon binding; α-amylase/α-amylase inhibitor drawn at ±0.5 k_BT/e; blue is positive and red is negative. The signaling protein, α-amylase, where all PRS determined residues reside (**Table 2**), maintains its electrostatic potential distribution while α-amylase inhibitor displays altered electrostatic potential distribution. **(A)** α-amylase in the unbound form, with overall negative electrostatic potential distribution along the surface. **(B)** α-amylase inhibitor in its unbound form with a mixed pattern of negative and positive negative electrostatic potential distribution along the surface. **(C)** α-amylase/α-amylase inhibitor complex with overall negative electrostatic potential distribution. The spatial orientation of the proteins in the complex is kept the same as the respective unbound forms. Dashed circle indicates positioning of the inhibitor in the complex.

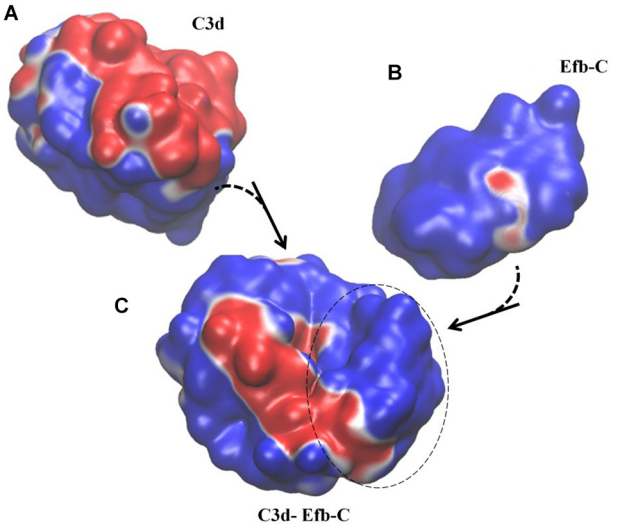


FIGURE 2 | Example of Type II electrostatic isocontour shifts upon binding; Efb-C and its complement target C3d drawn at ±0.5 k_BT/e. Blue is positive and red is negative. **(A)** C3d in its unbound form, with a mix of negative/positive electrostatic potential distribution along the surface. **(B)** Efb-C in its unbound form with predominantly positive electrostatic potential distribution along the surface. **(C)** Efb-C/C3d complex. Dashed circle displays Efb-C protein in the complex and the rest of the surface belongs to C3d protein. Both proteins display a mixture of negative/positive electrostatic potential distribution along the surface. In particular, positive surface of the Efb-C displays increased negative areas upon complex formation, while C3d loses negative patches. The spatial orientation of the proteins in the complex is kept the same as that presented in unbound forms.

To determine if these long range effects are controlled by electrostatics, we obtain the electrostatic potential distributions on the biomolecular surface using the APBS package (Baker et al., 2001). In APBS calculations, parameters are set to their default values; i.e., biomolecular and solvent dielectric constant are set to 2 and 78.54, respectively, the radius of the solvent molecules is 1.4 Å and the temperature is 298.15 K; finally, the cubic B-spline discretization method is used to grid biomolecular point charges. Electrostatic effects play a major role in the functionality of this group of protein complexes. Monitoring the electrostatic potential distribution along the protein surfaces in their bound and unbound forms reveals that any given protein in the complex whose electrostatic potential distribution state is stable in their bound and unbound form is also the protein controlling dissociation. Conversely, the binding partner that does not have any PRS determined residue displays considerable change in its charge distribution. In **Figure 1A**, we exemplify how the electrostatic potential distribution on the surface of the protein changes from the free to the bound form.

Type II group represents another set of protein complexes with remotely controlled conformation changes from the bound to the unbound form. In this group, there is cross-controlled dissociation; i.e., residues that lead to the conformational change upon dissociation on one protein are located on the partner in the complex. We observe that perturbations in a stretch of consecutive resdues is required to trigger the interconversion between two conformational endpoints (see **Table 2**, Type II). In addition, analysis of the electrostatic potential distribution shows that the proteins interacting with each other possess a similar state of charge distribution. Thus, if the unbound forms had a different electrostatic potential distribution, they reorient themselves to the same state upon protein complex formation. In **Figure 2**, we illustrate electrostatic potential surfaces before and after complex formation in Efb-C and its complement target C3d (pdb code: 3D5S; Haspel et al., 2008). In fact, it has been reported that formation and stability of Efb-C binding to C3d is electrostatic in nature (Haspel et al., 2008). Kinetic experiments in salt concentrations of 75–600 mM indicate the sensitivity of association/dissociation phases of wild type and various mutants to ionicity (Haspel et al., 2008). This suggests overall electrostatic contribution to be of importance in the initial complex formation and further in stabilizing the complex under the prevailing conditions. Our PRS analysis of Types I and II is further improved by these observations such that under various environmental perturbations, disruption of long-range and short-range electrostatic complementarity seem to impair stability and affect complex formation of binding partners.

Remarkably, electrostatic effects provide an excellent description for the observed pattern in both Type I and Type II complexes. Electrostatic interactions are the primary factors of pH dependent processes in biochemical reactions. Particularly, we find that among the 16 protein complexes

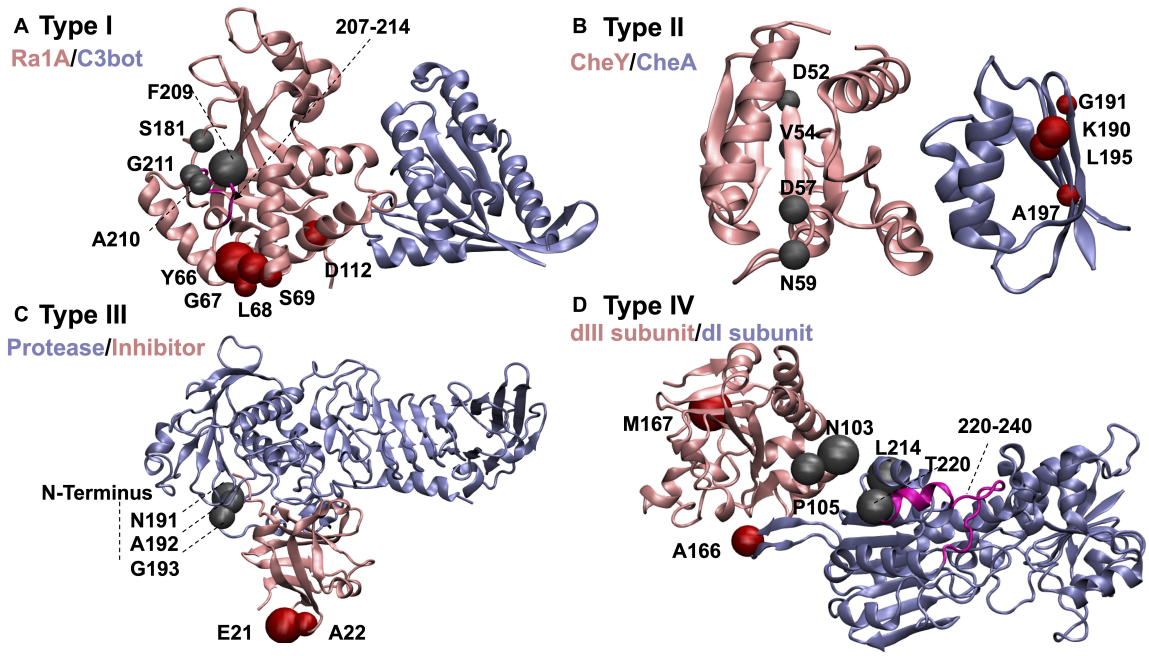


FIGURE 3 | Example complexes demonstrating dissociation scenarios discussed in the text. The color of each subunit is associated with its name. Residues controlling dissociation of the pink colored proteins are displayed as red beads, and those for ice blue ones as gray beads: **(A)** C3bot-RalA complex as an example of Type I proteins. PRS selected residues that control the dissociation are confined to RalA compartment. Substrate recognition site of C3bot (residues 207–214) is shown in magenta. **(B)** Chemotaxis CheA:CheY complex as an example of Type II proteins. **(C)** Alkaline protease and its cognate inhibitor as an example of Type III proteins. **(D)** dI_2dIII_1 complex of proton-translocating transhydrogenase as an example of Type IV proteins. Loop residues that become less mobile upon ligand binding due to surface closing of the protein are displayed in magenta.

included into Type I and II groups, 13 of them belong to enzymes. Enzymatic activities are known to be pH dependent and protonation state of catalytic and active site residues are effective in potential distribution of the binding region. Consequently, charge distribution of these regions will modulate the interactions between the proteins and the reaction products. It has been reported that, enzymes make use of their preoriented environment to stabilize the transition state and the reduction in catalytic energy is accomplished by electrostatic stabilization of the active site of the enzyme (Warshel, 1998; Warshel et al., 2006).

In the same vein, we propose that the electrostatic characteristics of residues top-rated by PRS might be found essential to specificity and ligand binding properties in enzymatic reactions. For example, in the Type I complex C3bot-RalA (Ras-related protein Ral-A:Mono-ADP-ribosyltransferase C3, pdb code:2A9K; Pautsch et al., 2005), displayed in **Figure 3A**, residues 207–214 of C3bot (displayed in magenta in **Figure 3A**) are part of the substrate recognition site and important in catalytic activity. PRS selects residues 209–211 as candidate positions to provoke the dissociation of the complex. Another key region involved in dissociation upon force application is residue 112 of C3bot, which also contributes to complex formation. In addition, residue 181 is located on a loop responsible in substrate recognition. Accordingly, perturbation of specific sites found by PRS analysis and experimentally verified to be functionally important might reorient the enzyme/substrate dipoles that organize the catalysis and destabilize the charged transition state (Thomas et al., 1985). These changes might thus prevent the binding event or promote dissociation after the chemical reaction has terminated.

We also find that the effect of the local charge distribution on enzyme functions is not limited to the active site and that the remote locations on the protein are effectively involved in the dissociation process. Mutations of charged patches remote from either the protein or ligand binding site might alter the binding kinetic rates, shift pK_as and lead to weak molecular recognition (Thomas et al., 1985). In response to a particular perturbation, exposure to a different environment, reorganization of charged atomic groups and dielectric relaxation of the protein affects the electrostatic potential distribution of the interface or active site region considerably, creating the steering forces that guide the dissociation reaction. Thus, a local change of interactions at a remote site leads to a global structural change that modifies the organization of the interface contact network and leads to dissociation of the two proteins. The contribution of this distal perturbation on enzyme/inhibitor activity basically may be viewed as leading to a cooperative conformational transition. In both Type I and Type II, the information transmission between remote functional sites on one protein and the entire structure of the interacting partner naturally occurs via the interface linkage. In such proteins, the network of contacts in the interface could form a so-called "conductive" region so that the signal from

one protein is transferred through the interface to control the functionality of the second binding protein.

As one of the test beds, we have studied chemotaxis CheA:CheY complex (pdb code:1FFW; Gouet et al., 2001). PRS finds residues 52, 54, 57, and 59 of CheY to play a role in selecting the unbound over bound conformation in the presence of an external force (**Figure 3B**). This complex is an example of remote communication between a two-component signal-transducer. pH dependent catalytic activity has also been observed in chemotaxis CheA:CheY, in which ligand binding on CheA is conducted to CheY. Phosphorylation reaction on H48 of CheA subsequently transfers to D57 of CheY and the complex dissociates (Silversmith et al., 1997). CheY itself is incapable of providing an acidic residue during the phosphorylation event and the complex formation with CheA results in a conformational change on CheY as an acidic residue (D57) donates a proton to a phosphodonor in an optimal orientation and the protein–protein phosphotransfer occurs (Silversmith et al., 1997). The pH dependence of the phosphotransfer kinetics in the pH range of 7.5–10, studied through two mutants of CheY active site residues shows simply a moderate decrease in rate constants compared to the wild type CheY. This observation suggests that conserved active site residues do not have an essential and direct role in catalysis. Thus, the loss of activity throughout this range for phosphotransfer to wild type CheY is attributed not to deprotonation events on CheY; rather it is likely due to deprotonations in CheA (Silversmith et al., 1997). Variability of position 59 of CheY as a non-conserved and indirect active site residue in modulating the autophosphorylation of CheY with small molecule phosphodonors shows no detectable binding between the phosphodonor and CheY, validating the significant impact of position 59 on autophosphorylation kinetics (Thomas et al., 2013). Mutation of N59 to R, K, M, L, A, D, and E results in both increase and decrease in autophosphorylation rate constants. Substitution with positively charged residues increases the kinetic rates while substitution with negatively charged residues decreases the rate, implying how local electrostatic interactions at position 59 modulate the CheY autophosphorylation (Thomas et al., 2013).

Auto-Controlled Conformational Transitions Are Correlated to Mechanical Organization of the Protein in Type III Complexes

In Type III group of protein complexes, dissociations of either constituent of the complexes are governed by a local perturbation on the respective protein. We label this behavior as auto-controlled dissociation. In this category, all the residues involved in global transformation of bound to unbound forms are located out of the interface region, trypsinogen and its Bowman-Birk proteinase inhibitor precursor (pdb code:1D6R; Koepke et al., 2000) being an exception. We observe that in Type III complexes, the global RMSD between bound and unbound forms of each protein varies between 0.6 and 1.4 Å. However, regions of high mobility exist in which the local RMSD may take values as high as 7.3 Å (**Table 1**). Such regions belong to unstructured surface exposed loops where they display large deviations upon complex formation. The amino acids identified by PRS in this group belong to these regions and very specific locations may be perturbed so that the local fluctuations of the amino acids choose the conformational switch to the target structure. Thus, mechanical motions of the loops produce essential conformational transitions under such point perturbations.

Contrary to Type I and II complexes, the electrostatic potential distribution along the protein surfaces of Type III reveals that each protein maintains the same electrostatic potential distribution state in their respective bound and unbound forms. This means that for any protein complex in this group, orientation of the negatively and positively charged surfaces of the constituents has the same distribution in their respective free forms. Thus, the analysis identifies an "insulating" interface area that prevents the allosteric communication between interacting proteins and each protein functions independently under various perturbations. We note that while there are also stretches with large RMSDs in Type II protein complexes, those regions cross-control the dissociation of the binding proteins remotely and their own shape changes are not mechanically controlled.

As an example case, for the alkaline protease and its cognate inhibitor (pdb code: 1JIW; Hege et al., 2001) classified as Type III, PRS finds residues 191–193 of protease which are in direct contact with the N-terminus of the inhibitor (**Figure 3C**). The latter has been shown to coordinate the catalytic zinc anion (Hege et al., 2001). Such an interaction adds to the structural stability and leads to a low dissociation constant. Upon any kind of perturbation applied to residues 191–193 of the protease, disruption of the interactions in this region would change the extended conformation of the N-terminus and modify the zinc coordination and thus might facilitate the dissociation process.

Combined Perturbations on Both Partners Is Essential for Dissociation in Type IV Complexes

In Type IV group of protein complexes, dissociation of the two constituents might be triggered by point perturbations on either subunit. This contrasts Type III complexes where perturbation of any subunit will mediate the disruption of interactions in the protein pair, facilitating dissociation. Different functional subdomains may exist on these proteins which contribute to binding to diverse set of ligands in their functional pathway and promote dissociation. Additionally, complexes in our test set involving transmembrane proteins also fall into this group. Presence of each fragment of the protein pair in different compartments of the cell environment assist their exposure to different perturbation scenarios and support the idea of simultaneous perturbation of both partners of the protein complex in the dissociation process. As an example, we focus on dI_2dIII_1 complex of proton-translocating transhydrogenase (pdb code:2OOR; Bhakta et al., 2007). The complex is found in the membrane compartment of the bacteria or animal cells. Proton transfer across the membranes is facilitated by conformational changes of transhydrogenase. PRS identifies residues A166, L214 and T220 of dI subunit and residues N103, P105 and M167

TABLE 5 | Summary of types of complexes and properties of their dissociation mechanisms.

Complex Type (# of complexes out of total 25)	Control	Range	Electrostatic change upon binding
I (11)	Centralized on one protein; residues on one protein controls conformational changes on both proteins	Long-range	Yes
II (5)	Cross-control; residues on one protein controls the conformational change in the other partner	Long-range	Yes
III (4)	Auto-control; residues on the same protein control their own conformational change	Local	No
IV (5)	Scattered control; residues on both proteins control the conformational changes	Local	No

of dIII subunit to be effective in dissociation of the complex (**Figure 3D**). dI and dIII protrude from the membrane while a third compartment dII spans the membrane. Thus, each part of the protein is exposed to a different part of the cell, making the protein susceptible to alternative perturbation scenarios. Residue T220 of dI is part of a loop (residues 220–240; **Figure 3D** magenta) that becomes less mobile upon ligand binding due to surface closing of the protein (Bhakta et al., 2007). Furthermore, M167 of dIII is in the neighboring site of H_2NADH and any perturbation on this site might alter the proton pump reactions due to changes made to interaction network of ligand binding region. Accordingly, any perturbation on H_2NADH binding region in dIII subdomain may alter the structural features of dI subdomain through remote communication.

CONCLUSION

There is plethora of work addressing association of proteins partaking in complexes, and the consensus is to focus on the interface to determine the major features of binding events (Gainza et al., 2020), concentrating on, e.g., pockets formed upon complexation (Li et al., 2004), prediction of binding energies based on the interface (Moreira et al., 2007), close-range electrostatic interactions (Kumar and Nussinov, 2002), and conserved residues along the interface (Kumar and Nussinov, 2002; Li et al., 2004). However, to alter protein functions, e.g., for therapeutic applications, it is also essential to understand the mechanisms affecting their dissociation, a question that has not been thoroughly explored, to the best of our knowledge. In this work, we have studied the characteristics of residues responsible for the dissociation of a set of 25 non-redundant protein complexes, using PRS as the predictor of residues whose perturbation encourages the unbound conformations. Significance of the residues identified by PRS are discussed in detail for four sample cases (**Figure 3**).

In a statistical analysis of the PRS identified residues, we find that in terms of residue types, the only significant enhancement is in glycine residues, up from 8% of all residues found in the protein set to 13% in the subset of residues implicated in protein dissociation. This is in contrast to the studies reporting on hotspots on the interaction surface, whereby tryptophan (21%), arginine (13.3%), and tyrosine (12.3%) have the highest probabilities of occurrence (Moreira et al., 2007). Moreover, PRS identifies residues labeled as controlling dissociation are also significantly enhanced on loops, and are predominantly located on the complex surface, remote from the interface. This finding is plausible, since in contrast to an association event whereby interface compatibility is a major determinant, exposed residues are expected to partake in disrupting the complex.

We find that dissociation events disclosed by PRS analysis may be classified into four main groups as summarized in **Table 5**. The nature of the events leading to dissociation are either expected to be due to mechanical perturbations arriving at certain locations on the surface, or due to environmental triggers that interfere with the electrostatic potential distribution of the complex. In the latter case, a signature event is in the change of the electrostatic potential distributions of one of the binding partners (see Type I in **Tables 1**, **2** and **Figure 1**) or both of them (Type II complexes; **Figure 2**). The physics of these observations are resolved by an electro-mechanical coupled ENM proposed by our group (Sensoy et al., 2017). Accordingly, we find that even in the absence of an external force, positional displacements may still be obtained, provided there is electromechanical coupling. The role of PRS is to identify residues where such equivalent forces are focused on and are relevant to the observed conformational change. In the absence of coupling, the conformational change may indeed be triggered by an external force, e.g., upon binding of a ligand. Therefore, PRS either identifies the residues that facilitate displacements by mechanical perturbations (Type III and IV), or those which are mechanical mimics to the effects expedited by changes in electric displacement (Type I and II). In the latter situation, a coupling term that links electrical drive to mechanical response survives.

This study is a step toward developing descriptors aimed at disrupting protein complexes with the aim of developing therapeutic approaches to alter the function of proteins working in tandem. In particular, targeting remote sites to destabilize interacting proteins using unique approaches will aid in the emerging field of allosteric drug design (Guarnera and Berezovsky, 2020).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

HA designed the research, conducted the computer experiments, analyzed the results, interpreted the data, wrote the manuscript, and constructed the figures. FJ wrote the manuscript, interpreted the data, and constructed the figures. AA and CA designed the research, guided the computer experiments and analyses, interpreted the data, guided the structure and contents of the manuscript, and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Scientific and Technological Research Council of Turkey (Grant Number 116F229).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00210/full#supplementary-material

REFERENCES

Abdizadeh, H., Atilgan, A. R., and Atilgan, C. (2015a). Detailed molecular dynamics simulations of human transferrin provide insights into iron release dynamics at serum and endosomal pH. *JBIC J. Biol. Inorgan. Chem.* 20, 705–718. doi: 10.1007/s00775-015-1256-4

Abdizadeh, H., Guven, G., Atilgan, A. R., and Atilgan, C. (2015b). Perturbation response scanning specifies key regions in subtilisin serine protease for both function and stability. *J. Enzyme Inhibit. Med. Chem.* 30, 867–873. doi: 10.3109/14756366.2014.979345

Abdizadeh, H., and Atilgan, C. (2016). Predicting long term cooperativity and specific modulators of receptor interactions in human transferrin from dynamics within a single microstate. *Phys. Chem. Chem. Phys.* 18, 7916–7926. doi: 10.1039/c5cp05107j

Adcock, S. A., and McCammon, J. A. (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* 106, 1589–1615. doi: 10.1002/chin.200630297

Ahmed, A., and Gohlke, H. (2006). Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins Struct. Funct. Bioinform.* 63, 1038–1051. doi: 10.1002/prot.20907

Atilgan, A. R., Aykut, A. O., and Atilgan, C. (2011). Subtle p H differences trigger single residue motions for moderating conformations of calmodulin. *J. Chem. Phys.* 135:10B613. doi: 10.1063/1.3651807

Atilgan, A. R., Durell, S., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80, 505–515. doi: 10.1016/s0006-3495(01)76033-x

Atilgan, C. (2018). Computational methods for efficient sampling of protein landscapes and disclosing allosteric regions. *Adv. Protein Chem. Struct. Biol.* 113, 33–63. doi: 10.1016/bs.apcsb.2018.06.001

Atilgan, C., and Atilgan, A. R. (2009). Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput. Biol.* 5:e000544. doi: 10.1371/journal.pcbi.1000544

Atilgan, C., Gerek, Z., Ozkan, S., and Atilgan, A. (2010a). Manipulation of conformational change in proteins by single-residue perturbations. *Biophys. J.* 99, 933–943. doi: 10.1016/j.bpj.2010.05.020

Atilgan, C., Okan, O. B., and Atilgan, A. R. (2010b). How orientational order governs collectivity of folded proteins. *Proteins Struct. Funct. Bioinform.* 78, 3363–3375. doi: 10.1002/prot.22843

Atilgan, C., Okan, O. B., and Atilgan, A. R. (2012). Network-based models as tools hinting at nonevident protein functionality. *Annu. Rev. Biophys.* 41, 205–225. doi: 10.1146/annurev-biophys-050511-102305

Aykut, A. O., Atilgan, A. R., and Atilgan, C. (2013). Designing molecular dynamics simulations to shift populations of the conformational states of calmodulin. *PLoS Comput. Biol.* 9:1003366. doi: 10.1371/journal.pcbi.1003366

Bahar, I., Atilgan, A. R., and Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Design* 2, 173–181. doi: 10.1016/s1359-0278(97)00024-2

Bahar, I., Lezon, T. R., Yang, L.-W., and Eyal, E. (2010). Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.* 39, 23–42. doi: 10.1146/annurev.biophys.093008.131258

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10037–10041. doi: 10.1073/pnas.181342398

Baysal, C., and Atilgan, A. R. (2001a). Coordination topology and stability for the native and binding conformers of chymotrypsin inhibitor 2. *Proteins Struct. Funct. Bioinform.* 45, 62–70. doi: 10.1002/prot.1124

Baysal, C., and Atilgan, A. R. (2001b). Elucidating the structural mechanisms for biological activity of the chemokine family. *Proteins Struct. Funct. Bioinform.* 43, 150–160. doi: 10.1002/1097-0134(20010501)43:2<150::aid-prot1027>3.0.co;2-m

Ben-Avraham, D. (1993). Vibrational normal-mode spectrum of globular proteins. *Phys. Rev. B* 47:14559. doi: 10.1103/physrevb.47.14559

Bhakta, T., Whitehead, S. J., Snaith, J. S., Dafforn, T. R., Wilkie, J., Rajesh, S., et al. (2007). Structures of the dI2dIII1 complex of proton-translocating transhydrogenase with bound, inactive analogues of NADH and NADPH reveal active site geometries. *Biochemistry* 46, 3304–3318. doi: 10.1021/bi061843r.s001

Bolia, A., Gerek, Z. N., and Ozkan, S. B. (2014). BP-Dock: a flexible docking scheme for exploring protein–ligand interactions based on unbound structures. *J. Chem. Inform. Model.* 54, 913–925. doi: 10.1021/ci4004927

Bradshaw, R. T., Patel, B. H., Tate, E. W., Leatherbarrow, R. J., and Gould, I. R. (2011). Comparing experimental and computational alanine scanning techniques for probing a prototypical protein–protein interaction. *Protein Eng. Design Select.* 24, 197–207. doi: 10.1093/protein/gzq047

Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P. E., Oberhauser, A. F., and Fernandez, J. M. (2003). The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Mol. Biol.* 10, 738–743. doi: 10.1038/nsb965

Chen, R., and Weng, Z. (2002). Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins Struct. Funct. Bioinform.* 47, 281–294. doi: 10.1002/prot.10092

Chennubhotla, C., Rader, A., Yang, L.-W., and Bahar, I. (2005). Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.* 2:S173. doi: 10.1088/1478-3975/2/4/s12

De Las Rivas, J., and Fontanillo, C. (2010). Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput. Biol.* 6:807. doi: 10.1371/journal.pcbi.1000807

Delarue, M., and Dumas, P. (2004). On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6957–6962. doi: 10.1073/pnas.0400301101

Dietz, H., Berkemeier, F., Bertz, M., and Rief, M. (2006). Anisotropic deformation response of single protein molecules. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12724–12728. doi: 10.1073/pnas.0602995103

Doruker, P., Jernigan, R. L., and Bahar, I. (2002). Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.* 23, 119–127. doi: 10.1002/jcc.1160

Durand, P., Trinquier, G., and Sanejouand, Y. H. (1994). A new approach for determining low-frequency normal modes in macromolecules. *Biopolym. Orig. Res. Biomol.* 34, 759–771. doi: 10.1002/bip.360340608

Filipek, R., Rzychon, M., Oleksy, A., Gruca, M., Dubin, A., Potempa, J., et al. (2003). The Staphostatin-Staphopain complex a forward binding inhibitor in complex with its target cysteine protease. *J. Biol. Chem.* 278, 40959–40966. doi: 10.2210/pdb1pxv/pdb

Gainza, P., Sverrisson, F., Monti, F., Rodola, E., Boscaini, D., Bronstein, M., et al. (2020). Deciphering interaction fingerprints from protein molecular surfaces

using geometric deep learning. *Nat. Methods* 17, 184–192. doi: 10.1038/s41592-019-0666-6

Gao, Y., Wang, R., and Lai, L. (2004). Structure-based method for analyzing protein–protein interfaces. *J. Mol. Model.* 10, 44–54. doi: 10.1007/s00894-003-0168-3

Gouet, P., Chinardet, N., Welch, M., Guillet, V., Cabantous, S., Birck, C., et al. (2001). Further insights into the mechanism of function of the response regulator CheY from crystallographic studies of the CheY–CheA124–257 complex. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 57, 44–51. doi: 10.1107/s090744490001492x

Guarnera, E., and Berezovsky, I. N. (2020). Allosteric drugs and mutations: chances, challenges, and necessity. *Curr. Opin. Struct. Biol.* 62, 149–157. doi: 10.1016/j.sbi.2020.01.010

Guven, G., Atilgan, A. R., and Atilgan, C. (2014). Protonation states of remote residues affect binding–release dynamics of the ligand but not the conformation of apo ferric binding protein. *J. Phys. Chem. B* 118, 11677–11687. doi: 10.1021/jp5079218

Han, R., Leo-Macias, A., Zerbino, D., Bastolla, U., Contreras-Moreira, B., and Ortiz, A. R. (2008). An efficient conformational sampling method for homology modeling. *Proteins Struct. Funct. Bioinform.* 71, 175–188. doi: 10.1002/prot.21672

Haspel, N., Ricklin, D., Geisbrecht, B. V., Kavraki, L. E., and Lambris, J. D. (2008). Electrostatic contributions drive the interaction between Staphylococcus aureus protein Efb-C and its complement target C3d. *Protein Sci.* 17, 1894–1906. doi: 10.1110/ps.036624.108

Hege, T., Feltzer, R. E., Gray, R. D., and Baumann, U. (2001). Crystal structure of a complex between *Pseudomonas aeruginosa* alkaline protease and its cognate inhibitor inhibition by a zinc–NH2 coordinative boND. *J. Biol. Chem.* 276, 35087–35092. doi: 10.1074/jbc.m104020200

Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. *Proteins Struct. Funct. Bioinform.* 33, 417–429. doi: 10.1002/(sici)1097-0134(19981115)33:3<417::aid-prot10<3.0.co;2-8

Hinsen, K., Reuter, N., Navaza, J., Stokes, D. L., and Lacapère, J.-J. (2005). Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. *Biophys. J.* 88, 818–827. doi: 10.1529/biophysj.104.050716

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Ibarra, A. A., Bartlett, G. J., Hegedüs, Z. F., Dutt, S., Hobor, F., Horner, K. A., et al. (2019). Predicting and experimentally validating hot-spot residues at protein–protein interfaces. *ACS Chem. Biol.* 14, 2252–2263. doi: 10.1021/acschembio.9b00560.s001

Ikeguchi, M., Ueno, J., Sato, M., and Kidera, A. (2005). Protein structural change upon ligand binding: linear response theory. *Phys. Rev. Lett.* 94:078102. doi: 10.1103/physrevlett.94.078102

Jacobs, D. J., Livesay, D. R., Mottonen, J. M., Vorov, O. K., Istomin, A. Y., and Verma, D. (2012). Ensemble properties of network rigidity reveal allosteric mechanisms. *Methods Mol. Biol.* 796, 279–304. doi: 10.1007/978-1-61779-334-9_15

Jalalypour, F., Sensoy, O., and Atilgan, C. (2020). Perturb-Scan-Pull: a novel method facilitating conformational transitions in proteins. *J. Chem. Theory Comput.* 16, 3825–3841. doi: 10.1021/acs.jctc.9b01222

Kenneth Morrow, J., and Zhang, S. (2012). Computational prediction of protein hot spot residues. *Curr. Pharm. Design* 18, 1255–1265. doi: 10.2174/138161212799436412

Keskin, O., Bahar, I., Flatow, D., Covell, D., and Jernigan, R. (2002). Molecular mechanisms of chaperonin GroEL- GroES function. *Biochemistry* 41, 491–501. doi: 10.1021/bi011393x

Kim, M. K., Jernigan, R. L., and Chirikjian, G. S. (2002). Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J.* 83, 1620–1630. doi: 10.1016/s0006-3495(02)73931-3

Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., and Flecker, P. (2000). Crystal structure of cancer chemopreventive Bowman-Birk inhibitor in ternary complex with bovine trypsin at 2.3 Å resolution. Structural basis of Janus-faced serine protease inhibitor specificity. *J. Mol. Biol.* 298, 477–491. doi: 10.2210/pdb1d6r/pdb

Kong, Y., Ming, D., Wu, Y., Stoops, J. K., Zhou, Z. H., and Ma, J. (2003). Conformational flexibility of pyruvate dehydrogenase complexes: a computational analysis by quantized elastic deformational model. *J. Mol. Biol.* 330, 129–135. doi: 10.1016/s0022-2836(03)00555-2

Krebs, W. G., Alexandrov, V., Wilson, C. A., Echols, N., Yu, H., and Gerstein, M. (2002). Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins Struct. Funct. Bioinform.* 48, 682–695. doi: 10.1002/prot.10168

Kumar, S., and Nussinov, R. (2002). Close-range electrostatic interactions in proteins. *Chembiochem* 3, 604–617.

Kurkcuoglu, O., Turgut, O. T., Cansu, S., Jernigan, R. L., and Doruker, P. (2009). Focused functional dynamics of supramolecules by use of a mixed-resolution elastic network model. *Biophys. J.* 97, 1178–1187. doi: 10.1016/j.bpj.2009.06.009

Leo-Macias, A., Lopez-Romero, P., Lupyan, D., Zerbino, D., and Ortiz, A. R. (2005). An analysis of core deformations in protein superfamilies. *Biophys. J.* 88, 1291–1299. doi: 10.1529/biophysj.104.052449

Li, X., Keskin, O., Ma, B., Nussinov, R., and Liang, J. (2004). Protein–protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. *J. Mol. Biol.* 344, 781–795. doi: 10.1016/j.jmb.2004.09.051

Lichtarge, O., and Sowa, M. E. (2002). Evolutionary predictions of binding surfaces and interactions. *Curr. Opin. Struct. Biol.* 12, 21–27. doi: 10.1016/s0959-440x(02)00284-1

Liu, Q., Chen, P., Wang, B., Zhang, J., and Li, J. (2018). Hot spot prediction in protein-protein interactions by an ensemble system. *BMC Syst. Biol.* 12:132. doi: 10.1186/s12918-018-0665-8

Liu, S., Liu, C., and Deng, L. (2018). Machine learning approaches for protein–protein interaction hot spot prediction: progress and comparative assessment. *Molecules* 23:2535. doi: 10.3390/molecules23102535

Ming, D., and Wall, M. E. (2005). Quantifying allosteric effects in proteins. *Proteins Struct. Funct. Bioinform.* 59, 697–707. doi: 10.1002/prot.20440

Miyashita, O., Onuchic, J. N., and Wolynes, P. G. (2003). Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12570–12575. doi: 10.3410/f.1016237.200300

Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2007). Hot spots—A review of the protein–protein interface determinant amino-acid residues. *Proteins Struct. Funct. Bioinform.* 68, 803–812. doi: 10.1002/prot.21396

Mottonen, J. M., Jacobs, D. J., and Livesay, D. R. (2010). Allosteric response is both conserved and variable across three CheY orthologs. *Biophys. J.* 99, 2245–2254. doi: 10.1016/j.bpj.2010.07.043

Ni, D., Liu, N., and Sheng, C. (2019). Allosteric modulators of protein–protein interactions (PPIs). *Protein Allostery Drug Discov.* 1163, 313–334. doi: 10.1007/978-981-13-8719-7_13

Orellana, L., Gustavsson, J., Bergh, C., Yoluk, O., and Lindahl, E. (2019). eBDIMS server: protein transition pathways with ensemble analysis in 2D-motion spaces. *Bioinformatics* 35, 3505–3507. doi: 10.1093/bioinformatics/btz104

Orozco, M. (2014). A theoretical view of protein dynamics. *Chem. Soc. Rev.* 43, 5051–5066. doi: 10.1002/chin.201437300

Ozdemir, E. S., Halakou, F., Nussinov, R., Gursoy, A., and Keskin, O. (2019). Methods for discovering and targeting druggable protein-protein interfaces and their application to repurposing. *Computat. Methods Drug Repurp.* 1903, 1–21. doi: 10.1007/978-1-4939-8955-3_1

Pautsch, A., Vogelsgesang, M., Tränkle, J., Herrmann, C., and Aktories, K. (2005). Crystal structure of the C3bot–RalA complex reveals a novel type of action of a bacterial exoenzyme. *EMBO J.* 24, 3670–3680. doi: 10.2210/pdb2a78/pdb

Pereira, P. J. B., Lozanov, V., Patthy, A., Huber, R., Bode, W., Pongor, S., et al. (1999). Specific inhibition of insect α-amylases: yellow meal worm α-amylase in complex with the Amaranth α-amylase inhibitor at 2.0 *Å resolution*. *Structure* 7, 1079–1088. doi: 10.1016/s0969-2126(99)80175-0

Qiao, Y., Xiong, Y., Gao, H., Zhu, X., and Chen, P. (2018). Protein-protein interface hot spots prediction based on a hybrid feature selection strategy. *BMC Bioinform.* 19:14. doi: 10.1186/s12859-018-2009-5

Ross, C., Nizami, B., Glenister, M., Sheik Amamuddy, O., Atilgan, A. R., Atilgan, C., et al. (2018). MODE-TASK: large-scale protein motion tools. *Bioinformatics* 34, 3759–3763. doi: 10.1093/bioinformatics/bty427

Ross, C. J., Atilgan, A. R., Bishop, ÖT., and Atilgan, C. (2018). Unraveling the motions behind enterovirus 71 uncoating. *Biophys. J.* 114, 822–838. doi: 10.1016/j.bpj.2017.12.021

Sensoy, O., Atilgan, A. R., and Atilgan, C. (2017). FbpA iron storage and release are governed by periplasmic microenvironments. *Phys. Chem. Chem. Phys.* 19, 6064–6075. doi: 10.1039/c6cp06961d

Shi, T.-L., Li, Y.-X., Cai, Y.-D., and Chou, K.-C. (2005). Computational methods for protein-protein interaction and their application. *Curr. Protein Peptide Sci.* 6, 443–449. doi: 10.5772/36716

Silversmith, R. E., Appleby, J. L., and Bourret, R. B. (1997). Catalytic mechanism of phosphorylation and dephosphorylation of CheY: kinetic characterization of imidazole phosphates as phosphodonors and the role of acid catalysis. *Biochemistry* 36, 14965–14974. doi: 10.1021/bi9715573

Siragusa, L., Cross, S., Baroni, M., Goracci, L., and Cruciani, G. (2015). BioGPS: navigating biological space to predict polypharmacology, off-targeting, and selectivity. *Proteins Struct. Funct. Bioinform.* 83, 517–532. doi: 10.1002/prot.24753

Tama, F., and Sanejouand, Y.-H. (2001). Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14, 1–6. doi: 10.1093/protein/14.1.1

Tan, K. P., Nguyen, T. B., Patel, S., Varadarajan, R., and Madhusudhan, M. S. (2013). Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic Acids Res.* 41, W314–W321. doi: 10.1093/nar/gkt503

Thomas, P. G., Russell, A. J., and Fersht, A. R. (1985). Tailoring the p H dependence of enzyme catalysis using protein engineering. *Nature* 318, 375–376. doi: 10.1038/318375a0

Thomas, S. A., Immormino, R. M., Bourret, R. B., and Silversmith, R. E. (2013). Nonconserved active site residues modulate CheY autophosphorylation kinetics and phosphodonor preference. *Biochemistry* 52, 2262–2273. doi: 10.1021/bi301654m

Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905. doi: 10.1103/physrevlett.77.1905

Tozzini, V. (2010). Multiscale modeling of proteins. *Acc. Chem. Res.* 43, 220–230. doi: 10.1021/ar9001476

Warshel, A. (1998). Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.* 273, 27035–27038. doi: 10.1074/jbc.273.42.27035

Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H., and Olsson, M. H. (2006). Electrostatic basis for enzyme catalysis. *Chem. Rev.* 106, 3210–3235. doi: 10.1007/978-94-009-5574-5_2

Whitford, P. C., Miyashita, O., Levy, Y., and Onuchic, J. N. (2007). Conformational transitions of adenylate kinase: switching by cracking. *J. Mol. Biol.* 366, 1661–1671. doi: 10.1016/j.jmb.2006.11.085

Xie, L., Li, J., Xie, L., and Bourne, P. E. (2009). Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comput. Biol.* 5:387. doi: 10.1371/journal.pcbi.1000387

Yilmaz, L. S., and Atilgan, A. R. (2000). Identifying the adaptive mechanism in globular proteins: fluctuations in densely packed regions manipulate flexible parts. *J. Chem. Phys.* 113, 4454–4464. doi: 10.1063/1.1288512

Zheng, W., Brooks, B. R., and Hummer, G. (2007). Protein conformational transitions explored by mixed elastic network models. *Proteins Struct. Funct. Bioinform.* 69, 43–57. doi: 10.1002/prot.21465

Zheng, W., and Doniach, S. (2003). A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13253–13258. doi: 10.1073/pnas.2235686100

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 15 September 2020
doi: 10.3389/fmolb.2020.574759

Investigating the Role of the N-Terminal Loop of PD-1 in Binding Process Between PD-1 and Nivolumab via Molecular Dynamics Simulation

Wenping Liu[1], Haoyu Jin[1], Ting Chen[1], Gangping Zhang[1], Shengsheng Lai[1] and Guangjian Liu[2]**

[1] Guangdong Food and Drug Vocational College, Guangzhou, China, [2] Clinical Data Center, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China

The blockade of immune checkpoints, such as programmed death receptor 1 (PD-1) and programmed death ligand 1 protein (PD-L1), is a promising therapeutic approach in cancer immunotherapy. Nivolumab, a humanized IgG4 antibody targeting PD-1, was approved by the US Food and Drug Administration for several cancers in 2014. Crystal structures of the nivolumab/PD-1 complex show that the epitope of PD-1 locates at the IgV domain (including the FG and BC loops) and the N-terminal loop. Although the N-terminal loop of PD-1 has been shown to play a dominant role in the complex interface of the static structure, its role in the dynamic binding process has not been illustrated clearly. Here, eight molecular systems were established for nivolumab/PD-1 complex, and long-time molecular dynamics simulations were performed for each. Results showed that the N-terminal loop of PD-1 prefers to bind with nivolumab to stabilize the interface between IgV and nivolumab. Furthermore, the binding of the N-terminal loop with nivolumab induces the rebinding between the IgV domain and nivolumab. Thus, we proposed a two-step binding model for the nivolumab/PD-1 binding, where the interface switches to a high-affinity state with the help of the N-terminal loop. This finding suggests that the N-terminal loop of PD-1 might be a potential target for anti-PD-1 antibody design, which could serve as an important gatekeeper for the anti-PD-1 antibody binding.

Keywords: PD-1, nivolumab, N-terminal loop, molecular dynamics simulation, two-step model

OPEN ACCESS

Edited by:
Ebru Demet Akten,
Kadir Has University, Turkey

Reviewed by:
Jianzhong Chen,
Shandong Jiaotong University, China
Carlo Camilloni,
University of Milan, Italy

****Correspondence:***
Guangjian Liu
liugj@gwcmc.org
Haoyu Jin
jinhy_gdyzy@163.com

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *21 June 2020*
Accepted: *14 August 2020*
Published: *15 September 2020*

Citation:
Liu W, Jin H, Chen T, Zhang G, Lai S and Liu G (2020) Investigating the Role of the N-Terminal Loop of PD-1 in Binding Process Between PD-1 and Nivolumab via Molecular Dynamics Simulation. Front. Mol. Biosci. 7:574759. doi: 10.3389/fmolb.2020.574759

INTRODUCTION

Cancer, the leading cause of death worldwide, constitutes a considerable burden to society (Bray et al., 2018; Miller et al., 2019). Cancer immunotherapy, that is, harnessing the immune system to battle tumors, has attained remarkable achievement in cancer treatment. Cancer immunotherapy comprises various treatment approaches, including antitumor monoclonal antibodies, cancer vaccines, and antibodies that block immune inhibitory pathways (Mellman et al., 2011; Couzin-Frankel, 2013). Among these treatments, blockade

of immune checkpoints is the most promising approach to activate therapeutic antitumor immunity (Ribas and Wolchok, 2018). The 2018 Nobel Prize in Physiology or Medicine was awarded to James P. Allison and Tasuku Honjo for their pioneering discoveries that led to the development of immune checkpoint inhibitors, which block the inhibitory action of T cell molecules, including programmed death receptor-1 (PD-1) and cytotoxic T lymphocyte-associated protein 4 (CTLA-4) (Pardoll, 2012; Guo, 2018; Zaidi and Jaffee, 2018).

PD-1 is an immune checkpoint receptor of the CD28 family expressed in tumor and immune cells (Keir et al., 2008). It is a 288 amino acid type I transmembrane receptor, and its ectodomain consists of four domains, namely, signal peptide, N-terminal loop, extracellular immunoglobulin variable (IgV) domain, and stalk region. The blockade of the interaction between PD-1 and its ligand programmed death ligand 1 protein (PD-L1) was observed to restore the attenuated immune response and lead to increased antitumor and antiviral activities (Hirano et al., 2005; Zitvogel and Kroemer, 2012; Sharma and Allison, 2015). Several PD-1/PD-L1 pathway inhibitors, including pembrolizumab, nivolumab targeting PD-1 and atezolizumab, durvalumab, and avelumab targeting PD-L1, have been approved by the US Food and Drug Administration (FDA) for the treatment of multiple cancers to date (Callahan et al., 2016; Ivashko and Kolesar, 2016; Leventakos and Mansfield, 2016; Bellmunt et al., 2017; Kim, 2017; Muller et al., 2017; Rittmeyer et al., 2017; Sidaway, 2017; Syed, 2017). Humanized IgG4 antibody nivolumab received the most attention and was approved for the treatment of melanoma, metastatic non-small-cell lung cancer, renal cell carcinoma, and Hodgkin lymphoma in 2014.

Two crystal structures of the nivolumab/PD-1complex have been reported in 2015 and 2016, respectively, providing interface information at the atomic level (**Figures 1A,E**; Lee et al., 2016; Tan et al., 2017). Three loops of PD-1 provide a flexible platform for nivolumab binding, including the N-terminal loop, the FG loop, and the BC loop. The FG and BC loops locate on the IgV domain. Previous studies showed that the FG loop (i.e., $_{PD-1}$PRO130 and $_{PD-1}$LYS131) is a binding site for PD-L1 as well as a "hot spot" for several immune checkpoint blockade monoclonal antibodies, such as GY-5 and GY-14 (Zak et al., 2015; Liu and Liu, 2017; Chen et al., 2019a). They clearly suggest a steric clash blockade mechanism of nivolumab. Unexpectedly, the N-terminal loop of PD-1 is far from the interface of PD-1/PD-L1 complex but contributes the majority of hydrogen bonds (H-bonds) to the binding of nivolumab and PD-1. Experiments further proved that the truncation of the N-terminal loop of PD-1 would abolish the nivolumab binding (Tan et al., 2017). Thus, the N-terminal loop of PD-1 plays a dominant role in the complexation of PD-1 and nivolumab. In spite of this, how the N-terminal loop regulates the dynamic binding process has not been answered clearly. Molecular recognition is a dynamic process, and the binding of a ligand to its receptor is regarded not as a single, frozen structure but rather a macromolecule in constant motion (Moroni et al., 2015). Considering the two-site mode (the N-terminal loop and the IgV domain) at the interface of the nivolumab/PD-1 complex, we assumed a two-step binding model, where the N-terminal loop will help switch the binding to a stronger state whether it comes across nivolumab firstly or not.

Molecular dynamics (MD) simulation is well suited for studying the dynamics of proteins (Liu et al., 2018). We have

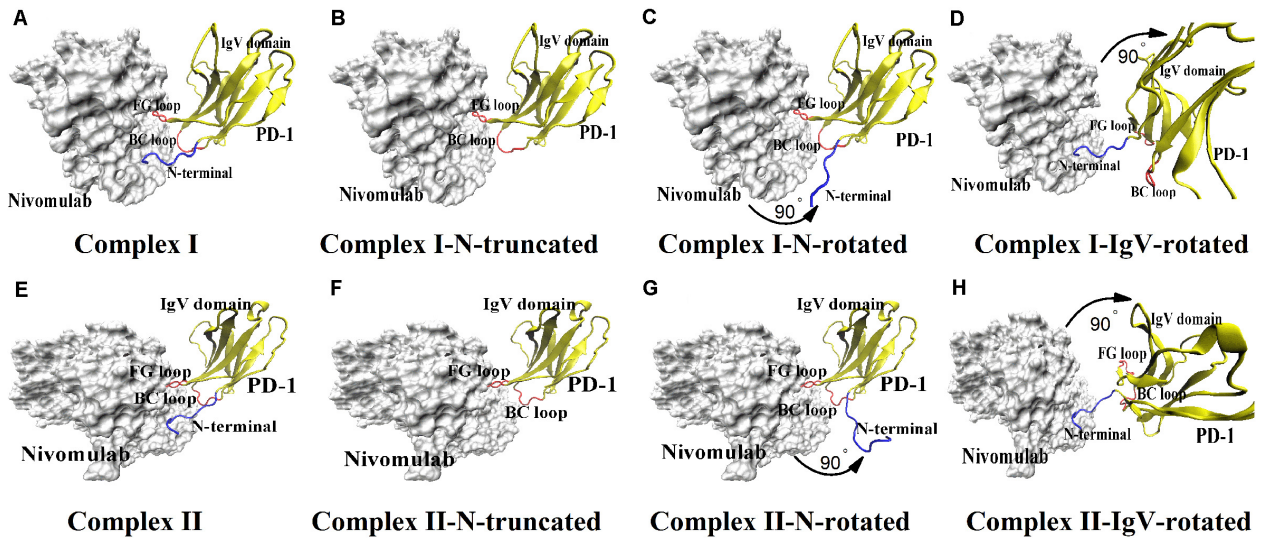


FIGURE 1 | Eight molecular systems of nivolumab/PD-1 complex. **(A)** Complex I. It is downloaded from the PDB with accession code of 5GGR. **(B)** Complex I-N-truncated. It is built through cutting off the N-terminal loop of PD-1 of Complex I. **(C)** Complex I-N-rotated. It is built through the rotation of the N-terminal loop of PD-1 of Complex I. **(D)** Complex I-IgV-rotated. It is built through the rotation of the IgV domain of PD-1 of Complex I. **(E)** Complex II. It is downloaded from the PDB with accession code of 5WT9. **(F)** Complex II-N-truncated. It is built through cutting off the N-terminal loop of PD-1 of Complex II. **(G)** Complex II-N-rotated. It is built through the rotation of the N-terminal loop of PD-1 of complex II. **(H)** Complex II-IgV-rotated. It is built through the rotation of the IgV domain of PD-1 of Complex II. The nivolumab is shown in silver (Surface mode). The N-terminal loop of PD-1 is shown in blue (New Cartoon mode). The FG loop and the BC loop of the IgV domain are shown in red, and the rest parts of the IgV domain are shown in yellow (New Cartoon mode).

used MD simulations to elucidate conformational selection and induced fit mechanisms in the binding of PD-1 and PD-L1 via MD simulations, and found that the CC' loop of PD-1 is flexible and switches from an open form to a close one to stabilize the PD-1/PD-L1 complex (Liu et al., 2017). MD simulation is also proven to be a useful tool to detect the "hot-spots" in the complex interface, such as the 6B4/GPIba complex, the 10B12/GPVI complex, the PD-1/PD-L1 complex, and the PD-1/pembrolizumab complex (Fang et al., 2012; Liu et al., 2016; Liu and Liu, 2017).

Therefore, MD simulations were adopted here to investigate the role of the N-terminal loop of PD-1 in the dynamic binding process between PD-1 and nivolumab. Two crystal structures of the nivolumab/PD-1 complex were used to build eight molecular systems with different binding states to mimic the scenarios with or without the N-terminal loop, and the N-terminal loop binds firstly or not (**Figure 1**). The results show that the N-terminal loop of PD-1 prefers to bind with nivolumab to stabilize the complex interface between the IgV domain (i.e., FG loop and BC loop) and nivolumab. The binding of the N-terminal loop with nivolumab also induces the rebinding between the IgV domain and nivolumab. These findings suggest a two-step binding model, in which the interface of nivolumab/PD-1complex switches to a stronger binding state with the help of the N-terminal loop of PD-1.

MATERIALS AND METHODS

System Setup

Eight nivolumab/PD-1 complex structures were set up for MD simulations (**Figure 1**). First, two crystal structures of the nivolumab/PD-1 complex with accession codes of 5GGR and 5WT9 were downloaded from the PDB, and were designated as Complex I and Complex II respectively (**Figures 1A,E**). Both structures include a long N-terminal loop at the complex interface, but in different lengths ($_{PD-1}$SER27-$_{PD-1}$ASN33 for complex I, $_{PD-1}$LEU25-$_{PD-1}$ASN33 for complex II). Actually, Complex II has an intact N-terminal loop and Complex I only lacks two residues, because the residues before $_{PD-1}$LEU25 belong to the signal peptide, which will be post-translationally removed and cannot be secreted. Second, the N-terminal loops of Complex I and II were cut off, and the remaining structures were designated as Complex I-N-truncated and Complex II-N-truncated (**Figures 1B,F**). Third, the N-terminal loop of Complex I and Complex II was rotated backward at the interface with 90° to dissociate from nivolumab, with the IgV domain of PD-1 and nivolumab fixed. These two structures were used to mimic the scenario where the IgV domain of PD-1 binds to nivolumab at the first step and designated as Complex I-N-rotated and Complex II-N-rotated (**Figures 1C,G**). Finally, the IgV domain of Complex I ($_{PD-1}$PRO34-$_{PD-1}$ARG143) and Complex II ($_{PD-1}$PRO34-$_{PD-1}$LEU142) was rotated backward at the interface with 90° to dissociate from nivolumab, with the N-terminal loop and nivolumab fixed. These two structures were used to mimic the scenario where the N-terminal loop of PD-1 binds to nivolumab at the first step and designated as Complex I-IgV-rotated and Complex II-IgV-rotated (**Figures 1D,H**).

The N- and C-terminal residues of each complex were acetylated and amidated, respectively, to mimic the continuation of the protein chains. The missing residues of PD-1 in each complex structure were modeled by the SWISS-MODEL server, with the free PD-1 structures with PDB accession codes 3RRQ and 2M2D as templates (Bordoli et al., 2008; Biasini et al., 2014). The protonation state of each protein residue at neutral pH was determined with the software PROPKA (Bas et al., 2008). Each complex was first solvated with TIP3P water molecules in a rectangular box with walls at least 15Å away from any protein atom. Then, Na^+ and Cl^- ions were added to neutralize the systems at a 150 mM salt concentration. The final system each consists of ∼35,500 water molecules, ∼100 Na^+ and ∼100 Cl^- ions (**Supplementary Table S1**).

MD Simulations

VMD 1.9.3 program was used for visualization, modeling, data analysis, and conformation presentation (Humphrey et al., 1996). NAMD 2.11 program with CHARMM36 all-atom force field was used for simulations (Phillips et al., 2005; Best et al., 2012). The cMAP correction for protein backbone, a time step of 2 fs, and a periodic boundary condition were applied in the simulations. The particle mesh Ewald method and a smooth cutoff of 12 Å were employed to calculate the full electrostatic and van der Waals forces.

First, each system was energy-minimized for 5,000 steps with all protein atoms fixed and for another 5,000 steps with all atoms free. Next, each system was heated gradually from 0 to 310 K in 1 ns. Then, equilibrium simulation of 100 ns was performed thrice (named Equ1, Equ2, and Equ3) for each system. A 310 K heat bath was manipulated using the Langevin thermostat, and a 1 atm pressure was controlled using the Langevin piston method during equilibriums.

Data Analysis

Transient complex formation usually relies on H-bonds (Kar et al., 2012), and they are the dominant linkers at the interface of nivolumab/PD-1 complex (Lee et al., 2016; Tan et al., 2017). Therefore, H-bonds across the interface in simulations were detected using VMD software with in-house scripts. An H-bond was defined if the donor–acceptor distance and bonding angle were smaller than 3.5 Å and 30°, respectively. The survival ratio of an H-bond was defined as the percentage of bond survival time.

The root mean square deviation (RMSD) of heavy atoms was used to illustrate the stability of the structures as well as the conformational changes of the N-terminal loop and IgV domain of PD-1. When analyzing the RMSD for the N-terminal loop and IgV domain of PD-1, the structures of nivolumab were aligned. Buried solvent-accessible surface area (SASA) of each complex, representing the interface area, was calculated using VMD software with handwritten scripts. The interaction energy, mainly including van der Waals and electrostatic energy, was calculated using NAMD Energy plugin (version 1.4) provided in VMD 1.9.3.

RESULTS

Interface Analysis of the Nivolumab/PD-1 Complex

To describe the dynamic picture of the interface of the nivolumab/PD-1 complex, two crystal structures of the complex with accession codes 5GGR and 5WT9 were downloaded from the PDB database, and designated as Complex I and Complex II (**Figures 1A,E**), respectively. They were solvated with TIP3P water molecules in a rectangular box. Equilibrium simulation of 100 ns was performed thrice (named Equ1, Equ2 and Equ3) for each complex after an energy minimization of 10,000 steps. The RMSD of heavy atoms showed that these two complexes had reached a local minimum after 20 ns (**Supplementary Figures S1, S2**).

Variations of buried SASA, interaction energy, and number of H-bonds were recorded for each complex to evaluate the stability of the complex interface (**Figures 2**, **3**), and their distributions are shown in **Figure 4**. The buried SASA of Complex I fluctuated around 1,600 Å^2, and its interaction energy and number of H-bonds fluctuated around −260 kcal/mol and 6, respectively (**Figures 2A–C**, **4A–C**). However, the interface of Complex II was larger and stronger, with buried SASA, interaction energy and number of H-bonds fluctuating around 1,800 Å^2, −320 kcal/mol and 6, respectively (**Figures 3A–C**, **4D–F**).

H-bonds with a maximum survival ratio of above 0.2 were sorted out to recognize the interaction residues because they are proposed as the dominant linkers across the interface (Lee et al., 2016; Tan et al., 2017; **Table 1**). In total 15 H-bonds were detected for Complex I, and 20 for Complex II. Interaction residues on PD-1 of these two complexes all located on three loops, including the N-terminal loop ($_{PD-1}$ASP26, $_{PD-1}$PRO28, $_{PD-1}$ASP29, and $_{PD-1}$ARG30), the BC loop ($_{PD-1}$THR59, $_{PD-1}$SER60, and $_{PD-1}$GLU61), and the FG loop ($_{PD-1}$ALA129, $_{PD-1}$PRO130, and $_{PD-1}$LYS131) (**Figures 2D–F**, **3D–F**). The FG loop formed five H-bonds in both the interfaces of Complex I and Complex II with the highest survival ratio of 0.68. The BC loop formed four and six H-bonds at the interface of Complex I and Complex II, respectively, with the highest survival ratio of 0.53. Although the numbers of H-bonds were similar to those of the FG loop, the interface formed by the BC loop was more vulnerable because it only appeared in two of three runs for both Complex I and Complex II. The average survival ratios of the H-bonds involved with the FG loop were higher than those with the BC loop, while the standard deviations with the FG loop were lower than those with the BC loop (**Table 1**). As the FG loop is a binding region for PD-L1, these results clearly suggest a stable steric clash blockade mechanism of nivolumab.

In the crystal structures, the N-terminal loop of PD-1 is not a binding site for PD-L1 but forms a major interface with nivolumab. In our simulations, it involved in six H-bonds at the interface of Complex I and contributed nine H-bonds to the interface of Complex II, with the highest survival ratio of 0.82. The standard deviations of the survival ratios of the H-bonds formed by the N-terminal loop were a little high. This might be due to the high flexibility of the long N-terminal loop, which consists of eight residues ($_{PD-1}$SER27 to $_{PD-1}$PRO34) in Complex I and 10 residues ($_{PD-1}$LEU25 to $_{PD-1}$PRO34) in Complex II, and can only be constrained

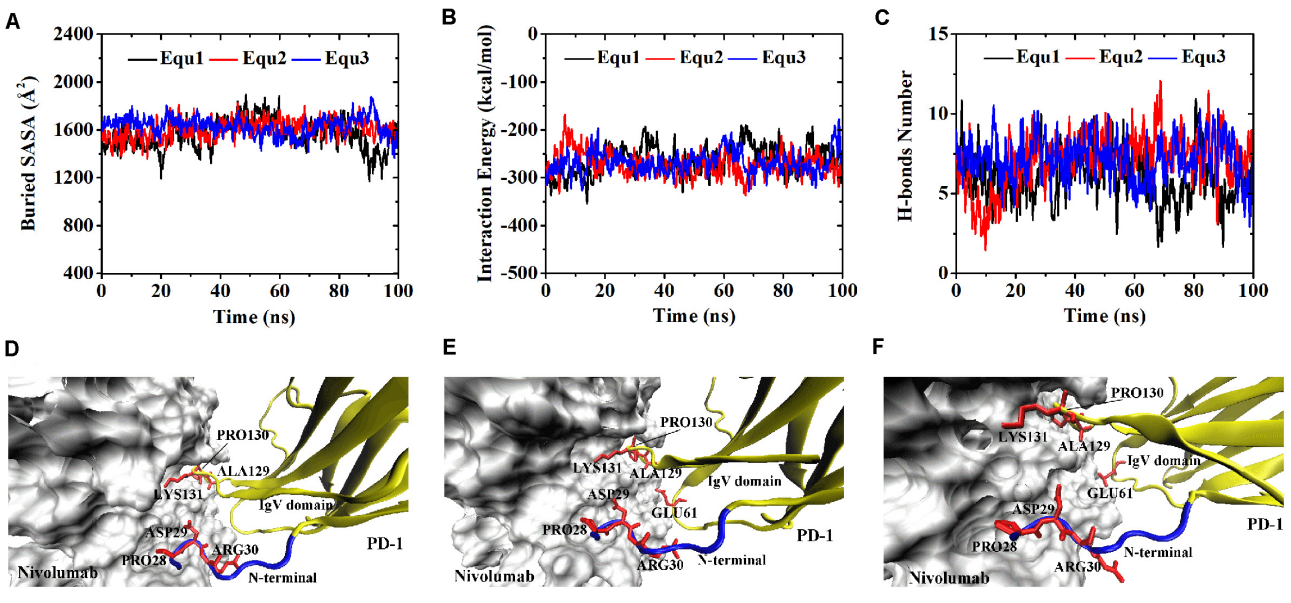


FIGURE 2 | Buried SASA **(A)**, interaction energy **(B)**, and number of H-bonds **(C)** of Complex I in three runs (Equ1, Equ2, and Equ3). **(D–F)** Show the last frame of Complex I in Equ1, Equ2, and Equ3, respectively. The nivolumab is shown in silver (Surface mode). The N-terminal loop of PD-1 is shown in blue and the IgV domain of PD-1 is shown in yellow (New Cartoon mode). The interaction residues of PD-1 are shown in red (Licorice mode).

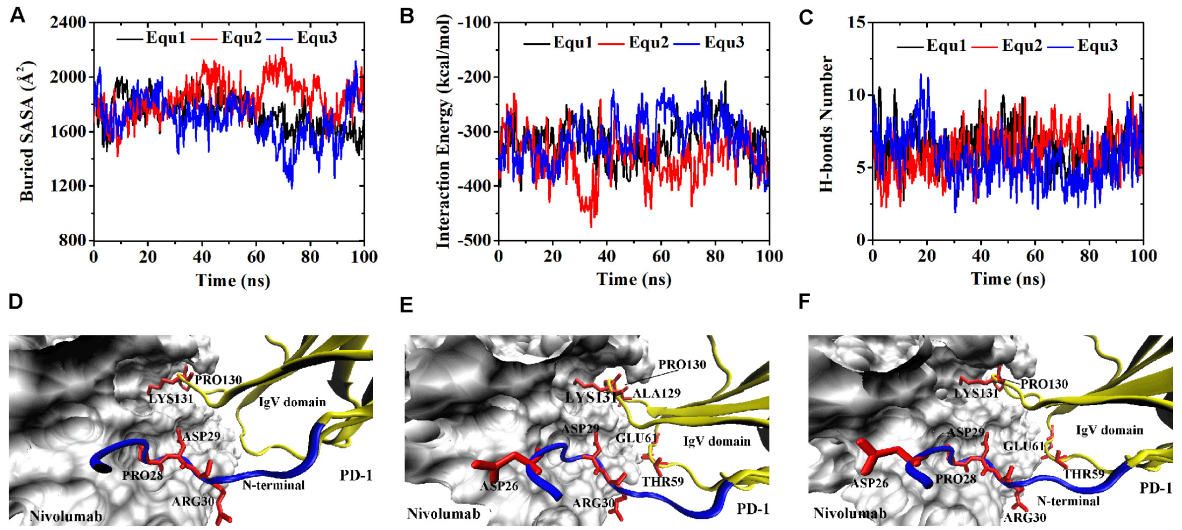


FIGURE 3 | Buried SASA **(A)**, interaction energy **(B)**, and number of H-bonds **(C)** of Complex II in three runs (Equ1, Equ2, and Equ3). **(D–F)** Show the last frame of Complex II in Equ1, Equ2, and Equ3, respectively. The nivolumab is shown in silver (Surface mode). The N-terminal loop of PD-1 is shown in blue and the IgV domain of PD-1 is shown in yellow (New Cartoon mode). The interaction residues of PD-1 are shown in red (Licorice mode).

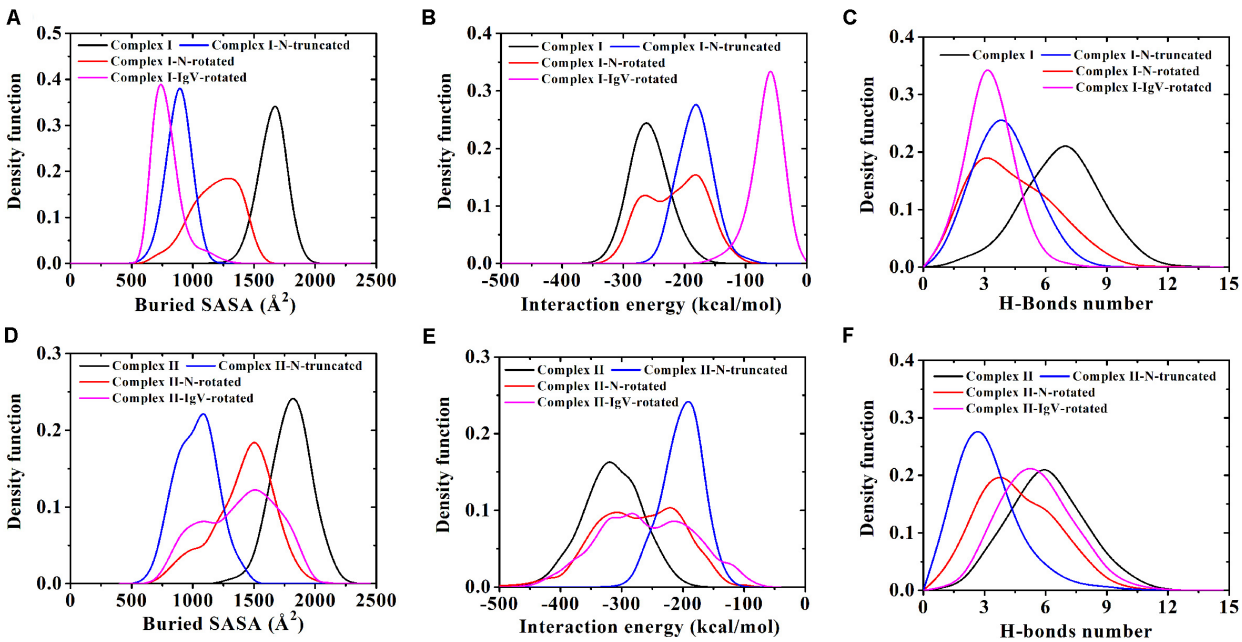


FIGURE 4 | (A–C) Demonstrate distributions of buried SASA, interaction energy, and number of H-bonds for Complex I, Complex I-N-truncated, Complex I-N-rotated, and Complex I-IgV-rotated. **(D–F)** Demonstrate distributions of buried SASA, interaction energy, and number of H-bonds for Complex II, Complex II-N-truncated, Complex II-N-rotated, and Complex II-IgV-rotated.

by one side. The$_{PD-1}$ASP26 located at the N-terminal loop formed two H-bonds with nivolumab in Complex II, but it was missing in Complex I. Thus, the N-terminal loop in Complex II formed a larger interface with nivolumab. These MD results showed that interactions between the N-terminus of PD-1 and nivolumab are definite and stable on the nanosecond time scale we simulated.

Truncation of the N-Terminal Loop of PD-1 Impairs the Interface Between PD-1 and Nivolumab

The N-terminal loop of PD-1 greatly contributes to the interface of the nivolumab/PD-1 complex, and mutagenesis study revealed that the cut-off of the N-terminal loop abolished the binding

TABLE 1 | Summary of survival ratio and involved residues of H-bonds in the interface of complex I and complex II.

Complex	Bond No.	PD-1			Nivolumab		Survival ratio[#]					
		Secondary structure	Residue	Atom	Residue*	Atom	Equ1	Equ2	Equ3	Max	Ave	Std
Complex I	1	N-terminal loop	ASP29	OD1	$GLY33_H$	N	0.41	0.72	0.80	0.80	0.76	0.06
	2		ARG30	N	$ASN31_H$	O	0.52	0.70	0.76	0.76	0.66	0.12
	3		ASP29	OD2	$ASN99_H$	ND2	0.32	0.61	0.66	0.66	0.53	0.18
	4		ASP29	OD1	$ASN99_H$	ND2	0.37	0.15	0.08	0.37	0.20	0.15
	5		ASP29	OD2	$GLY33_H$	N	0.36	0.02	0.00	0.36	0.13	0.20
	6		PRO28	O	$TYR53_H$	N	0.28	0.16	0.21	0.28	0.22	0.06
	7	FG loop	ALA129	O	$THR56_L$	OG1	0.30	0.57	0.64	0.64	0.50	0.18
	8		PRO130	O	$THR56_L$	N	0.50	0.62	0.56	0.62	0.56	0.06
	9		LYS131	NZ	$ASP101_H$	OD2	0.36	0.27	0.29	0.36	0.31	0.05
	10		LYS131	N	$ASP100_H$	O	0.35	0.28	0.32	0.35	0.32	0.04
	11		LYS131	NZ	$ASP101_H$	OD1	0.26	0.32	0.31	0.32	0.30	0.03
	12	BC loop	GLU61	OE2	$THR28_H$	OG1	0.01	0.43	0.39	0.43	0.28	0.23
	13		GLU61	OE2	$THR28_H$	N	0.00	0.41	0.31	0.41	0.24	0.21
	14		GLU61	OE1	$THR28_H$	OG1	0.01	0.32	0.31	0.32	0.21	0.18
	15		GLU61	OE1	$THR28_H$	N	0.00	0.24	0.28	0.28	0.17	0.15
Complex II	16	N-terminal loop	ARG30	N	$ASN31_H$	O	0.82	0.76	0.25	0.82	0.61	0.31
	17		ASP29	OD2	$GLY33_H$	N	0.59	0.00	0.11	0.59	0.23	0.31
	18		ASP29	OD1	$ASN99_H$	ND2	0.52	0.23	0.33	0.52	0.36	0.15
	19		ARG30	NH1	$ASN31_H$	OD1	0.45	0.03	0.00	0.45	0.16	0.25
	20		ASP29	OD2	$ASN99_H$	ND2	0.29	0.14	0.42	0.42	0.28	0.14
	21		ASP29	OD1	$GLY33_H$	N	0.36	0.01	0.28	0.36	0.22	0.18
	22		PRO28	O	$TYR53_H$	N	0.21	0.09	0.27	0.27	0.19	0.09
	23		ASP26	OD2	$LYS57_H$	NZ	0.09	0.10	0.24	0.24	0.14	0.08
	24		ASP26	OD1	$LYS57_H$	NZ	0.06	0.24	0.17	0.24	0.16	0.09
	25	FG loop	PRO130	O	$THR56_L$	N	0.28	0.68	0.60	0.68	0.52	0.21
	26		LYS131	NZ	$ASP101_H$	OD2	0.50	0.18	0.28	0.50	0.32	0.16
	27		LYS131	NZ	$ASP101_H$	OD1	0.35	0.38	0.42	0.42	0.38	0.04
	28		ALA129	O	$THR56_L$	OG1	0.08	0.29	0.06	0.29	0.14	0.13
	29		LYS131	NZ	$ASN99_H$	O	0.25	0.15	0.16	0.25	0.19	0.06
	30	BC loop	THR59	O	$THR28_H$	N	0.00	0.53	0.05	0.53	0.19	0.29
	31		GLU61	N	$GLY26_H$	O	0.00	0.42	0.04	0.42	0.15	0.23
	32		GLU61	N	$THR28_H$	OG1	0.01	0.00	0.32	0.32	0.11	0.18
	33		THR59	O	$ASN31_H$	ND2	0.05	0.01	0.27	0.27	0.11	0.14
	34		SER60	OG	$GLY26_H$	O	0.00	0.27	0.01	0.27	0.09	0.15
	35		THR59	OG1	$ASN31_H$	ND2	0.01	0.21	0.03	0.21	0.08	0.11

**The name of the residues with H or L indicating that the residues are on the heavy or the light chain of nivolumab. [#]Equ1, Equ2, and Equ3 donate three runs. Max represents the maximum value of three survival ratios of a bond. Ave represents the average value of three survival ratios of a bond. Std represents the standard deviation of three survival ratios of a bond.*

between PD-1 and nivolumab. How will the truncation of the N-terminal loop impair the interfaces? To answer this question, we cut off the N-terminal loops in Complex I and II, and designated them as Complex I-N-truncated and Complex II-N-truncated (**Figures 1B,F**). These two systems were simulated with the same scenario as before. The RMSD of heavy atoms showed that these two complexes reached a local minimum after 20 ns (**Supplementary Figures S3, S4**).

Variations of buried SASA, interaction energy, and number of H-bonds along the simulation time were analyzed, as shown in **Figures 5A–C**, **6A–C**, and their distributions are shown in **Figure 4**. The buried SASA of Complex I-N-truncated greatly decreased to around 900 $Å^2$, and its interaction energy and number of H-bonds dropped to -180 kcal/mol and 3, respectively. The buried SASA of Complex II-N-truncated greatly decreased to around 1,000 $Å^2$, and its interaction energy and number of H-bonds dropped to -180 kcal/mol and 3, respectively. These results indicate that the truncation of the N-terminal loop seriously impairs the binding between PD-1 and nivolumab (**Figures 5D–F**, **6D–F**). Moreover, the binding strength of Complex I-N-truncated is similar to that of Complex II-N-truncated, implying that Complex II is more stable than Complex I because it has a longer N-terminal loop ($_{PD-1}$LEU25 to $_{PD-1}$PRO34) than Complex I ($_{PD-1}$SER27 to $_{PD-1}$PRO34). This conclusion was further proved by the H-bonds with a survival ratio of above 0.2, where nine H-bonds

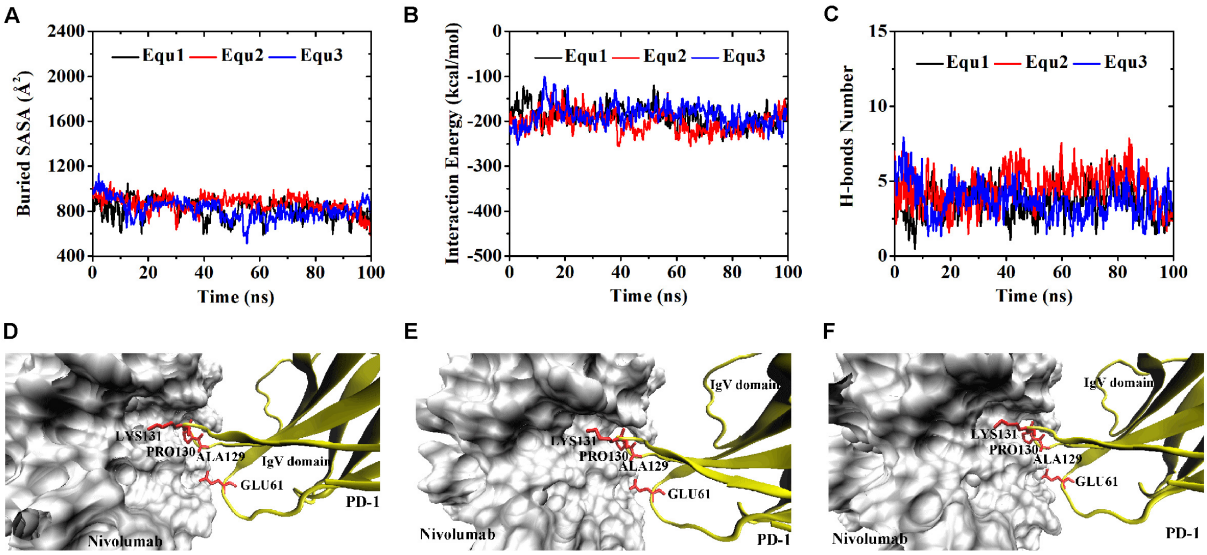


FIGURE 5 | Buried SASA **(A)**, interaction energy **(B)**, and number of H-bonds **(C)** of Complex I-N-truncated in three runs (Equ1, Equ2, and Equ3). **(D–F)** Show the last frame of Complex I-N-truncated in Equ1, Equ2, and Equ3, respectively. The IgV domain of PD-1 is shown in yellow (New Cartoon mode). The interaction residues of PD-1 are shown in red (Licorice mode).

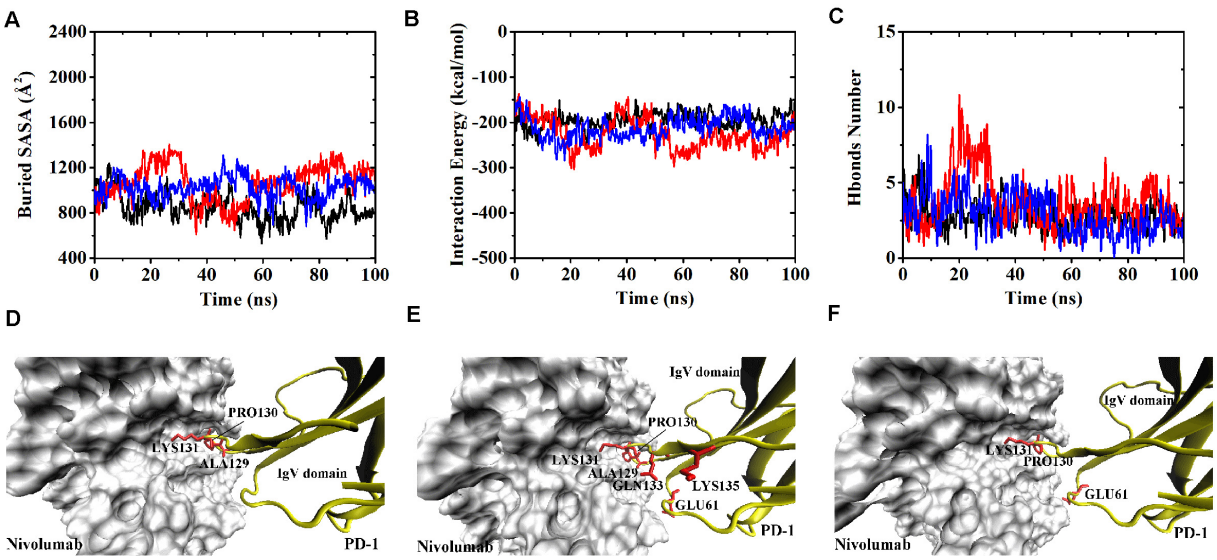


FIGURE 6 | Buried SASA **(A)**, interaction energy **(B)**, and number of H-bonds **(C)** of Complex II-N-truncated in three runs (Equ1, Equ2, and Equ3). **(D–F)** Show the last frame of Complex II-N-truncated in Equ1, Equ2, and Equ3, respectively. The IgV domain of PD-1 is shown in yellow (New Cartoon mode). The interaction residues of PD-1 are shown in red (Licorice mode).

for Complex I-N-truncated, and ten for Complex II-N-truncated were found with the interaction residues contributed by the FG and BC loops of PD-1, similar to those in Complex I and Complex II (**Table 2**).

Although the interface of the nivolumab/PD-1 was seriously impaired by cutting off the N-terminal loop of PD-1, the dissociation was not observed. A possible reason is that the energy barrier involved in the FG and BC loops of PD-1 was too high to overcome within our simulation time. However, analysis of accessibility of water molecules around the FG and BC loops revealed one major difference before and after deletion of the N-terminal loop. As shown in **Figures 7A,B**, the difference lies in extra water accessibility near the BC loop, where about nine and ten more water molecules entered within 4 Å of the BC loops of Complex I-N-truncated and Complex II-N-truncated. For the FG loop, the water accessibility showed no significant change. This result suggests the interface involved in the BC loop is protected from water attack by the N-terminal loop. Deleting

TABLE 2 | Summary of survival ratio and involved residues of H-bonds in the interface of complex I-N-truncated and complex II-N-truncated.

Complex	Bond No.	PD-1			Nivolumab		Survival ratio[#]					
		Secondary structure	Residue	Atom	Residue*	Atom	Equ1	Equ2	Equ3	Max	Ave	Std
Complex I-N-truncated	1	FG loop	PRO130	O	$THR56_L$	N	0.77	0.74	0.71	0.77	0.74	0.03
	2		ALA129	O	$THR56_L$	OG1	0.60	0.57	0.61	0.61	0.59	0.02
	3		LYS131	NZ	$ASP101_H$	OD1	0.34	0.42	0.39	0.42	0.38	0.04
	4		LYS131	NZ	$ASP101_H$	OD2	0.40	0.29	0.31	0.40	0.33	0.06
	5	BC loop	GLU61	OE1	$TYR102_H$	OH	0.21	0.38	0.42	0.42	0.34	0.11
	6		GLU61	OE2	$TYR102_H$	OH	0.30	0.25	0.38	0.38	0.31	0.07
	7		GLU61	OE1	$THR28_H$	OG1	0.15	0.29	0.01	0.29	0.15	0.14
	8		GLU61	OE2	$THR28_H$	OG1	0.08	0.26	0.14	0.26	0.16	0.09
	9		GLU61	N	$GLY26_H$	O	0.18	0.23	0.20	0.23	0.20	0.03
Complex II-N-truncated	10	FG loop	PRO130	O	$THR56_L$	N	0.75	0.47	0.71	0.75	0.64	0.15
	11		LYS131	NZ	$ASP101_H$	OD1	0.52	0.46	0.33	0.52	0.44	0.10
	12		ALA129	O	$THR56_L$	OG1	0.52	0.32	0.10	0.52	0.31	0.21
	13		LYS131	NZ	$ASP101_H$	OD2	0.29	0.27	0.33	0.33	0.30	0.03
	14		LYS131	N	$ASP100_H$	O	0.03	0.34	0.06	0.34	0.14	0.17
	15		LYS135	NZ	$ASP100_H$	OD1	0.00	0.32	0.00	0.32	0.11	0.18
	16		LYS131	NZ	$ASN99_H$	O	0.09	0.26	0.02	0.26	0.12	0.12
	17		GLN133	OE1	$TYR49_L$	OH	0.03	0.20	0.05	0.20	0.09	0.09
	18	BC loop	GLU61	N	$GLY26_H$	O	0.03	0.28	0.05	0.28	0.12	0.14
	19		GLU61	N	$THR28_H$	OG1	0.00	0.02	0.22	0.22	0.08	0.12

**The name of the residues with H or L indicating that the residues are on the heavy or the light chain of nivolumab. [#]Equ1, Equ2, and Equ3 donate three runs. Max represents the maximum value of three survival ratios of a bond. Ave represents the average value of three survival ratios of a bond. Std represents the standard deviation of three survival ratios of a bond.*

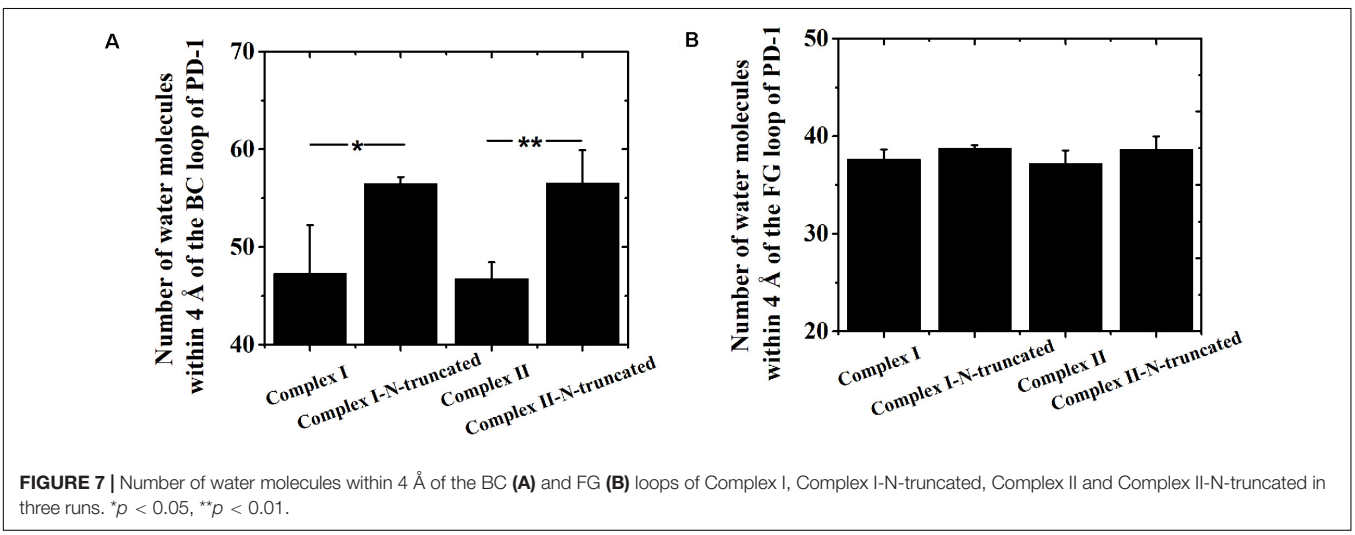


FIGURE 7 | Number of water molecules within 4 Å of the BC **(A)** and FG **(B)** loops of Complex I, Complex I-N-truncated, Complex II and Complex II-N-truncated in three runs. $^*p < 0.05$, $^{**}p < 0.01$.

the N-terminal loop might lead to a fast dissociation of the BC loop from nivolumab.

The N-Terminal Loop of PD-1 Prefers to Interact With Nivolumab to Stabilize the Complex Interface Further

Two binding regions were mapped on PD-1 for nivolumab, namely, the N-terminal loop and the IgV domain (including the FG loop and the BC loop), implying the possibility of a two-step binding process. Therefore, four additional complexes were built to verify this hypothesis. First, the N-terminal loop of PD-1 in Complex I and Complex II was rotated backward against the interface to dissociate from nivolumab to mimic the scenario where the IgV domain of PD-1 binds to nivolumab at the first step, designated as Complex I-N-rotated and Complex II-N-rotated, respectively (**Figures 1C,G**). Second, the IgV domain of PD-1 in Complex I and Complex II was rotated backward against the interface to dissociate from nivolumab to mimic the scenario where the N-terminal loop of PD-1 binds to nivolumab at the first step, designated as Complex I-IgV-rotated and Complex II-IgV-rotated, respectively (**Figures 1D,H**).

Similarly, Complex I-N-rotated and Complex II-N-rotated were simulated for 100 ns thrice after an energy minimization of 10,000 steps. The RMSD of heavy atoms showed that these two complexes reached a local minimum after 20 ns (**Supplementary Figures S5, S6**). Buried SASA, interaction energy, and number of H-bonds of Complex I-N-rotated and Complex II-N-rotated are shown in **Figures 8A–C**, **9A–C**, respectively, and their distributions are demonstrated in **Figure 4**. For Complex I-N-rotated, its buried SASA decreased to around 800 Å^2 during Equ1 and Equ2, and its interaction energy and number of H-bonds decreased to around -180 kcal/mol and 3, respectively, which was close to the binding strength of Complex I-N-truncated. However, its buried SASA increased to around 1,400 Å^2 during Equ3 with interaction energy and number of H-bonds fluctuating around −280 kcal/mol and 7, respectively, which was close to the binding strength of the initial Complex I.

Were these changes induced by the N-terminal loop or the IgV domain of PD-1? To answer this question, the number of intermolecular H-bonds formed by the N-terminal loop and the IgV domain of PD-1 was calculated (**Figures 8D,E**). H-bonds with a survival ratio of above 0.2 are listed in **Table 3**. The results clearly demonstrated that the increase of the complex interface was mainly caused by the N-terminal loop. The RMSD of the N-terminal loop in relative to its initial conformation in Complex I further confirmed that it returned back toward nivolumab in Equ3 (**Figure 8F**). The N-terminal loop bound to nivolumab with six H-bonds (Bond No. 1–6 in **Table 3**) formed by $_{PD-1}$ARG30 and $_{PD-1}$PRO28 after 10 ns in Equ3, but it kept free in Equ1 and Equ2 until the end of simulations (**Figures 8G–I**).

The buried SASA of Complex II-N-rotated increased to around 1500 Å^2 in all three runs (**Figures 9A–C**). Its interaction energy decreased to around −350 kcal/mol in Equ1 and Equ3, and to around −250 kcal/mol in Equ2. The number of H-bonds of Complex II-N-rotated increased to 4 in Equ1 and Equ2, but to 6 in Equ3. Next, the number of H-bonds formed by the N-terminal loop as well as its RMSD in relative to its

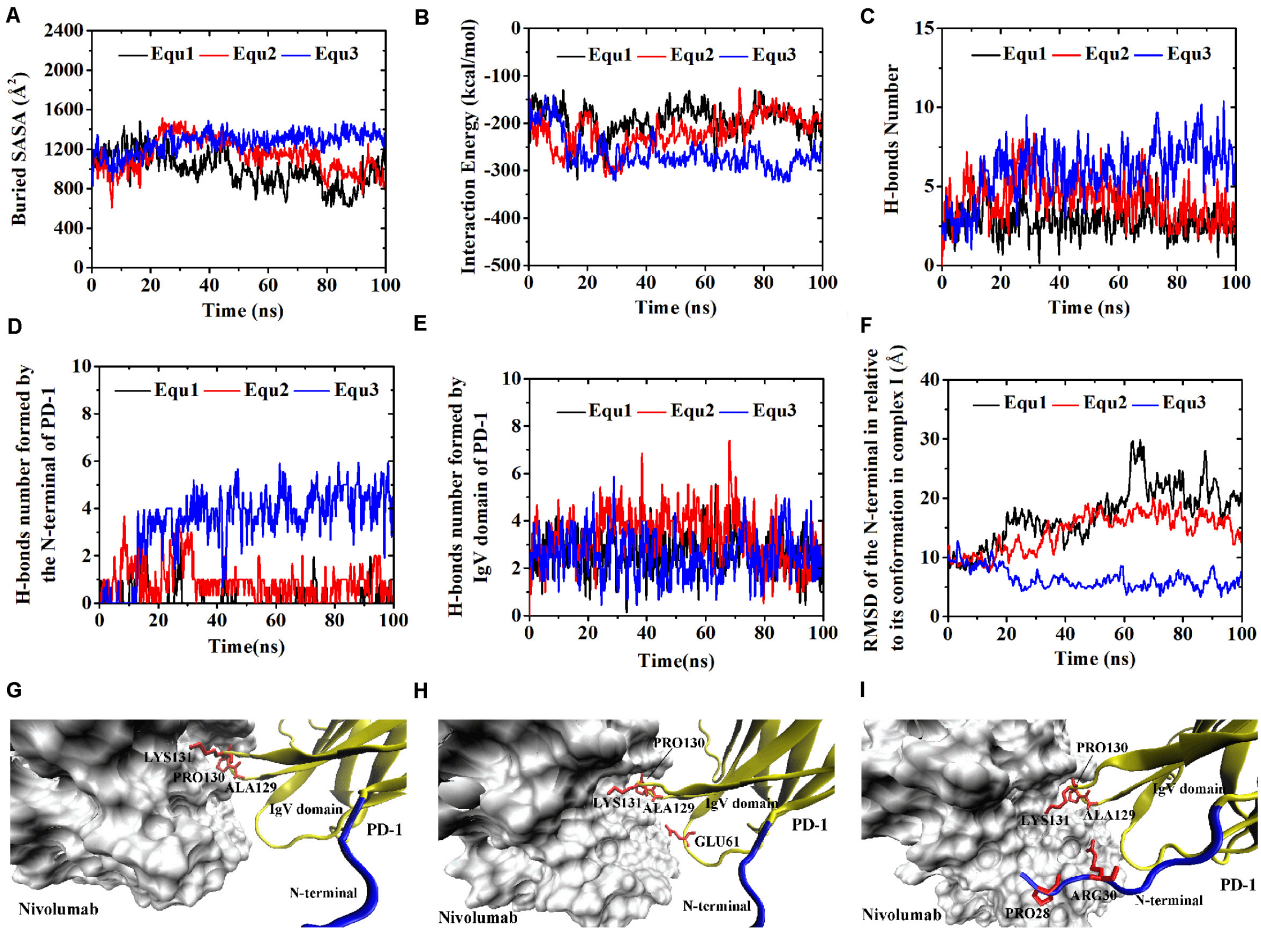


FIGURE 8 | Buried SASA **(A)**, interaction energy **(B)**, and number of H-bonds **(C)** of Complex I-N-rotated in three runs (Equ1, Equ2, and Equ3). **(D,E)** Show the number of H-bonds formed by the N-terminal loop and the IgV domain of PD-1, respectively. **(F)** shows the RMSD of the N-terminal loop of PD-1 of Complex I-N-rotated in relative to its initial conformation in Complex I in three runs. **(G–I)** Show the last frame of Complex I-N-rotated in Equ1, Equ2, and Equ3, respectively. The N-terminal loop of PD-1 is shown in blue and the IgV domain of PD-1 is shown in yellow (New Cartoon mode). The interaction residues of PD-1 are shown in red (Licorice mode).

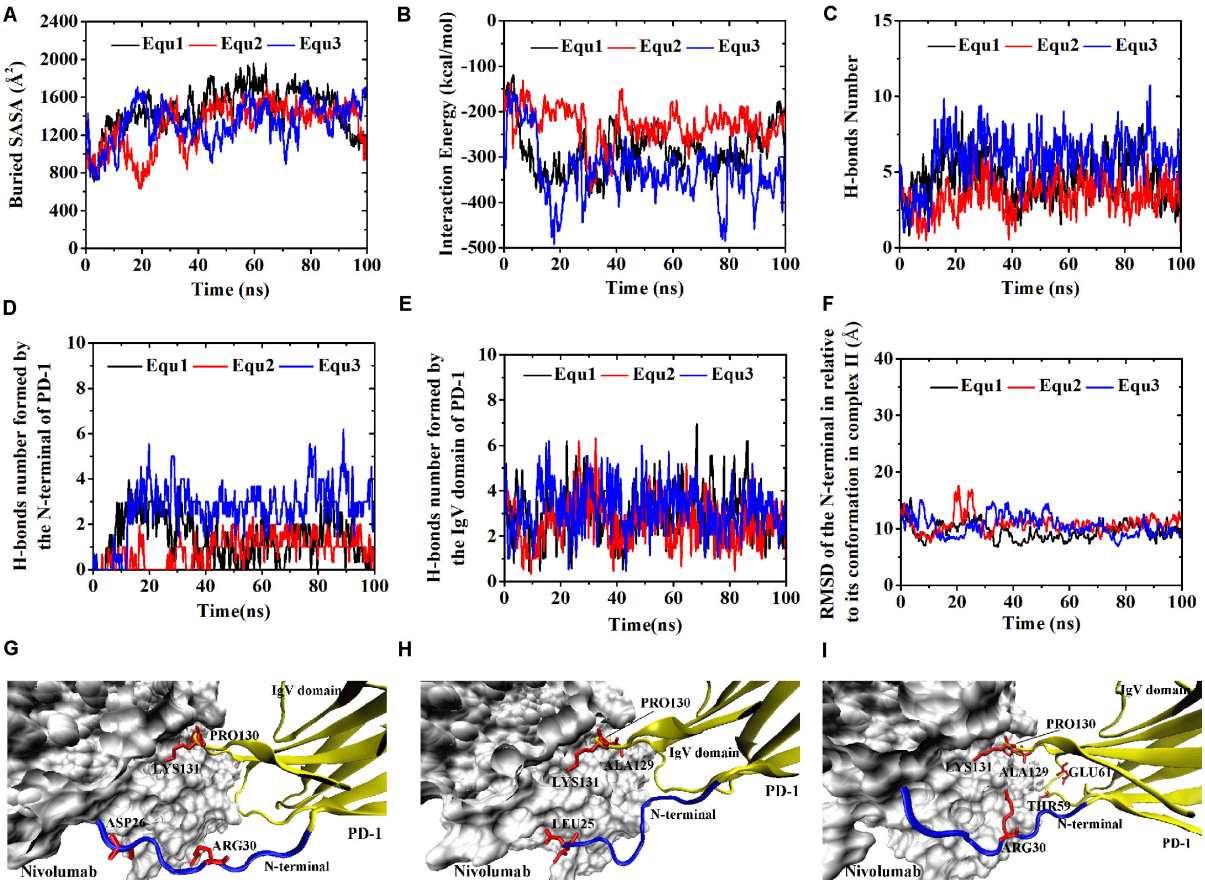


FIGURE 9 | Buried SASA **(A)**, interaction energy **(B)**, and number of H-bonds **(C)** of Complex II-N-rotated in three runs (Equ1, Equ2, and Equ3). **(D,E)** Show the number of H-bonds formed by the N-terminal loop and the IgV domain of PD-1, respectively. **(F)** Shows the RMSD of the N-terminal loop of PD-1 of Complex II-N-rotated in relative to its initial conformation in Complex II in three runs. **(G–I)** Show the last frame of Complex II-N-rotated in Equ1, Equ2, and Equ3, respectively. The N-terminal loop of PD-1 is shown in blue and the IgV domain of PD-1 is shown in yellow (New Cartoon mode). The interaction residues of PD-1 are shown in red (Licorice mode).

initial conformation in Complex II were calculated, as shown in **Figures 9D–F** and H-bonds with a survival ratio of above 0.2 are listed in **Table 3**. The results reveal that the N-terminal loop rebuilt the complex interface and formed stable H-bonds with nivolumab in all three runs after 20 ns, especially in Equ3. $_{PD-1}$ARG30 and $_{PD-1}$ASP26 on the N-terminal loop formed two H-bonds with ASN31$_H$ and LYS57$_H$ of nivolumab (Bond No. 16 and 22 in **Table 3**) in Equ1, whereas $_{PD-1}$LEU25 formed one bond with TYR53$_H$ in Equ2 (Bond No. 17 in **Table 3**). The interface between the N-terminal loop of PD-1 and the nivolumab in Equ3 was most stable, with five H-bonds formed by $_{PD-1}$ARG30 with ASN31$_H$ and ASP100$_H$ of nivolumab (Bond No. 15, 18–21 in **Table 3** and **Figures 9G–I**).

Overall, on the nanosecond time scale, the N-terminal loop of PD-1 prefers to interact with nivolumab to stabilize the complex interface further. Interfaces of Complex I-N-rotated and Complex II-N-rotated tend to be stronger with the help of the N-terminal loop of PD-1. The binding strength indexes showed bimodal distributions, especially for the interaction energy of Complex II-N-rotated (red lines in **Figure 4**).

Binding of the N-Terminal Loop With Nivolumab Could Induce the Rebinding of the IgV Domain With Nivolumab

For the IgV-rotated complexes, the RMSD of heavy atoms showed that Complex I-IgV-rotated fluctuated more violently than Complex II-IgV-rotated in three runs (**Supplementary Figures S7, S8**). Buried SASA, interaction energy, and number of H-bonds of Complex I-IgV-rotated and Complex II-IgV-rotated are shown in **Figures 10A–C**, **11A–C**, respectively, and their distributions are demonstrated in **Figure 4**. For all three runs of Complex I-IgV-rotated, the buried SASA fluctuated around 700 Å^2, the interaction energy fluctuated around −70 kcal/mol, and the number of H-bonds fluctuated around 3 (**Figures 10A–C**). The interface of complex I-IgV-rotated was mainly contributed by the N-terminal loop of the PD-1 within our simulation time (**Figures 10D,E**), with seven H-bonds formed by $_{PD-1}$PRO28, $_{PD-1}$ASP29, and $_{PD-1}$ARG30 with ASN31$_H$, GLY33$_H$, TYR53$_H$, and ASN99$_H$ of nivolumab (Bond No. 1–7 in **Table 4** and **Figures 10G–I**). The RMSD of the IgV domain in relative to its

TABLE 3 | Summary of survival ratio and involved residues of H-bonds in the interface of complex I-N-rotated and complex II-N-rotated

Complex	Bond No.	PD-1			Nivolumab		Survival ratio[#]					
		Secondary structure	Residue	Atom	Residue*	Atom	Equ1	Equ2	Equ3	Max	Ave	Std
Complex I-N-rotated	1	N-terminal loop	ARG30	NH1	$ASN31_H$	O	0.00	0.00	0.75	0.75	0.25	0.43
	2		PRO28	O	$TYR53_H$	OH	0.00	0.01	0.60	0.60	0.20	0.34
	3		ARG30	NH1	$ASP100_H$	OD1	0.02	0.03	0.54	0.54	0.20	0.30
	4		ARG30	NH2	$ASP100_H$	OD2	0.02	0.09	0.48	0.48	0.20	0.25
	5		ARG30	NH2	$ASP100_H$	OD1	0.00	0.05	0.24	0.24	0.10	0.13
	6		ARG30	NH1	$ASP100_H$	OD2	0.01	0.02	0.23	0.23	0.09	0.12
	7	FG loop	PRO130	O	$THR56_L$	N	0.65	0.71	0.65	0.71	0.67	0.03
	8		ALA129	O	$THR56_L$	OG1	0.41	0.53	0.43	0.53	0.46	0.06
	9		LYS131	NZ	$ASP101_H$	OD1	0.27	0.44	0.49	0.49	0.40	0.12
	10		LYS131	NZ	$ASP101_H$	OD2	0.48	0.21	0.25	0.48	0.31	0.15
	11		LYS131	NZ	$ASN99_H$	O	0.14	0.13	0.23	0.23	0.17	0.06
	12	BC loop	GLU61	N	$GLY26_H$	O	0.00	0.43	0.00	0.43	0.14	0.25
	13		GLU61	OE1	$TYR102_H$	OH	0.01	0.39	0.00	0.39	0.13	0.22
	14		GLU61	OE2	$TYR102_H$	OH	0.03	0.25	0.00	0.25	0.09	0.14
Complex II-N-rotated	15	N-terminal loop	ARG30	NH1	$ASN31_H$	O	0.00	0.00	0.62	0.62	0.21	0.36
	16		ARG30	NE	$ASN31_H$	OD1	0.52	0.00	0.00	0.52	0.17	0.30
	17		LEU25	O	$TYR53_H$	OH	0.00	0.52	0.00	0.52	0.17	0.30
	18		ARG30	NH1	$ASP100_H$	OD2	0.00	0.00	0.41	0.41	0.14	0.24
	19		ARG30	NH2	$ASP100_H$	OD1	0.00	0.00	0.40	0.40	0.13	0.23
	20		ARG30	NH1	$ASP100_H$	OD1	0.00	0.01	0.38	0.38	0.13	0.22
	21		ARG30	NH2	$ASP100_H$	OD2	0.00	0.00	0.38	0.38	0.13	0.22
	22		ASP26	OD1	$LYS57_H$	NZ	0.24	0.02	0.05	0.24	0.10	0.12
	23	FG loop	PRO130	O	$THR56_L$	N	0.62	0.78	0.81	0.81	0.74	0.10
	24		LYS131	NZ	$ASP101_H$	OD2	0.43	0.32	0.50	0.50	0.42	0.09
	25		ALA129	O	$THR56_L$	OG1	0.14	0.32	0.40	0.40	0.29	0.13
	26		LYS131	NZ	$ASP101_H$	OD1	0.19	0.20	0.32	0.32	0.24	0.07
	27		LYS131	N	$ASP100_H$	O	0.26	0.11	0.03	0.26	0.13	0.12
	28	BC loop	GLU61	N	$GLN1_H$	OY	0.00	0.00	0.27	0.27	0.09	0.16
	29		THR59	OG1	$GLY26_H$	O	0.00	0.00	0.22	0.22	0.07	0.13

**The name of the residues with H or L indicating that the residues are on the heavy or the light chain of nivolumab. [#]Equ1, Equ2, and Equ3 donate three runs. Max represents the maximum value of three survival ratios of a bond. Ave represents the average value of three survival ratios of a bond. Std represents the standard deviation of three survival ratios of a bond.*

initial conformation in Complex I showed that the IgV domain of PD-1 remained free in all three runs, which caused great fluctuations (**Figure 10F**).

By contrast, for all three runs of Complex II-IgV-rotated, the buried SASA increased to nearly 1,600 $Å^2$, the interaction energy decreased to nearly −300 kcal/mol, and the number of H-bonds increased to around 6. The number of H-bonds formed by the N-terminal loop and the IgV domain of PD-1, as well as the RMSD of the IgV domain in relative to its initial conformation in Complex II are shown in **Figures 11D–F**. It can be seen that the N-terminal loop of PD-1 maintained a stable interface with nivolumab on the nanosecond time scale, and changes were mainly caused by the IgV domain, which got close to and formed stable interface with nivolumab in all three runs. The FG loop ($_{PD-1}$LYS131) and BC loop ($_{PD-1}$THR59) of PD-1 interacted firmly with nivolumab ($ASN31_H$, $ASN99_H$, and $ASP101_H$) by forming four bonds in Equ1 (Bond No. 17–19, 26 in **Table 4** and **Figure 11G**). $_{PD-1}$LYS131 of FG loop and $_{PD-1}$GLU61 of BC loop formed five H-bonds with $ASP50_L$ and $THR28_H$ of nivolumab in Equ2 (Bond No. 21–25 in **Table 4** and **Figure 11H**). The FG loop (i.e., $_{PD-1}$LYS131 and $_{PD-1}$PRO130) formed two H-bonds with $ASP101_H$ and $THR56_L$ of nivolumab in Equ3 (Bond No. 19–20, **Figure 11I**). Thus, in our simulations, binding of the N-terminal loop with nivolumab could further induce the interaction of the IgV domain with nivolumab, which would switch the interface of nivolumab/PD-1 complex to a stronger binding state (purple lines in **Figures 4D–F**).

DISCUSSION

The N-terminal loop of PD-1 has not attracted much attention in recent years because it is not a binding region for PD-L1. However, it was proven critical for the binding between PD-1 and nivolumab, which is a humanized IgG4 antibody approved by the FDA for several cancers. Crystal structures of nivolumab/PD-1 complex showed that the interaction residues of PD-1 locate on the N-terminal loop and IgV domain. The

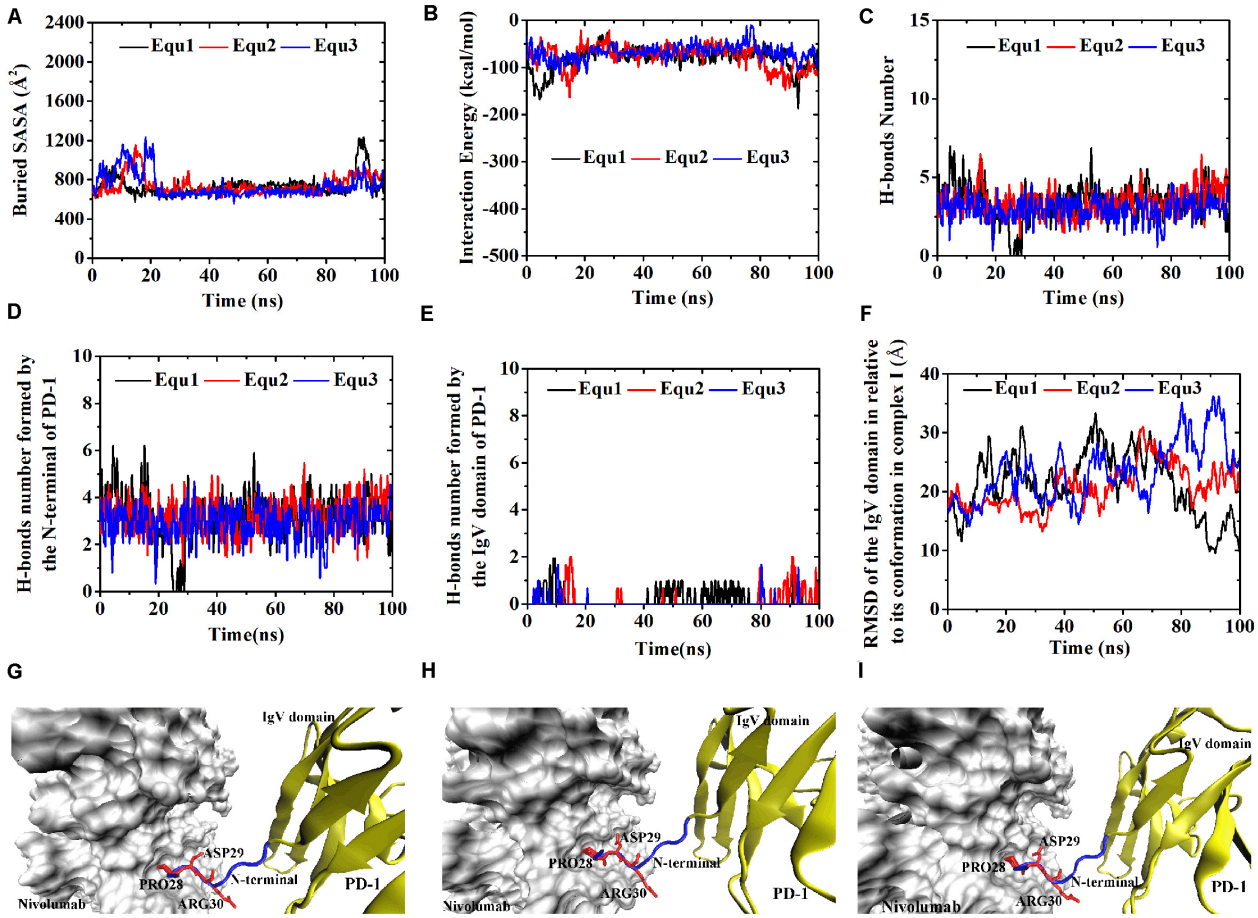


FIGURE 10 | Buried SASA **(A)**, interaction energy **(B)**, and number of H-bonds **(C)** of Complex I-IgV-rotated in three runs (Equ1, Equ2, and Equ3). **(D,E)** Show the number of H-bonds formed by the N-terminal loop and the IgV domain of PD-1, respectively. **(F)** Shows the RMSD of the IgV domain of PD-1 of Complex I-IgV-rotated in relative to its initial conformation in Complex I in three runs. **(G–I)** Show the last frame of Complex I-IgV-rotated in Equ1, Equ2, and Equ3, respectively. The N-terminal loop of PD-1 is shown in blue and the IgV domain of PD-1 is shown in yellow (New Cartoon mode). The interaction residues of PD-1 are shown in red (Licorice mode).

N-terminal loop of PD-1 greatly contributes to the complex formation, and we believe that it should also play an important role in the dynamic molecular recognition process. As dynamics of terminal loops are hard to predict based on crystal structure alone, eight molecular systems of nivolumab/PD-1 complex with different binding states were established, and long-time MD simulations of three replicas were performed for each of them, with the total simulated time of 2.4 μs. Our results proposed a two-step binding mode, in which the N-terminal loop of PD-1 could switch the complex interface into a stronger binding state. When the IgV domain binds to nivolumab first, the N-terminal loop of PD-1 prefers to interact with nivolumab to stabilize the complex interface. When the N-terminal loop is occupied by nivolumab, it could further induce the binding between the IgV domain (i.e., the FG and BC loops) of PD-1 and nivolumab.

The present results provided a detailed picture on the dynamic properties of the N-terminal loop of PD-1 in molecular interactions. Although this is the first time to systematically study the function of the N-terminal loop of PD-1, the N-terminal loops of other proteins have been revealed to have similar regulatory functions. For example, the N-terminus of model protein thaumatin serves as a major binding for cisplatin fragments (Russo Krauss et al., 2016). The N-terminal loop residues of beta-amyloid plays a key role in its interactions with integrin receptor and cell surface (Venkatasubramaniam et al., 2014). The N-terminal loop region of A1 domain in von Willebrand factor could stabilize A1A2A3 complex and modulate platelet activation under shear stress (Ju et al., 2013). Surface-exposed loops, generally as the most flexible parts of a protein structure, are not mere connectors but also have the potential to interact with solvent, ligands, and other biomolecules (Papaleo et al., 2016).

Since loop regions are too flexible to be resolved by crystallography, our simulations pave the way for investigating the binding mechanism between PD-1 and nivolumab. With the proposed two-step binding mode, nivolumab might be at least twice as likely to bind PD-1 as other antibodies with only one

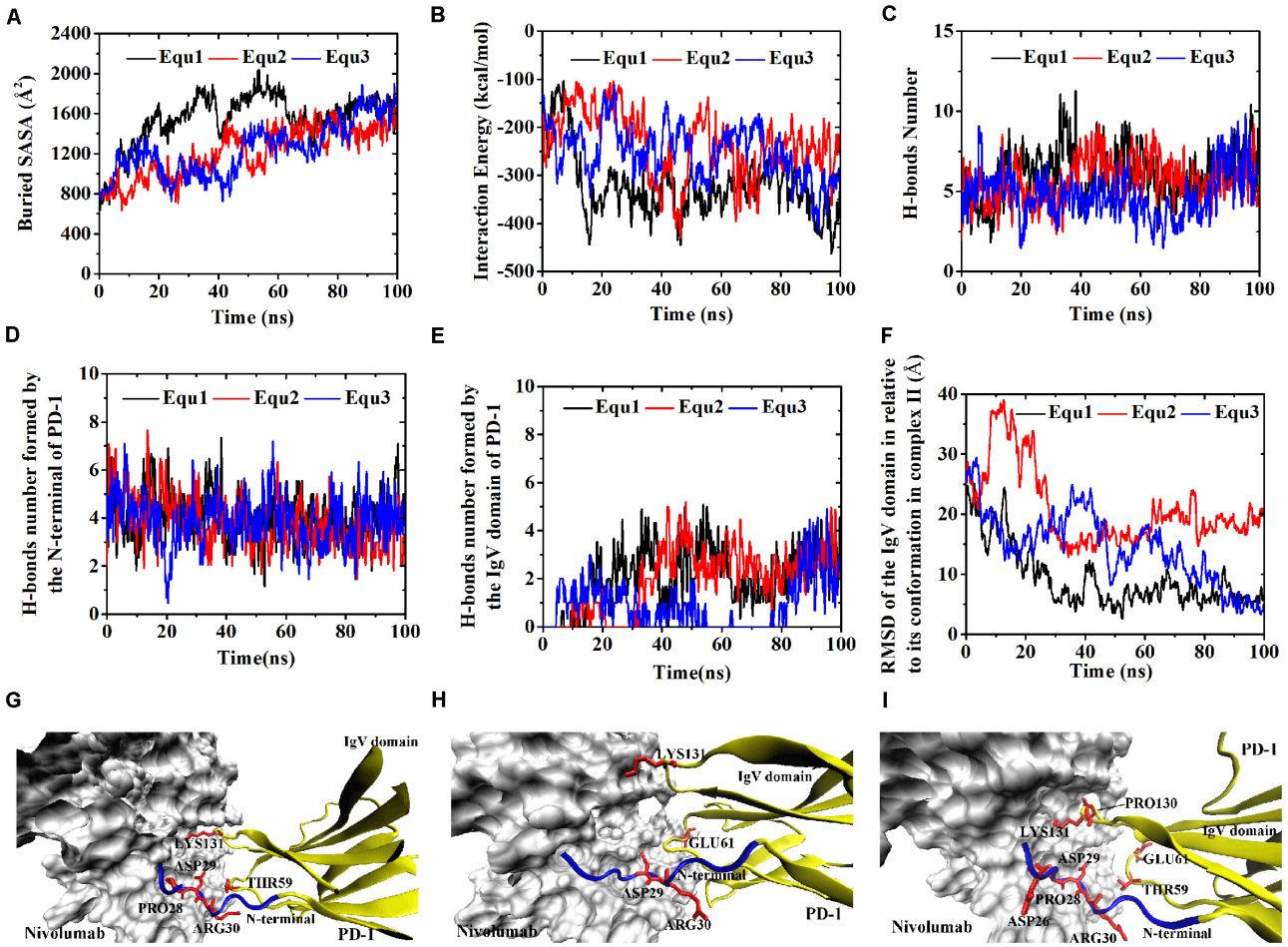


FIGURE 11 | Buried SASA **(A)**, interaction energy **(B)**, and number of H-bonds **(C)** of Complex II-IgV-rotated in three runs (Equ1, Equ2, and Equ3). **(D,E)** Show the number of H-bonds formed by the N-terminal loop and the IgV domain of PD-1, respectively. **(F)** Shows the RMSD of the IgV domain of PD-1 of Complex II-IgV-rotated in relative to its initial conformation in Complex II in three runs. **(G–I)** Show the last frame of Complex II-IgV-rotated in Equ1, Equ2, and Equ3, respectively. The N-terminal loop of PD-1 is shown in blue and the IgV domain of PD-1 is shown in yellow (New Cartoon mode). The interaction residues of PD-1 are shown in red (Licorice mode).

binding site. Furthermore, due to the high flexibility and mobility of the N-terminal loop, it can greatly facilitate the scanning efficiency and thus increase the probability of PD-1-nivolumab binding. This is of great importance for molecular interactions in the crowded intracellular environment.

Besides the binding process, our work also revealed the role of the N-terminal loop in augmenting the PD-1-nivolumab residence time, which is defined as the reciprocal of the dissociation rate constant. An abundance of experimental data suggests that high-affinity drug interactions with macromolecular targets generally rely on multistep binding and dissociation described by the two-step, induced-fit model (Copeland, 2016). Here, we show that the dissociation trajectory for the PD-1/nivolumab complex probably involves a retrograde induced-fit mechanism, that the N-terminal loop of PD-1 is able to rebind the dissociated nivolumab and IgV domain before nivolumab is completely released from PD-1 (**Figure 11**). Thus, the retrograde induced-fit mechanism creates multiple kinetic and structural barriers to nivolumab dissociation. This might partially explain why nivolumab has a nearly 10-fold higher affinity than that of pembrolizumab (Kd = 3.06 vs. 29 pM) (Fessas et al., 2017), another humanized anti-PD-1 monoclonal antibody approved by FDA, although the epitope of nivolumab is distinctly smaller than that of pembrolizumab (buried surface areas = 1,487 vs. 2,126 Å^2). This model suggests that the N-terminal loop of PD-1 may be viewed as a "kinetic gatekeeper" that guides the docking of nivolumab onto the IgV domain and prevents nivolumab from dissociating.

Apart from the role of pulling the IgV domain back to nivolumab, the N-terminal loop may also increase the nivolumab residence time by shielding the IgV-nivolumab H-bonds from water. In previously reported studies, the relationship between long residence time and accessibility of water has been established (Schmidtke et al., 2011; Magarkar et al., 2019). In this study, binding of the N-terminal loop with nivolumab can create an environment of lower dielectric

constant around the BC loop, shielded from bulk solvent (**Figure 7**). Such a solvent-shielded environment might result in the strengthening of non-covalent forces between the BC loop and nivolumab, such as hydrogen bonds, hydrophobic forces and van der Waals forces.

The function of the N-terminal loop of PD-1 for nivolumab requires its structure to be as complete as possible. The N-terminal loop of PD-1 in Complex I-N-rotated only bound with nivolumab in one run (Equ3 for complex I-N-rotated, **Figure 8I**), and it did not induce the rebinding of the IgV domain with nivolumab in the simulations (**Figure 10**). By contrast, the N-terminal loop of PD-1 interacted with nivolumab in all three runs of Complex II-N-rotated (**Figure 9**) and successfully induced the binding of the IgV domain with nivolumab in the simulations of Complex II-IgV-rotated (**Figure 11**). The reason is that the N-terminal loop of PD-1 contained 10 residues ($_{PD-1}$LEU25 to $_{PD-1}$PRO34) in Complex II, but it only eight residues ($_{PD-1}$SER27 to $_{PD-1}$PRO34) in Complex I. The missing residues in Complex I, such as $_{PD-1}$LEU25 and $_{PD-1}$ASP26, could also form stable H-bonds with nivolumab (Bond No. 23–24 in **Table 1**, Bond No. 17 and 22 in **Table 3**, and Bond No. 16 in **Table 4**). Therefore, the N-terminal loop of PD-1 in Complex II-N-rotated is more likely to be captured by nivolumab. Moreover, the interface between the N-terminal loop of PD-1 and nivolumab in Complex II-IgV-rotated (around 800 Å^2 and -140 kcal/mol) was stronger than that in Complex I-IgV-rotated (around 700 Å^2 and −70 kcal/mol), which is more beneficial in pulling the IgV domain back to nivolumab.

Despite the use of massive computational resources and highly precise models (full atomic representation and detailed force field), plain all-atom MD simulation is still insufficient for adequately exploring biomolecular structural dynamics. Multiple evidences indicate that long simulations cannot address how to catch the possible transition paths, which are still rare during μs-long MD due to inherent stochasticity and high-energy barriers. Nevertheless, we may predict some approximate behavior from simulations that suffer from sampling inefficiencies, in certain conditions e.g., upon introducing mutations or relaxation after removing ligands (Orellana, 2019). Here, we employed different starting configurations and multiple short simulations to enhances the configuration space sampling to better probe the conformational changes (Knapp et al., 2018; Chen et al., 2019b,c;

TABLE 4 | Summary of survival ratio and involved residues of H-bonds in the interface of complex I-IgV-rotated and complex II-IgV-rotated.

Complex	Bond No.	PD-1			Nivolumab		Survival ratio[#]					
		Secondary structure	Residue	Atom	Residue*	Atom	Equ1	Equ2	Equ3	Max	Ave	Std
Complex I-IgV-rotated	1	N-terminal loop	ASP29	OD1	$GLY33_H$	N	0.22	0.41	0.95	0.95	0.53	0.38
	2		ARG30	N	$ASN31_H$	O	0.84	0.82	0.71	0.84	0.79	0.07
	3		ASP29	OD2	$GLY33_H$	N	0.68	0.49	0.00	0.68	0.39	0.35
	4		ASP29	OD2	$ASN99_H$	ND2	0.23	0.44	0.64	0.64	0.44	0.21
	5		ASP29	OD1	$ASN99_H$	ND2	0.55	0.41	0.22	0.55	0.39	0.17
	6		PRO28	O	$TYR53_H$	N	0.25	0.29	0.29	0.29	0.28	0.02
	7		ARG30	NH1	$ASN31_H$	OD1	0.20	0.15	0.05	0.20	0.13	0.08
Complex II-IgV-rotated	8	N-terminal loop	ASP29	OD1	$GLY33_H$	N	0.95	0.06	0.24	0.95	0.42	0.47
	9		ASP29	OD2	$GLY33_H$	N	0.00	0.90	0.70	0.90	0.53	0.47
	10		ARG30	N	$ASN31_H$	O	0.84	0.82	0.80	0.84	0.82	0.02
	11		ASP29	OD1	$ASN99_H$	ND2	0.17	0.67	0.49	0.67	0.44	0.25
	12		ASP29	OD2	$ASN99_H$	ND2	0.63	0.17	0.31	0.63	0.37	0.24
	13		ARG30	NH1	$ASN31_H$	OD1	0.19	0.04	0.36	0.36	0.20	0.16
	14		PRO28	O	$TYR53_H$	N	0.26	0.19	0.29	0.29	0.25	0.05
	15		ARG30	NE	$ASN31_H$	OD1	0.28	0.14	0.06	0.28	0.16	0.11
	16		ASP26	OD1	$LYS57_H$	NZ	0.07	0.09	0.20	0.20	0.12	0.07
	17	FG loop	LYS131	NZ	$ASP101_H$	OD1	0.49	0.00	0.01	0.49	0.17	0.28
	18		LYS131	NZ	$ASN99_H$	O	0.28	0.00	0.09	0.28	0.12	0.14
	19		LYS131	NZ	$ASP101_H$	OD2	0.22	0.00	0.26	0.26	0.16	0.14
	20		PRO130	O	$THR56_L$	N	0.18	0.00	0.21	0.21	0.13	0.11
	21		LYS131	NZ	$ASP50_L$	OD1	0.00	0.20	0.01	0.20	0.07	0.11
	22	BC loop	GLU61	OE1	$THR28_H$	OG1	0.08	0.32	0.00	0.32	0.13	0.17
	23		GLU61	OE2	$THR28_H$	N	0.03	0.31	0.00	0.31	0.11	0.17
	24		GLU61	OE1	$THR28_H$	N	0.02	0.25	0.00	0.25	0.09	0.14
	25		GLU61	OE2	$THR28_H$	OG1	0.13	0.23	0.01	0.23	0.12	0.11
	26		THR59	O	$ASN31_H$	ND2	0.22	0.06	0.13	0.22	0.14	0.08

**The name of the residues with H or L indicating that the residues are on the heavy or the light chain of nivolumab. [#]Equ1, Equ2, and Equ3 donate three runs. Max represents the maximum value of three survival ratios of a bond. Ave represents the average value of three survival ratios of a bond. Std represents the standard deviation of three survival ratios of a bond.*

Chen et al., 2020). Fortunately, the conformational transitions of the nivolumab/PD-1 complex we concerned have been observed in a simulation time of 100 ns. For example, the N-terminal loop of PD-1 rotated back toward the interface and interacted with nivolumab in all three runs of Complex I-N-rotated (**Figure 9**). Moreover, the IgV domain of PD-1 also rotated back toward the interface and interacted with nivolumab with the help of the N-terminal loop of PD-1 in the simulations of Complex II-IgV-rotated (**Figure 11**). This work provides useful dynamics information on the role of the N-terminal loop in the molecular recognition process between PD-1 and nivolumab.

Previous research usually focused on the "hot-spot" of PD-1, such as the FG loop and the C'D loop. Our results suggest that the N-terminal loop of PD-1, which acts as an important gatekeeper for the binding of nivolumab and PD-1, should also be considered in the anti-PD-1 blockade antibody design. We are hopeful that the results presented in this study will ultimately provide a theoretical framework to understand the structural landscape of N-terminal loop of PD-1. In general, this opens a new opportunity for medicinal biologists or chemists to optimize affinity for antibodies, if such gatekeepers can be identified.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

GL and WL conceived and designed the experiments and wrote the manuscript. WL, HJ, and TC conducted modeling and simulation. WL, SL, and GZ analyzed the results. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Science and Technology Program of Guangzhou (Project No. 201904010207), Young Innovative Talents Program in Universities and Colleges of Guangdong Province (Project No. 2018GkQNCX002) and GDHVPS 2016, as well as Guangdong Province Medical Scientific Research Foundation (No. A2019142), Natural Science Foundation of Guangdong Province (Grant No. 2015A030310106) and Research Gant of Guangdong Food and Drug Vocational College (No. 2018ZR028).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.574759/full#supplementary-material

REFERENCES

Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008). Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins Struct. Funct. Bioinformatics* 73, 765–783. doi: 10.1002/prot.22102

Bellmunt, J., Powles, T., and Vogelzang, N. J. (2017). A review on the evolution of PD-1/PD-L1 immunotherapy for bladder cancer: the future is now. *Cancer Treat. Rev.* 54, 58–67. doi: 10.1016/j.ctrv.2017.01.007

Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., et al. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* 8, 3257–3273. doi: 10.1021/ct300400x

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, W252–W258. doi: 10.1093/nar/gku340

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2008). Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* 4, 1–13. doi: 10.1038/nprot.2008.197

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Callahan, M. K., Postow, M. A., and Wolchok, J. D. (2016). Targeting T Cell Co-receptors for Cancer therapy. *Immunity* 44, 1069–1078. doi: 10.1016/j.immuni.2016.04.023

Chen, D., Tan, S., Zhang, H., Wang, H., He, W., Shi, R., et al. (2019a). The FG Loop of PD-1 Serves as a "Hotspot" for therapeutic monoclonal antibodies in tumor immune checkpoint therapy. *iScience* 14, 113–124. doi: 10.1016/j.isci.2019.03.017

Chen, J., Wang, X., Pang, L., Zhang, J. Z. H., and Zhu, T. (2019b). Effect of mutations on binding of ligands to guanine riboswitch probed by free energy perturbation and molecular dynamics simulations. *Nucleic Acids Res.* 47, 6618–6631. doi: 10.1093/nar/gkz499

Chen, J., Wang, J., Yin, B., Pang, L., Wang, W., and Zhu, W. (2019c). Molecular mechanism of binding selectivity of inhibitors toward BACE1 and BACE2 Revealed by multiple short molecular dynamics simulations and free-energy predictions. *ACS Chem. Neurosci.* 10, 4303–4318. doi: 10.1021/acschemneuro.9b00348

Chen, J., Liu, X., Zhang, S., Sun, H., Zhang, L., and Zhang, Q. (2020). Molecular mechanism with regard to the binding selectivity of inhibitors toward FABP5 and FABP7 explored by multiple short molecular dynamics simulations and free energy analyses. *Phys. Chem. Chem. Phys.* 22, 2262–2275. doi: 10.1039/c9cp05704h

Copeland, R. A. (2016). The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.* 15, 87–95. doi: 10.1038/nrd.2015.18

Couzin-Frankel, J. (2013). Breakthrough of the year 2013. Cancer Immunotherapy. *Science* 342, 1432–1433. doi: 10.1126/science.342.6165.1432

Fang, X., Fang, Y., Liu, L., Liu, G., and Wu, J. (2012). Mapping paratope on antithrombotic antibody 6B4 to epitope on platelet glycoprotein Ibalpha via molecular dynamic simulations. *PLoS One* 7:e42263. doi: 10.1371/journal.pone.0042263

Fessas, P., Lee, H., Ikemizu, S., and Janowitz, T. (2017). A molecular and preclinical comparison of the PD-1-targeted T-cell checkpoint inhibitors nivolumab and pembrolizumab. *Semin. Oncol.* 44, 136–140. doi: 10.1053/j.seminoncol.2017.06.002

Guo, Z. S. (2018). The 2018 Nobel Prize in medicine goes to cancer immunotherapy (editorial for BMC cancer). *BMC Cancer* 18:1086. doi: 10.1186/s12885-018-5020-3

Hirano, F., Kaneko, K., Tamura, H., Dong, H., Wang, S., Ichikawa, M., et al. (2005). Blockade of B7-H1 and PD-1 by monoclonal antibodies potentiates cancer therapeutic immunity. *Cancer Res.* 65, 1089–1096.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38, 27–38. doi: 10.1016/0263-7855(96)00018-5

Ivashko, I. N., and Kolesar, J. M. (2016). Pembrolizumab and nivolumab: PD-1 inhibitors for advanced melanoma. *Am. J. Health Syst. Pharm.* 73, 193–201. doi: 10.2146/ajhp140768

Ju, L., Dong, J. F., Cruz, M. A., and Zhu, C. (2013). The N-terminal flanking region of the A1 domain regulates the force-dependent binding of von Willebrand factor to platelet glycoprotein Ibalpha. *J. Biol. Chem.* 288, 32289–32301. doi: 10.1074/jbc.M113.504001

Kar, G., Kuzu, G., Keskin, O., and Gursoy, A. (2012). Protein-protein interfaces integrated into interaction networks: implications on drug design. *Curr. Pharm. Des.* 18, 4697–4705. doi: 10.2174/138161212802651643

Keir, M. E., Butte, M. J., Freeman, G. J., and Sharpe, A. H. (2008). PD-1 and its ligands in tolerance and immunity. *Annu. Rev. Immunol.* 26, 677–704. doi: 10.1146/annurev.immunol.26.021607.090331

Kim, E. S. (2017). Avelumab: first global approval. *Drugs* 77, 929–937. doi: 10.1007/s40265-017-0749-6

Knapp, B., Ospina, L., and Deane, C. M. (2018). Avoiding false positive conclusions in molecular simulation: the importance of replicas. *J. Chem. Theory Comput.* 14, 6127–6138. doi: 10.1021/acs.jctc.8b00391

Lee, J. Y., Lee, H. T., Shin, W., Chae, J., Choi, J., Kim, S. H., et al. (2016). Structural basis of checkpoint blockade by monoclonal antibodies in cancer immunotherapy. *Nat. Commun.* 7:13354. doi: 10.1038/ncomms13354

Leventakos, K., and Mansfield, A. S. (2016). Advances in the treatment of non-small cell lung cancer: focus on nivolumab. Pembrolizumab, and Atezolizumab. *BioDrugs* 30, 397–405. doi: 10.1007/s40259-016-0187-0

Liu, W., Huang, B., Kuang, Y., and Liu, G. (2017). Molecular dynamics simulations elucidate conformational selection and induced fit mechanisms in the binding of PD-1 and PD-L1. *Mol. Biosyst.* 13, 892–900. doi: 10.1039/c7mb00036g

Liu, W., Liu, G., Zhou, H., Fang, X., Fang, Y., and Wu, J. (2016). Computer prediction of paratope on antithrombotic antibody 10B12 and epitope on platelet glycoprotein VI via molecular dynamics simulation. *Biomed. Eng. Online* 15(Suppl. 2):152. doi: 10.1186/s12938-016-0272-0

Liu, X., Shi, D., Zhou, S., Liu, H., and Yao, X. (2018). Molecular dynamics simulations and novel drug discovery. *Expert Opin. Drug Discov.* 13, 23–37. doi: 10.1080/17460441.2018.1403419

Liu, W., and Liu, G. (2017). "Mapping paratope and epitope residues of antibody pembrolizumab via molecular dynamics simulation," in *Bioinformatics Research and Applications. ISBRA 2017. Lecture Notes in Computer Science*, Vol. 10330, eds Z. Cai, O. Daescu, and M. Li (Berlin: Springer). doi: 10.1007/978-3-319-59575-7_11

Magarkar, A., Schnapp, G., Apel, A. K., Seeliger, D., and Tautermann, C. S. (2019). Enhancing drug residence time by shielding of intra-protein hydrogen bonds: a case study on CCR2 antagonists. *ACS Med. Chem. Lett.* 10, 324–328. doi: 10.1021/acsmedchemlett.8b00590

Mellman, I., Coukos, G., and Dranoff, G. (2011). Cancer immunotherapy comes of age. *Nature* 480, 480–489. doi: 10.1038/nature10673

Miller, K. D., Nogueira, L., Mariotto, A. B., Rowland, J. H., Yabroff, K. R., Alfano, C. M., et al. (2019). Cancer treatment and survivorship statistics, 2019. *CA Cancer J. Clin.* 69, 363–385. doi: 10.3322/caac.21565

Moroni, E., Paladino, A., and Colombo, G. (2015). The dynamics of drug discovery. *Curr. Top. Med. Chem.* 15, 2043–2055. doi: 10.2174/1568026615666150519102950

Muller, M., Schouten, R. D., De Gooijer, C. J., and Baas, P. (2017). Pembrolizumab for the treatment of non-small cell lung cancer. *Expert Rev. Anticancer Ther.* 17, 399–409. doi: 10.1080/14737140.2017.1311791

Orellana, L. (2019). Large-scale conformational changes and protein function: breaking the in silico barrier. *Front. Mol. Biosci.* 6:117. doi: 10.3389/fmolb.2019.00117

Papaleo, E., Saladino, G., Lambrughi, M., Lindorff-Larsen, K., Gervasio, F. L., and Nussinov, R. (2016). The role of protein loops and linkers in conformational dynamics and allostery. *Chem. Rev.* 116, 6391–6423. doi: 10.1021/acs.chemrev.5b00623

Pardoll, D. M. (2012). The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* 12, 252–264. doi: 10.1038/nrc3239

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289

Ribas, A., and Wolchok, J. D. (2018). Cancer immunotherapy using checkpoint blockade. *Science* 359, 1350–1355. doi: 10.1126/science.aar4060

Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., von Pawel, J., et al. (2017). Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* 389, 255–265. doi: 10.1016/S0140-6736(16)32517-X

Russo Krauss, I., Ferraro, G., and Merlino, A. (2016). Cisplatin-protein interactions: unexpected drug binding to N-Terminal amine and lysine side chains. *Inorg. Chem.* 55, 7814–7816. doi: 10.1021/acs.inorgchem.6b01234

Schmidtke, P., Luque, F. J., Murray, J. B., and Barril, X. (2011). Shielded hydrogen bonds as structural determinants of binding kinetics: application in drug design. *J. Am. Chem. Soc.* 133, 18903–18910. doi: 10.1021/ja207494u

Sharma, P., and Allison, J. P. (2015). Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell* 161, 205–214. doi: 10.1016/j.cell.2015.03.030

Sidaway, P. (2017). Urological cancer: atezolizumab: an alternative to cisplatin? *Nat. Rev. Clin. Oncol.* 14:139. doi: 10.1038/nrclinonc.2016.222

Syed, Y. Y. (2017). Durvalumab: first global approval. *Drugs* 77, 1369–1376. doi: 10.1007/s40265-017-0782-5

Tan, S., Zhang, H., Chai, Y., Song, H., Tong, Z., Wang, Q., et al. (2017). An unexpected N-terminal loop in PD-1 dominates binding by nivolumab. *Nat. Commun.* 8:14369. doi: 10.1038/ncomms14369

Venkatasubramaniam, A., Drude, A., and Good, T. (2014). Role of N-terminal residues in Abeta interactions with integrin receptor and cell surface. *Biochim. Biophys. Acta* 1838, 2568–2577. doi: 10.1016/j.bbamem.2014.06.011

Zaidi, N., and Jaffee, E. M. (2018). Immunotherapy transforms cancer treatment. *J. Clin. Invest.* 129, 46–47. doi: 10.1172/JCI126046

Zak, K. M., Kitel, R., Przetocka, S., Golik, P., Guzik, K., Musielak, B., et al. (2015). Structure of the complex of human programmed death 1, PD-1, and Its Ligand PD-L1. *Structure* 23, 2341–2348. doi: 10.1016/j.str.2015.09.010

Zitvogel, L., and Kroemer, G. (2012). Targeting PD-1/PD-L1 interactions for cancer immunotherapy. *Oncoimmunology* 1, 1223–1225. doi: 10.4161/onci.21335

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 25 September 2020
doi: 10.3389/fmolb.2020.586075

Exploring Allosteric Signaling in the Exit Tunnel of the Bacterial Ribosome by Molecular Dynamics Simulations and Residue Network Model

*Pelin Guzel[1,2], Hatice Zeynep Yildirim[3], Merve Yuce[1] and Ozge Kurkcuoglu[1]**

[1] Department of Chemical Engineering, Istanbul Technical University, Istanbul, Turkey, [2] Science and Advanced Technology Research and Application Center, Istanbul Medeniyet University, Istanbul, Turkey, [3] Polymer Research Center and Graduate Program in Computational Science and Engineering, Bogazici University, Istanbul, Turkey

The bacterial ribosomal tunnel is equipped with numerous sites highly sensitive to the course of the translation process. This study investigates allosteric pathways linking distant functional sites that collaboratively play a role either in translation regulation or recruitment of chaperones. We apply perturbation response scanning (PRS) analysis to 700 ns long and 500 ns long coarse-grained molecular dynamics simulations of *E. coli* and *T. thermophilus* large subunits, respectively, to reveal nucleotides/residues with the ability to transmit perturbations by dynamic rationale. We also use the residue network model with the *k*-shortest pathways method to calculate suboptimal pathways based on the contact topology of the ribosomal tunnel of *E. coli* crystal structure and 101 ClustENM generated conformers of *T. thermophilus* large subunit. In the upper part of the tunnel, results suggest that A2062 and A2451 can communicate in both directions for translation stalling, mostly through dynamically coupled C2063, C2064, and A2450. For a similar purpose, U2585 and U2586 are coupled with A2062, while they are also sensitive to uL4 and uL22 at the constriction region through two different pathways at the opposite sides of the tunnel wall. In addition, the constriction region communicates with the chaperone binding site on uL23 at the solvent side but through few nucleotides. Potential allosteric communication pathways between the lower part of the tunnel and chaperone binding site mostly use the flexible loop of uL23, while A1336–G1339 provide a suboptimal pathway. Both species seem to employ similar mechanisms in the long tunnel, where a non-conserved cavity at the bacterial uL23 and 23S rRNA interface is proposed as a novel drug target.

Keywords: bacterial ribosome, ribosomal tunnel, allostery, signal relay, trigger factor, translation arrest, antibiotics

OPEN ACCESS

Edited by:
Guang Hu,
Soochow University, China

Reviewed by:
Angelo Nicola Felline,
University of Modena and Reggio Emilia, Italy
Yong Wang,
University of Copenhagen, Denmark

***Correspondence:**
Ozge Kurkcuoglu
olevitas@itu.edu.tr

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *22 July 2020*
Accepted: *08 September 2020*
Published: *25 September 2020*

Citation:
Guzel P, Yildirim HZ, Yuce M and Kurkcuoglu O (2020) Exploring Allosteric Signaling in the Exit Tunnel of the Bacterial Ribosome by Molecular Dynamics Simulations and Residue Network Model. Front. Mol. Biosci. 7:586075. doi: 10.3389/fmolb.2020.586075

INTRODUCTION

Ribosomal complexes synthesize proteins according to the genetic information on mRNA across all kingdoms of life. The ribosome complex called as 70S in bacteria is formed by the association of two subunits, small subunit 30S, and large subunit 50S through numerous inter-subunit bridges (Liu and Fredrick, 2016). Each subunit is formed of ribosomal RNAs (16S, 5S, and 23S rRNA)

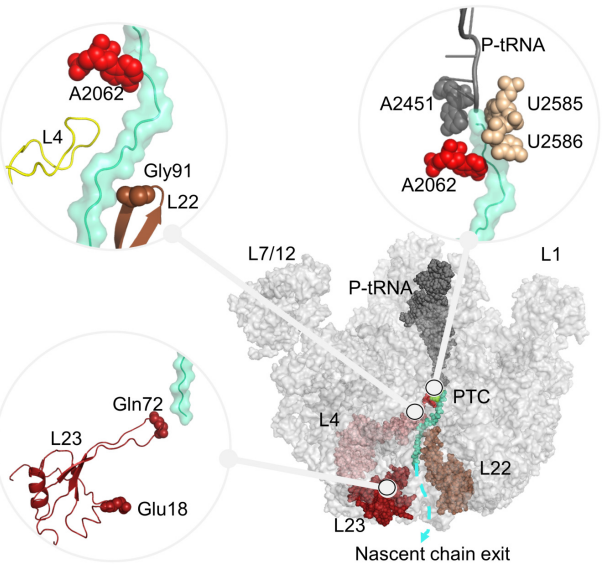


FIGURE 1 | Large subunit 50S, including P-tRNA (gray), polyAla chain (turquoise) and ribosomal proteins protruding to the ribosomal tunnel, namely uL4 (pink), uL22 (brown), uL23 (red) are shown. Nucleotides A2451 (PTC), A2062, U2585–U2586 (ribosomal tunnel), and residues Gly91 (tip of uL22 loop), Glu18 (trigger factor binding site on uL23), Gln72 (tip of uL23 loop), which are investigated in this study are also indicated. In all figures, PyMol (DeLano Scientific LLC., 2002) is used for the molecular visualization.

and around 50 ribosomal proteins. The subunits have different functional properties in translation, while they function together as a complex (Ramakrishnan, 2002; Schmeing and Ramakrishnan, 2009). The large subunit 50S catalyzes peptide bond synthesis at the highly conserved catalytic cavity peptidyl transferase center (PTC), where nucleotides G2251, G2252, A2451, C2452, U2506, U2585, A2602 play critical roles in the translation process (Polacek et al., 2003; Youngman et al., 2004; Erlacher et al., 2005; Martin Schmeing et al., 2005; Long et al., 2006; Selmer et al., 2006; Amort et al., 2007; Deutsch, 2014). The nascent polypeptide chain attached to the peptidyl-tRNA (P-tRNA) grows through the ∼100 Å long ribosomal tunnel. The ribosomal tunnel wall is mainly formed of 23S rRNA nucleotides. Nucleotides close to the PTC are highly conserved while nucleotides toward the exit site exhibit variations in bacteria and eukaryotes (Liutkute et al., 2020). Few ribosomal proteins, namely uL4, uL22, and bacteria-specific extension of uL23 also reside on the ribosomal tunnel. The extended loops of these proteins reach from the solvent side into the ribosome exit tunnel as shown in **Figure 1**. Approximately 25 Å far from the PTC, the loops of uL4 and uL22 form the narrowest part of the ribosomal tunnel, also referred to as the constriction region. Toward to its exit, the ribosomal tunnel accommodates a vestibule, where the long loop of uL23 protrudes.

The ribosomal tunnel is not a passive passageway but is actively taking a role in translation regulation (Wilson and Beckmann, 2011; Ito and Chiba, 2014; Liutkute et al., 2020). Several polypeptides with arrest sequences of up to ∼20 amino acids are known to stall the translation process at the elongation or termination steps for a variety of biological outputs. Some nascent chains require cofactors like amino acids as in TnaC (Cruz-Vera et al., 2005) and antibiotics as in ErmCL (Vazquez-Laslop et al., 2008; Ramu et al., 2011) to stall the protein synthesis in bacteria. Cofactor-dependent translation arrest usually serves to regulate the gene expression related to the cofactor itself. On the other hand, SecM (Yap and Bernstein, 2009; Bhushan et al., 2011) and MifM (Chiba et al., 2009) control their translation without necessitating cofactors. SecM-mediated translation arrest is used to regulate protein export, whereas MifM-mediated translation arrest optimizes both the quality and quantity of membrane proteins under changing physiological conditions. The arrest sequence recognition in these cases realizes due to specific interactions between the nascent chain and the ribosomal constituents at the upper part of the ribosomal tunnel, limited with the PTC and the constriction region. Not far from the PTC, the flexible nucleotide A2062 can trigger a conformational change at the PTC after sensing the arrest sequence on the nascent chain, such as by making contacts with Asp21 on TnaC or Arg163 on SecM, then stall the protein synthesis (Cruz-Vera et al., 2005; Bhushan et al., 2011; Ito and Chiba, 2013). Polypeptide stalling mechanisms also involve direct interactions with nascent chain and the ribosomal tunnel elements A2058, A2059, G2061, A2503, U2504, G2583, U2584, U2585, U2609 (close to the PTC), as well as A751, A752 (close to the constriction region) and flexible loops of uL4 and uL22 (at the constriction region) (Seidelt et al., 2009; Ito and Chiba, 2013; Deutsch, 2014; **Figure 1**). For the antibiotic-dependent arrest of ErmCL, a signal relay mechanism is suggested between the flexible nucleotide A2062 and nucleotides A2451 and C2452 at the A- site crevice of the PTC, assisted by nucleotides A2503, G2061 and U2504 (Vazquez-Laslop et al., 2008; Ramu et al., 2011). This network of nucleotides is also supported by graph and elastic network studies on *T. thermophilus* ribosome complex structures at different translation states (Guzel and Kurkcuoglu, 2017). Similarly, signal relay mechanisms proposed for the SecM include nucleotides A2062 and A2503 (Gumbart et al., 2012) as well as U2585, U2586 and U2506 (Zhang et al., 2015).

Nascent polypeptide chains can compact to adopt secondary structures in the narrower parts of the ribosomal tunnel, and their tertiary structures at the wider regions (Liutkute et al., 2020). Here, the dynamics of the large subunit (Kurkcuoglu et al., 2009), the ribosomal tunnel geometry (Trylska, 2010; Trovato and O'Brien, 2016) together with its electrostatic potential seems to play an important role on complexity and production rate of small folded proteins (Kudva et al., 2018). During its passage through the ribosomal tunnel, compacted chain interacts with the ribosomal tunnel elements and affects the recruitment of chaperones to the exit of the tunnel in bacteria (Trabuco et al., 2010a; Lin et al., 2012; Deutsch, 2014; Denks et al., 2017). This suggests conformational crosstalk not only within the tunnel but also outside the tunnel at the solvent side (Lu et al., 2011; Lin et al., 2012). Here, recruitment of signal recognition particle and trigger factor (TF), both binding uL23 at the solvent side is driven by the nascent chain at the early stages of the translation process. While the nascent chain containing a specific sequence can promote the binding of the signal recognition particle, a

compacted nascent chain can lessen the recruitment of the TF to the ribosome complex. These are possibly driven by a network of nucleotides/residues between the extension of uL23 into the ribosomal tunnel and chaperone binding site again on uL23 (Bornemann et al., 2008; Lin et al., 2012) (marked by Glu18 on uL23 in **Figure 1**). More interestingly, the degree of TF binding is shown to be dependent on the location of the compacted chain in the ribosomal tunnel (Lin et al., 2012). Interactions between the compacted nascent chain and flexible loop of uL23 have a high effect on TF binding, while interactions at the middle parts of the ribosomal tunnel slightly reduce TF recruitment. However, the upper part of the ribosomal tunnel does not affect the recruitment of the chaperone.

Evidently, allostery is an important mechanism at the ribosomal tunnel during translation. The key components that play in regulating the translation process are dispersed along the exit pathway. However, the molecular details of the allosteric communication pathways between these distinct sites remain elusive. At this point, the network of nucleotides and residues on the ribosomal tunnel taking a role in constant communication of the distant functional regions can be considered as targets to eliminate bacterial activity. Indeed, the region marked by the sensor A2062 is an attractive site for macrolides and ketolides in bacteria (Wilson, 2014; Arenz and Wilson, 2016), where most of these antibiotics allosterically stop the catalytic activity of the PTC. To reveal details of allosteric networks and suggest more plausible druggable sites, computational approaches focusing on the contact topology of the ribosomal tunnel can be employed for relatively fast and efficient screening.

In this study, we use two different methods to reveal potential allosteric communication pathways along the ribosomal tunnel: coarse-grained molecular dynamics simulations (Górecki et al., 2009) and residue network model (Guzel and Kurkcuoglu, 2017). Previous coarse-grained molecular dynamics simulations of length 500 ns (Trylska et al., 2005) enabled to observe functional motions of the ribosomal complex. Here, 700 ns long coarse-grained molecular dynamics simulations of the ribosomal complex of *E. coli* with a polyAla chain in the ribosomal tunnel (PDB ID 4v5h; Seidelt et al., 2009) is performed. Then, the perturbation response scanning (PRS) method (Bakan et al., 2011; General et al., 2014) is applied to the resulting covariance matrix to identify effectors and sensors at the ribosomal tunnel. We also calculate *k*-shortest pathways on the residue network representation of ribosomal complex of *E. coli* (PDB ID 4v5h). To reveal any similarities in potential allosteric communication pathways between bacterial species, the ribosomal complex of *T. thermophilus* (PDB ID: 4v5d) is studied with the PRS using 500 ns long coarse-grained molecular dynamics simulations. Then, *k*-shortest pathways of 101 conformers of the ribosomal complex of *T. thermophilus* previously generated by ClustENM (Kurkcuoglu et al., 2016) using PDB ID 4v9m (Zhou et al., 2013) are calculated. Although the specific interactions between the nascent chain and the ribosomal tunnel are critical in the sequence-dependent arrest of translation, the dynamical traits of nucleotides for this task must strongly rely on the topology of the structure. In this line, we aim to reveal pathways of nucleotides/residues that maintain constant communication through tertiary interactions, which can be commonly used in bacteria to regulate the translation of specific nascent chains or the recruitment of chaperones.

We first assess our computational approach by investigating allosteric communication pathways between the flexible A2062 and the PTC A-site A2451, which is previously studied in detail (Vazquez-Laslop et al., 2008; Ramu et al., 2011). Then, we focus on SecM interacting nucleotides U2585–U2586 (*E. coli* numbering), investigate signal relaying in the upper part of the tunnel and discuss our results in the light of previous studies (Nakatogawa and Ito, 2002; Seidelt et al., 2009; Wilson and Beckmann, 2011; Gumbart et al., 2012; Tsai et al., 2014). Finally, allosteric communication of Glu18 on uL23 with two different sites, namely Gln72 on uL23 (lower part of the tunnel, **Figure 1**), and Gly91 on uL22 β-hairpin (constriction region of the tunnel, **Figure 1**) is explored. These two residues represent the distinct zones that are reported to play a role in the recruitment of TF (Lin et al., 2012). Each case is discussed in detail while seeking a consensus of the two different methods. Findings for *E. coli* and *T. thermophilus* are also compared, where a common mechanism for allostery in the bacterial ribosomal tunnel as well as a novel drug binding region is proposed.

MATERIALS AND METHODS

To reveal potential allosteric communication pathways between distant functional sites, we use two different approaches and two different species. 700 ns long coarse-grained molecular dynamics simulations of *E. coli* ribosomal complex 70S are employed in PRS analysis. Sensitivity profiles of A2062, U2585–U2586 on 23S rRNA, and Glu18 on uL23 are analyzed to determine nucleotides/residues highly coupled to these functional sites in their dynamics. Then, the *k*-shortest pathways method is used to predict suboptimal pathways between distant functional sites around the ribosomal tunnel of *E. coli*. In addition, 500 ns long coarse-grained molecular dynamics simulations of *T. thermophilus* ribosomal complex 70S are investigated with the PRS method. Then, 101 conformers of the *T. thermophilus* ribosomal complex 70S generated with ClustENM are studied with the *k*-shortest pathways method focusing on the same functional sites. Here, molecular dynamics simulations provide local fluctuations of the ribosomal tunnel elements at a time scale that can reflect experimental B-factors. The residue network model used here takes contact topology of the native structure as a basis and highlights the "wirings" between predetermined sites of the molecular machine using the *k*-shortest pathways method. ClustENM provides an effective sampling around the functionally relevant low-frequency motions and gives distinct and reasonable topologies to investigate with the *k*-shortest pathways method.

Data Set

The crystal structure of the ribosomal complex of *E. coli* with PDB ID 4v5h of resolution 5.8 Å includes 5S, 16S and 23S rRNAs, around 50 ribosomal proteins, a P-tRNA and polyAla chain in the ribosomal tunnel, as depicted in **Figure 1**

(Seidelt et al., 2009). The crystal structure of *T. thermophilus* ribosomal complex with PDB ID 4v5d of resolution 3.5 Å contains 70S with A-, P-, E-tRNAs, and mRNA (Voorhees et al., 2009). In addition, large subunit 50S from 101 different conformers of the *T. thermophilus* ribosomal complex (PDB ID 4v9m of resolution 4.0 Å) with elongation factor G previously generated by ClustENM (Kurkcuoglu et al., 2016) are used. ClustENM is an iterative algorithm, which generates plausible full-atom conformers by deformation along with the collective modes of the elastic network model. The generated conformers are then clustered, and a representative conformer from each cluster is energetically minimized in implicit solvent. Obtained conformers are taken as starting structures for another round, and this procedure is repeated for several generations. Construction of several generations of conformers at full atomic scale provides an accurate sampling of large conformational changes of biomolecules in large systems. Ribosomal complex conformers employed in the data set were generated using five low-frequency vibrational modes with two generation cycles, which corresponded to five different classes of structures. These structures include functional conformational states, such as the ratchet-like motion of subunits and correlated motion of the L1 stalk with the E-tRNA, as detailed in Kurkcuoglu et al. (2016).

Coarse-Grained Molecular Dynamics Simulations

Coarse-grained molecular dynamics (CGMD) simulations are performed using RedMD (Górecki et al., 2009), which is suitable to study ribosome dynamics. The full-atom ribosome complexes 70S with PDB IDs 4v5h and 4v5d, are described as a one-bead model, where pseudo-atoms are located at Cα and P atoms to represent residues and nucleotides, respectively. The total potential energy of the structure is given by,

$$E = E_{1-2} + E_{1-3} + E_{1-4} + E_{bp} + E_{non\text{-}bonded} \quad (1)$$

The harmonic E_{1-2}, E_{1-3}, and E_{1-4} account for pseudo-bond, pseudo-angle, and pseudo-dihedral interactions involving two, three, and four successive beads, respectively. E_{bp} indicates the harmonic interactions between the nucleic acid base-pairs, and $E_{non-bonded}$ energy term represents the Morse potential to determine non-bonded interaction energy considering anharmonicity as,

$$V(r) = A_{P,C_\alpha}(r_0)\left[1 - exp\left(-\alpha\left(r - r_0\right)\right)\right]^2 \quad (2)$$

V(r) is used for both local and non-local non-bonded interactions. The local terms are calculated within a cut-off distance $R_{cut-off}$, which is 12.0 Å for Cα and 20.0 Å for P atoms. For the non-local terms, a cut-off distance of 35.0 Å is taken for all nodes. For local interactions, r_0 is taken as the equilibrium distance in the native structure, while for non-local interactions it changes according to the node type. $A_{P,C\alpha}$ is an exponential function, which differs for P$\cdots$P, Cα$\cdots$Cα and P$\cdots$Cα interactions and decreases with increasing distance between pseudo-atoms. All parameters used in this study are listed in **Supplementary Table 1**. In order to account for the solvent-ribosomal complex 70S interactions, Langevin dynamics are applied by adding viscous and random forces to Newton's equation of motion. Here, for the *E. coli* ribosomal complex 70S, two independent simulations of 700 ns are performed. For the ribosomal structure 70S of *T. thermophilus*, two independent simulations of 500 ns are carried. Prior to simulations, each system is subjected to an energy minimization as implemented in RedMD. Each system is heated from 10 to 300 K, and then production simulations are run at 300 K with a collision frequency of 2 ps^{-1} for Langevin dynamics. RedMD describes a constraint between CCA end of P-tRNA and polyAla chain to fix the polypeptide from one end, where the remaining is allowed to fluctuate in the ribosomal tunnel.

Perturbation Response Scanning

CGMD simulations are used to reveal the effectors and the sensors in the dynamic large subunit 50S of the ribosome. The effectors propagate signals in response to external perturbations and the sensors have a high propensity to sense signals. These two different dynamic properties of nucleotides/residues can shed light on the allosteric mechanisms in the tunnel region of the supramolecule. We used ProDy to perform PRS analysis on the CGMD trajectories (Bakan et al., 2011). In the PRS module of ProDy, a perturbation (one nucleotide/residue at a time) is applied by employing a $3N$-dimensional force vector based on Hooke's law $F = H \bullet \Delta R$. Then, displacements of nucleotides/residues as a response to that perturbation is observed considering the overall network. An $N \times N$ PRS matrix (heat map) is generated to display the influence and sensitivity profiles of nucleotides/residues (Atilgan and Atilgan, 2009; General et al., 2014). The *j*th column of the PRS matrix represents the response of all nucleotides/residues to the perturbation at nucleotide/residue *j*, and the average of this column elements point to the signal transmission potential of nucleotide/residue *j* as a sensor. The *i*th row of the matrix describes the response of *i*th nucleotide/residue to perturbations at all other sites and the average of the elements along the row indicates the potential of that nucleotide/residue acting as a propagator or an effector (Dutta et al., 2015).

k-Shortest Pathways

Structures from the data set are represented as undirected weighted graphs, formed of nodes linked by edges. Here, each node is located at Cα (residue) or P atom (nucleotide). The neighboring nodes are linked by edges, where the edge lengths indicate the strength of interactions. In this line, the length of an edge between a node pair (*i,j*) is calculated based on their local interaction strengths or affinity a_{ij} as,

$$a_{i,j} = \frac{N_{ij}}{\sqrt{N_i.N_j}} \quad (3)$$

N_{ij} is the total number of heavy atom-atom neighboring of the (*i,j*) node pair within a cut-off distance of 4.5 Å. A weighting factor of $N_i.N_j$ overcomes any bias due to the different sizes of nucleotides/residues. The edge length between (*i,j*) node pair can be described as the inverse of the interaction strength,

a_{ij}^{-1}. With this approach, the edges representing both bonded and non-bonded interactions have comparable values. Here, the communication capability of a node pair is assumed to be proportional to its interaction strength, and thus strongly interacting nodes are close to each other having the ability to transmit information using conformational changes (Brinda and Vishveshwara, 2005; Chennubhotla and Bahar, 2007; Seeber et al., 2015; Guzel and Kurkcuoglu, 2017).

After constructing the residue network model of the ribosome structure, *k*-shortest pathways between the selected source and sink nodes are calculated using Dijskra's algorithm (Dijkstra, 1959) and Yen's algorithm (Yen, 1971). As the network is undirected, the source and sink nodes are interchangeable, i.e., *k*-shortest pathways from the source to the sink are identical to those from the sink to the source. The value of $k = 20$ was previously found sufficient to reveal suboptimal pathways on the ribosome complex at different translation states (Guzel and Kurkcuoglu, 2017). This value is controlled for this study as well, which is discussed later. The length (or cost) of each pathway is determined by summing node-pair edge lengths. As one node may be found on more than one pathway, the occurrences of the nodes are calculated. In this way, suboptimal pathways between two functional sites can be determined; moreover, pathways of nodes with high occurrences can be suggested as potential allosteric pathways.

RESULTS AND DISCUSSION

We aim to explore potential allosteric communication pathways between distant regions at the long exit tunnel, also the nucleotides/residues that form these pathways. For this purpose, we focus on three different sites: (1) A2062 at the upper part of the ribosomal tunnel; (2) U2585–U2586 at the upper part of the ribosomal tunnel; and (3) Glu18 on uL23, which marks the binding region of trigger factor (TF) at the solvent side. We discuss our findings following this sequence of locations, i.e., from the upper part of the tunnel to its lower part toward the polypeptide exit. Results from CGMD simulations and *k*-shortest pathways calculations complement each other by revealing dynamic and topological features of the ribosomal tunnel, respectively. Finally, conservation analysis is carried for uL23 sequences of *H. sapiens* and bacteria to reveal potential druggable regions to stop the bacterial activity.

CGMD simulations of 700 ns long are used to obtain the dynamics of the *E. coli* ribosomal complex including both small and large subunits. Root mean square deviation (rmsd) and energy profiles of two independent runs are given in **Supplementary Figure 1A**. As Run1 has smaller fluctuations in rmsd, this trajectory is analyzed and reported. Principal component analysis (PCA) of the trajectory is carried using Bio3d (Grant et al., 2006). The variance percentages in the scree plot indicate that the first five PCs describe the half of the motions (**Supplementary Figure 2A**). Here, the PC-one corresponds to the anti-correlated motions of the uL1 and uL11 stalks, while the rotational motion of the small subunit 30S, similar to the ratchet-like motion is also noted (**Supplementary Figure 3A**). In the PC-two, uL1 stalk makes an anti-correlated motion with respect to the remaining of the complex, and in the PC-three, the anti-correlated motion of the stalks and the small subunit 30S is depicted. The ratchet rotation of the subunits requires GTP hydrolysis on the elongation factor G for the translocation of tRNAs. However, the ribosome complex is able to do a similar motion during the course of the simulations. Other two PCs also correspond to different functional motions of the ribosome complex required for the translation process. In **Supplementary Figure 4**, normalized B-factors are displayed for the large subunit 50S, and ribosomal proteins uL4, uL22, uL23, which are investigated in this study. The crystal structure 4v5h lacks experimental B-factors, therefore these values are taken from another crystal structure with PDB ID 4v9d (Dunkle et al., 2011) to assess the findings. The Pearson product correlation is calculated to compare the experimental and calculated fluctuations. Correlation coefficients are found as 0.75 (high amplitude fluctuation of the L1 stalk is excluded), 0.71, 0.59, and 0.58 for 23S rRNA, uL4, uL22, and uL23, respectively, which indicate good agreement of the calculated values with experimental data.

We also perform 500 ns long CGMD simulations of the *T. thermophilus* ribosome complex. The rmsd and energy profiles of two independent runs are shown in **Supplementary Figure 1B**. The rmsd increases up to 5.0 Å due to the large displacement of bL9 extended to the solvent side. Based on smaller fluctuations in rmsd, Run1 is analyzed in this study. The scree plot for the variance percentages of the PCs is shown in **Supplementary Figure 2B**, where the first five PCs describe more than half of the motions. With a more focused look, the highly flexible bL9 is noted to dominate the motions in the first PCs (not shown). We then exclude bL9 from the PCA to clearly observe collective motions of the ribosomal complex (**Supplementary Figure 3B**). Accordingly, the PC-one corresponds to the ratchet-like motion of the subunits where two subunits rotate around the same axis in opposite directions. The PC-two shows the correlated motion of uL1 stalk and E-tRNA and the PC-three corresponds to an anti-correlated motion of the subunits such as to open/close the interface from the A-site. All these motions are critical in different steps of the translation. In **Supplementary Figure 5A**, normalized B-factors are given for the large subunit 50S, including bL9. The Pearson product correlation between the experimental and calculated fluctuations is determined as 0.76 for the 23S rRNA while excluding very high peaks of the calculated fluctuations. Correlation coefficients for the ribosomal proteins uL4, uL22, and uL23 are found as 0.63, 0.75, and 0.45, respectively, where the trends in both fluctuation curves highly agree (**Supplementary Figure 5**).

Then, perturbation response scanning (PRS) analysis using the covariance matrix from PCA of the large subunit 50S trajectories is carried to get insights into two groups of residues, "sensors," and "effectors," which are both important for long-range signal transmission in allostery. In **Supplementary Figure 6**, the strongest effectors and sensors in the large subunit of *E. coli* and *T. thermophilus* are given. The strongest effectors, which are the most influential nucleotides/residues, are mostly located in the core regions, where the PTC and the ribosomal tunnel

are located. The nucleotides A2062 and U2585–U2586, and the residue Glu18 on uL23, which are chosen as source nodes in *k*-shortest pathways calculations, are determined as moderate effectors in the *E. coli* structure (**Supplementary Figure 6A**). For the *T. thermophilus* structure, A2062 and C2586 are noted to have moderate effectivity when compared to the remaining of the structure (**Supplementary Figure 6C**). On the other hand, the sensors that are highly sensitive to external perturbations are located at the periphery sites (**Supplementary Figures 6B,D**). Here, uL1 and uL11 stalks are highly mobile parts of the large subunit (**Supplementary Figures 4A, 5A**), they have also high sensitivity. This finding is meaningful in the sense that regulation of critical translation steps including the exit of tRNAs and elongation factor-G turnover during protein synthesis by uL1 and uL11, respectively (Harms et al., 2008; Trabuco et al., 2010b). In addition, the ribosomal protein bL9 in *T. thermophilus* structure, which has a closed conformation in the *E. coli* structure, has high flexibility (**Supplementary Figure 5A**) and high sensitivity (**Supplementary Figure 6D**). This finding may have a functional significance since bL9 helps the regulation of stress response protein RelA for the survival of the cell under stress conditions (Pei et al., 2017).

Potential Allosteric Communication Pathways Between A2062 and the PTC

A2062 is a critical nucleotide that interacts with antibiotics and nascent chains, and its related stalling mechanisms include sensing, interpreting, and relaying of a signal to PTC (Vazquez-Laslop and Mankin, 2014). An allosteric communication mechanism for drug-dependent ribosomal stalling was previously suggested between A2062 and nucleotides A2451 and C2452 at the A-site crevice of the PTC (Vazquez-Laslop et al., 2008; Ramu et al., 2011). Here, we further explore this mechanism by investigating a data set including numerous conformers from long CGMD simulations and ClustENM, while comparing the results for the large subunit 50S of *E. coli* and *T. thermophilus*.

Figure 2A indicates locations of the most influential nucleotides on A2062, obtained from the CGMD simulations of *E. coli* ribosomal complex (also listed in **Supplementary Table 2**). These nucleotides can be classified as having the strongest dynamic coupling with A2062, and thus they have high potential to establish allosteric communication with A2062. Among these, A2450, A2451, A2503, G2061, C2063, C2064, and U2504 are depicted, which are previously proposed to involve in an allosteric network linking the flexible A2062 to the PTC (Ramu et al., 2011; Guzel and Kurkcuoglu, 2017). Here, the universally conserved non-Watson-Crick base-pair A2450–C2063 is highly coupled to A2062, which may help to increase the strength of long-range signal transmission, as was previously suggested (Guzel and Kurkcuoglu, 2017).

A2503 and U2504 are also noted as strongly coupled nucleotides with A2062, underlying their role on allosteric communication, as was previously shown for antibiotic-dependent stalling (Vazquez-Laslop et al., 2008, 2010; Seidelt et al., 2009). The presence of erythromycin restricts the passage of the nascent polypeptide in the tunnel, which in turn forces A2062 to adopt an orientation clashing with A2503. This restriction then stalls the protein synthesis of ErmCL (Vazquez-Laslop et al., 2010). Binding of tiamulin causes similar conformational rearrangements involving A2504 (Gürel et al., 2009). In the CGMD simulations, there is no antibiotic to trigger such a situation, and polyAla chain in the ribosomal tunnel has moderate coupling with A2062, especially from Ala24 and Ala25. Consequently, the coupling of A2062 and A2503–U2504 seems to be inherent to maintain constant communication.

We also note that A2062 and U2585 are coupled, where the latter is in close proximity with Ala24 of the polyAla chain. Indeed, U2585 is known to interact with Pro24 of SecM for ribosomal stalling (Wilson and Beckmann, 2014). In CGMD simulations, Ala24 is sandwiched between U2585 and C2063, which can relay signal from A2062. Another interesting finding is the coupling of A2062 with G2251 and G2252 at the P-site crevice of the PTC. Here, the dynamic coupling is plausibly maintained using the CCA end of P-tRNA and A2450–A2451 at the A-site of the PTC. Another possible route is provided by C2065 and C2066, which are neighboring G2252.

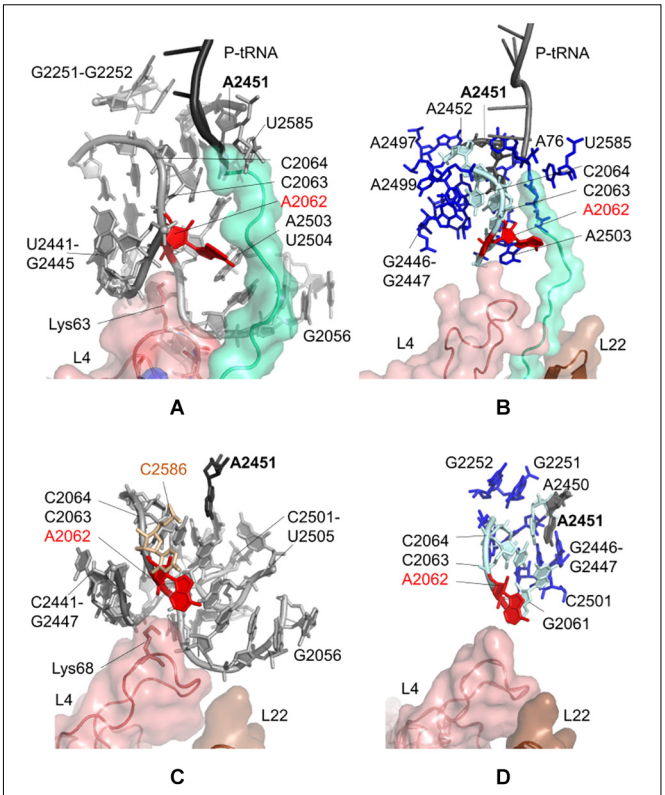


FIGURE 2 | Nucleotides/residues from the sensitivity profile of A2062 based on CGMD simulations of **(A)** *E. coli* ribosomal complex with PDB ID 4v5h, and **(C)** *T. thermophilus* ribosomal complex with PDB ID 4v5d. Nucleotides/residues forming the *k*-shortest pathways on **(B)** *E. coli* large subunit with PDB ID 4v5h, and **(D)** *T. thermophilus* conformers generated by ClustENM using crystal structure with PDB ID 4v9m. polyAla chain in the ribosomal tunnel is shown in turquoise, ribosomal protein uL4 in salmon and uL22 in brown. In **(B,D)** cyan and blue sticks represent the most and the least frequent nucleotides from the calculated pathways, respectively.

Moreover, sensitivity analysis highlights Lys63–Arg67 on the uL4 loop protruding to the ribosomal tunnel. Especially, the long side chain of Lys63 is oriented to A2062, which suggests the potential role of uL4 in allosteric signaling in this region.

The sensitivity profile of A2451 from CGMD simulations for *E. coli* is also investigated and given in **Supplementary Table 3**. PRS analysis stresses that the communication between A2451 and A2062 is in both directions; a perturbation on one nucleotide is sensed by the other and vice versa. Similarly, A2451 is also dynamically coupled to G2061–C2066, G2251–G2252 at the P-site crevice of the PTC and G2447–U2448, highlighting these nucleotides as elements of an allosteric network sharing information.

k = 20 shortest pathways are calculated between A2062 and A2451 based on the large subunit crystal structure 4v5h of *E. coli*. The cost of the pathways converges for all investigated cases of *E. coli* (**Supplementary Figure 7A**), indicating that the value of k = 20 is suitable for the analysis as was previously shown for the ribosome structures (Guzel and Kurkcuoglu, 2017). The analysis points to nucleotides known to be critical in ribosomal stalling (**Figure 2B**). The shortest pathway is determined as A2062 → C2063 → A2450 → A2451, where all these nucleotides have a high occurrence in the calculated 20 pathways (**Supplementary Table 4** and **Supplementary Figure 8A**). In addition, four sequential amino acids (Ala21–Ala24) on polyAla chain and A76 of P-tRNA are found on the shortest pathways linking A2062 and A2451, successfully capturing the role of a specific nascent chain – ribosomal tunnel interactions to trigger ribosomal stalling.

CGMD simulations taking a dynamic approach and k-shortest pathways method using a static crystal structure have high agreement on allosteric communication pathways at the upper part of the ribosomal tunnel of *E. coli*. Then, we investigate the CGMD simulations of *T. thermophilus* ribosomal complex, lacking the polypeptide chain in the tunnel. PRS analysis suggests that dynamic couplings of nucleotides in the upper part of the ribosomal tunnel (**Figure 2C**) highly agree with those in the *E. coli* case (**Figure 2A**), even in the absence of the nascent chain. The lists of nucleotides/residues with high sensitivity values for A2062 and A2451 are given in **Supplementary Tables 5, 6**, respectively.

In addition, we employ the k-shortest pathways method to a collection of 101 large subunit conformers previously generated from the crystal structure 4v9m of *T. thermophilus* using ClustENM (Kurkcuoglu et al., 2016). These conformers are generated around the low-frequency normal modes of the large subunit, which describe global functional motions of the structure, such as anti-correlated motions of the large stalks L1 and L7/L12 (Trylska et al., 2005), and reveal folding zones of the ribosomal tunnel (Kurkcuoglu et al., 2009). Therefore, they provide plausible structures to investigate allosteric communication pathways based on the conformational rearrangements around the ribosomal tunnel. A total of 2020 shortest pathways (k = 20 pathways/conformer × 101 conformers) are calculated, where the costs of all pathways converge at k = 20 (**Supplementary Figure 7B**). The analysis indicates that nucleotides with the highest occurrences highly agree with the findings from CGMD and k-shortest pathways of the *E. coli* crystal structure (**Figure 2D**, **Supplementary Table 7**, and **Supplementary Figure 8B**). While the nascent chain is missing from the conformers, the shortest pathway is determined as A2062 → C2063 → C2064 → A2450 → A2451. All these nucleotides are commonly determined from CGMD simulations, k-shortest pathways calculations of *E. coli*, and *T. thermophilus*. These results imply that signal relay mechanism between two relatively distant functional nucleotides A2062 and A2451 is the same in both species.

The similarity in the findings for both bacterial species stems from the contact topologies of their ribosomal structures. When the large subunits 50S of *E. coli* (4v5h) and *T. thermophilus* (4v9m) are structurally aligned, the rmsd is 2.5 Å over all atoms, and 1.6 Å when only phosphorous atoms are considered. The deviation is due to the flexible uL1 stalk. Then, a cylindrical region with a radius of 40.0 Å around the central axis of the tunnel is taken into account; the rmsd is found as 1.3 Å over all atoms. On the other hand, the rmsd values between the large subunits 50S of *E. coli* (4v5h) and *T. thermophilus* (4v5d) are calculated as 2.3 Å (all atoms), 1.3 Å (only phosphorous atoms), and 0.9 Å (all atoms, tunnel wall). We also calculate the number of contacts of the nucleotides investigated in this study (**Supplementary Figure 9**). Accordingly, the contact numbers are highly similar for *E. coli* and *T. thermophilus* structures.

Successful prediction of critical residues of the well-known signal relay mechanism at the upper part of the ribosomal tunnel motivates us to employ our approach for estimating allosterically predisposed nucleotides/residues between other distal functional sites in *E. coli* and *T. thermophilus*.

Potential Allosteric Communication Pathways Between U2585–U2586 and the PTC

During the synthesis of SecM, the ribosomal stalling process requires two components: a well-conserved stalling sequence and a ribosomal tunnel topology ready to detect this important detail, where U2585 and U2586 play a critical role (Zhang et al., 2015). Here, we focus on the contact topology of the ribosomal tunnel and investigate the sensitivity profiles for U2585–U2586 obtained from CGMD and PRS analyses. For the *E. coli* case, nucleotides A751–A753, A781–A782, U1782, A2062, A2439, and A2602 are found to be dynamically coupled to U2585–U2586, implying their role on long-range signal transmission between U2585–U2586 and the PTC (**Figure 3A** and **Supplementary Tables 8, 9**).

We note two apparent networks of nucleotides coupled to U2585–U2586 dynamics at the opposite sides of the ribosomal tunnel. The first contains U1782, U2609, and A751 neighboring the flexible β-hairpin of uL22. The other involves A2062, C2063, C2064, C2443, G2444, and A2059 next to the uL4 loop. The constriction region of the tunnel, where uL4 and uL22 loops protrude, is therefore linked to U2585–U2586. Moreover, closer to the tunnel entrance, A2439 and A2062 are coupled with U2585–U2586. These three networks of nucleotides agree well with the previous structural study on SecM mediated stalling (Seidelt et al., 2009). The analysis indicates that U2585–U2586

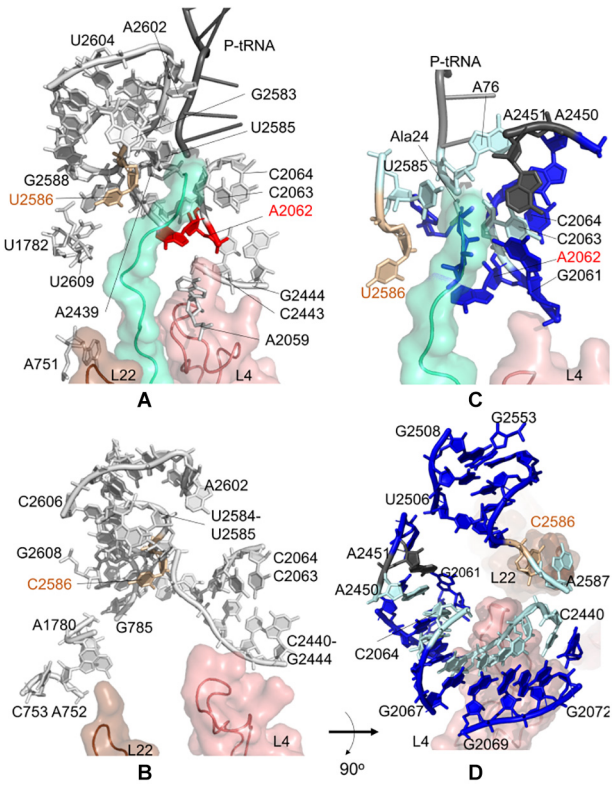


FIGURE 3 | Nucleotides/residues from the sensitivity profile of U2585–U2586 based on CGMD simulations of **(A)** *E. coli* ribosomal complex with PDB ID 4v5h, and **(B)** *T. thermophilus* with PDB ID 4v5d. Nucleotides forming the *k*-shortest pathways on **(C)** *E. coli* large subunit with PDB ID 4v5h, and **(D)** *T. thermophilus* conformers generated by ClustENM using crystal structure with PDB ID 4v9m. Coloring is as in **Figure 2**.

are not coupled with A2450–A2451 at the PTC, to which a signal/perturbation is plausibly directed through A2062–C2064 as previously discussed. As the polypeptide in the ribosomal tunnel does not contain a stalling sequence, we do not detect any significant coupling between the polyAla chain and nucleotides U2585–U2586. At this point, the results underline that at the upper part of the ribosomal tunnel, there exist multiple sites constantly monitoring and communicating during the translation of chains with or without stalling sequences.

Interestingly, we detect the same picture for the *T. thermophilus* ribosome tunnel: three different networks of nucleotides linking U2585–C2586 (i) to A2602 using C2441–C2442, (ii) to uL4 using C2063, C2064, C2443, G2444, and (iii) to uL22 using G785, A1780, A752, C753 (**Figure 3B** and **Supplementary Tables 10, 11**).

Closer to the PTC of *E. coli*, A2602 is coupled to U2585 and U2586. A2602 is known to be critical in nascent peptide release (Polacek et al., 2003) but not in drug-dependent translation arrest of ErmCL (Koch et al., 2017). Additionally, in all species, sparsomycin binds A2602 to change the PTC conformation (Porse et al., 1999). We also note the dynamic coupling of A2602 with U2586–C2586 in *T. thermophilus*. Considering the location and role of the highly conserved A2602 in the PTC, this nucleotide has a high potential to take a role in the translation arrest of other nascent chains that can employ different signal relay mechanisms.

Then, $k = 20$ shortest pathways between U2586 and A2451 are determined for *E. coli* large subunit structure (**Figure 3C**, **Supplementary Table 4** and **Supplementary Figure 10A**). The most frequently occurring nucleotides are determined as U2585, C2063, C2064, and A76 of P-tRNA, which are suggested to maintain distant communication. Here, as the method is based on the contact topology of the structure, residues of the polyAla chain also involve in suboptimal pathways. The shortest pathways are in good agreement with the CGMD results as well as with the previously reported signal relay mechanisms (Seidelt et al., 2009).

Shortest pathways calculations between U2586 and A2451 on *T. thermophilus* large subunit conformers point to C2063–C2066, C2440–G2446, A2450, and A2587 as the nucleotides with highest occurrences (**Figure 3D**, **Supplementary Table 7** and **Supplementary Figure 10B**). The shortest pathways calculated for these conformers involve more neighboring nucleotides when compared to *k*-shortest pathways for *E. coli* structure, due to lack of polypeptide in the ribosomal tunnel. Nonetheless, CGMD and *k*-shortest pathways of ClustENM conformers agree on the potential allosteric pathways. The contact topology points to functionally important nucleotides, such as G2251 at the P-site of PTC (**Supplementary Table 7**). Highly conserved flexible U2506 is another important nucleotide found from the calculations. This nucleotide plays a key role in peptide bond synthesis (Martin Schmeing et al., 2005) and contributes to pleuromutilin binding pocket together with A2058, A2059, and G2505 in *E. coli* and *T. thermophilus* (Long et al., 2006; Killeavy et al., 2020).

Potential Allosteric Communication Pathways Between uL23 and the Ribosomal Tunnel

The sensitivity profile of Glu18 on uL23 from CGMD simulations is visualized in **Figure 4A** and given in **Supplementary Table 12**. Nucleotides/residues dynamically coupled to Glu18 are mostly populated at the lower part of the ribosomal tunnel. On uL23, residues His15–Ser17, Lys33, Val63, Gly65, Lys81, Lys82 are highlighted, where Gly65 and Lys81 are located on the hinge of the flexible loop protruding to the tunnel. Moreover, 23S rRNA nucleotides G1339, G1395, A1610, A1616, have high potential to relay signal at the lower part of the ribosomal tunnel. Here, the polyAla chain from the crystal structure is 20 amino acids long and does not interact with the uL23 loop. Still, our findings highly agree with the experimental observations indicating that the interactions between the compacted nascent chain and the lower part of the tunnel strongly modulate the recruitment of TF and signal recognition particle (Lin et al., 2012).

While nucleotides of 23S rRNA and residues of uL23 are dynamically coupled at the lower part of the tunnel, this coupling seems to continue toward the constriction region of the tunnel through few nucleotides. As noted from **Figure 4A**, A1610–A1616 on 23S rRNA neighboring β-hairpin of uL22 can plausibly assist relaying signal between the inner wall of the ribosomal

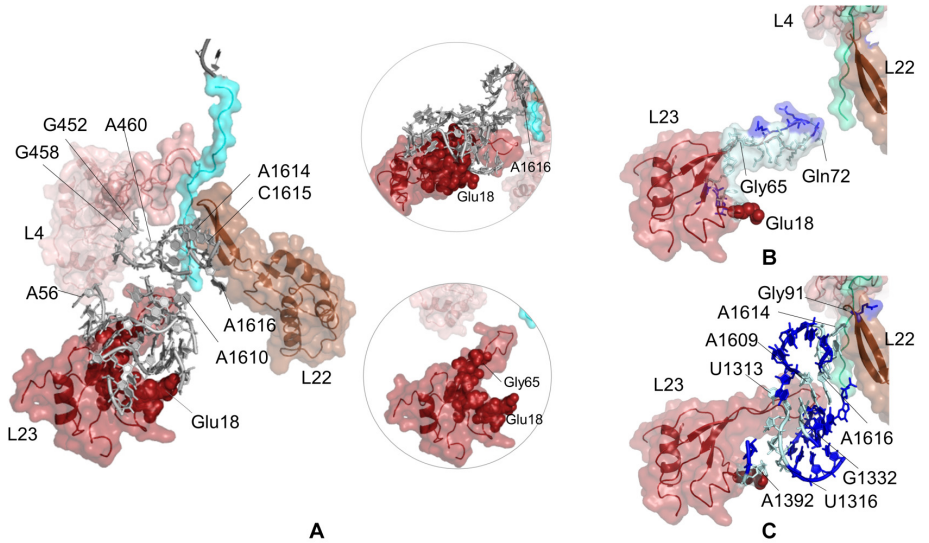


FIGURE 4 | (A) Nucleotides/residues from the sensitivity profile of Glu18 (uL23) based on CGMD simulations of *E. coli* ribosomal complex with PDB ID 4v5h. Insets show pathways from different perspectives. Nucleotides/residues forming the *k*-shortest pathways **(B)** between Glu18 (uL23) and Gln72 (uL23) and **(C)** between Glu18 (uL23) and Gly91 (uL22) on *E. coli* large subunit with PDB ID 4v5h. Coloring is as in **Figure 2**.

tunnel at the constriction region and solvent side. In addition, A56–U59, G452, and G458–A460 of 23S rRNA reaching the hinge of the uL4 loop are also involved in a network of coupled nucleotides linking the constriction region and the chaperone binding site. We do not detect any other nucleotide/residue near the upper regions of the tunnel coupled to Glu18. These findings support the previous FRET results (Lin et al., 2012); the chaperone binding site is weakly linked to the constriction region marked by flexible loops of uL4 and uL22, but not to upper regions close to the PTC. However, if an allosteric communication between the chaperone binding site and the PTC exists, approaches achieving higher time scales would be necessary to reveal the mechanism.

Potential allosteric communication pathways between Glu18 on uL23 and ribosomal tunnel of *E. coli* are further investigated using the *k*-shortest pathways method. Since the method requires a source and a sink node, we first calculate $k = 20$ shortest pathways between Glu18 and Gln72 (uL23). Twenty shortest pathways include only residues of uL23 based on contact topology of the crystal structure (**Figure 4B**, **Supplementary Table 4** and **Supplementary Figure 11A**). Here, tertiary interactions on uL23 trace a path using His70 (tip of uL23 loop) → Gly65 (hinge of uL23 loop) → Lys64 → Val63 → Asp79 → Trp80 → Lys33 → Ser17 → His15 → Glu18, between the inside of the tunnel and the solvent side, consistent with CGMD results. Then, $k = 20$ shortest pathways are calculated between Glu18 and Gly91 on uL22 β-hairpin (**Supplementary Figure 12A** and **Supplementary Table 4**). Gly91 is known to be a hot spot for the nascent chain stalling (Wilson and Beckmann, 2011). As displayed in **Figure 4C**, the shortest pathways involve mostly 23S rRNA nucleotides, where Lys19 (uL23) → A1392 → U1316 → C1315 → C1314 → G1332 → A1609 → A1616 → C1615 → A1614 is the shortest route between distant Glu18 (uL23) and Gly91 (uL22). These results highly agree with the PRS analysis of CGMD simulations and also suggest the suboptimal pathways between these distant sites.

Potential communication pathways between chaperone binding site on uL23 and both lower part and the constriction region of the tunnel are investigated for large subunit 50S of *T. thermophilus*. **Figure 5A** displays the nucleotides and residues, which are dynamically coupled to Glu15 of uL23. Similar to the findings for *E. coli* simulations using PRS analysis, nucleotides/residues with high sensitivity cluster on and around uL23, but interestingly they do not reach uL22. Here, CGMD simulations of the *T. thermophilus* ribosome complex are 500 ns long, whereas time length is 700 ns for the *E. coli* ribosome complex simulations. This implies that the signal transmission between these distant regions, Glu15 on uL23 and the loop of uL22, plausibly requires longer than 500 ns. Then, ClustENM conformers of *T. thermophilus* are analyzed with the *k*-shortest pathways method. **Figure 5B** shows the nucleotides and residues on a total of 2020 shortest pathways between Glu15Tt (*T. thermophilus* numbering) and Arg68Tt at the tip of the uL23 loop. Two suboptimal pathways are noted; one tracing uL23 residues similar to the results for *E. coli* (see **Figure 4B**) and the other using 23S rRNA nucleotides. Shortest pathways calculations point that allosteric communication between chaperone binding site and the lower part of the tunnel can employ tertiary interactions both on uL23 and 23S rRNA depending on the conformational rearrangements (**Supplementary Table 7** and **Supplementary Figure 11B**). It is worth to note here that ClustENM conformers provide conformations that reveal shortest pathways similar to those obtained from PRS analysis of CGMD simulations.

Suboptimal pathways between Glu15 on uL23 and Gly91 on uL22 determined from ClustENM conformers of *T. thermophilus*

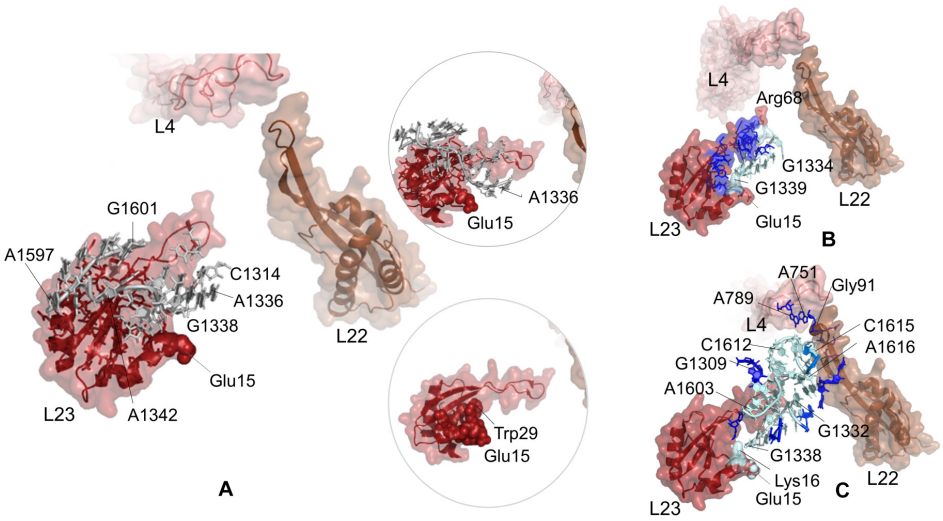


FIGURE 5 | (A) Nucleotides/residues from the sensitivity profile of Glu15 (uL23) based on CGMD simulations of *T. thermophilus* ribosomal complex with PDB ID 4v5d. Insets show pathways from different perspectives. Nucleotides/residues forming the *k*-shortest pathways **(B)** between Glu15 (uL23) and Arg68 (uL23) and **(C)** between Glu15 (uL23) and Gly91 (uL22) on ClustENM conformers of the *T. thermophilus* large subunit with PDB ID 4v9m. Coloring is as in **Figure 2**.

are shown in **Figure 5C** (also see **Supplementary Table 7** and **Supplementary Figure 12B**). These pathways employ similar nucleotides as in *E. coli*, where nucleotides G1332Tt-G1338Tt and C1612Tt-A1616Tt are highlighted as potential components of an allosteric network common to both ribosomal structures.

Finally, we perform conservation analysis of ribosomal protein uL23 by multiple sequence alignment for *E. coli*, *T. thermophilus*, and *H. sapiens* to explore suitable sites for drug design. Sequence alignments are done using the Clustal Omega program with the default settings on the UniProt.org server. **Supplementary Figure 13A** displays the results of the *E. coli* structure. Here, the hinge of the uL23 loop, which plausibly plays a critical role in signal relaying in trigger factor recruitment, is not conserved among human and *E. coli*/*T. thermophilus*. Especially, a small non-conserved cavity is detected around the hinge of the uL23 loop, contoured by residues Lys9, Arg12, His15, Lys33, Lys36, Ser78, and Trp80 in *E. coli*. Functional motions of long loops are often controlled by the hinge regions, highlighting this site attractive as a drug target. Moreover, electropositive side chains of these residues interact with the electronegative backbone of U59–U62 on 23S rRNA, which in turn holds uL23. Binding of a small molecule on this cavity while interacting with the 23S rRNA nucleotides can affect the functional motions of uL23 in the ribosome complex. Worth to note that conservation of the cavity is low among bacteria as well (**Supplementary Figure 13B**), which in turn suggests this region as a species-specific target site. Moreover, uL23 also hosts the signal recognition particle providing two binding sites; globular domain (Glu18 and Glu52) and the loop (Gly71) (Denks et al., 2017). Accordingly, interacting with the tip of the uL23 loop is suggested to enable the signal recognition particle to sense the arrival of the nascent chain. After sensing the nascent chain from the loop motions, the binding affinity of the chaperone apparently increases, then the chaperone proceeds with the standby or anticipatory mode and later with the recognition step. Consequently, blocking the motions of this loop can also affect the binding of the signal recognition particle.

CONCLUSION

The ribosomal tunnel can be considered as having three compartments, an upper part, a middle part and a lower part, similar to folding zones (Deutsch, 2014), where separate control elements regulate translation process. At the upper part, 23S rRNA nucleotides A2062, U2585, U2586 control co-factor dependent/independent translation arrest of specific sequences (Cruz-Vera and Yanofsky, 2014; Ito and Chiba, 2014; Vazquez-Laslop and Mankin, 2014). Our results indicate that, even in the absence of a specific stalling sequence or a co-factor, a network of inherently coupled nucleotides exists, which is ready to detect the sequence then stall the translation. Especially, CGMD simulations point out that the communication of A2062 and A2451 at the PTC is in both directions, dictated by the contact topology. On the other hand, critical nucleotides U2585–U2586 are not dynamically coupled to the PTC, yet they can communicate with A2451 through C2063–C2064. We determine two other distinct suboptimal pathways linking U2585–U2586 to uL4 and uL22 loops at the constriction region, which marks the middle part of the ribosomal tunnel.

At the lower part of the tunnel, other allosteric communication pathways plausibly exist to regulate the recruitment of chaperones to the ribosomal complex. Here, uL23 is acting as a bridge between the chaperone binding region at the solvent side and the vestibule, where compacted chains are waiting to emerge. We suggest that the chaperone binding site is strongly communicating with the ribosomal tunnel using the uL23 loop and His15, Ser17, Lys33, while nucleotides

A1336–G1339 also contribute. On the other hand, a weak signal relaying path from the chaperone binding site uses nucleotides G458–A460 and C1611–C1615 respectively reaching uL4 and uL22 loops at the constriction region. Based on 700 ns long CGMD simulations, we do not detect any dynamic coupling between the chaperone binding site and the upper part of the ribosomal tunnel.

As the contact topology of *E. coli* and *T. thermophilus* are highly similar, PRS analysis results and *k*-shortest pathways calculations point to similar suboptimal pathways implying similar signaling mechanisms at their ribosomal tunnels. The conservation analysis of uL23 using *H. sapiens*, *E. coli*, and *T. thermophilus* sequences reveals a non-conserved pocket contoured by polar amino acids as well as 23S rRNA nucleotides U59–U62, which is proposed as a novel site for drug design to disrupt the function of uL23.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

PG and OK designed the work. PG and HZY carried the molecular dynamics simulations and *k*-shortest pathways calculations, respectively. PG, HZY, MY, and OK analyzed the findings. PG and OK wrote and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by Istanbul Technical University Scientific Research Projects Foundation, Project No. 36110. PG acknowledges TUBITAK BIDEB 2211. Computing resources used in this work were partially provided by the National Center for High Performance Computing of Turkey (UHeM) under grant no. 4007722020 and TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

ACKNOWLEDGMENTS

OK thanks Pemra Doruker for stimulating discussions on the *k*-shortest pathways of the ribosomal complex. HZY thanks Pemra Doruker for the support from the Bogazici University Scientific Research Projects, Project No. 9360.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.586075/full#supplementary-material

REFERENCES

Amort, M., Wotzel, B., Bakowska-Zywicka, K., Erlacher, M. D., Micura, R., and Polacek, N. (2007). An intact ribose moiety at A2602 of 23S rRNA is key to trigger peptidyl-tRNA hydrolysis during translation termination. *Nucleic Acids Res.* 35, 5130–5140. doi: 10.1093/nar/gkm539

Arenz, S., and Wilson, D. N. (2016). Bacterial protein synthesis as a target for antibiotic inhibition. *Cold Spring Harb. Perspect. Med.* 6:a025361. doi: 10.1101/cshperspect.a025361

Atilgan, C., and Atilgan, A. R. (2009). Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput. Biol.* 5:e1000544. doi: 10.1371/journal.pcbi.1000544

Bakan, A., Meireles, L. M., and Bahar, I. (2011). ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* 27, 1575–1577. doi: 10.1093/bioinformatics/btr168

Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., et al. (2011). SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol.* 9:e1000581. doi: 10.1371/journal.pbio.1000581

Bornemann, T., Jöckel, J., Rodnina, M. V., and Wintermeyer, W. (2008). Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. *Nat. Struct. Mol. Biol.* 15, 494–499. doi: 10.1038/nsmb.1402

Brinda, K. V., and Vishveshwara, S. (2005). A network representation of protein structures: implications for protein stability. *Biophys. J.* 89, 4159–4170. doi: 10.1529/biophysj.105.064485

Chennubhotla, C., and Bahar, I. (2007). Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput. Biol.* 3:e0030172. doi: 10.1371/journal.pcbi.0030172

Chiba, S., Lamsa, A., and Pogliano, K. (2009). A ribosome-nascent chain sensor of membrane protein biogenesis in *Bacillus subtilis*. *EMBO J.* 28, 3461–3475. doi: 10.1038/emboj.2009.280

Cruz-Vera, L. R., Rajagopal, S., Squires, C., and Yanofsky, C. (2005). Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression. *Mol. Cell* 19, 333–343. doi: 10.1016/j.molcel.2005.06.013

Cruz-Vera, L. R., and Yanofsky, C. (2014). "Instructing the translating ribosome to sense l-tryptophan during synthesis of the TnaC nascent regulatory peptide," in *Regulatory Nascent Polypeptides*, ed. K. Ito (Tokyo: Springer).

Denks, K., Sliwinski, N., Erichsen, V., Borodkina, B., Origi, A., and Koch, H. G. (2017). The signal recognition particle contacts uL23 and scans substrate translation inside the ribosomal tunnel. *Nat. Microbiol.* 2:16265. doi: 10.1038/nmicrobiol.2016.265

Deutsch, C. (2014). "Tunnel vision: insights from biochemical and biophysical studies," in *Regulatory Nascent Polypeptides*, ed. K. Ito (Cham: Springer), 61–87.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Num. Math.* 1, 269–271. doi: 10.1007/BF01386390

Dunkle, J. A., Wang, L., Feldman, M. B., Pulk, A., Chen, V. B., Kapral, G. J., et al. (2011). Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* 332, 981–984. doi: 10.1126/science.1202692

Dutta, A., Krieger, J., Lee, J. Y., Garcia-Nafria, J., Greger, I. H., and Bahar, I. (2015). Cooperative dynamics of intact AMPA and NMDA glutamate receptors: similarities and subfamily-specific differences. *Structure* 23, 1692–1704. doi: 10.1016/j.str.2015.07.002

Erlacher, M. D., Lang, K., Shankaran, N., Wotzel, B., Hüttenhofer, A., Micura, R., et al. (2005). Chemical engineering of the peptidyl transferase center reveals an important role of the 2'-hydroxyl group of A2451. *Nucleic Acids Res.* 33, 1618–1627. doi: 10.1093/nar/gki308

General, I. J., Liu, Y., Blackburn, M. E., Mao, W., Gierasch, L. M., and Bahar, I. (2014). ATPase subdomain ia is a mediator of interdomain allostery in Hsp70 molecular chaperones. *PLoS Comput. Biol.* 10:e1003624. doi: 10.1371/journal.pcbi.1003624

Górecki, A., Szypowski, M., Długosz, M., and Trylska, J. (2009). RedMD-redcuded molecular dynamics. *J. Comput. Chem.* 30, 2364–2374.

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., and Caves, L. S. D. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 22, 2695–2696. doi: 10.1093/bioinformatics/btl461

Gumbart, J., Schreiner, E., Wilson, D. N., Beckmann, R., and Schulten, K. (2012). Mechanisms of SecM-mediated stalling in the ribosome. *Biophys. J.* 103, 331–341. doi: 10.1016/j.bpj.2012.06.005

Gürel, G., Blaha, G., Moore, P. B., and Steitz, T. A. (2009). U2504 determines the species specificity of the a-site cleft antibiotics: the structures of tiamulin, homoharringtonine, and bruceantin bound to the ribosome. *J. Mol. Biol.* 389, 146–156. doi: 10.1016/j.jmb.2009.04.005

Guzel, P., and Kurkcuoglu, O. (2017). Identification of potential allosteric communication pathways between functional sites of the bacterial ribosome by graph and elastic network models. *Biochim. Biophys. Acta Gen. Subj.* 1861, 3131–3141. doi: 10.1016/j.bbagen.2017.09.005

Harms, J. M., Wilson, D. N., Schluenzen, F., Connell, S. R., Stachelhaus, T., Zaborowska, Z., et al. (2008). Translational regulation via L11: molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. *Mol. Cell* 30, 26–38. doi: 10.1016/j.molcel.2008.01.009

Ito, K., and Chiba, S. (2013). Arrest peptides: cis-acting modulators of translation. *Annu. Rev. Biochem.* 82, 171–202. doi: 10.1146/annurev-biochem-080211-105026

Ito, K., and Chiba, S. (2014). "Biological significance of nascent polypeptides that stall the ribosome," in *Regulatory Nascent Polypeptides*, ed. K. Ito (Tokyo: Springer).

Killeavy, E. E., Jogl, G., and Gregory, S. T. (2020). Tiamulin-resistant mutants of the thermophilic bacterium thermus thermophilus. *Antibiotics* 9, 1–10. doi: 10.3390/antibiotics9060313

Koch, M., Willi, J., Pradere, U., Hall, J., and Polacek, N. (2017). Critical 23S rRNA interactions for macrolide-dependent ribosome stalling on the ErmCL nascent peptide chain. *Nucleic Acids Res.* 45, 6717–6728. doi: 10.1093/nar/gkx195

Kudva, R., Tian, P., Pardo-Avila, F., Carroni, M., Best, R. B., Bernstein, H. D., et al. (2018). The shape of the bacterial ribosome exit tunnel affects cotranslational protein folding. *eLife* 7:e36326. doi: 10.7554/eLife.36326

Kurkcuoglu, O., Kurkcuoglu, Z., Doruker, P., and Jernigan, R. L. (2009). Collective dynamics of the ribosomal tunnel revealed by elastic network modeling. *Proteins* 75, 837–845. doi: 10.1002/prot.22292

Kurkcuoglu, Z., Bahar, I., and Doruker, P. (2016). ClustENM: ENM-based sampling of essential conformational space at full atomic resolution. *J. Chem. Theory Comput.* 12, 4549–4562. doi: 10.1021/acs.jctc.6b00319

Lin, K. F., Sun, C. S., Huang, Y. C., Chan, S. I., Koubek, J., Wu, T. H., et al. (2012). Cotranslational protein folding within the ribosome tunnel influences trigger-factor recruitment. *Biophys. J.* 102, 2818–2827. doi: 10.1016/j.bpj.2012.04.048

Liu, Q., and Fredrick, K. (2016). Intersubunit Bridges of the Bacterial Ribosome. *J. Mol. Biol.* 428, 2146–2164. doi: 10.1016/j.jmb.2016.02.009

Liutkute, M., Samatova, E., and Rodnina, M. V. (2020). Cotranslational folding of proteins on the ribosome. *Biomolecules* 10:97. doi: 10.3390/biom10010097

Long, K. S., Hansen, L. H., Jakobsen, L., and Vester, B. (2006). Interaction of pleuromutilin derivatives with the ribosomal peptidyl transferase center. *Antimicrob. Agents Chemother.* 50, 1458–1462. doi: 10.1128/AAC.50.4.1458-1462.2006

Lu, J., Hua, Z., Kobertz, W. R., and Deutsch, C. (2011). Nascent peptide side chains induce rearrangements in distinct locations of the ribosomal tunnel. *J. Mol. Biol.* 411, 499–510. doi: 10.1016/j.jmb.2011.05.038

Martin Schmeing, T., Huang, K. S., Strobel, S. A., and Steitz, T. A. (2005). An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* 438, 520–524. doi: 10.1038/nature04152

Nakatogawa, H., and Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. *Cell* 108, 629–636. doi: 10.1016/S0092-8674(02)00649-9

Pei, H., Han, S., Yang, S., Lei, Z., Zheng, J., and Jia, Z. (2017). Phosphorylation of bacterial L9 and its functional implication in response to starvation stress. *FEBS Lett.* 591, 3421–3430. doi: 10.1002/1873-3468.12840

Polacek, N., Gomez, M. J., Ito, K., Xiong, L., Nakamura, Y., and Mankin, A. (2003). The critical role of the universally conserved A2602 of 23S ribosomal RNA in the release of the nascent peptide during translation termination. *Mol. Cell* 11, 103–112. doi: 10.1016/S1097-2765(02)00825-0

Porse, B. T., Kirillov, S. V., Awayez, M. J., Ottenheijm, H. C. J., and Garrett, R. A. (1999). Direct crosslinking of the antitumor antibiotic sparsomycin, and its derivatives, to A2602 in the peptidyl transferase center of 23S-like rRNA within ribosome-tRNA complexes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9003–9008. doi: 10.1073/pnas.96.16.9003

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* 108, 557–572. doi: 10.1016/S0092-8674(02)00619-0

Ramu, H., Vazquez-Laslop, N., Klepacki, D., Dai, Q., Piccirilli, J., Micura, R., et al. (2011). Nascent peptide in the ribosome exit tunnel affects functional properties of the A-site of the peptidyl transferase center. *Mol. Cell* 41, 321–330. doi: 10.1016/j.molcel.2010.12.031

Schmeing, T. M., and Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461, 1234–1242. doi: 10.1038/nature08403

Seeber, M., Felline, A., Raimondi, F., Mariani, S., and Fanelli, F. (2015). WebPSN: a web server for high-throughput investigation of structural communication in biomacromolecules. *Bioinformatics* 31, 779–781. doi: 10.1093/bioinformatics/btu718

Seidelt, B., Innis, C. A., Wilson, D. N., Gartmann, M., Armache, J. P., Villa, E., et al. (2009). Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* 326, 1412–1415. doi: 10.1126/science.1177662

Selmer, M., Dunham, C. M., Murphy, F. V. IV, Weixlbaumer, A., Petry, S., Kelley, A. C., et al. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935–1942. doi: 10.1126/science.1131127

Trabuco, L. G., Harrison, C. B., Schreiner, E., and Schulten, K. (2010a). Recognition of the regulatory nascent chain TnaC by the ribosome. *Structure* 18, 627–637. doi: 10.1016/j.str.2010.02.011

Trabuco, L. G., Schreiner, E., Eargle, J., Cornish, P., Ha, T., Luthey-Schulten, Z., et al. (2010b). The role of L1 Stalk-tRNA interaction in the ribosome elongation cycle. *J. Mol. Biol.* 402, 741–760. doi: 10.1016/j.jmb.2010.07.056

Trovato, F., and O'Brien, E. P. (2016). Insights into cotranslational nascent protein behavior from computer simulations. *Annu. Rev. Biophys.* 45:345. doi: 10.1146/annurev-biophys-070915-094153

Trylska, J. (2010). Coarse-grained models to study dynamics of nanoscale biomolecules and their applications to the ribosome. *J. Phys. Condens. Matter* 22:453101. doi: 10.1088/0953-8984/22/45/453101

Trylska, J., Tozzini, V., and McCammon, J. A. (2005). Exploring global motions and correlations in the ribosome. *Biophys. J.* 89, 1455–1463. doi: 10.1529/biophysj.104.058495

Tsai, A., Kornberg, G., Johansson, M., Chen, J., and Puglisi, J. D. (2014). The dynamics of SecM-induced translational stalling. *Cell Rep.* 7, 1521–1533. doi: 10.1016/j.celrep.2014.04.033

Vazquez-Laslop, N., and Mankin, A. S. (2014). "Triggering peptide-dependent translation arrest by small molecules: ribosome stalling modulated by antibiotics," in *Regulatory Nascent Polypeptides*, ed. K. Ito (Tokyo: Springer).

Vazquez-Laslop, N., Ramu, H., Klepacki, D., Kannan, K., and Mankin, A. S. (2010). The key function of a conserved and modified rRNA residue in the ribosomal response to the nascent peptide. *EMBO J.* 29, 3108–3117. doi: 10.1038/emboj.2010.180

Vazquez-Laslop, N., Thum, C., and Mankin, A. S. (2008). Molecular mechanism of drug-dependent ribosome stalling. *Mol. Cell* 30, 190–202. doi: 10.1016/j.molcel.2008.02.026

Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C., and Ramakrishnan, V. (2009). Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat. Struct. Mol. Biol.* 16, 528–533. doi: 10.1038/nsmb.1577

Wilson, D. N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* 12, 35–48. doi: 10.1038/nrmicro3155

Wilson, D. N., and Beckmann, R. (2011). The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Curr. Opin. Struct. Biol.* 21, 274–282. doi: 10.1016/j.sbi.2011.01.007

Wilson, D. N., and Beckmann, R. (2014). "Structures of nascent polypeptide chain-dependent-stalled ribosome complexes," in *Regulatory Nascent Polypeptides*, ed. K. Ito (Berlin: Springer), 45–60.

Yap, M. N., and Bernstein, H. D. (2009). The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel. *Mol. Cell* 34, 201–211. doi: 10.1016/j.molcel.2009.04.002

Yen, J. Y. (1971). Finding the K shortest loopless paths in a network. *Manag. Sci.* 17, 712–716. doi: 10.1287/mnsc.17.11.712

Youngman, E. M., Brunelle, J. L., Kochaniak, A. B., and Green, R. (2004). The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell* 117, 589–599. doi: 10.1016/S0092-8674(04)00411-8

Zhang, J., Pan, X., Yan, K., Sun, S., Gao, N., and Sui, S. F. (2015). Mechanisms of ribosome stalling by SecM at multiple elongation steps. *eLife* 4:e09684. doi: 10.7554/eLife.09684

Zhou, J., Lancaster, L., Donohue, J. P., and Noller, H. F. (2013). Crystal structures of EF-G-Ribosome complexes trapped in intermediate states of translocation. *Science* 340:1236086. doi: 10.1126/science.1236086

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 25 September 2020
doi: 10.3389/fmolb.2020.586540

Molecular Dynamics Investigations of Binding Mechanism for Triazoles Inhibitors to CYP51

Na Shi[1], Qingchuan Zheng[1,2] and Hongxing Zhang[1]**

[1] *Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, International Joint Research Laboratory of Nano-Micro Architecture Chemistry, College of Chemistry, Jilin University, Changchun, China,* [2] *Key Laboratory for Molecular Enzymology and Engineering of the Ministry of Education, Jilin University, Changchun, China*

The sterol 14α demethylase enzyme (CYP51) is an important target of fungal infections. However, the molecular mechanism between triazoles inhibitors and CYP51 remains obscure. In this study, we have investigated the binding mechanism and tunnel characteristic upon four triazoles inhibitors with CYP51 based on the molecular docking and molecular dynamics simulations. The results indicate the four inhibitors stabilize in the binding cavity of CYP51 in a similar binding mode. We discover a hydrophobic cavity (F58, Y64, Y118, L121, Y132, L376, S378, S506, S507, and M508) and the hydrophobic interaction is the main driving force for inhibitors binding to CYP51. The long-tailed inhibitors (posaconazole and itraconazole) have stronger binding affinities than short-tailed inhibitors (fluconazole and voriconazole) because long-tailed inhibitors can form more hydrophobic interactions with CYP51. The tunnel 2f is the predominant pathway for inhibitors ingress/egress protein, which is similar to the other works of CYP51. This study could provide the theoretical basis for the development of efficient azoles inhibitors and may lead a better insight into structure–function relationships of CYP51.

Keywords: molecular dynamics simulations, MM-GB/SA, CYP51, triazoles, tunnels

OPEN ACCESS

Edited by:
Guang Hu,
Soochow University, China

Reviewed by:
Weiwei Xue,
Chongqing University, China
Jinyu Li,
Fuzhou University, China

****Correspondence:***
Qingchuan Zheng
zhengqc@jlu.edu.cn
Hongxing Zhang
zhanghx@jlu.edu.cn

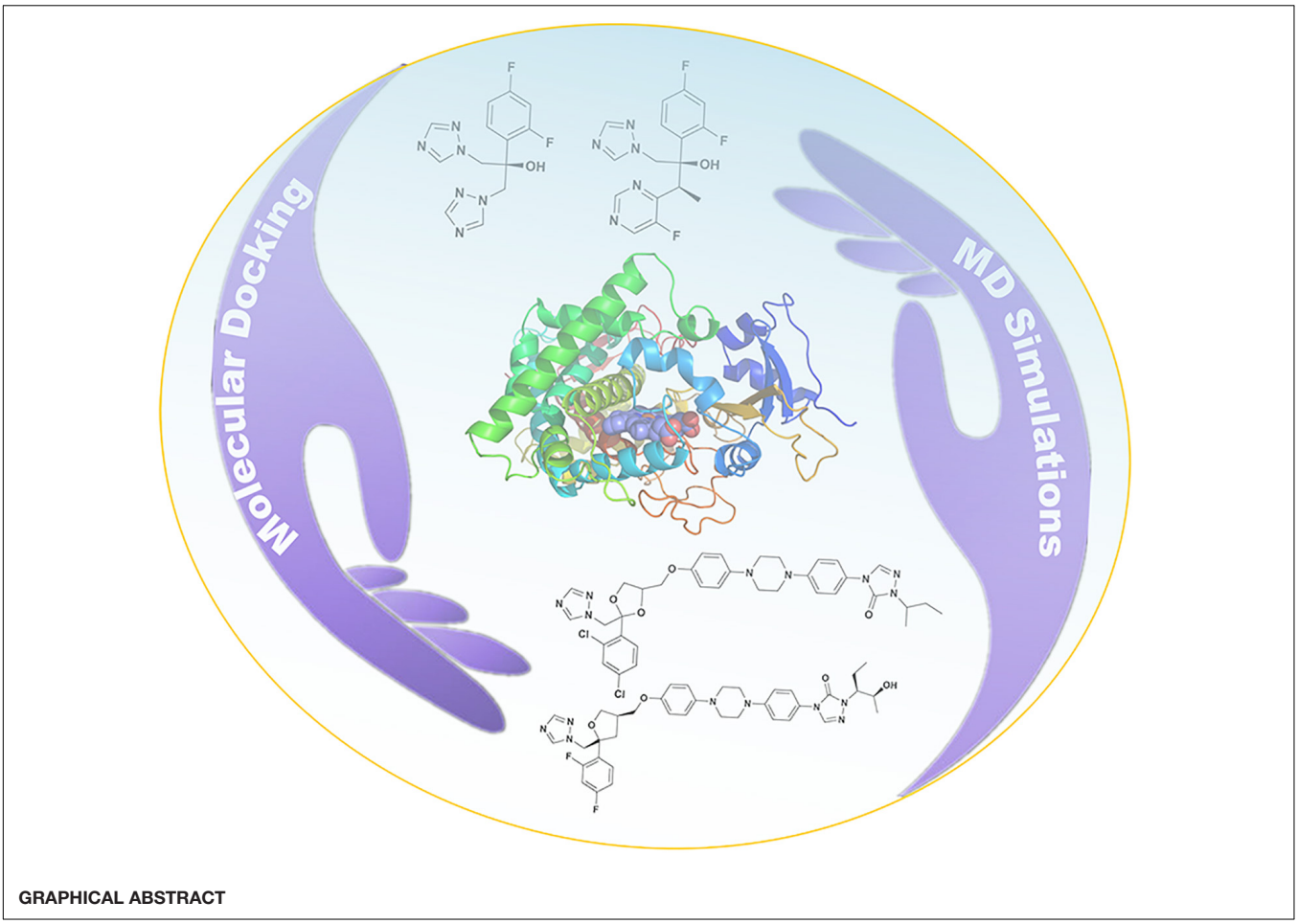
Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *23 July 2020*
Accepted: *02 September 2020*
Published: *25 September 2020*

Citation:
Shi N, Zheng Q and Zhang H (2020) Molecular Dynamics Investigations of Binding Mechanism for Triazoles Inhibitors to CYP51. Front. Mol. Biosci. 7:586540. doi: 10.3389/fmolb.2020.586540

INTRODUCTION

Life-threatening infections caused by fungi have increased rapidly, especially for patients that have immunocompromised diseases, such as AIDS, cancer, and organ-transplantation (Bongomin et al., 2017; Lee et al., 2020). It is reported that fungal diseases kill 1.5 million people per year, whose number is almost equal to the death of tuberculosis, and nearly three times as that of malaria (Brown et al., 2012; Bongomin et al., 2017). Among all kinds of fungal pathogens, *Candida albicans* is the most general fungi, which leads to candidemia (Antinori et al., 2016; Lee et al., 2020). To deal with the serious effects of life-threatening fungal infections, the development of antifungal agents has become a widespread concern (Lepesheva et al., 2018). In fungal cell membrane, the sterol 14α demethylase enzyme (CYP51) is responsible for catalyzing the lanosterol 14α methylation to produce ergosterol which can regulate the integrity, fluidity and permeability of the cell membrane (Balding et al., 2008; Lass-Floerl, 2011; Ene et al., 2012; Hargrove et al., 2016). Thus, influencing the growth and replication of fungi by inhibiting CYP51 has become a strategy for the development of antifungal agents (Lepesheva et al., 2008; Choi and Roush, 2017).



GRAPHICAL ABSTRACT

As we all know, there are four clinical triazoles agents targeting CYP51 for the treatment of systemic fungal infections: fluconazole (Flu), itraconazole (Itc), voriconazole (Vor), and posaconazole (Pos) (Lepesheva et al., 2018; Lee et al., 2020). The structures of four inhibitors are shown in **Figure 1**, they can be divided into short-tailed agents (ST: Flu and Vor) and long-tailed agents (LT: Itc and Pos) (Keniya et al., 2018). As the drug resistance mutations in CYP51 of *Candida albicans* increased, the effectiveness of four existing inhibitors is limited (Warrilow et al., 2019; Nishimoto et al., 2020). Therefore, elaborating the molecular mechanism of the existing drugs is very positive for further design and development of new drugs.

The successful resolution of the crystal structure of *Candida albicans* CYP51 by Lepesheva group inspired us to explore the molecular mechanisms between inhibitors and CYP51 (Hargrove et al., 2017). Molecular dynamics (MD) simulations are widely used to study the molecular mechanisms of inhibitors and enzyme (Watson et al., 2017; Sun et al., 2018; Xiao et al., 2020). Thus, we employed MD simulations and molecular docking to explore the binding mechanism between inhibitors and CYP51 in the present study. The results might offer insights into the structure–function relationships of CYP51 and provide the molecular basis for the rational design of new azoles inhibitors.

MATERIALS AND METHODS

Preparation of Molecular Systems

The three-dimensional structure of *Candida albicans* CYP51 enzyme was obtained from the Protein Data Bank (PDB code: 5FSA) (Hargrove et al., 2017). Moreover, the structures of the four inhibitors: Flu (Compound CID: 3365), Vor (Compound CID: 71616), Itc (Compound CID: 55283), and Pos (Compound CID: 468595) were obtained from the PubChem database. After removing the ligand from the complex, the apo protein (**Supplementary Figure S1**) was saved by the Discovery Studio 3.1 (Studio, 2011). The CDOCKER protocol of Discovery Studio 3.1 (Studio, 2011) was employed to build the four complex structures, including Flu-CYP51 complex, Vor-CYP51 complex, Itc-CYP51 complex, and Pos-CYP51 complex.

Molecular Docking

CDOCKER (Studio, 2011) is a grid-based molecular docking method by CHARMm force field. The geometry optimization of four ligands was performed by Gaussian 09 (Frisch et al., 2009) at the B3LYP/6-31G (d) level. The CHARMm force field was applied to ligands and receptors. The receptor was held rigid, whereas ligands were allowed to flex during the refinement. The

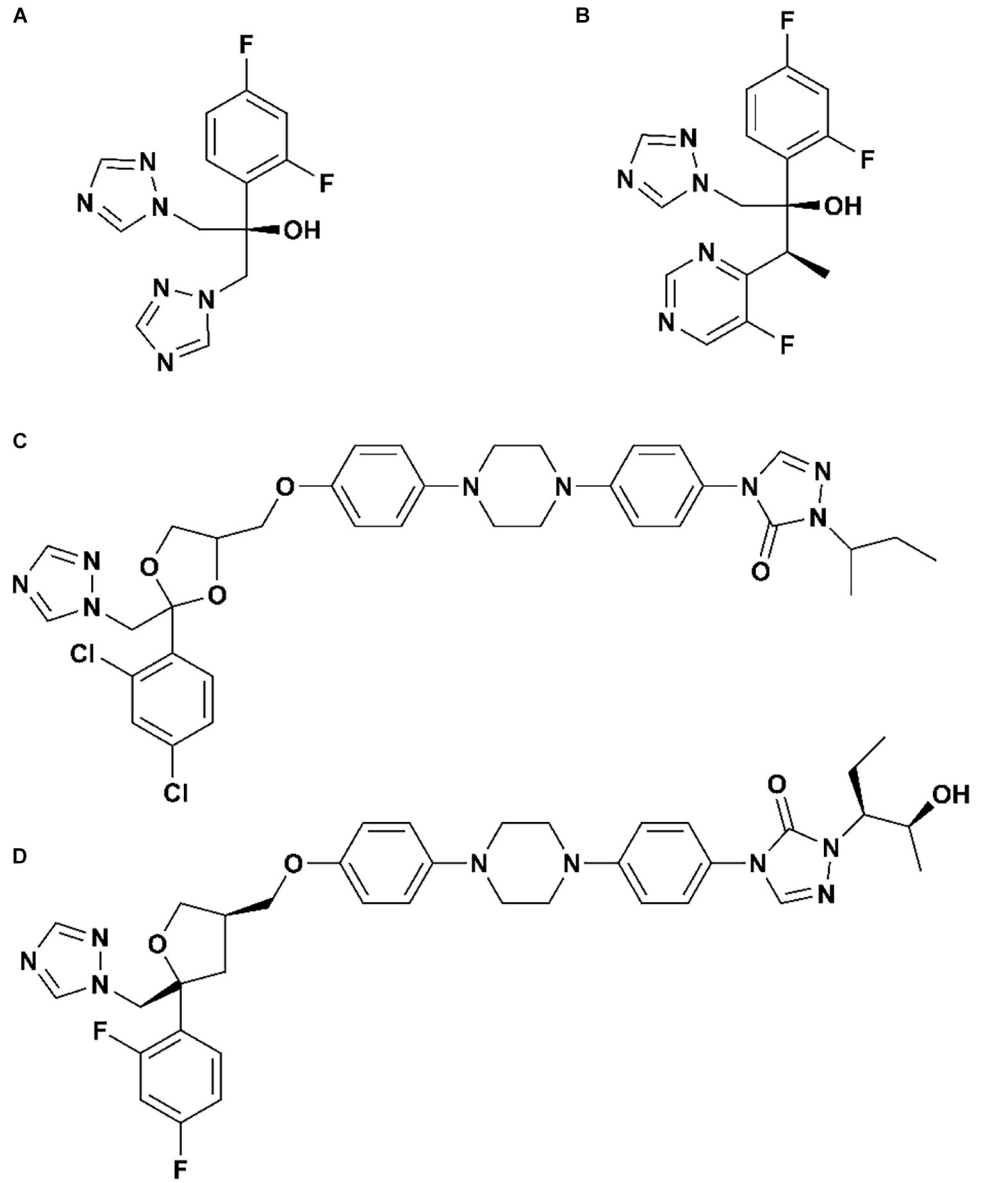


FIGURE 1 | The structures of **(A)** fluconazole, **(B)** voriconazole, **(C)** itraconazole, and **(D)** posaconazole.

input site sphere was defined as a radius of 12 and 18 Å for short-tailed inhibitors (Flu and Vor) and long-tailed inhibitors (Itc and Pos), respectively. The other docking information was set as the default value. A conformational search of the ligands was performed by a grid-based simulated annealing method. The ligands were firstly heated to 700 K (2000 steps) and then annealed to 300 K (5000 steps). The value of the grid extension was set as 8 Å. The top 20 poses of each complex were saved for comparison and analysis. Finally, combined with the information of binding site in the literature (Nair et al., 2016; Hargrove et al., 2017) and docking score, the bested conformation of each system was chosen as the initial structure for the subsequent MD simulations.

Molecular Dynamics Simulations

For each system, geometry optimization of four inhibitors was performed by Gaussian 09 (Frisch et al., 2009) with the *ab initio* calculation method at the B3LYP/6-31G (d) level (Lee et al., 1988; Andersson and Uvdal, 2005). For charge derivation, the restrained electrostatic potential (RESP) fitting procedure was used (Bayly et al., 1993). The force field parameters of inhibitors were supplied by the general AMBER force field (GAFF) in the Antechamber module of AMBER 16 package (Wang et al., 2004, 2006; Case et al., 2016). The force field parameters developed by Shahrokh et al. were assigned to heme (Shahrokh et al., 2012). Finally, all missing atoms and hydrogen atoms were added using the t-leap procedure in the AMBER

16 package (Case et al., 2016). MD simulations were performed by the AMBER 16 package using the ff14SB force field (Maier et al., 2015; Case et al., 2016). To ensure the overall neutrality of the systems, counterions were added. All systems were subjected to MD simulations in explicit solvent, and all systems were solvated with TIP3P water box with 10 Å between the solute boundary (Yoo and Xantheas, 2011). First, protein and inhibitor were fixed with a 500 kcal/mol/Å^2, and minimized the energy of all water molecules and counterions for 10000 steps of steepest descent (SD) followed by 10000 steps of the conjugate gradient. Subsequently, to remove conflicting contacts, the entire system was repeated for 12000 steps of SD minimization and 8000 steps of CG minimization. Next, the system was gradually heated up to 310 K in the NVT ensemble, thereby applying harmonic restraints with a force constant of 10.0 kcal/mol/Å^2 on the solute atoms, and equilibration was performed three times with 3000 ps using a force constant of 5.0 kcal/mol/Å^2 (Loncharich et al., 1992). Finally, 200 ns MD simulations were performed using the NPT ensemble without restraints. We used the Particle mesh Ewald (Darden et al., 1993) technique with a non-bonded cutoff of 12.0 Å to limit the direct space sum to treat the long-range electrostatic interactions. The SHAKE (Krautler et al., 2001) algorithm was used to constrain bonds involving hydrogen atoms. The time step of MD simulation was set to 2 fs, and sampling was performed every 10 ps into the MD file. Cluster analysis was performed by employing average linkage as the clustering algorithm (Shao et al., 2007). All figures in this contribution were generated by PyMOL (DeLano, 2002). The hydrogen bonds and hydrophobic interactions between the inhibitors and CYP51 were studied using LigPlot + v.2.2 (Laskowski and Swindells, 2011).

Free Energy Calculations

The Molecular Mechanics Generalized Born Surface Area (MM-GB/SA) method implemented in AMBER 16 package was performed to calculate the binding free energy (Sun et al., 2014; Case et al., 2016), as well as the free energy decomposition of the four inhibitors systems. For each system, 2000 snapshots were selected from the last 100 ns MD trajectories to calculate the relevant energies. The binding free energy (ΔG_{bind}) in MM-GB/SA between enzyme and ligand was summarized as the follows:

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand}) \quad (1)$$

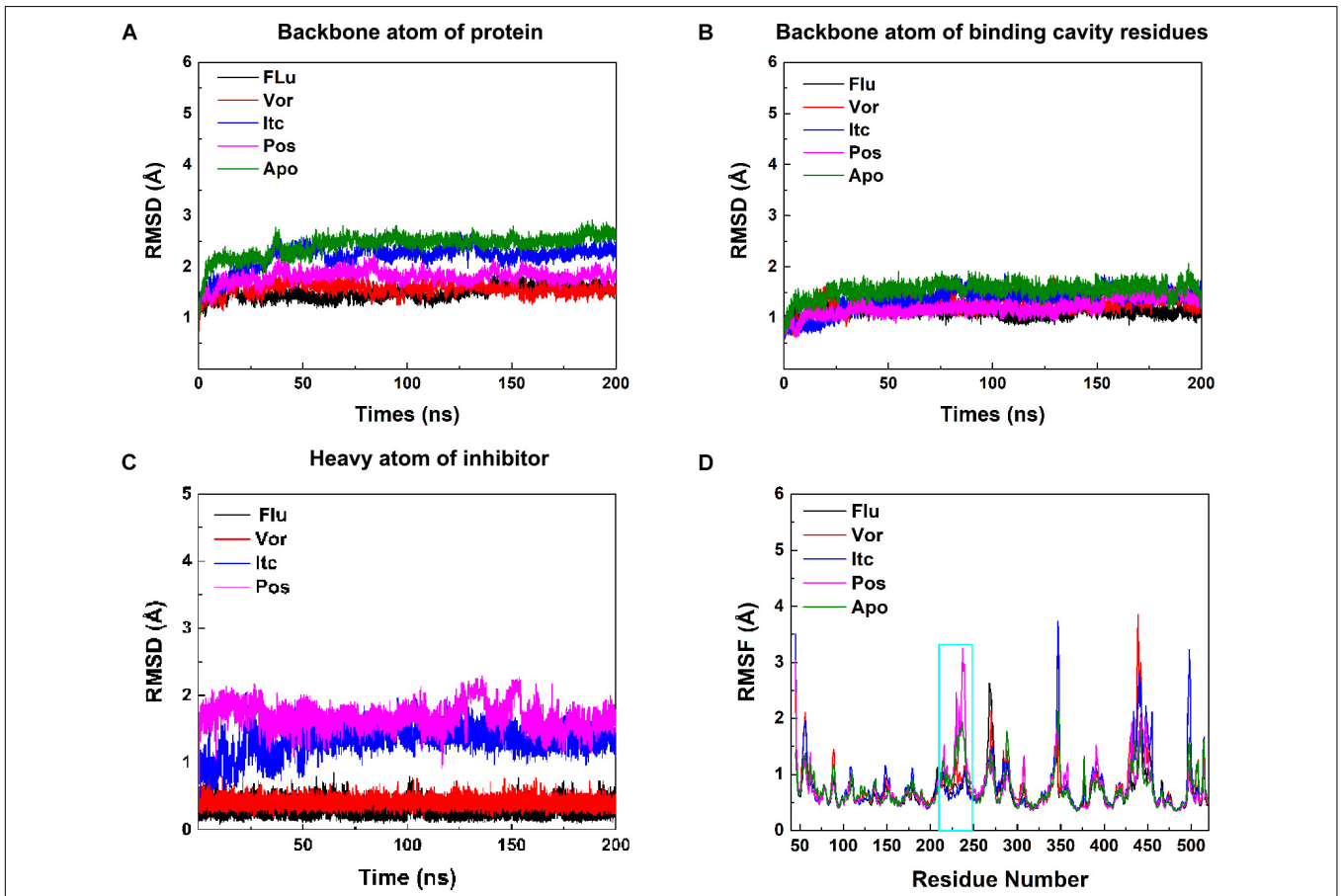


FIGURE 2 | (A) RMSDs of backbone atoms of protein, **(B)** backbone atoms of binding cavity residues, and **(C)** heavy atoms of the inhibitor as a function of time. Black, red, blue, pink, and green for Flu, Vor, Itc, Pos, and Apo system, respectively. **(D)** RMSF values of the backbone Cα atom of each residue. The region of F-F′′ loop, F′′ helix, and F′′-G loop is shown in cyan rectangle.

$$\Delta G = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (2)$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW} \quad (3)$$

$$\Delta G_{sol} = \Delta G_{GB} + \Delta G_{SA} \quad (4)$$

In equation (2), ΔE_{MM}, ΔG_{sol}, and TS represent molecular mechanics components in the gas phase, the stabilization energy due to solvation, and a vibrational entropy term, respectively. ΔE_{MM} represents the summation of ΔE_{int}, ΔE_{ele}, and ΔE_{vdW} which are the internal, coulomb, and van der Waals interaction terms, respectively. ΔG_{sol} represents the solvation energy, which is divided into the electrostatic solvation free energy (ΔG_{GB}) (Hawkins et al., 1996) and the non-polar solvation free energy (ΔG_{SA}). ΔG_{GB} can be obtained by using the generalized Born method, and ΔG_{SA} is calculated as follows:

$$\Delta G_{SA} = \gamma SASA + \beta \quad (5)$$

Here, γ and β, two empirical constants, were set as 0.0072 and 0.00 kcal/mol/Å^2, respectively, and *SASA* (Weiser et al., 1999) represents the solvent accessible surface area determined by a probe radius of 1.4 Å. To estimate the change in conformational entropy ($T\Delta S$) for all atoms, the normal-mode analysis was performed using the nmode module of AMBER 16 package (Case et al., 2016). 100 snapshots from the last 100 ns MD trajectories were used to calculate the entropic contribution.

Energy decompositions were performed to identify the important residues within the systems. Here, only per-residue decomposition was included, which was used to separate the energy contribution of each residue from the combination of enzyme with the inhibitor into three terms: van der Waals contribution (ΔE_{vdW}), electrostatic contribution (ΔE_{ele}), and solvation contribution ($\Delta G_{GB} + \Delta G_{SA}$).

Analysis of Access Tunnels

CAVER (Chovancova et al., 2012) is a famous software to explore routes leading from buried cavities (active sites) to enzyme surfaces. The starting point for the tunnel search was located in the position between heme and the inhibitor. The CAVER algorithm (Petrek et al., 2006) divides three-dimensional space into a grid and calculations are based on grid points. During calculations, the probe radius and the clustering threshold were set to 0.8 and 4.5 Å, respectively. A total of 200 frames of each system were extracted from the last 100 ns MD simulations trajectories. Other parameters were maintained at their default settings. Subsequently, tunnels were visualized by using PyMOL (DeLano, 2002).

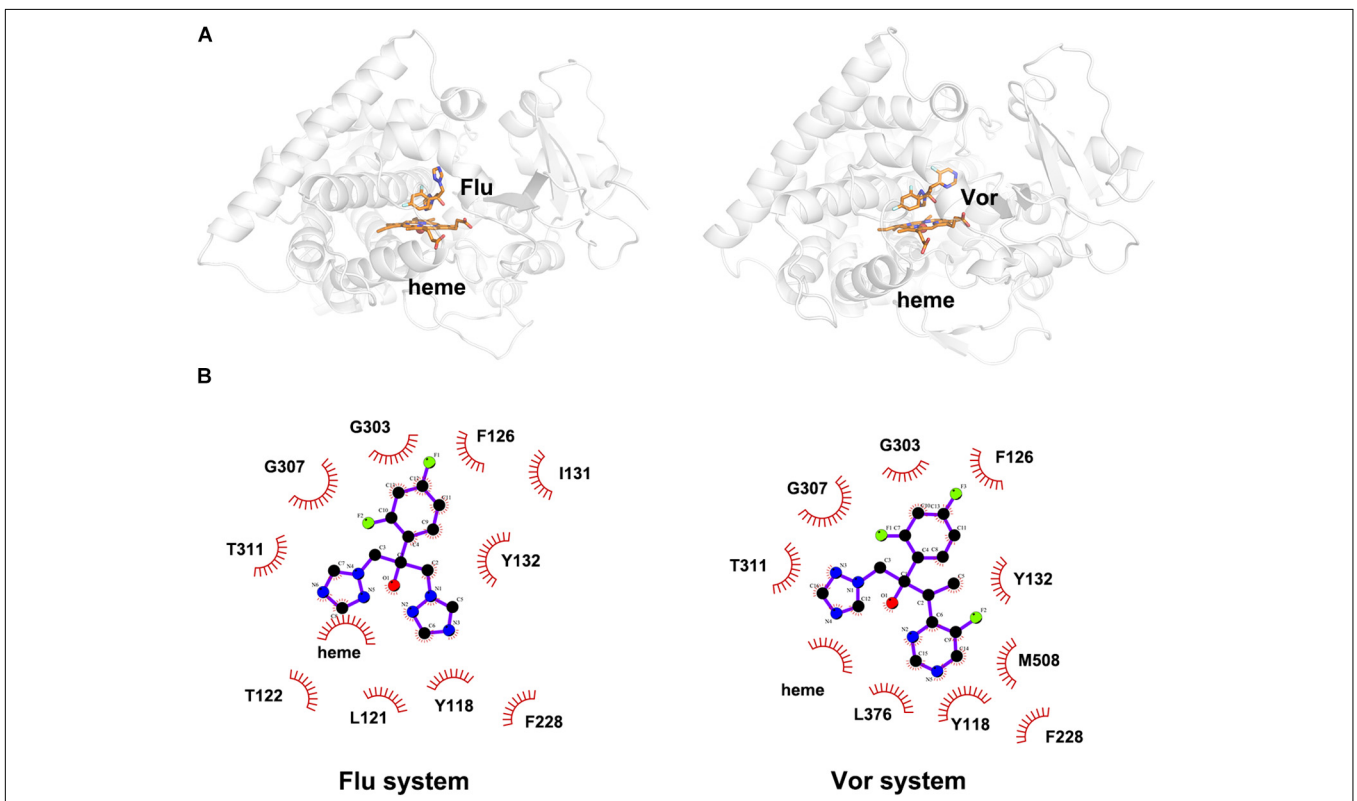


FIGURE 3 | (A) The representative structures of Flu system and Vor system. **(B)** The 2D diagrams of the detailed binding information of Flu system and Vor system. The protein is shown in white cartoon, the inhibitor and heme are displayed in orange sticks. The molecular interactions show hydrophobic interactions as semi-arcs with red eyelashes.

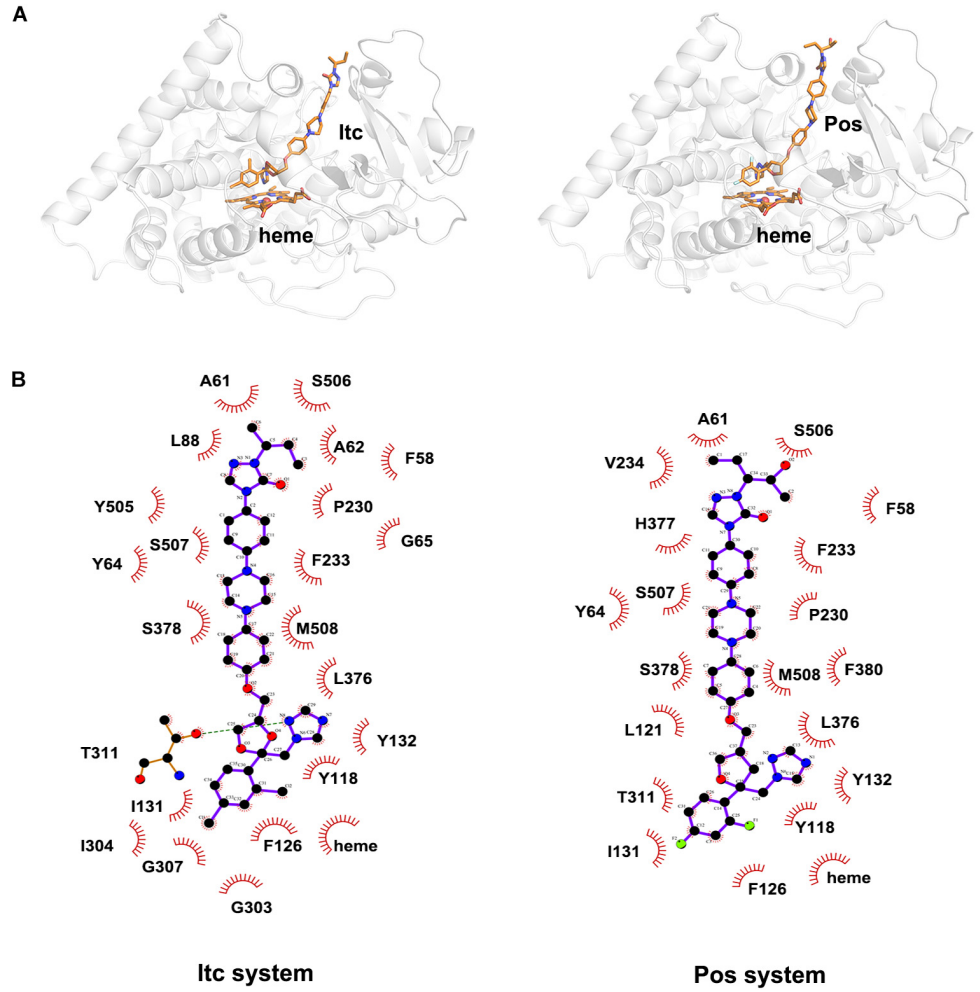


FIGURE 4 | (A) The representative structures of Itc system and Pos system. **(B)** The 2D diagrams of the detailed binding information of Itc system and Pos system. The protein is shown in white cartoon, the inhibitor and heme are displayed in orange sticks. The molecular interactions show hydrogen bonds as green dashed lines and hydrophobic interactions as semi-arcs with red eyelashes.

RESULTS AND DISCUSSION

Determination of the Optimal Binding Pose of Inhibitor by Docking Analysis

The CDOCKER protocol (Studio, 2011) is a CHARMm-based docking method, which was carried out to obtain an optimum initial model of the complex. To determine the reliability of this docking method, co-crystallized ligand (Pos) was firstly re-docked into defined cavity with the CDOCKER protocol. It has been reported that the distance between the nitrogen atoms on the triazole ring of ligands and the iron atom of heme ($N6_{Flu}$-Fe, $N1_{Vor}$-Fe, $N4_{Pos}$-Fe; $N7_{Itc}$-Fe) is less than 5 Å (Nair et al., 2016; Hargrove et al., 2017). The optimal conformation was the one with the best score among the structures that satisfy the above distance condition. The root-mean-square deviation (RMSD) value between the docking and initial conformation of Pos was 1.18 Å, which suggested that the CDOCKER protocol was suitable for docking in this work. The ligands Flu, Vor, and Itc were successively docked into the receptor, and the optimal binding pose was selected for the further MD analyses according to the above criteria. For the sake of clarity, Flu-CYP51 complex, Vor-CYP51 complex, Itc-CYP51 complex, and Pos-CYP51 complex was referred to as Flu system, Vor system, Itc system, and Pos system, respectively."

The Structural Stability and Dynamics Properties of the Inhibitor-CYP51 Systems

In the 200 ns MD simulations of five systems, the root-mean-square deviation (RMSD) value of backbones atom of protein, binding cavity residues, and heavy atom of inhibitors were calculated to investigate the structural stability of CYP51. As shown in **Figures 2A–C**, each system gradually reached equilibrium, which remained quite stable during the last 100 ns. Thus, all subsequent analysis was performed on the last 100 ns of the simulation trajectories. Comprehensively considered these

TABLE 1 | Binding free energies (kcal/mol) and its components of four inhibitors systems.

	Flu	Vor	Itc	Pos
ΔE_{vdW}	−42.98	−49.12	−90.33	−91.32
ΔE_{ele}	−6.48	−1.31	−7.00	−12.81
ΔG_{GB}	25.78	21.62	23.53	35.44
ΔG_{SA}	−4.99	−5.71	−9.76	−10.02
ΔE_{polar} [a]	19.30	20.31	16.53	22.63
$\Delta E_{non-polar}$ [b]	−47.97	−54.83	−100.09	−101.34
ΔG_{mmgbsa} [c]	−28.67	−34.51	−83.56	−78.71
$T\Delta S$	−20.17	−19.43	−31.04	−30.18
ΔG_{bind} [d]	−8.50	−15.08	−52.52	−48.53

[a] $\Delta E_{polar} = \Delta E_{ele} + \Delta G_{GB}$. [b] $\Delta E_{non-polar} = \Delta E_{vdW} + \Delta G_{SA}$. [c] $\Delta G_{mmgbsa} = \Delta E_{vdW} + \Delta E_{ele} + \Delta G_{GB} + \Delta G_{SA}$. [d] $\Delta G_{bind} = \Delta G_{mmgbsa} - T\Delta S$.

TABLE 2 | Decomposition of binding free energy (kcal/mol) on per residue basis for Flu system, Vor system, Pos system, and Itc system.

Residue	Flu	Vor	Itc	Pos
Y118	−1.84	−2.90	−1.82	−2.03
F126	−0.92	−0.53	−1.05	−0.71
Y132	−1.39	−0.82	−0.71	−0.85
F228	−1.02	−0.74	−0.88	−0.24
T311	−0.27	−0.99	−2.78	−0.98
L376	−1.91	−2.19	−1.90	−2.11
M508	−0.76	−2.40	−3.63	−3.41
F380	−0.23	−0.74	−1.02	−1.09
A61	0.01	0.01	−0.70	−1.37
Y64	0.01	−0.02	−1.07	−1.40
L87	0.01	0.04	−1.19	−1.08
L88	0.01	0.02	−1.08	−1.33
P230	0.01	−0.02	−2.38	−1.25
F233	−0.06	−0.24	−1.04	−1.47
H310	−0.74	−0.62	−2.16	−0.54
S506	0.02	0.06	−1.21	−1.36
S507	−0.05	−0.07	−1.00	−1.01

RMSD values of systems, the binding of the inhibitors reduced the perturbation of the protein to some extent. To further explore the effect of inhibitors binding on fluctuations of a certain residue, the root-mean-square-fluctuation (RMSF) of backbones Cα atoms in CYP51 was calculated (**Figure 2D**). The comparison of RMSF between the Apo system and four inhibitors systems showed that the fluctuations were mostly similar except for the structural elements of F-F″ loop, F″ helix, and F″-G loop (**Supplementary Figure S1**). This region is also called F-G loop in P450 enzyme, which may affect the channel characteristics of the enzyme (Cojocaru et al., 2007). The F- F″ loop, F″ helix, and F″-G loop showed large RMSF values in Apo system (**Figure 2D**). Compared with Apo system, the RMSF values of Flu, Vor, and Itc systems were reduced, while that of Pos system was increased. Further analysis showed that due to the 2-hydroxypentan of Pos is close to the F-G loop, the instability of 2-hydroxypentan (**Supplementary Figure S2**) caused the F-G loop to change greatly, thus increasing the RMSF value of the F-G loop. These results indicated that inhibitors binding may affect protein tunnel characteristics by influencing the conformations of F- F″ loop, F″ helix, and F″-G loop. These results suggested that inhibitors' binding increased the stability of CYP51. Local conformational changes of F-G loop caused by inhibitors binding may affect the protein tunnel characteristics.

Analysis Inhibitor Binding Mode

To explore the binding mode of inhibitor in the binding cavity of CYP51, clustering analysis was used to extract the representative conformation. As shown in **Figures 3A**, **4A**, the shared triazole ring of four inhibitors located above the heme was coordinated with the heme. The halogenated phenyl was pointed toward to the crack between the I helix and the B′-C loop. The rest side chain oriented toward the entrance of the binding cavity of CYP51. These results indicated the four inhibitors maintained a similar binding pattern. The 2D diagrams displayed the interaction between inhibitor and protein. As shown in **Figures 3B**, **4B**, the hydrophobic interaction was the main driving force for inhibitors binding to CYP51. Based on molecular shapes and scaffolds, the four inhibitors were divided into two kinds of inhibitors: ST inhibitors and LT inhibitors (Keniya et al., 2018). ST inhibitors (Flu and Vor) formed hydrophobic interaction with the shared residues Y118, F126, Y132, F228, G303, G307, and T311 (**Figure 3B**). For LT inhibitors (Itc and Pos), they formed hydrophobic interaction with the common residues F58, Y64, Y118, L121, Y132, L376, S378, S506, S507, and M508 (**Figure 4B**). The important and obvious discrepancy of two kinds of inhibitors was the length of the side chain. Comparing with ST inhibitors, the extended side chains of LT inhibitors (Itc and Pos) provided additional points in contact with the azole target CYP51. Thus, LT inhibitors can form more hydrophobic interactions with CYP51 than ST inhibitors, which may demonstrate LT inhibitors have stronger binding affinities with CYP51.

The analysis of inhibitor binding mode suggested four inhibitors hold a similar binding pattern and the hydrophobic interactions were the dominant driving force for inhibitors' binding to CYP51. For both types of inhibitors, LT inhibitors can form hydrophobic interactions with more residues due to the characteristics of their long side chains. This may indicate that LT types of inhibitors are more suitable for targeted CYP51.

Rational Ranking of Binding Ability by Binding Free Energy Calculations

To gain energic information about the four inhibitors systems, the binding free energy calculations were performed by the MM-GB/SA method, and the entropy contributions were also considered. As presented in **Table 1**, the total ΔG_{bind} values of Flu, Vor, Itc, and Pos systems were −8.50, −15.08, −52.52, and −48.53 kcal/mol, respectively. LT inhibitors (Pos and Itc) had stronger binding affinities when compared with ST inhibitors (Flu and Vor), which confirmed our previous speculation. Our results showed the rational ordering of binding affinities in different systems which were consistent with that of the experimental inhibitory effects (Hargrove et al., 2017). Further analysis of the binding free energy

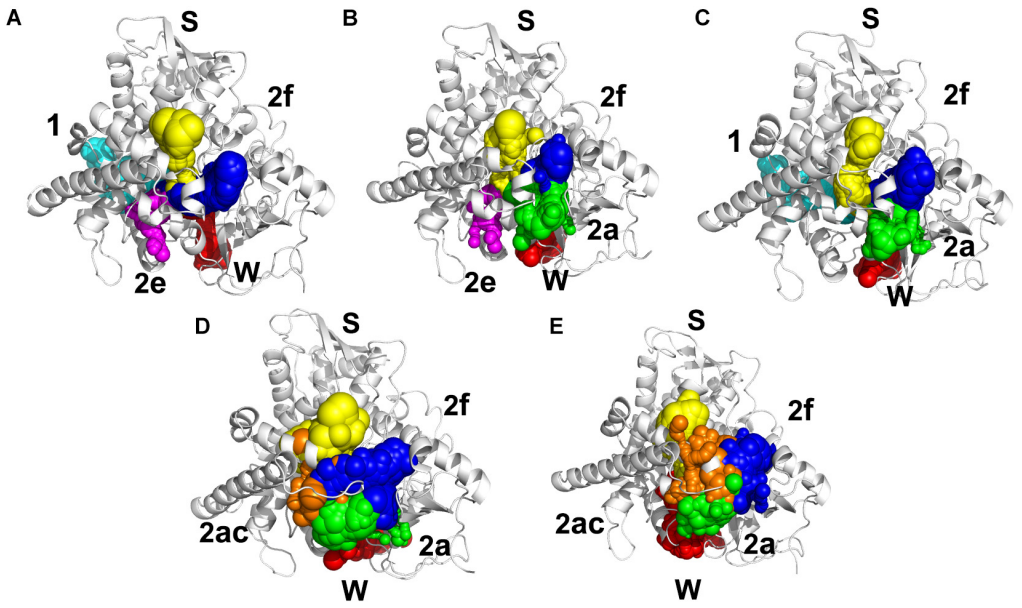


FIGURE 5 | Access tunnel identified from the average structures of **(A)** Flu system, **(B)** Vor system, **(C)** Itc system, **(D)** Pos system, and **(E)** Apo system, respectively. The protein backbone is shown in cartoon. Tunnels are shown in sphere for blue (2f), green (2a), magenta (2e), orange (2ac), yellow (s), red (w), and cyan (1).

indicated that the contributions of van der Waals interactions (ΔE_{vdW}), electrostatic energy (ΔE_{ele}), and non-polar solvation energy ($\Delta G_{non-polar}$) were favorable for the formation of the inhibitors' complexes. As listed in **Table 1**, non-polar interaction ($\Delta G_{non-polar}$) was mainly responsible for the formation of the Flu system (−47.97 kcal/mol), Vor system (−54.83 kcal/mol), Pos system (−101.34 kcal/mol), and Itc system (−100.09 kcal/mol). In comparison to the non-polar interaction, polar interaction of the four systems had an unfavorable contribution. The entropy change values of the four systems were less than zero, which were detrimental to the binding of the inhibitors to the CYP51. The results of free energy analysis elaborated that non-polar interaction was the key factor for the binding of inhibitors and CYP51.

The total binding free energy was decomposed into residues to identify key residues for inhibitors binding to CYP51. Essential residues with the binding free energy value below −1.0 kcal/mol were listed in **Table 2**. The number of residues meeting to the criterion were 4, 3, 14, and 12 in four systems, respectively, which also indicated that the LT inhibitors were tightly bound to CYP51. We found that Y118 and L376 had significant contributions in inhibitors binding of all four inhibitors systems, Y64, L87, L88, P230, F233, F380, and M508 made outstanding contributions during the LT inhibitors binding.

Tunnel Analysis

Illustrating the tunnel characteristic of CYP51 is beneficial to develop new inhibitors and understand the structure–function relationships of CYP51 (Yu et al., 2016; Fischer et al., 2018). In this work, 200 frames were extracted from the last 100 ns trajectories to classify and analyze the characteristics of access pathways in four inhibitors systems and Apo system. The tunnels were clustered by the average-link algorithm according to the pairwise distances of tunnels. On the basis of spatial and secondary structure, the nomenclature of these tunnels is defined systematically by Wade group (Cojocaru et al., 2007). The five highest ranked tunnels of five systems were all displayed in **Figure 5**, and the characteristics of these tunnels were summarized in **Table 3**. As shown in **Figure 5**, the locations of five tunnels were marked with different color spheres (Flu system: 2f, W, S, 1, and 2e; Vor system: 2f, 2a, W, S, and 2e; Itc system: 2f, S, W, 1, and 2a; Pos system: 2f, 2a, W, S, and 2ac; Apo system: 2f, S, W, 2a, and 2ac). Tunnels 2f, 2a, 2e, and 2ac are subclasses of tunnel 2. Tunnel 2f locates between the F-G loop, Pro-rich loop, and A helix, whereas tunnel 2a locates between the F-G loop and B′-C loop (**Supplementary Figure S1**). Tunnel 2e egresses through the B′-C loop, and tunnel 2ac egresses between the B′ helix and the G helix (**Supplementary Figure S1**). Tunnel W (water tunnel) egresses at the base of the B′-C loop near the C-terminus of the B helix, and tunnel S (solvent tunnel) runs between the E, F, and I helices and β5 sheet (**Supplementary Figure S1**). As listed in **Table 3**, tunnel 2f was the most frequently identified collective pathway and had the highest bottlenecks radius in five systems. Thus, tunnel 2f was regarded as the predominant tunnel for inhibitors ingress/egress from the active site to the surface of CYP51, which was similar as the other works of CYP51 (Monk et al., 2014; Yu et al., 2016; Gao et al., 2018). The occurrence of tunnel 2f was different slightly in five systems, which may be related to the size of the inhibitors. In the LT systems (Itc and Pos), the inhibitor is long and bulky, and its binding mode analysis showed that its long side chain extended to the entrance of the tunnel 2f, resulting in a fully opened tunnel 2f. In the ST systems (Flu and Vor), the inhibitor is in small size, and the inhibitor was submerged in the binding cavity of CYP51,

TABLE 3 | Characteristics of the five top ranked tunnels of Flu system (A), Vor system (B), Itc system (C), Pos system (D), and Apo system, respectively.

(A) Characteristics of Flu system

Tunnel	2f	W	S	1	2e
Occurrence	93%	82%	34%	25%	11%
Mean bottleneck radius [Å]	1.26	0.99	0.96	0.91	0.84
Max. bottleneck radius [Å]	1.74	1.26	1.45	1.19	0.97
Mean pathway length [Å]	23.86	23.50	21.14	29.70	21.40

(B) Characteristics of Vor system

Tunnel	2f	2a	W	S	2e
Occurrence	98%	84%	77%	27%	15%
Mean bottleneck radius [Å]	1.26	1.08	0.92	0.97	0.83
Max. bottleneck radius [Å]	2.08	1.83	1.18	1.47	0.93
Mean pathway length [Å]	25.88	28.70	24.08	21.80	2

(C) Characteristics of Itc system

Tunnel	2f	S	W	1	2a
Occurrence	100%	89%	85%	51%	34%
Mean bottleneck radius [Å]	2.00	1.09	0.95	0.93	0.93
Max. bottleneck radius [Å]	2.46	1.56	1.30	1.31	1.34
Mean pathway length [Å]	16.95	20.23	23.69	33.58	20.95

(D) Characteristics of Pos system

Tunnel	2f	2a	W	S	2ac
Occurrence	100%	86%	81%	38%	16%
Mean bottleneck radius [Å]	1.86	1.78	0.93	1.03	1.00
Max. bottleneck radius [Å]	2.25	2.39	1.23	1.89	1.69
Mean pathway length [Å]	24.65	24.20	22.24	19.56	24.60

(E) Characteristics of Apo system

Tunnel	2f	S	W	2a	2ac
Occurrence	88%	80%	69%	65%	38%
Mean bottleneck radius [Å]	1.69	1.34	0.91	1.44	1.40
Max. bottleneck radius [Å]	2.37	1.72	1.40	2.03	2.10
Mean pathway length [Å]	20.92	19.93	23.31	22.53	22.82

causing decreased slightly in the opening frequency of tunnel 2f, which was in line with the RMSF analysis that inhibitors' binding affected the tunnel characteristic. Further, we determined the essential residues lining the dominant tunnel 2f, and all residues located within the 3 Å distance from the tunnel surface will be regarded as tunnel-lining residues (**Supplementary Table S1**). The key residues (F58, Y64, Y118, L121, Y132, L376, S378, S506, S507, and M508) determined based on the binding mode analysis and per-residue binding free energy decomposition analysis also belong to the tunnel-lining residues. Most of the tunnel-lining residues were hydrophobic residues, which formed a stable hydrophobic cavity and provided hydrophobic interactions that play an indispensable role in inhibitor stabilization. Thus, when designing more efficient inhibitors, the interactions between inhibitors and these residues should be rationally increased and the new inhibitors should be hydrophobic ligands.

CONCLUSION

The sterol 14α-demethylase enzyme (CYP51) belongs to cytochrome P450 family essential in sterol biosynthesis, which is the target for fungal infections. In this work, molecular docking and molecular dynamics simulations were employed to investigate the binding mechanism and tunneling characteristics between four inhibitors and CYP51, so as to provide useful information for inhibitors design. The results show that four inhibitors bind CYP51 in similar binding mode and hydrophobic interactions are the main driving force for inhibitors binding to CYP51. Due to long-tailed inhibitors (posaconazole and itraconazole) can contact with more residues of CYP51 by hydrophobic interactions than short-tailed inhibitors (fluconazole, voriconazole), long-tailed inhibitors have stronger binding affinities. Tunnel analysis showed that tunnel 2f is the predominant pathway for inhibitors ingress/egress from the active site to the surface of CYP51. We discover a hydrophobic cavity and identify the key residues (F58, Y64, Y118, L121, Y132, L376, S378, S506, S507, and M508) which are responsible for anchoring the inhibitors binding to CYP51. Therefore, in order to enhance the binding affinity of inhibitors to CYP51, we should focus on strengthening hydrophobic interactions of inhibitors and these residues, while longer inhibitors are probably best suited to target CYP51. Taken together, the results obtained in this study will be valuable for designing potent azoles inhibitors and improve the understanding of the structure–function relationships of CYP51.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

NS performed the research, analyzed the data, and wrote the manuscript. QZ and HZ designed the research and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 21273095) and the Fundamental Research Funds for the Central Universities, JLU.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.586540/full#supplementary-material

REFERENCES

Andersson, M. P., and Uvdal, P. (2005). New scale factors for harmonic vibrational frequencies using the B3LYP density functional method with the triple-xi basis set 6-311+G(d,p). *J. Phys. Chem. A* 109, 2937–2941. doi: 10.1021/jp045733a

Antinori, S., Milazzo, L., Sollima, S., Galli, M., and Corbellino, M. (2016). Candidemia and invasive candidiasis in adults: a narrative review. *Eur. J. Intern. Med.* 34, 21–28. doi: 10.1016/j.ejim.2016.06.029

Balding, P. R., Porro, C. S., McLean, K. J., Sutcliffe, M. J., Marechal, J. D., Munro, A. W., et al. (2008). How do azoles inhibit cytochrome P450 enzymes? A density functional study. *J. Phys. Chem. A* 112, 12911–12918. doi: 10.1021/jp802087w

Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97, 10269–10280. doi: 10.1021/j100142a004

Bongomin, F., Gago, S., Oladele, R. O., and Denning, D. W. (2017). Global and multi-national prevalence of fungal diseases-estimate precision. *J. Fungi* 3:57. doi: 10.3390/jof3040057

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden killers: human fungal infections. *Sci. Transl. Med.* 4:165rv13. doi: 10.1126/scitranslmed.3004404

Case, D., Betz, R., Cerutti, D. S., Cheatham, T., Darden, T., Duke, R., et al. (2016). *Amber 2016*. San Francisco, CA: University of California.

Choi, J. Y., and Roush, W. R. (2017). Structure based design of CYP51 inhibitors. *Curr. Top. Med. Chem.* 17, 30–39. doi: 10.2174/1568026616666161607191 64933

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., et al. (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* 8:e1002708. doi: 10.1371/journal.pcbi.1002708

Cojocaru, V., Winn, P. J., and Wade, R. C. (2007). The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta Gen. Subj.* 1770, 390–401. doi: 10.1016/j.bbagen.2006.07.005

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an N · log (N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397

DeLano, W. (2002). *PyMOL Molecular Graphics System*. San Carlos, CA: DeLano Scientific.

Ene, I. V., Adya, A. K., Wehmeier, S., Brand, A. C., MacCallum, D. M., Gow, N. A. R., et al. (2012). Host carbon sources modulate cell wall architecture, drug resistance and virulence in a fungal pathogen. *Cell. Microbiol.* 14, 1319–1335. doi: 10.1111/j.1462-5822.2012.01813.x

Fischer, A., Don, C. G., and Smiesko, M. (2018). Molecular dynamics simulations reveal structural differences among allelic variants of membrane-anchored cytochrome P450. *J. Chem. Inform. Model.* 58, 1962–1975. doi: 10.1021/acs.jcim.8b00080

Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., et al. (2009). *Gaussian 09*. Wallingford, CT: Gaussian Inc.

Gao, P., Cui, Y. L., and Wu, R. L. (2018). Molecular dynamic modeling of CYP51B in complex with azole inhibitors. *J. Biomol. Struct. Dyn.* 36, 1511–1519. doi: 10.1080/07391102.2017.1328315

Hargrove, T. Y., Friggeri, L., Wawrzak, Z., Qi, A., Hoekstra, W. J., Schotzinger, R. J., et al. (2017). Structural analyses of *Candida albicans* sterol 14 alpha-demethylase complexed with azole drugs address the molecular basis of azole-mediated inhibition of fungal sterol biosynthesis. *J. Biol. Chem.* 292, 6728–6743. doi: 10.1074/jbc.M117.778308

Hargrove, T. Y., Friggeri, L., Wawrzak, Z., Sivakumaran, S., Yazlovitskaya, E. M., Hiebert, S. W., et al. (2016). Human sterol 14 alpha-demethylase as a target for anticancer chemotherapy: towards structure-aided drug design. *J. Lipid Res.* 57, 1552–1563. doi: 10.1194/jlr.M069229

Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996). Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* 100, 19824–19839. doi: 10.1021/jp961710n

Keniya, M. V., Sabherwal, M., Wilson, R. K., Woods, M. A., Sagatova, A. A., Tyndall, J. D. A., et al. (2018). Crystal structures of full-length lanosterol 14 alpha-demethylases of prominent fungal pathogens *Candida albicans* and *Candida glabrata* provide tools for antifungal discovery. *Antimicrob. Agents Chemother.* 62:e01134-18. doi: 10.1128/aac.01134-18

Krautler, V., Van Gunsteren, W. F., and Hunenberger, P. H. (2001). A fast SHAKE: algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* 22, 501–508. doi: 10.1002/1096-987x(20010415)22:5<501::aid-jcc1021>3.0.co;2-v

Laskowski, R. A., and Swindells, M. B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inform. Model.* 51, 2778–2786. doi: 10.1021/ci200227u

Lass-Floerl, C. (2011). Triazole antifungal agents in invasive fungal infections a comparative review. *Drugs* 71, 2405–2419. doi: 10.2165/11596540-000000000-00000

Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condensed Matter* 37, 785–789. doi: 10.1103/PhysRevB.37.785

Lee, Y., Puumala, E., Robbins, N., and Cowen, L. E. (2020). Antifungal drug resistance: molecular mechanisms in *Candida albicans* and beyond. *Chem. Rev.* [Epub ahead of print]. doi: 10.1021/acs.chemrev.0c00199

Lepesheva, G. I., Friggeri, L., and Waterman, M. R. (2018). CYP51 as drug targets for fungi and protozoan parasites: past, present and future. *Parasitology* 145, 1820–1836. doi: 10.1017/s0031182018000562

Lepesheva, G. I., Hargrove, T. Y., Kleshchenko, Y., Nes, W. D., Villalta, F., and Waterman, M. R. (2008). CYP51: a major drug target in the cytochrome P450 superfamily. *Lipids* 43, 1117–1125. doi: 10.1007/s11745-008-3225-y

Loncharich, R. J., Brooks, B. R., and Pastor, R. W. (1992). Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* 32, 523–535. doi: 10.1002/bip.360320508

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255

Monk, B. C., Tomasiak, T. M., Keniya, M. V., Huschmann, F. U., Tyndall, J. D. A., O'Connell, J. D. I. I. I., et al. (2014). Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3865–3870. doi: 10.1073/pnas.1324245111

Nair, P. C., McKinnon, R. A., and Miners, J. O. (2016). Cytochrome P450 structure-function: insights from molecular dynamics simulations. *Drug Metab. Rev.* 48, 434–452. doi: 10.1080/03602532.2016.1178771

Nishimoto, A. T., Sharma, C., and Rogers, P. D. (2020). Molecular and genetic basis of azole antifungal resistance in the opportunistic pathogenic fungus *Candida albicans*. *J. Antimicrob. Chemother.* 75, 257–270. doi: 10.1093/jac/dkz400

Petrek, M., Otyepka, M., Banas, P., Kosinova, P., Koca, J., and Damborsky, J. (2006). CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* 7:316. doi: 10.1186/1471-2105-7-316

Shahrokh, K., Orendt, A., Yost, G. S., and Cheatham, T. E. III (2012). Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* 33, 119–133. doi: 10.1002/jcc.21922

Shao, J., Tanner, S. W., Thompson, N., and Cheatham, T. E. III (2007). Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* 3, 2312–2334. doi: 10.1021/ct700119m

Studio, D. (2011). *Version 3.1*. San Diego, CA: Accelrys.

Sun, D. R., Wang, Z. J., Zheng, Q. C., and Zhang, H. X. (2018). Exploring the inhibition mechanism on HIF-2 by inhibitor PT2399 and 0X3 using molecular dynamics simulations. *J. Mol. Recogn.* 31:e2730. doi: 10.1002/jmr.2730

Sun, H., Li, Y., Shen, M., Tian, S., Xu, L., Pan, P., et al. (2014). Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys. Chem. Chem. Phys.* 16, 22035–22045. doi: 10.1039/c4cp03179b

Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247–260. doi: 10.1016/j.jmgm.2005.12.005

Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035

Warrilow, A. G., Nishimoto, A. T., Parker, J. E., Price, C. L., Flowers, S. A., Kelly, D. E., et al. (2019). The evolution of azole resistance in *Candida albicans* sterol 14 alpha-demethylase (CYP51) through incremental amino acid substitutions. *Antimicrob. Agents Chemother.* 63:e02586-18. doi: 10.1128/aac.02586-18

Watson, G. M., Lucas, W. A. H., Gunzburg, M. J., and Wilce, J. A. (2017). Insight into the selectivity of the G7-18NATE inhibitor peptide for the Grb7-SH2 domain target. *Front. Mol. Biosci.* 4:64. doi: 10.3389/fmolb.2017.00064

Weiser, J., Shenkin, P. S., and Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* 20, 217–230. doi: 10.1002/(sici)1096-987x(19990130)20:2<217::aid-jcc4>3.0.co;2-a

Xiao, F., Song, X., Tian, P., Gan, M., Verkhivker, G. M., and Hu, G. (2020). Comparative dynamics and functional mechanisms of the CYP17A1 tunnels regulated by ligand binding. *J. Chem. Inform. Model.* 60, 3632–3647. doi: 10.1021/acs.jcim.0c00447

Yoo, S., and Xantheas, S. S. (2011). Communication: the effect of dispersion corrections on the melting temperature of liquid water. *J. Chem. Phys.* 134:121105. doi: 10.1063/1.3573375

Yu, X., Nandekar, P., Mustafa, G., Cojocaru, V., Lepesheva, G. I., and Wade, R. C. (2016). Ligand tunnels in T-brucei and human CYP51: insights for parasite-specific drug design. *Biochim. Biophys. Acta Gen. Subj.* 1860, 67–78. doi: 10.1016/j.bbagen.2015.10.015

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 25 September 2020
doi: 10.3389/fmolb.2020.575196

Probing the Structural Dynamics of the *Plasmodium falciparum* Tunneling-Fold Enzyme 6-Pyruvoyl Tetrahydropterin Synthase to Reveal Allosteric Drug Targeting Sites

*Afrah Khairallah, Caroline J. Ross and Özlem Tastan Bishop**

Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Grahamstown, South Africa

OPEN ACCESS

Edited by:
Guang Hu,
Soochow University, China

Reviewed by:
Zhongjie Liang,
Soochow University, China
Sophie Sacquin-Mora,
UPR9080 Laboratoire de Biochimie Théorique (LBT), France

***Correspondence:**
Özlem Tastan Bishop
O.TastanBishop@ru.ac.za

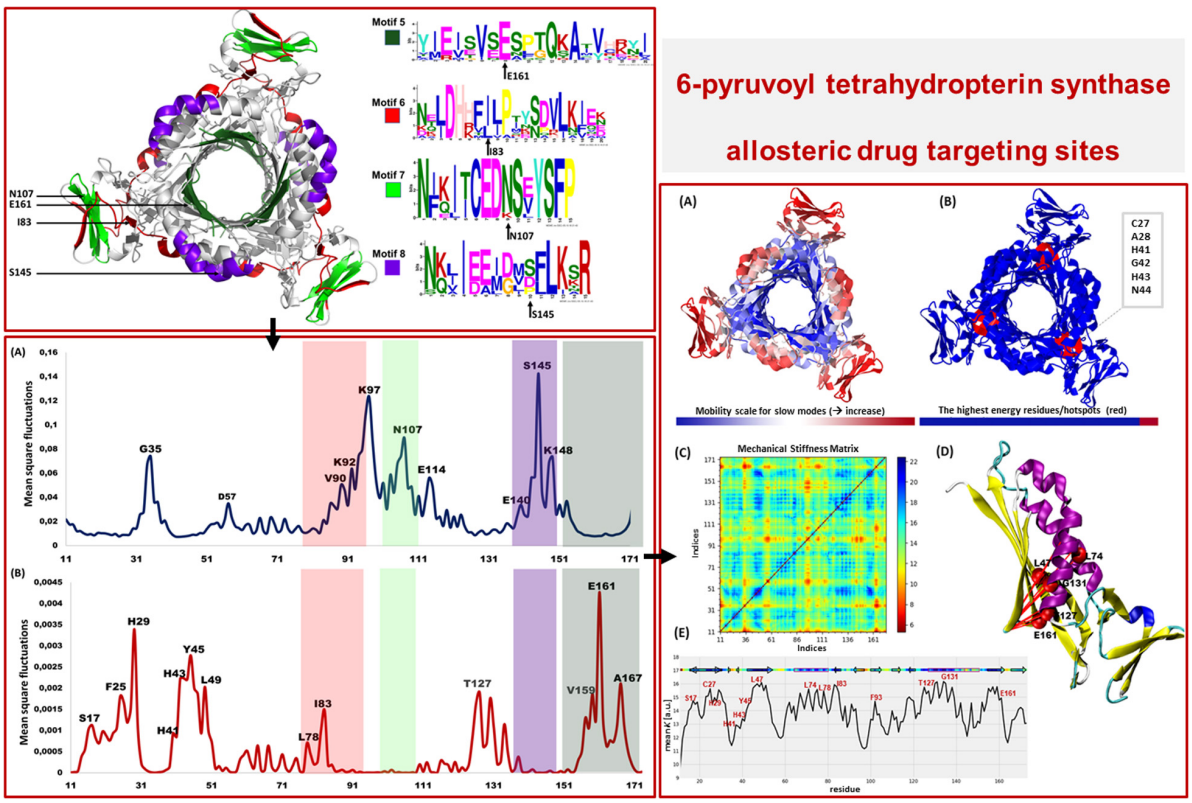
Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *22 June 2020*
Accepted: *20 August 2020*
Published: *25 September 2020*

Citation:
Khairallah A, Ross CJ and Tastan Bishop Ö (2020) Probing the Structural Dynamics of the Plasmodium falciparum Tunneling-Fold Enzyme 6-Pyruvoyl Tetrahydropterin Synthase to Reveal Allosteric Drug Targeting Sites. Front. Mol. Biosci. 7:575196. doi: 10.3389/fmolb.2020.575196

The *de novo* folate synthesis pathway is a well-established drug target in the treatment of many infectious diseases. Antimalarial antifolate drugs have proven to be effective against malaria, however, rapid drug resistance has emerged on the two primary targeted enzymes: dihydrofolate reductase and dihydroptoreate synthase. The need to identify alternative antifolate drugs and novel metabolic targets is of imminent importance. The 6-pyruvol tetrahydropterin synthase (PTPS) enzyme belongs to the tunneling fold protein superfamily which is characterized by a distinct central tunnel/cavity. The enzyme catalyzes the second reaction step of the parasite's *de novo* folate synthesis pathway and is responsible for the conversion of 7,8-dihydroneopterin to 6-pyruvoyl-tetrahydropterin. In this study, we examine the structural dynamics of *Plasmodium falciparum* PTPS using the anisotropic network model, to elucidate the collective motions that drive the function of the enzyme and identify potential sites for allosteric modulation of its binding properties. Based on our modal analysis, we identified key sites in the N-terminal domains and central helices which control the accessibility to the active site. Notably, the N-terminal domains were shown to regulate the open-to-closed transition of the tunnel, via a distinctive wringing motion that deformed the core of the protein. We, further, combined the dynamic analysis with motif discovery which revealed highly conserved motifs that are unique to the Plasmodium species and are located in the N-terminal domains and central helices. This provides essential structural information for the efficient design of drugs such as allosteric modulators that would have high specificity and low toxicity as they do not target the PTPS active site that is highly conserved in humans.

Keywords: *de novo* folate synthesis pathway, malaria, antifolate drugs, structural dynamics, motif discovery, allosteric modulation



GRAPHICAL ABSTRACT |

INTRODUCTION

Despite the major progress achieved to eradicate malaria, the mosquito-borne disease remains a major public health problem (World Health Organization, 2018). Malaria is transmitted through the bite of an infected female Anopheles mosquito. Five species from the genus Plasmodium cause malaria in humans, namely: *Plasmodium falciparum (P. falciparum), P. vivax, P. ovale, P. malariae and P. knowlesi*. Among these five species, *P. falciparum* is the most pathogenic (Snow, 2015).

Tetrahydrofolate derivatives are essential for the one-carbon unit transfer during nucleotide biosynthesis and amino acid metabolism (Nzila et al., 2005). This has led to the *de novo* folate synthesis pathway being recognized as an attractive metabolic target for the treatment of numerous infectious diseases, including malaria (Swarbrick et al., 2009). The antimalarial antifolate drugs target the *de novo* folate synthesis pathway of the parasite to limit the production of folate derivatives and thereby prevent the growth and reproduction of the parasite. The two main targeted enzymes are dihydrofolate reductase (DHFR) and dihydropteroate synthase (DHPS), which are inhibited by the use of pyrimethamine and sulfadoxine antifolate drugs. The two drugs have also been used in synergy to target both enzymes simultaneously. However, resistance against the available antifolate drugs has emerged rapidly and has been observed even with the use of higher drug dosages, rendering these drugs ineffective in many cases. The combination therapy was previously used to overcome the resistance, however, in many cases, this treatment also failed as the parasite has also developed tolerance to a combination of drugs (Alker et al., 2008). The emerging resistance to antimalarial drugs drives a continuous need to develop drugs that have a novel mechanism of action. In this study, we explore an alternative drug target and drug targeting sites in the *de novo* folate synthesis pathway.

The first reaction of the *de novo* folate synthesis pathway is catalyzed by guanosine-5'-triphosphate (GTP) cyclohydrolase (GCH1), which converts the GTP moiety to the 7,8-dihydroneopterin (DHNP) (Colloc'h et al., 2000). The enzyme 6-pyruvol tetrahydropterin synthase (PTPS) then converts DHNP to 6-pyruvoyl-tetrahydropterin (PTP) via an internal redox transfer and final elimination of the DHNP phosphate tail (Bürgisser et al., 1995). The PTP product is then processed further by the downstream enzymes of the *de novo* folate synthesis pathway, including 6-hydroxymethyldihydropterin pyrophosphokinase (HPPK), DHPS, dihydrofolate synthase (DHFS) and DHFR (Kümpornsin et al., 2014).

PTPS is a hexameric lyase enzyme that has 3-fold symmetry (Bürgisser et al., 1995). The protein is composed of six identical monomers that assemble via tight hydrogen bonds to form two trimer structures (Nar et al., 1994). The two trimers join via a head-to-head association to form the functional hexameric unit (**Figure 1**). The trimers are 60 Å in diameter and have a height of

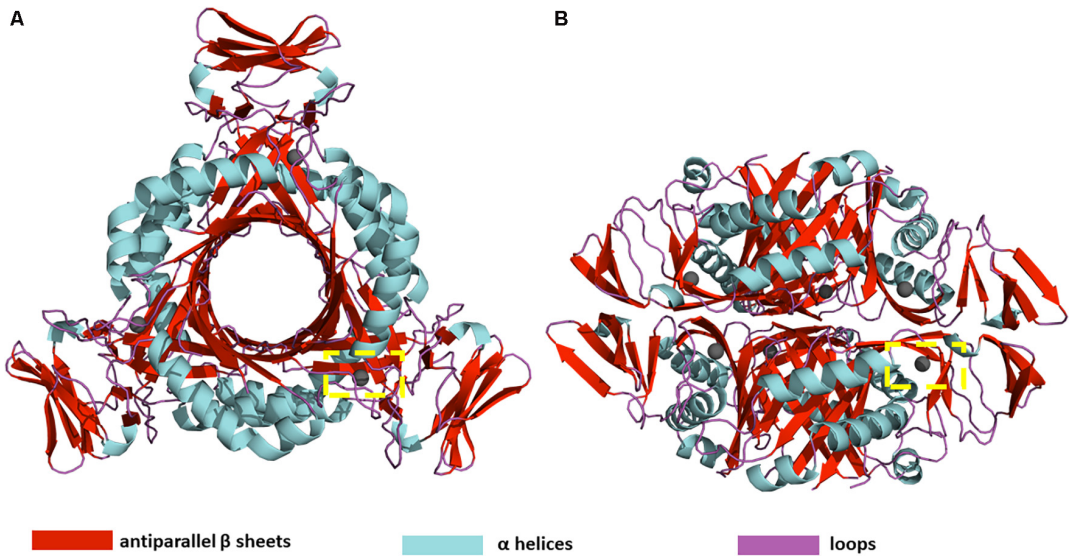


FIGURE 1 | (A) Top view **(B)** Side view of *P. falciparum* PTPS 3D structure (PDB: 1Y13). The 3D structure is composed of a dimer of trimers with D3 symmetry (Nar et al., 1994). The yellow dashed box shows the location of one of the six active sites.

30 Å (Nar et al., 1994). The structure of PTPS is characterized by a conically shaped central barrel that accumulates a cluster of basic and aromatic residues (Nar, 2011). PTPS has six zinc-containing active sites, and each site is buried in a deep cavity of 12 Å formed between every three adjacent monomers (**Figure 1**). Three histidine residues (H29, H41, H43) coordinate the metal ion through their NE2 atoms during the catalysis of the substrate (Khairallah et al., 2020). The deeply buried active sites of PTPS are accessible to the substrate through the central opening along the axis of the trimer (Bürgisser et al., 1995). Previous studies suggested that the barrel is necessary for the stabilization of the multimeric association and therefore incorporates a sophisticated use of functional allostery (Colloc'h et al., 2000). The PTPS enzyme belongs to the tunneling fold (T-fold) protein superfamily (Colloc'h et al., 2000). Enzymes of the T-fold superfamily exhibit a conserved structural topology and, with the exception of the active residues, have low sequence similarity (Colloc'h et al., 2000). The structural conservation that exists despite the low sequence similarity, points to its importance in maintaining protein function. Therefore, the characterized structural features of this protein superfamily, including its distinct central cavity, may encompass undiscovered allosteric sites that can be exploited for drug discovery.

Allostery is defined as the regulation of a protein's structure and activity by the binding of an effector molecule at a site other than the conserved active site (Nussinov and Tsai, 2013). Allosteric drugs are highly specific to sites other than the active sites, and therefore induce the desired effect of either activating or inhibiting a protein via a mechanism that does not rely on targeting a site that is highly conserved in humans. As a result, they are considered to be far less toxic to the host. However, discovering allosteric sites that have a dominant effect on protein conformation and respective compounds that modulate these sites is far more challenging than orthosteric drug discovery (Lu et al., 2019; Amamuddy et al., 2020). As in many cases, the location of allosteric sites is often unknown and their effects on the intrinsic motion of the protein are difficult to determine experimentally (Suplatov and Švedas, 2015). Our recent review article proposes a number of integrated computational approaches to identify allosteric sites (Amamuddy et al., 2020).

Protein structures are dynamic in that they undergo large-scale domain changes in response to the binding of ligands or other compounds, thereby assuming different conformational states that allow them to perform certain functions (Bahar et al., 2007; Henzler-Wildman and Kern, 2007; Teilum et al., 2009; Nussinov and Ma, 2012; Loutchko and Flechsig, 2020; Zhang et al., 2020). Studying large-scale structural changes can elucidate key sites that modulate the functional mechanisms of a protein (Penkler et al., 2017, 2018; Guarnera and Berezovsky, 2020; Shrivastava et al., 2020). Several experimental and computational approaches are used to measure quantitatively the structure, dynamics, and function of macromolecules. A comprehensive review of different approaches is provided by Orozco (2014); Palamini et al. (2016), and Maximova et al. (2016). In this study, we have applied Normal Mode Analysis (NMA), calculated on the Anisotropic Network Model (ANM) to uncover collective motions that are essential for the tunnel gating mechanism of PTPS, and thus pinpoint potential sites that modulate the allostery of the enzyme to be targeted in future anti-malarial drug discovery. The NMA analysis was combined with protein motif discovery to elucidate motifs that are located in regions that are essential to the dynamics of PTPS and that are uniquely conserved in Plasmodium PTPS enzymes. The identified motifs further support our finding of regions responsible for the PTPS conformational

transitions and therefore establish guidelines toward the selective inhibition of this enzyme.

MATERIALS AND METHODS

Structural and Sequence Data Retrieval

The PTPS crystal structure of the *P. falciparum* was retrieved from the online Protein Data Bank (PDB ID: 1Y13). The protein functional unit consists of six identical chains, a total of 978 residues. This structure was used to construct an ENM on the C_β atomic coordinates of the protein, as found in the PDB file. The C_β atoms were selected because it has been shown that they provide a better representation of the side chains orientation. A harmonic potential within a cut-off distance (*Rc)* of 15 Å was used to account for the pairwise interactions between all of the C_β atoms.

For motif analysis, the protein sequences of PTPS from 20 different species, including 9 Plasmodium species, four mammalian species, four bacteria species and three fungi species (**Supplementary Table S1**) were retrieved using the *P. falciparum* PTPS sequence as a query in a BLAST search (Altschul et al., 1990) to identify other Plasmodium homologs in the PlasmoDB (Aurrecoechea et al., 2009) and mammalian, bacterial and fungi homologs in UniProt (Bateman, 2019). The BLAST search tool was used with default alignment parameters.

Calculation of the Normal Modes

NMA was predominantly performed using the MODE-TASK tool suite (Ross et al., 2018) which employs the ANM (Atilgan et al., 2001) to construct an ENM of the protein structure. Although the ANM has been described previously (Atilgan et al., 2001), we provide a summary here. In the ANM, each residue of the *P. falciparum* PTPS structure was represented by a node placed at its C_β atom coordinate and all interactions between each pair of residues separated within a defined cutoff distance, Rc = 15 Å, was modeled as a set of springs with a uniform force constant γ. This yielded a network that contained N nodes and M springs representing the total number of interactions defined in the network, such that any given pair of nodes within Rc of each other will interact in accord with a conventional harmonic potential. The normal modes of this network are calculated in the absence of an external force, under an equilibrium condition. In the general case of N residues connected by M springs, the Hessian Matrix $\mathbf{H}$ is a $3N \times 3N$ super-matrix that may be derived from the second derivatives of the overall potential V, with respect to the components of $\mathbf{R}_i$, where $\mathbf{R}_j$ are the fluctuation vectors of the individual residues. Therefore, the Hessian matrix describes the force constant of the system. $\mathbf{H}$ is composed of $N \times N$ super-elements, i.e.,

$$H = \begin{bmatrix} H_{11} & H_{12} \cdots & H_{1N} \\ H_{21} & \cdots & H_{2N} \\ \vdots & \cdots & \vdots \\ H_{N1} & \cdots & H_{NN} \end{bmatrix} \quad (1)$$

where each super-element $\mathbf{H}_{ij}$ is a 3×3 matrix that holds the anisotropic information regarding the orientation of nodes *i,j* The ij^{th} super-element ($i \neq j$) of $\mathbf{H}$ is defined as:

$$H_{ij} = \begin{bmatrix} \partial^2 V/\partial\, X_i \partial\, X_j & \partial^2 V/\partial\, X_i \partial\, Y_j & \partial^2 V/\partial\, X_i \partial\, Z_j \\ \partial^2 V/\partial\, Y_i \partial\, X_j & \partial^2 V/\partial\, Y_i \partial\, Y_j & \partial^2 V/\partial\, Y_i \partial\, Z_j \\ \partial^2 V/\partial\, Z_i \partial\, X_j & \partial^2 V/\partial\, Z_i \partial\, Y_j & \partial^2 V/\partial\, Z_i \partial\, Z_j \end{bmatrix} \quad (2)$$

At equilibrium, the second derivatives may be calculated for the ANM using the β-carbon position vectors of PDB structures such that the elements of the off-diagonal $\mathbf{H}_{ij}$ are given by the equation:

$$\partial^2 V/\partial\, X_i \partial\, Y_j = -\gamma\, (X_j - X_i)(Y_j - Y_i)/S_{ij}^2 \quad (3)$$

And the elements of the diagonal super-elements $\mathbf{H}_{ii}$ are given by the equations:

$$\partial^2 V/\partial\, X_i^2 = \gamma \sum_j (X_j - X_i)^2/S_{ij}^2 \quad (4)$$

For the diagonal elements of $\mathbf{H}_{ii}$ and

$$\partial^2 V/\partial\, X_i \partial\, Y_j = \gamma \sum_j (X_j - X_i)(Y_j - Y_i)/S_{ij}^2 \quad (5)$$

For the off-diagonal elements of $\mathbf{H}_{ii}$

$\mathbf{H}$ can be decomposed into $3N$-6 eigenvalues and $3N$-6 eigenvectors that correspond to the respective frequencies and directions of the individual non-trivial modes. The modes with the lowest frequencies are termed the slowest modes and define the most collective, or global, motions of the protein. The highest frequency modes describe the more localized motions of the protein. The VMD program (Humphrey et al., 1996) was used to visualize the eigenvectors describing the structural change in each mode. Lastly, mechanical stiffness calculations (Eyal and Bahar, 2008) were performed on the ANM of the PTPS using the ProDy Python package (Bakan et al., 2011, 2014) and the obtained results were visualized using Matplotlib library and VMD.

In addition, the Gaussian Network Model (GNM) calculations were performed using the DynOmics portal (Li et al., 2017), specifically to identify hinge regions within the structure of PTPS. The GNM calculations were performed using a cutoff distance of 15 Å, with a spring constant scaling factor of 10 for contact distances ≤4.0 Å and a distance scaling exponent of 2. The Dynomics portal was further used to validate the high-frequency vibrating residues and to analyze the mechanical properties of PTPS, by obtaining a color-coded representation of the PTPS structure based on the mobility of the residues from the resultant low and high GNM modes.

Calculation of the Residue Mean Square Fluctuations

The inverse of $\mathbf{H}$ is equivalent to the covariance matrix C that is composed of $N \times N$ super-elements. Each off-diagonal, *ij*th, super-element of $\mathbf{H}^{-1}$ contains the 3×3 matrix of correlations between the x-, y- and z-components of fluctuation vectors of residues *i* and *j*, while the *i*th super-element of $\mathbf{H}^{-1}$ describes the self-correlation between the components of fluctuation

vectors of residue *i*. For any given mode, the mean square fluctuations (MSF) of individual residues can be obtained by summing the components of fluctuation. The MSF profiles from individual normal modes were calculated using MODE-TASK (Ross et al., 2018).

Residue Cross-Correlation Analysis

The BIO3D R package for the exploratory analysis of the structure and sequence data (Grant et al., 2006) was used to compute the deformation energy and residue cross-correlation. The BIO3D cross-correlation $S(i, j)$ is given by:

$$S\left(i,j\right) = (\Delta_{ri}.\Delta_{rj})/(\Delta r_i^2)^{\frac{1}{2}}(\Delta_{rj}^2)^{\frac{1}{2}} \quad (6)$$

where Δr_i and Δr_j are the displacement vectors for atoms *i* and *j*, respectively. The elements $S(i, j)$ are stored in matrix form and displayed as a three-dimensional dynamical cross-correlation map. If $Sij = 1$ the fluctuations of atoms *i* and *j* are completely correlated (same period and same phase), if $Sij = -1$ the fluctuations of atoms *i* and *j* are completely anti-correlated (same period and opposite phase), and if $Sij = 0$ the fluctuations of *i* and *j* are not correlated.

The deformation energy signifies the energy density due to deformation as a function of position (Hinsen, 1998) given by Eq. 7. It provides a measurement of the individual atom's energy contribution toward the structural deformation. PTPS deformation energy was derived from the eigen energies and vectors of the first not trivial 20 normal modes according to the equation:

$$E = \frac{1}{2}\sum_{1}^{N} K\left(R_{ij}^{(0)}\right)\frac{\left|\left(r_i - r_j\right).R_{ij}^{(0)}\right|^2}{\left|R_{ij}^{(0)}\right|^2} \quad (7)$$

where r_i, r_j donate the displacement of atom *i* and *j* in the mode to be analyzed, $Rij^{(0)}$ is the pair distance vector (R*i*–R*j*) in the input arrangement and K is the pair force constant.

Motif Discovery

Motif discovery was performed using Multiple Expectation Maximisation for Motif Elicitation (MEME) vs 4.11 (Bailey et al., 2015) to identify highly conserved motifs in the Plasmodium PTPS enzymes. A fasta file containing PTPS protein sequences was parsed to the MEME analysis software. The motifs search size was set to the range between 6 and 20 residues. The Motif alignment search tool (MAST) was then used to identify overlapping motifs (Bailey and Gribskov, 1998). A Python script was then used to analyze MAST files and MEME log files. Motif conservation was calculated as a number of sites per the total number of sequences, and the results were displayed as heatmaps. Motifs that were uniquely conserved in Plasmodium species were mapped onto the 3D structure of *P. falciparum* PTPS and visualized using the PyMOL Molecular Graphics System (DeLano, 2014).

RESULTS AND DISCUSSION

Notable Conformational Changes Were Captured by the Lowest Frequency Non-degenerate Modes of *P. falciparum* PTPS

In this study, the ANM was applied to the structure of PTPS to classify its collective motions that have the propensity to lead the enzyme from one conformational state to another, and thus modulate its function. Further investigation of the residue fluctuations within the modes was performed to identify highly active regions that may drive these motions. The nature of allosteric modulation requires a high degree of collectivity which is often well-described by the low-frequency modes (Bahar et al., 2010). The NMA of PTPS (a 3-fold symmetry structure) yielded a total of 2,934 modes. The first 20 slowest non-trivial modes were selected for the characterization of the global intrinsic motions of the protein structure.

Normal modes obtained from symmetrical structures are highly susceptible to degeneracy, thus producing degenerate and non-degenerate modes. The degenerate modes share the same frequency and consequently any orthogonal transformation. In contrast, non-degenerate modes characterize unique directions of motions that often capture global meaningful motions that account for large conformational changes. Previous studies showed that dominant conformational changes of complexes were captured within the slowest non-degenerate modes (Atilgan et al., 2001; Chennubhotla et al., 2005; Shrivastava and Bahar, 2006; Wako and Endo, 2011; Isin et al., 2012; Lee et al., 2017; Ross et al., 2018). Here, we identified eight non-degenerate modes of PTPS. These non-degenerate modes exhibited unique eigenvalues (i.e., frequencies) and displayed unique global motions. **Table 1** shows the first 20 non-trivial normal modes of PTPS, their associated frequencies, degeneracy levels and the contribution of each mode to the overall motion of the protein. Although the contribution of the individual modes was low and did not reveal a single dominant motion, we suspect that this may be a consequence of the large size of the protein and the extensive degeneracy of the normal modes. Here we have analyzed the first 20 slowest modes which only represent 0.68% of the total modes, yet when combined they account for 11.08% of total motion of the protein. Similarly, the 8 non-degenerate modes only represent 0.27% of the total modes but account for 3.64% of total motion. The displacement vectors of the individual non-degenerate modes were studied as well as the MSF average over the 20 slowest and 20 fastest frequency normal modes.

To visualize the atomic displacements during the collective motions of PTPS, the respective eigenvectors of each of the eight non-degenerate modes were projected onto the structure of PTPS (**Figure 2**). Movies were constructed by projecting the eigenvectors onto the structure as a set of frames in which the vectors were added to the original atomic coordinates in increasing steps and then visualized using VMD. The atomic displacement during these modes was then examined to identify distinctive motions that are associated with the enzyme's tunnel gating such as opening, closing or rotation as well as any other

TABLE 1 | PTPS first 20 non-trivial normal modes, associated eigenvalues, and level of degeneracy.

Mode	Eigenvalue	Degeneracy	Contribution%
Mode 1	0.33	2	1.38
Mode 2	0.33	2	1.37
Mode 3	0.48	1	0.96
Mode 4	0.55	2	0.84
Mode 5	0.55	2	0.84
Mode 6	0.89	1	0.52
Mode 7	0.92	1	0.50
Mode 8	0.95	2	0.48
Mode 9	0.96	2	0.48
Mode 10	0.96	2	0.48
Mode 11	0.97	1	0.47
Mode 12	1.02	2	0.45
Mode 13	1.03	2	0.44
Mode 14	1.21	1	0.38
Mode 15	1.34	2	0.34
Mode 16	1.35	2	0.34
Mode 17	1.63	1	0.28
Mode 18	1.68	1	0.27
Mode 19	1.79	1	0.26
Mode 20	1.95	2	0.24

The first six trivial modes were excluded, therefore mode 1 represents mode 7.

notable movements that resulted in structural changes around the tunnel or active site. The eight non-degenerate modes captured coupling movements between the tunnel gating, the N-terminal β-strands, and the central helices thus demonstrating that these regions primarily regulate the global dynamics of the protein. Furthermore, the observed structural changes of the tunnel were often accompanied by structural deformation around the active site region.

PTPS Tunnel Displayed Several Distinct Movements That Involved the Fluctuations of the N-Terminal β-Strands

The modal analysis captured four distinctive motions that resulted in structural changes around the tunnel and the active site of PTPS. Mode 3 captured an asymmetric global twist of the terminal domains of the two PTPS trimers, characterized by their rotation in opposite directions (**Supplementary Movie 1**). From visual inspection, it appeared that the twisting of the terminal domains controlled the expansion and narrowing of the tunnel. Most notably, the diameter of the tunnel was smaller when the terminal domains followed a wringing motion with respect to the principal axis. We, therefore, suggest that the fluctuations of the terminal domains modulate the tunnel gating and consequently induces structural changes around the active site region. In mode 6, a notable tilt of the entire terminal domains was observed, leading to a side to side motion of the protein structure (**Supplementary Movie 2**). A stretching motion that resulted in the lateral expansion and contraction of the tunnel was captured in mode 7. From visual inspection, it appeared that this expansion and contraction was largely driven by the extensive movement of

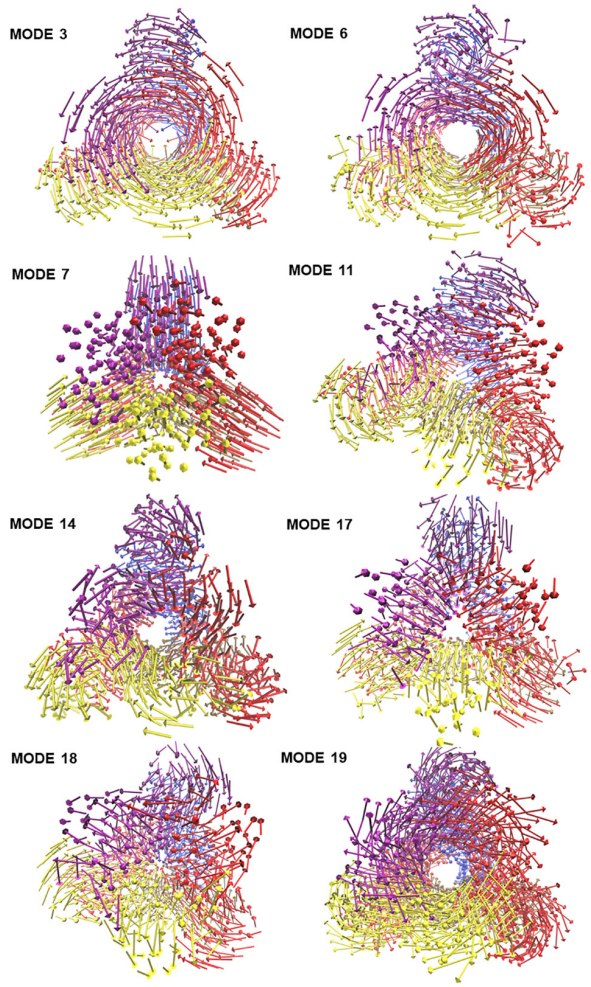


FIGURE 2 | The projection of the eight slowest non-degenerate modes onto the structure of PTPS. Eigenvectors have been projected as a set of arrows that denote the direction of motion of the protein C_β atoms. The 3D structure is PTPS *P. falciparum* (PDB ID: 1Y13). Each of the protein chains is colored differently for illustration purposes.

the terminal and central helices (**Supplementary Movie 3**). Mode 11 displayed a tilting motion of two central helices in an opposite direction to the third central helix (**Supplementary Movie 4**).

The identified motions within the tunnel were often accompanied by structural changes in the active site region, in which the tunnel acted as a connecting vessel to allow the entry of the substrate and provide the active site with flexibility. This further demonstrates how the tunnel gating mechanism exerts such control on the active site and illustrates the functional relevance of this cavity in the catalytic mechanism of the protein.

The N-Terminal Domain Wringing Motion Modulated the Exposure of the Protein Tunnel

Visualization of the non-degenerate modes revealed three modes in which the wringing of the N-terminal domains promoted

the surface exposure of buried regions within the tunnel. In particular, Mode 14 captured the outward and inward movement of the PTPS tunnel (likened to an engulfing movement) (**Supplementary Movie 5**). Mode 17 presented a prominent bending of the central helices from side to side, resulting in the protein tightening in the same direction (**Supplementary Movie 6**). Mode 18 featured a breathing motion that is characterized by the upward and downward movement of the central β-sheets (**Supplementary Movie 7**). Mode 19 displayed a clockwise rotation of the protein core that is associated with the terminal regions twisting in the opposite direction (**Supplementary Movie 8**). Within the eight identified non-degenerate modes, residue 76 to 114 of the N-terminal β-strands and residue 136 to 150 of the central helices exhibited the highest mobility (**Figure 9A**), which propelled the global motions of the protein. The expansion of the tunnel may promote the binding of PTPS to accessory proteins. Other proteins of this family are known to associate with other proteins, for example, the T-fold enzyme GCH1 is known to physically associate with another pentameric enzyme via the central opening of its tunnel in order to regulate its activity (Maita et al., 2004; Higgins and Gross, 2011).

Based on our modal analysis, we suggest that the flexibility of the tunnel is essential for the efficiency of the active site. The tunnel appears to modulate entry of the substrate via expansion and contraction which is primarily driven by the rotation of the N-terminal regions, thus allowing the substrate to pass through the tunnel into the active site for catalysis. Furthermore, we hypothesize that the twist and shear motions propel the substrate through the tunnel. Taken together, these results further the understanding of the functional dynamics of the PTPS enzyme, which will serve as significant guidelines toward the design of allosteric modulators that target structural regions that are pertinent to the function of the enzyme, and therefore the life cycle of the malarial parasite.

MSF of the Individual Normal Modes Showed That the N-Terminal β-Strands and Central Helices Displayed the Highest Fluctuation

The atomic MSF profiles of the eight non-degenerate low-frequency modes were calculated and plotted (**Figure 3**). In each of the individual profiles, the N-terminal β-strands and the central helices have shown a notable fluctuation when compared to the protein core. Sharp peaks on the MSF profile distinguished these regions (**Figure 3**). The calculated MSF of the slowest 20 normal modes and the experimental B-factor were compared to validate the fluctuations that were predicted in the normal modes. **Figure 4B** shows the atomic fluctuation over the first 20 non-trivial low-frequency modes mapped onto the PTPS structure. The combination of the slowest 20 modes identifies the PTPS terminal domains as the most mobile (red) regions of the structure (**Figure 4B**). The simplified models of the ANM were shown to agree with the experimental B-factor results, as presented in **Figure 4**.

Deformation Analysis Demonstrated a Significant Build-Up of Deformation Energy in the PTPS Tunnel

Deformation is defined as any structural change caused by either an external force or a change in temperature (Bao, 2002). When the applied force is sufficient, a notable deformation can be observed; otherwise, the structure will resist the applied force and revert to its original state. After obtaining the normal modes, an analysis of the deformation energy distributed across the structure was performed to observe which regions of the structure were deformed during these global motions. The obtained results showed that the extensive fluctuations of the N-terminal domains, as seen in the MSF profiles, caused a significant build-up of deformation energy in the tunnel region of the enzyme. This demonstrates that the fluctuation of N-terminal domains has a long-distance effect on the tunnel core, which is located ∼30 Å away. The deformation energy values were calculated as the sum of the contributions of the first 20 non-trivial modes. The derived energy values were mapped onto the PTPS structure (**Figure 5**). We observed low deformation energy in the fluctuating terminal domains (blue), which was accompanied by high deformation energy within the tunnel region (red).

Residue Cross-Correlation Analysis Presented the Movement of the N-Terminal Domains in a Concerted Manner

A heatmap showing the cross-correlation of the C_β atom pairs is shown in **Figure 6**. Only motions in the first 20 non-trivial modes were selected in the calculation to highlight the residues involved in the protein collective motions. The residue cross-correlation analysis highlighted regions moving in a concerted manner, which therefore illustrates their involvement in the structural dynamics of the enzyme. The off-diagonal elements presented a positive correlation coefficient within the same chain, more specifically the residues of the terminal β-strands displayed correlation motion. Given the fact that PTPS is a multimeric protein, chains within one trimer exhibited a similar direction of motion. Anticorrelated motions were identified across the two dimers which can be understood due to the twisting or wringing of the helices across the trimers. The observed anticorrelation across the two trimers designates the opposing twisting or wringing of the terminals, therefore allowing the tunnel to open and close accordingly.

Motif Analysis Revealed Uniquely Conserved Sites in Plasmodium PTPS Enzymes

Protein motifs are evolutionarily conserved sequence patterns that might unearth a biological function. Identifying the motifs and their structural locations can, therefore, provide insight into regions that modulate protein function. It can further reveal key residues that are unique and typically required to retain the protein function and structural stability (Mackenzie and Grigoryan, 2017; Ross et al., 2017; Zheng and Grigoryan, 2017).

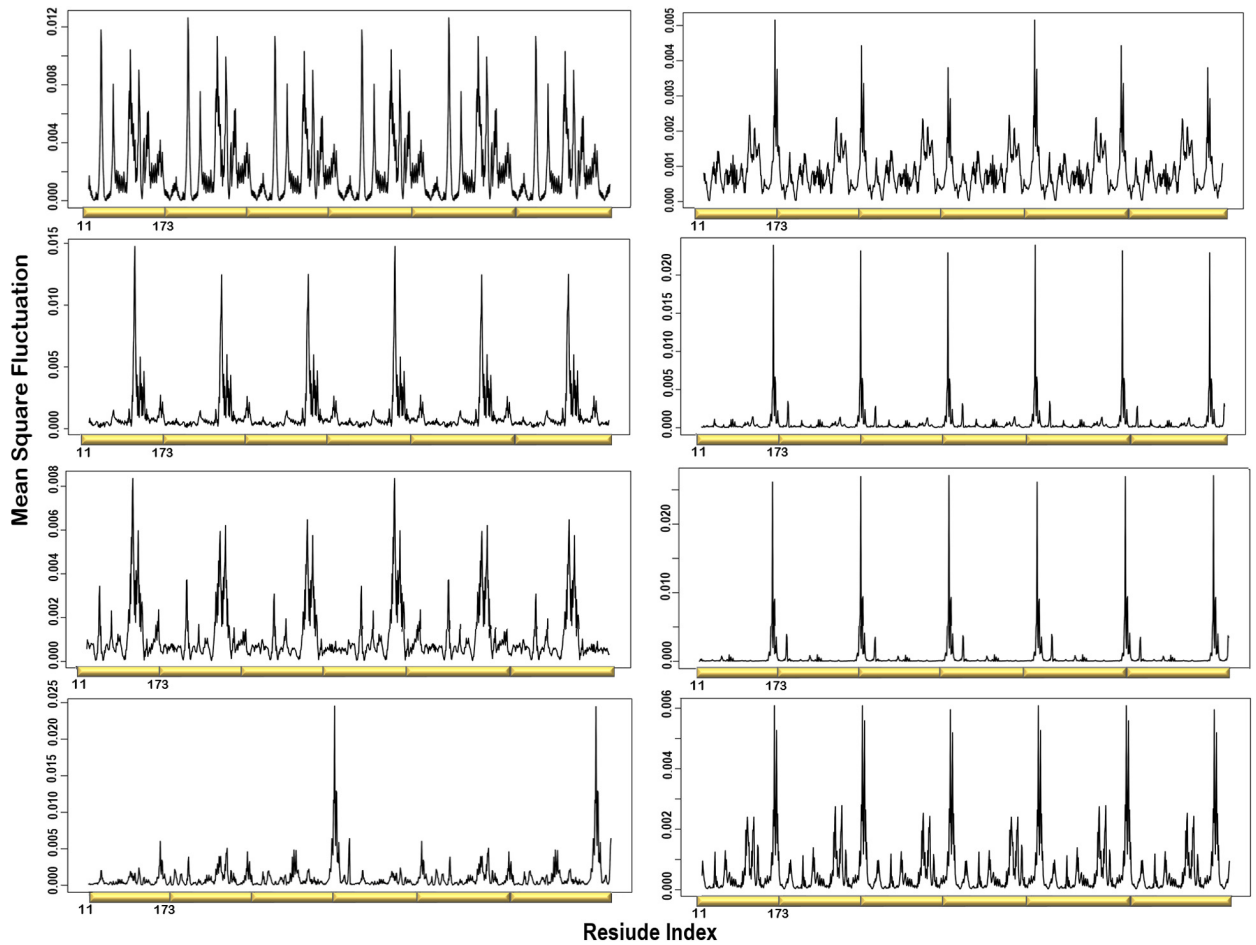


FIGURE 3 | Individual MSF profiles of the eight non-degenerate modes. The yellow blocks represent each chain of the protein structure. The residues index is labeled for one chain along the lower abscissa and according to the *P. falciparum* (PDB ID: 1Y13) file.

Following the analysis of the dynamics of the enzyme, we performed motif discovery to identify highly conserved motifs in *P. falciparum* PTPS and residues that are potentially involved in maintaining the secondary structure of the enzyme. The motif discovery was performed on the PTPS homologue sequences from four mammalian species including human, four bacterial species, three fungal species, and nine other Plasmodium species. A total of 27 motifs were identified. The analysis revealed four motifs (motifs 5–8) that were conserved in the Plasmodium PTPS enzyme sequences but were not detected in any of the mammalian species. Motifs 6–8 were uniquely conserved in the Plasmodium species, while motif 5 was also found among the bacterial and fungal PTPS enzymes. A heat map illustrating the occurrence of the motifs is shown in **Figure 7**.

The conserved Plasmodium motifs were mapped onto the crystal structure to identify their location (**Figure 8**). Motif 5 was located in the central β-sheet strands forming the tunnel cavity, motif 6 and 7 were located in the N-terminal antiparallel β-strands, as well as a loop linking the β-sheet strands and motif 8 was located in the central α-helices region. Motif 7 and 8 were shorter and more conserved relative to other motifs, which further suggests functional importance. As shown by the NMA, these regions were responsible for modulating the PTPS conformational transitions of the tunnel. The results presented significant structural and sequence differences between the Plasmodium PTPS enzyme and the human PTPS. As all of the identified motifs were not found in the human PTPS, this vital difference can be exploited for the attainment of drug selectivity and future antimalarial drug design.

The Lowest and Highest-Frequency Normal Modes Revealed Residues of Notable Mobility That Were Also Located in the Conserved Structural Motifs

The low-frequency normal modes are often associated with large amplitude conformational changes which are essential for function (Mahajan and Sanejouand, 2015). The first 20 modes illustrate the substantial contribution of the PTPS helices to the protein's global motion, in which they regularly control the tunnel movement via wringing or bending motions. In the MSF profile, the most substantial

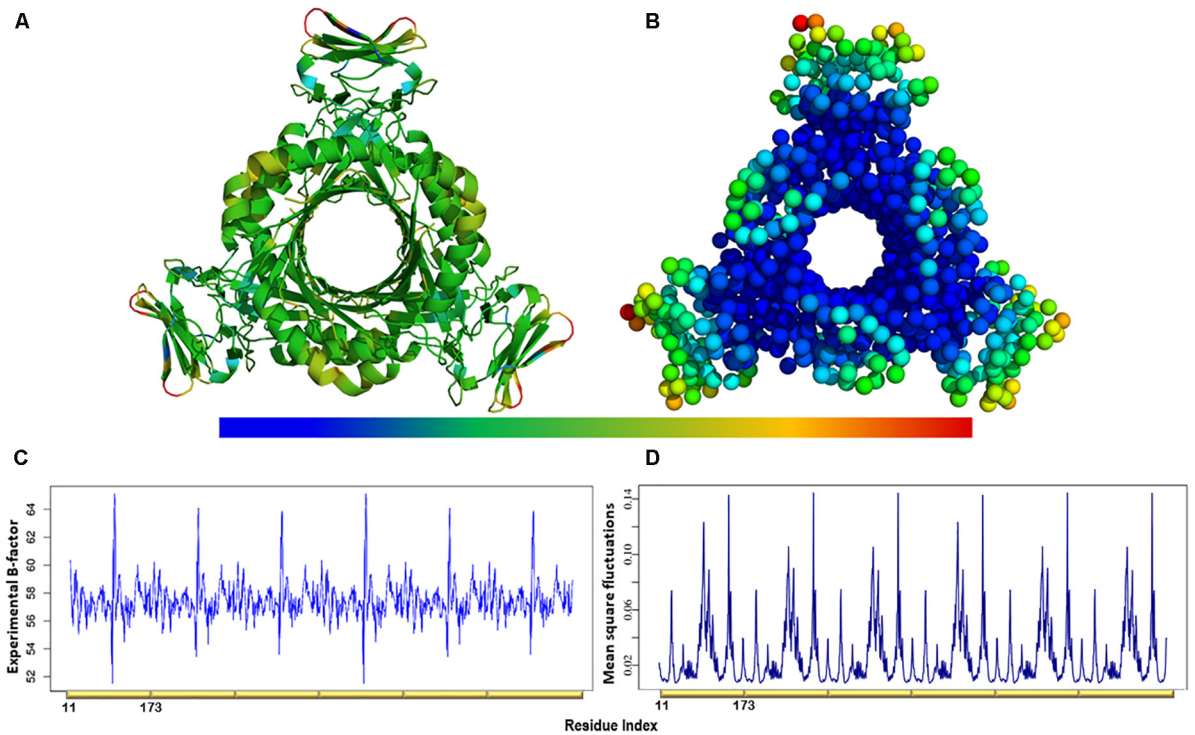


FIGURE 4 | (A) The experimental B-factors mapped onto the PTPS 3D structure. **(B)** Atomic mobility over the slowest 20 modes. Regions of the lowest fluctuation are shown in blue, higher mobility in green, yellow, and red. The fluctuations color scale is from blue to red (High/ flexible to low/rigid) atomic fluctuation. The figure demonstrates the high mobility of the terminal regions, central helices, and rigidity of the protein core, indicating that the NMA predicted motions have captured most of the experimentally determined motions. **(C)** Experimental B-factor plot and **(D)** Lowest 20 modes MSF profile. The residues index is labeled for one chain along the lower abscissa and according to the *P. falciparum* (PDB ID: 1Y13) file.

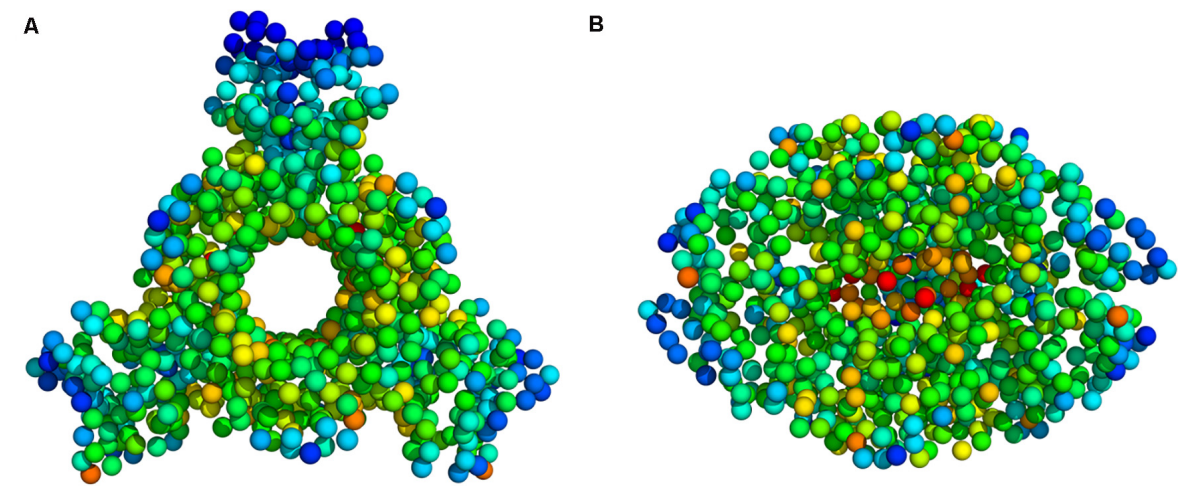


FIGURE 5 | Deformation energy values mapped onto the PTPS 3D structure **(A)** Top view **(B)** Side view. The colors of the atoms indicate the amount of deformation. The dark blue regions are the least deformed, whereas red areas are strongly deformed.

contributions to the atomic fluctuations emerged from the terminal and central helices (**Figure 9A**). The residues G35, K97, N107, and S145 showed notable fluctuations in the MSF profile. This demonstrates their key role in the global dynamics of the enzyme. Notably, residue N107 was found in motif 7 (position 9) and S145 was found in motif 8 (position 10) (**Figure 8**), further highlighting residues that are highly conserved in the Plasmodium species and which were active in modes that captured notable structural changes in the enzyme.

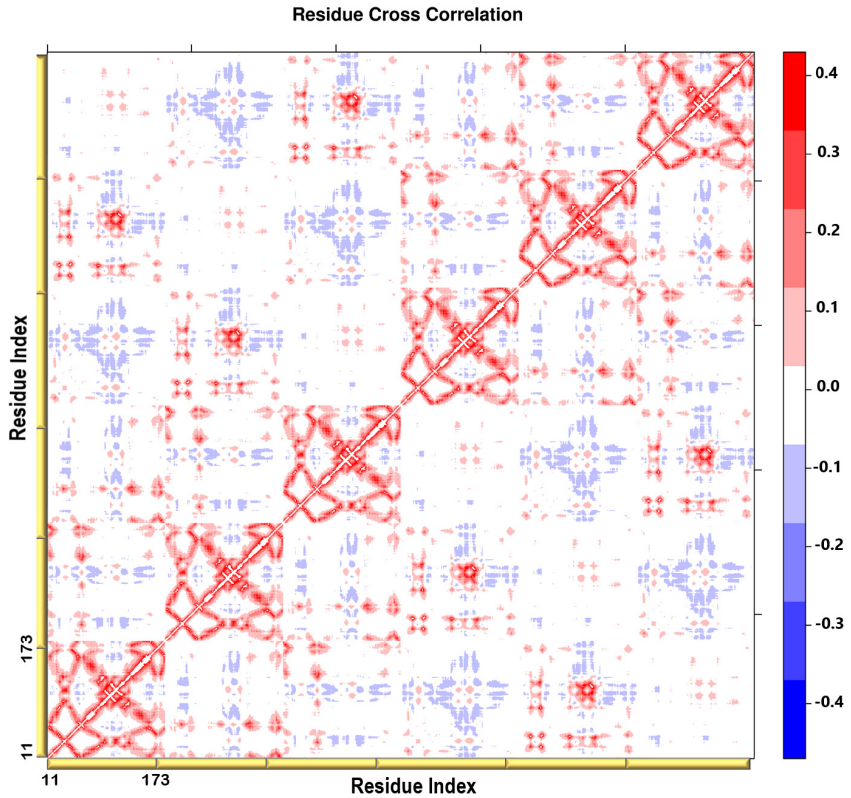


FIGURE 6 | A heatmap representation showing the C_β atoms pairwise cross-correlation. The correlation matrix ranges from −0.4 to +0.4. The scale color bar on the right indicates the extent of the correlation in which the red color highlights correlated motions (residue pairs moving together in the same direction), while the blue color highlights the anti-correlated movements (residue pairs moving in opposite direction). The yellow bars along the bottom represent each chain of the protein.

As the frequency increases, the modes become more localized and are accompanied by fast vibration of individual residues. It has previously been reported that high-frequency vibrating residues overlap with residues that are highly conserved and have an essential role in stabilizing the protein, and thus maintaining its function (Haliloglu et al., 2005). High-frequency vibrating residues in PTPS were identified from the average MSF calculated over the 20 highest frequency modes (**Figure 9B**). These included the active site residues H29, H41, and H43 as well as other important residues for the enzyme catalysis. The residues L78, I83, V159, E161, and A167 showed high fluctuation in the MSF profile of the fastest modes. Motif analysis showed that these residues do in fact overlap with sites of high conservation, with L78 and I83 found in motif 6 at (position 3 and 8), V159, E161, and A167 in motif 5 (position 6, 8, and 14), respectively (**Figure 8**).

Previous studies established that a cluster of hydrophobic residues surrounds the active site pocket and interacts with the substrate ring (Bürgisser et al., 1995; Colloc'h et al., 2000). Here we show that several of these hydrophobic residues also fluctuated in the high-frequency modes (**Figure 10B**). Residues E161 and T127, which are located at the bottom of the active site pocket (**Figure 10A**), showed significant fluctuation of the fast frequency modes, with E161 displaying the highest residue fluctuation (**Figure 9B**). These two residues were previously reported for their key role in substrate recognition and binding, as they both act as proton donors and acceptors during catalysis (Nar et al., 1994; Bürgisser et al., 1995; Ploom et al., 1999; Nar, 2011). Furthermore, Nar and colleagues reported that T105, T106, and E107, located around the active site pocket, constitute an acceptor site for the substrate ring during catalysis in the *Rattus norvegicus*. PTPS structure (PDB ID: 1B6Z). In the *P. falciparum* PTPS structure these residues are equivalent to S126, T127, E129, all three of which corresponded to high-frequency vibrating residues. Overall, the combined results obtained from our NMA and motif analysis have revealed conserved residues in the Plasmodium species that have key structural significance, with strong potential to modulate the stability and function of the enzyme. Thus, the characterized regions can then be proposed as alternatives that can be targeted and kept in consideration for future drug design efforts.

Hinge Residues Were Identified and Found to Overlap With High-Frequency Vibrating Residues Located in the Protein Core

Hinge regions are often found between domains in a protein to allow flexibility, in which they permit the domains to move relative to one another or clamp down on a substrate

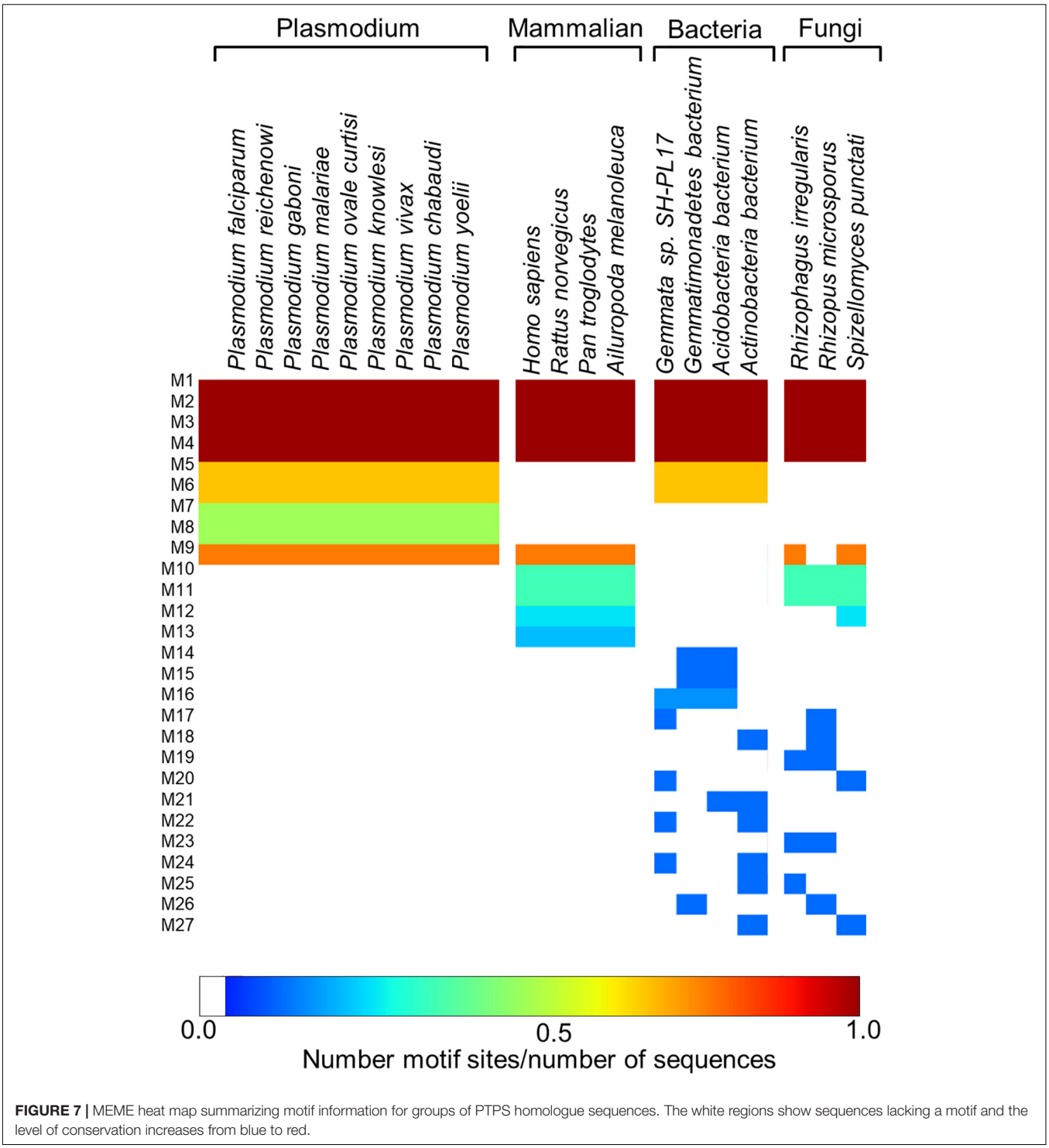


FIGURE 7 | MEME heat map summarizing motif information for groups of PTPS homologue sequences. The white regions show sequences lacking a motif and the level of conservation increases from blue to red.

(Towler et al., 2004; Yang and Bahar, 2005; Amusengeri and Tastan Bishop, 2019). The hinge residues, act as anchoring points and are involved in the propagation of large-scale conformational changes. The ANM provides a 3D description of motions identified in the modes, whereas in GNM the motion is projected to a mode space of N dimensions and therefore provides a description of atoms' mean squared displacements. GNM calculations are considered to be the preferred method for predicting the magnitude of motions at the cost of losing directions (Atilgan et al., 2001) and thus easily allow the identification of hinge sites. The DynOmics portal was used to recognize hinge residues (Li et al., 2017). The server provided information about key sites involved in the collective mechanics and allostery of the protein by mapping

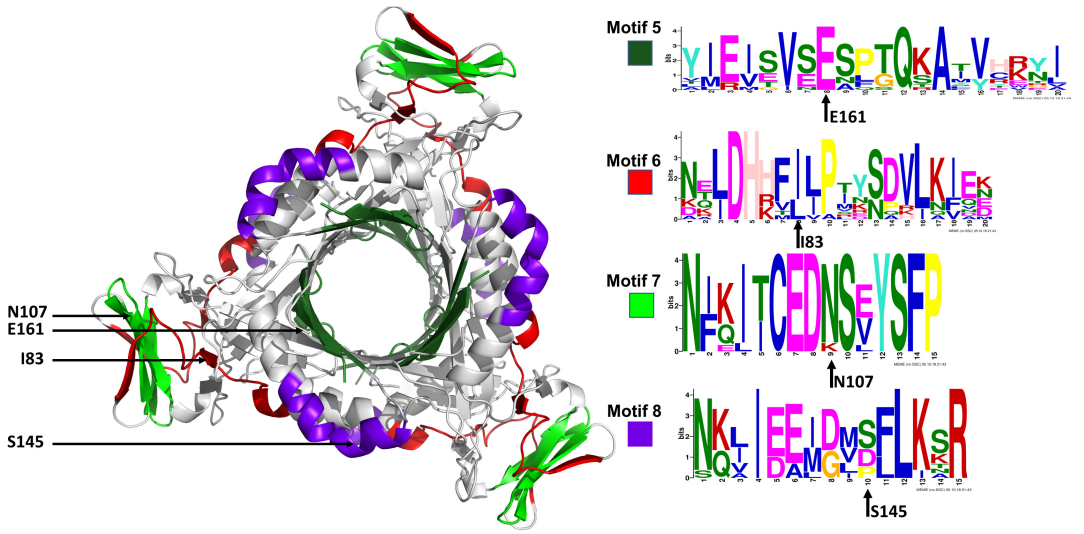


FIGURE 8 | The MEME sequence logo contains a stack of letters at every position in the motif. The height of the letters represents the probability (in bits) of the letter occurring at that position multiplied by the number of times that residue occurs within that site in each motif site in the total dataset (Bailey et al., 2015).

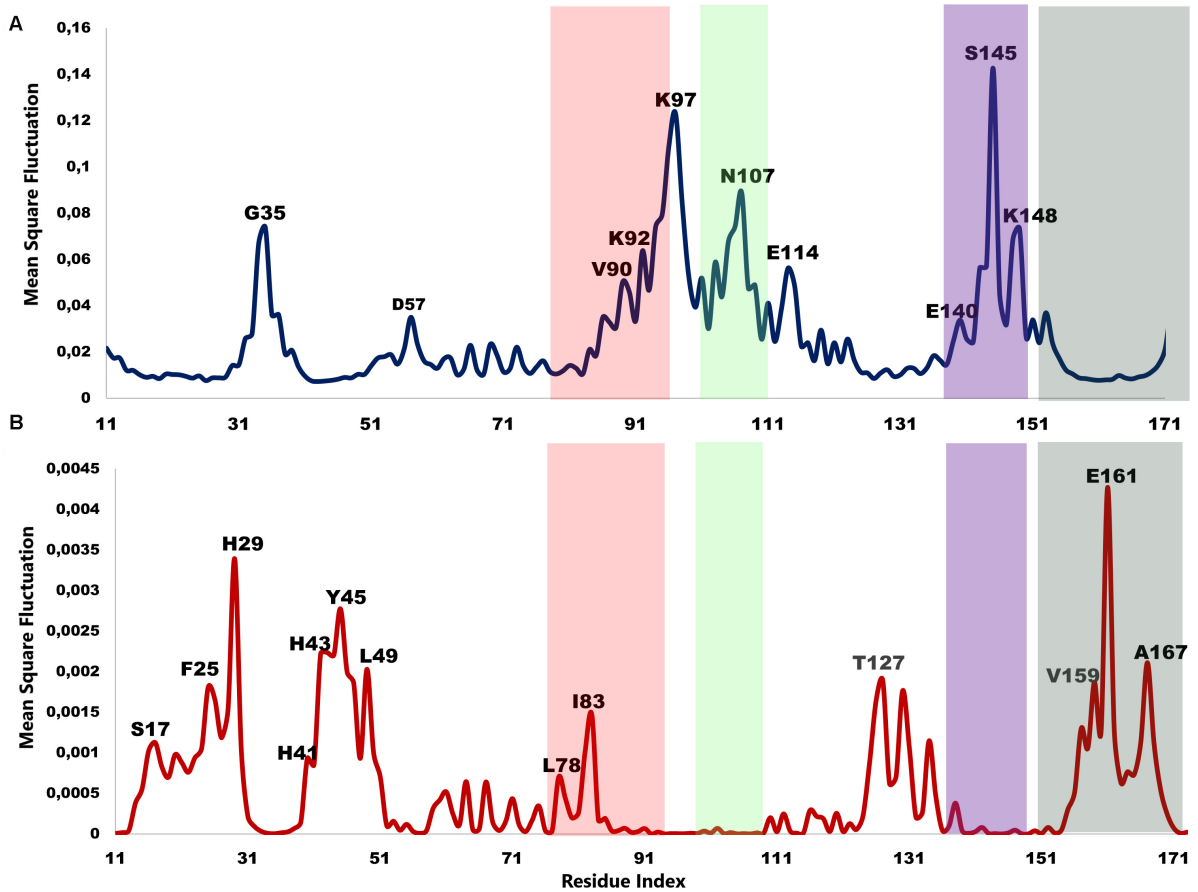


FIGURE 9 | MSF profiles of **(A)** Low-frequency and **(B)** High-frequency normal modes. Residues with the highest fluctuations of the slowest modes and the fastest modes are labeled. The dark green, purple, red, and lime green bars highlight the location of motif 5, 6, 7, and 8 respectively. N107 is located in motif 7 (position 9), S145 in motif 8 (position 10), L78, and I83 are located in motif 6 (position 3 and 8) and V159, E161, and A167 are located in motif 5 (position 6, 8, and 14) respectively.

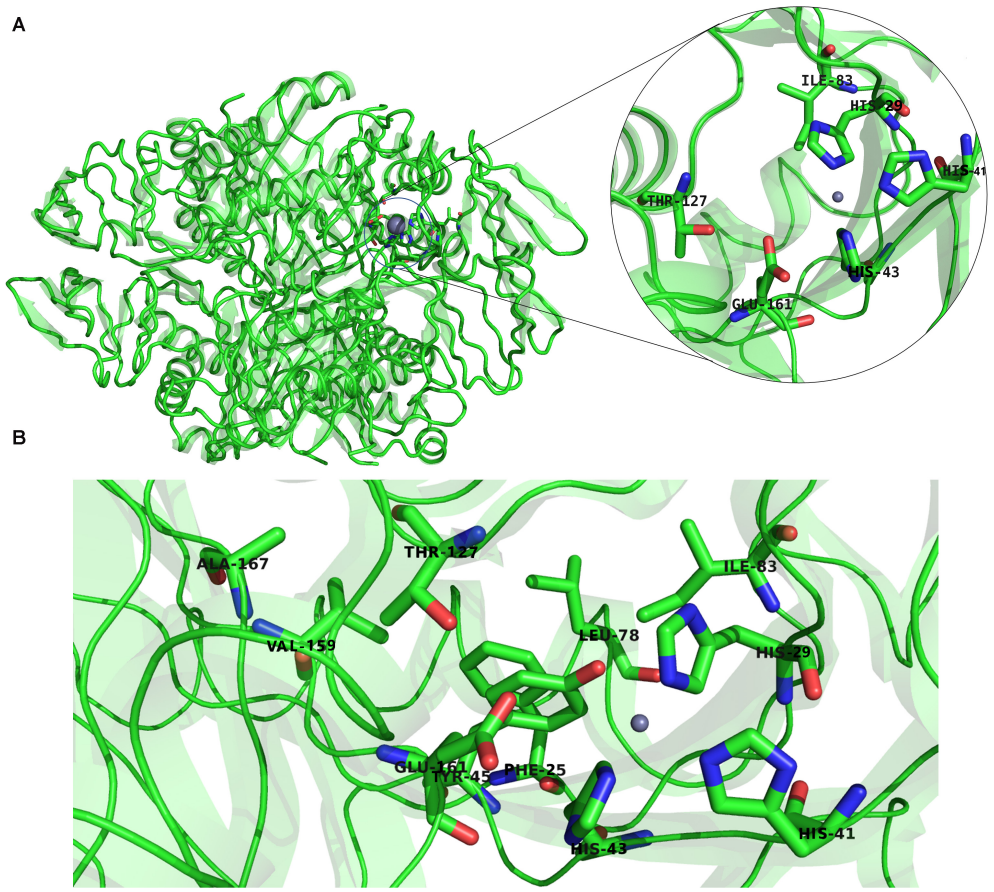


FIGURE 10 | (A) The structure of *P. falciparum* PTPS and the metal center (active site) showing in the circled zoom with the active site residues: H29, H41, H43, and the two catalytically important residues T127 and E161. **(B)** Boxed zoom showing the location of the identified high vibrating residues.

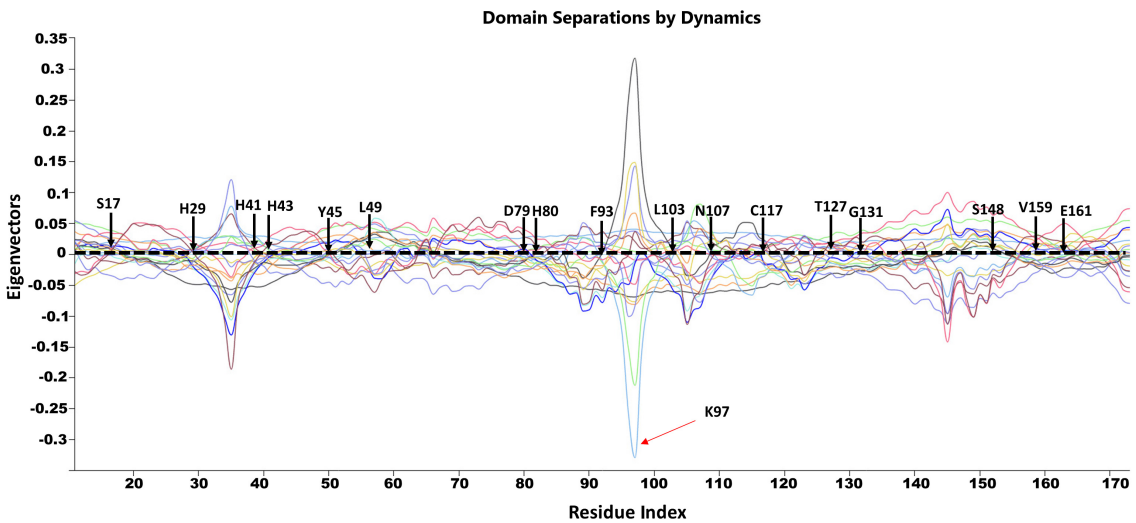


FIGURE 11 | GNM based identification of global hinge sites over the low-frequency modes. Hinge residues are located at the crossover line where the eigenvectors' values are equal to zero. Residues surrounding the hinge residues showed significant fluctuation in the positive and negative direction of motion. The identified hinge residues include the active site residues, H29, H41 and H43; the catalytically important residues, H80, T127 and E161; and the active site neighboring residues, Y45, L49, D89, F93, L103, N107, C117, S148, V159 which are illustrated by arrows.

of the coarse-grained conformations driven by collective modes to their full-atomic representations. The domain separations analysis based on modes of the GNM disclosed hinge sites (residues that exhibit minimal displacements in the softest two modes). The labeled residues in **Figure 11** correspond to the interfacial residues whose neighboring residues have a different sign in the eigenvector relative to themselves. Thus, these interfacial residues act as hinges about which their neighboring residues are displaced in opposite directions. Specifically, we identified hinge sites that overlapped with residues of the active site: H29, H41, and H43 as well as E161. The functional importance of these residues was supported by previous studies in which the mutation of these residues resulted in either a complete or a dramatic loss of enzymatic activity (Bürgisser et al., 1995; Nar, 2011). Furthermore, the active residues were responsible for the coordination of the catalytically important

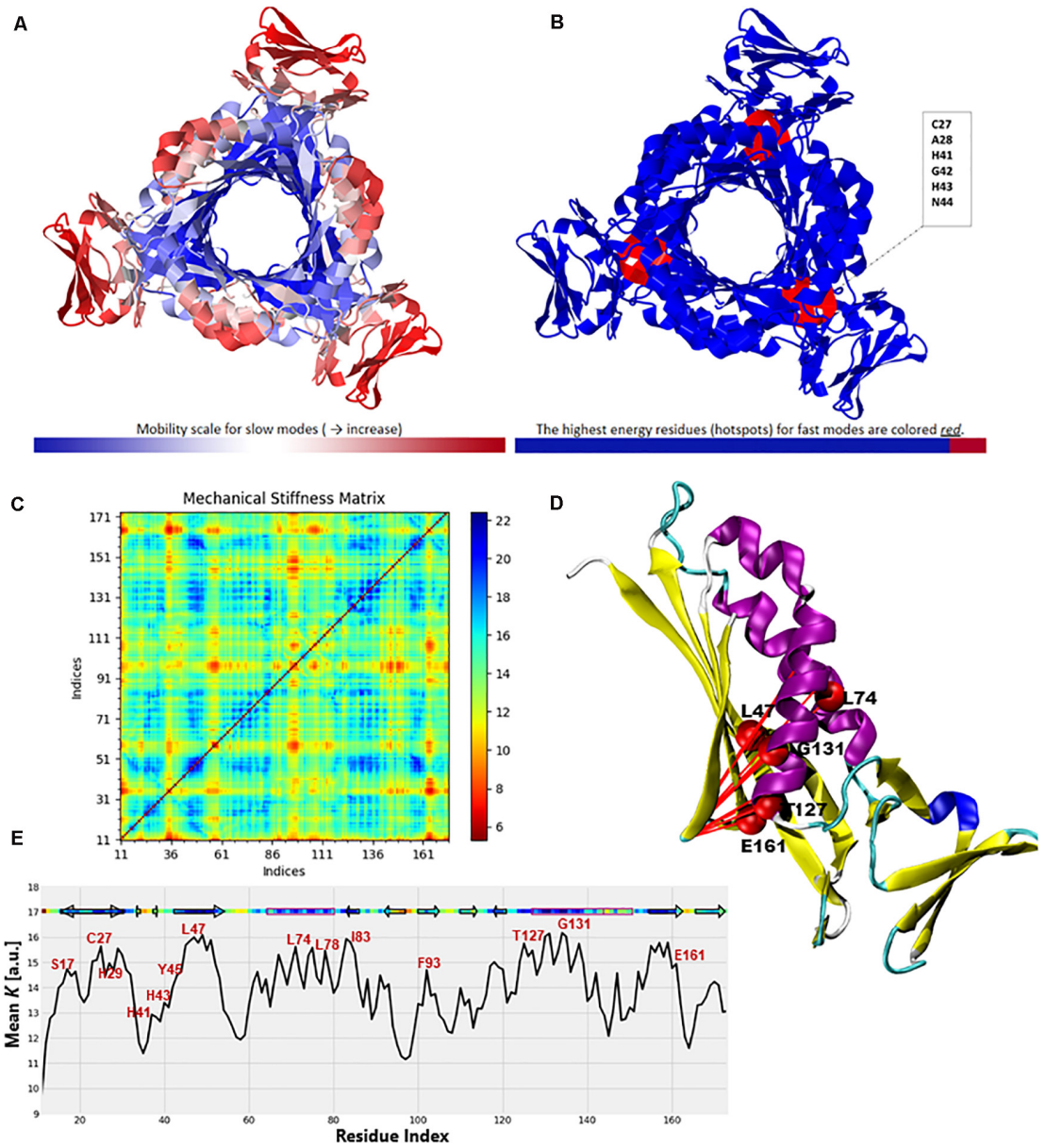


FIGURE 12 | Color-coded representation of the PTPS structure based on the mobility of the residues within the **(A)** Low-frequency and **(B)** High-frequency GNM modes. The color scale varies from blue as the most rigid to red as most mobile sites. The figures were obtained from the DynOmics portal (Li et al., 2017). **(C)** A mechanical resistance matrix obtained by calculating the effective force constant in response to uniaxial extensional forces exerted at each pair of residues. The scale varies from blue as residue pairs of the strongest interaction/rigid to red residue pairs of the weakest interaction/flexible. **(D)** A cartoon representation of a single chain of *P. falciparum* PTPS with the residues of strongest interaction mapped onto the structure and shown as red spheres. **(E)** The mean value of the effective spring constant over all pairs for each residue, with the secondary structures shown along the upper abscissa. The color bar shows strong regions/rigid in blue and weak regions/mobile in red. The figures were generated using the Matplotlib library and VMD.

metal ion (Khairallah et al., 2020). The residues Y45, L49, T127, S148, V159, E161 (as well as their neighboring residues) also corresponded to high frequency vibrating residues that were identified in the ANM (**Figure 9B**), suggesting that high-frequency vibrating residues correspond to rigid points about which significant conformational changes occur.

To further validate the high-frequency vibrating residues that were identified in the ANM, we used the DynOmics portal to analyze the mechanical properties of PTPS to identify the most rigid residues in the GNM (**Figure 12**). The results obtained from the DynOmics portal were in agreement with our findings from the ANM. Firstly, the N-terminal and central helices were predicted to have high mobility in the low-frequency modes, while the most rigid residues were located in the active site region (**Figure 12A**). Secondly, the GNM calculations performed using DynOmics located high energy hotspots in the active site region which directly overlapped with the high-frequency vibrating residues that were identified in the ANM (**Figure 12B**).

The ProDy Python package for protein structural dynamics analysis (Bakan et al., 2011, 2014) was also used, in particular, the ProDy MechStiff function to evaluate the mechanical stiffness of the PTPS protein (Eyal and Bahar, 2008). A mechanical stiffness matrix was then produced by calculating the effective force constants in response to uniaxial extensional forces exerted at each pair of residues (Eyal and Bahar, 2008; **Figure 12C**). The obtained results illustrated that pairs located in the active site and the surrounding region exhibited strong pairs of interactions. This suggests that the residues of the active site H29, H41 and H43 as well as its surrounding catalytic residues including T127 and E161 bear a relatively strong resistance to deformation. **Figure 12E** shows the results averaged over all pairs for each residue, which provides a profile of the mechanical resistance of individual residues to deformation. Some residues, especially those located in the active site and surrounding area were more rigid than others and more resistant to deformation, as indicated by this profile. More specifically, these residues involved the active site residues H29, H41, and H43 as well as S17, L78, I83, F93, T127, G131, and E161 which were also among the highly conserved, high frequency vibrating and hinge residues. A cartoon representation of a single chain of the *P. falciparum* PTPS is shown in **Figure 12D**. The figure shows pairs of residues with the strongest effective force constant and their location within the protein structure. The identified pairs of T127, E16, G131 were located around the active site area and were also among the high vibrating and hinge residues.

CONCLUSION

Antifolate drug resistance is a major challenge in the fight against malaria and the need to develop new drugs with a unique mechanism of action has become more crucial than ever. In this study, we classify the dynamics of the parasite's *de novo* folate synthesis pathway enzyme PTPS, in an attempt to uncover key regions that modulate conformational transitions which are imperative to its function. Notable global motions of functional significance were captured within the low-frequency non-degenerate modes. In particular, we showed that the opening and closing of the PTPS central tunnel was driven by the distinctive twisting and wringing of the terminal regions. Furthermore, the displacements observed in the PTPS N-terminal domains appeared to have a long-distance regulatory effect on the rigid core of the protein, located more than 30 Å away. Motif analysis further validated our findings with the identification of structural motifs that are uniquely conserved in the Plasmodium PTPS enzymes. These conserved motifs were located in the N-terminal domains, the central helices as well as the protein tunnel, and point toward the functional importance of these regions in the Plasmodium PTPS enzyme. Collectively these results suggest an opportunity for selective inhibition as these regions are not conserved in Human PTPS. Specifically, these regions can be proposed as potential allosteric sites for future antimalarial drug discovery attempts.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

ÖTB conceived the project. AK did the calculations and analysis of data under the supervision of CR and ÖTB. AK wrote the first draft. All authors edited it to the final version.

FUNDING

This work is supported by the Organization for Women in Science for the Developing World (OWSD) and the Swedish International Development Cooperation Agency (SIDA). Fund Reservation No. 3240287275. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

ACKNOWLEDGMENTS

AK would like to acknowledge the Organization for Women in Science for the Developing World (OWSD) fellowship support, the Swedish International Development Cooperation Agency (SIDA), and the Centre for High-Performance Computing (CHPC), South Africa.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.575196/full#supplementary-material

REFERENCES

Alker, A. P., Kazadi, W. M., Kutelemeni, A. K., Bloland, P. B., Tshefu, A. K., and Meshnick, S. R. (2008). dhfr and dhps genotype and sulfadoxine-pyrimethamine treatment failure in children with falciparum malaria in the Democratic Republic of Congo. *Trop. Med. Int. Health* 13, 1384–1391. doi: 10.1111/j.1365-3156.2008.02150.x

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Amamuddy, O. S., Veldman, W., Manyumwa, C., Khairallah, A., Agajanian, S., Oluyemi, O., et al. (2020). Integrated computational approaches and tools for allosteric drug discovery. *Int J. Mol. Sci.* 21, 847. doi: 10.3390/ijms21030847

Amusengeri, A., and Tastan Bishop, Ö. (2019). Discorhabdin N, a South African natural compound, for Hsp72 and Hsc70 allosteric modulation: combined study of molecular modeling and dynamic residue network analysis. *Molecules* 24, 188. doi: 10.3390/molecules24010188

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80, 505–515. doi: 10.1016/S0006-3495(01)76033-X

Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., et al. (2009). PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* 37(Suppl. 1), D539–D543. doi: 10.1093/nar/gkn814

Bahar, I., Chennubhotla, C., and Tobi, D. (2007). Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* 17, 633–640. doi: 10.1016/j.sbi.2007.09.011

Bahar, I., Lezon, T. R., Yang, L.-W., and Eyal, E. (2010). Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.* 39, 23–42. doi: 10.1146/annurev.biophys.093008.131258

Bailey, T. L., and Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14, 48–54. doi: 10.1093/bioinformatics/14.1.48

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416

Bakan, A., Dutta, A., Mao, W., Liu, Y., Chennubhotla, C., Lezon, T. R., et al. (2014). Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* 30, 2681–2683. doi: 10.1093/bioinformatics/btu336

Bakan, A., Meireles, L. M., and Bahar, I. (2011). ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* 27, 1575–1577. doi: 10.1093/bioinformatics/btr168

Bao, G. (2002). Mechanics of biomolecules. *J. Mech. Phys. Solids* 50, 2237–2274. doi: 10.1016/S0022-5096(02)00035-2

Bateman, A. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049

Bürgisser, D. M., Thöny, B., Redweik, U., Hess, D., Heizmann, C. W., Huber, R., et al. (1995). 6-pyruvoyl tetrahydropterin synthase, an enzyme with a novel type of active site involving both zinc binding and an intersubunit catalytic triad motif; site-directed mutagenesis of the proposed active center, characterization of the metal binding site and modelling of substrate binding. *J. Mol. Biol.* 253, 358–369. doi: 10.1006/jmbi.1995.0558

Chennubhotla, C., Rader, A. J., Yang, L. W., and Bahar, I. (2005). Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.* 2, S173–S180. doi: 10.1088/1478-3975/2/4/S12

Colloc'h, N., Poupon, A., and Mornon, J.-P. (2000). Sequence and structural features of the T-fold, an original tunnelling building unit. *Proteins* 39, 142–154. doi: 10.1002/(SICI)1097-0134(20000501)39:2<142::AID-PROT4<3.0.CO;2-X

DeLano, W. L. (2014). *The PyMOL Molecular Graphics System, Version 1.8*. Cambridge, MA: Schrödinger LLC, doi: 10.1038/hr.2014.17

Eyal, E., and Bahar, I. (2008). Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophys. J.* 94, 3424–3435. doi: 10.1529/biophysj.107.120733

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., and Caves, L. S. D. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 22, 2695–2696. doi: 10.1093/bioinformatics/btl461

Guarnera, E., and Berezovsky, I. N. (2020). Allosteric drugs and mutations: chances, challenges, and necessity. *Curr. Opin. Struct. Biol.* 62, 149–157. doi: 10.1016/j.sbi.2020.01.010

Haliloglu, T., Keskin, O., Ma, B., and Nussinov, R. (2005). How similar are protein folding and protein binding nuclei? Examination of vibrational motions of energy hot spots and conserved residues. *Biophys. J.* 88, 1552–1559. doi: 10.1529/biophysj.104.051342

Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. *Nature* 450, 964–972. doi: 10.1038/nature06522

Higgins, C. E., and Gross, S. S. (2011). The N-terminal peptide of mammalian GTP cyclohydrolase I is an autoinhibitory control element and contributes to binding the allosteric regulatory protein GFRP. *J. Biol. Chem.* 286, 11919–11928. doi: 10.1074/jbc.M110.196204

Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. *Proteins* 33, 417–429. doi: 10.1002/(SICI)1097-0134(19981115)33:3<417::AID-PROT10<3.0.CO;2-8

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Isin, B., Tirupula, K. C., Oltvai, Z. N., Klein-Seetharaman, J., and Bahar, I. (2012). Identification of motions in membrane proteins by elastic network models and their experimental validation. *Methods Mol. Biol.* 914, 285–317. doi: 10.1007/978-1-62703-23-6_17

Khairallah, A., Tastan Bishop, Ö, and Moses, V. (2020). AMBER force field parameters for the Zn (II) ions of the tunneling-fold enzymes GTP cyclohydrolase I and 6−pyruvoyl tetrahydropterin synthase. *J. Biomol. Struct. Dyn.* 1–18. doi: 10.1080/07391102.2020.1796800

Kümpornsin, K., Kotanan, N., Chobson, P., Kochakarn, T., Jirawatcharadech, P., Jaru-ampornpan, P., et al. (2014). Biochemical and functional characterization of *Plasmodium falciparum* GTP cyclohydrolase I. *Malaria Journal* 13, 150. doi: 10.1186/1475-2875-13-150

Lee, J. Y., Feng, Z., Xie, X. Q., and Bahar, I. (2017). Allosteric modulation of intact γ-secretase structural dynamics. *Biophys. J.* 113, 2634–2649. doi: 10.1016/j.bpj.2017.10.012

Li, H., Chang, Y. Y., Lee, J. Y., Bahar, I., and Yang, L. W. (2017). DynOmics: dynamics of structural proteome and beyond. *Nucleic Acids Res.* 45, W374–W380. doi: 10.1093/nar/gkx385

Loutchko, D., and Flechsig, H. (2020). Allosteric communication in molecular machines via information exchange: what can be learned from dynamical modeling. *Biophys. Rev.* 12, 443–452. doi: 10.1007/s12551-020-00667-8

Lu, S., Shen, Q., and Zhang, J. (2019). Allosteric methods and their applications: facilitating the discovery of allosteric drugs and the investigation of allosteric mechanisms. *Acc. Chem. Res.* 52, 492–500. doi: 10.1021/acs.accounts.8b00570

Mackenzie, C. O., and Grigoryan, G. (2017). Protein structural motifs in prediction and design. *Curr. Opin. Struct. Biol.* 44, 161–167. doi: 10.1016/j.sbi.2017.03.012

Mahajan, S., and Sanejouand, Y. H. (2015). On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. *Arch. Biochem. Biophys.* 567, 59–65. doi: 10.1016/j.abb.2014.12.020

Maita, N., Hatakeyama, K., Okada, K., and Hakoshima, T. (2004). Structural basis of biopterin-induced inhibition of GTP Cyclohydrolase I by GFRP, its feedback regulatory protein. *J. Biol. Chem.* 279, 51534–51540. doi: 10.1074/jbc.M409440200

Maximova, T., Moffatt, R., Ma, B., Nussinov, R., and Shehu, A. (2016). Principles and overview of sampling methods for modeling macromolecular structure and dynamics. *PLoS Comput. Biol.* 12:e1004619. doi: 10.1371/journal.pcbi.1004619

Nar, H. (2011). *6-Pyruvoyl-Tetrahydropterin Synthase: Encyclopedia of Inorganic and Bioinorganic Chemistry*. Hoboken, NJ: John Wiley and Sons, Ltd. doi: 10.1002/9781119951438.eibc0475

Nar, H., Huber, R., Heizmann, C. W., Thöny, B., and Bürgisser, D. (1994). Three-dimensional structure of 6-pyruvoyl tetrahydropterin synthase, an enzyme involved in tetrahydrobiopterin biosynthesis. *EMBO J.* 13, 1255–1262. doi: 10.1002/j.1460-2075.1994.tb06377.x

Nussinov, R., and Ma, B. (2012). Protein dynamics and conformational selection in bidirectional signal transduction. *BMC Biol.* 10:2. doi: 10.1186/1741-7007-10-2

Nussinov, R., and Tsai, C. J. (2013). Allostery in disease and in drug discovery. *Cell* 153, 293–305. doi: 10.1016/j.cell.2013.03.034

Nzila, A., Ward, S. A., Marsh, K., Sims, P. F. G., and Hyde, J. E. (2005). Comparative folate metabolism in humans and malaria parasites (part II): activities as yet untargeted or specific to *Plasmodium*. *Trends Parasitol.* 21, 334–339. doi: 10.1016/j.pt.2005.05.008

Orozco, M. (2014). A theoretical view of protein dynamics. *Chem. Soc. Rev.* 43, 5051–5066. doi: 10.1039/c3cs60474h

Palamini, M., Canciani, A., and Forneris, F. (2016). Identifying and visualizing macromolecular flexibility in structural biology. *Front. Mol. Biosci.* 3:47. doi: 10.3389/fmolb.2016.00047

Penkler, D., Sensoy, Ö, Atilgan, C., and Tastan Bishop, Ö (2017). Perturbation-response scanning reveals key residues for allosteric control in Hsp70. *J. Chem. Inform. Model.* 57, 1359–1374. doi: 10.1021/acs.jcim.6b00775

Penkler, D. L., Atilgan, C., and Tastan Bishop, Ö (2018). Allosteric modulation of human Hsp90α conformational dynamics. *J. Chem. Inform. Model.* 58, 383–404. doi: 10.1021/acs.jcim.7b00630

Ploom, T., Thöny, B., Yim, J., Lee, S., Nar, H., Leimbacher, W., et al. (1999). Crystallographic and kinetic investigations on the mechanism of 6-pyruvoyl tetrahydropterin synthase. *J. Mol. Biol.* 286, 851–860. doi: 10.1006/jmbi.1998.2511

Ross, C., Knox, C., and Tastan Bishop, Ö (2017). Interacting motif networks located in hotspots associated with RNA release are conserved in *Enterovirus capsids*. *FEBS Lett.* 591, 1687–1701. doi: 10.1002/1873-3468.12663

Ross, C., Nizami, B., Glenister, M., Sheik Amamuddy, O., Atilgan, A. R., Atilgan, C., et al. (2018). MODE-TASK: large-scale protein motion tools. *Bioinformatics* 34, 3759–3763. doi: 10.1093/bioinformatics/bty427

Shrivastava, I. H., and Bahar, I. (2006). Common mechanism of pore opening shared by five different potassium channels. *Biophys. J.* 90, 3929–3940. doi: 10.1529/biophysj.105.080093

Shrivastava, I. H., Liu, C., Dutta, A., Bakan, A., and Bahar, I. (2020). "Allostery as structure−encoded collective dynamics," in *Structural Biology in Drug Discovery*, ed. J. P. Renaud (Hoboken, NJ: Wiley). doi: 10.1002/9781118681121.ch6

Snow, R. W. (2015). Global malaria eradication and the importance of Plasmodium falciparum epidemiology in Africa. *BMC Med.* 13:23. doi: 10.1186/s12916-014-0254-7

Suplatov, D., and Švedas, V. (2015). Study of functional and allosteric sites in protein superfamilies. *Acta Nat.* 7, 34–45. doi: 10.32607/20758251-2015-7-4-34-54

Swarbrick, J., Iliades, P., Simpson, J. S., and Macreadie, I. (2009). Folate biosynthesis – reappraisal of old and novel targets in the search for new antimicrobials. *New Dev. Med. Chem.* 1, 12–33. doi: 10.2174/1874940200801010012

Teilum, K., Olsen, J. G., and Kragelund, B. B. (2009). Functional aspects of protein flexibility. *Cell. Mol. Life Sci.* 66, 2231–2247. doi: 10.1007/s00018-009-0014-6

Towler, P., Staker, B., Prasad, S. G., Menon, S., Tang, J., Parsons, T., et al. (2004). ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *J. Biol. Chem.* 279, 17996–18007. doi: 10.1074/jbc.M311191200

Wako, H., and Endo, S. (2011). Ligand-induced conformational change of a protein reproduced by a linear combination of displacement vectors obtained from normal mode analysis. *Biophys. Chem.* 159, 257–266. doi: 10.1016/j.bpc.2011.07.004

World Health Organization (2018). *The World Malaria Report 2018*. Geneva: Who.

Yang, L. W., and Bahar, I. (2005). Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure* 13, 893–904. doi: 10.1016/j.str.2005.03.015

Zhang, Y., Doruker, P., Kaynak, B., Zhang, S., Krieger, J., Li, H., et al. (2020). Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Curr. Opin. Struct. Biol.* 62, 14–21. doi: 10.1016/j.sbi.2019.11.002

Zheng, F., and Grigoryan, G. (2017). Sequence statistics of tertiary structural motifs reflect protein stability. *PLoS One* 12:e0178272. doi: 10.1371/journal.pone.0178272

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

REVIEW
published: 19 October 2020
doi: 10.3389/fmolb.2020.583826

Enzyme Kinetics by Isothermal Titration Calorimetry: Allostery, Inhibition, and Dynamics

*Yun Wang, Guanyu Wang, Nicolas Moitessier and Anthony K. Mittermaier**

Department of Chemistry, McGill University, Montreal, QC, Canada

Isothermal titration calorimetry (ITC) involves accurately measuring the heat that is released or absorbed in real time when one solution is titrated into another. This technique is usually used to measure the thermodynamics of binding reactions. However, there is mounting interest in using it to measure reaction kinetics, particularly enzymatic catalysis. This application of ITC has been steadily growing for the past two decades, and the method is proving to be sensitive, generally applicable, and capable of providing information on enzyme activity that is difficult to obtain using traditional biochemical assays. This review aims to give a broad overview of the use of ITC to measure enzyme kinetics. It describes several different classes of ITC experiment, their strengths and weaknesses, and recent methodological advancements. A summary of applications in the literature is given and several examples where ITC has been used to investigate challenging aspects of enzyme behavior are presented in more detail. These include examples of allostery, where small-molecule binding outside the active site modulates activity. We describe the use of ITC to measure the strength, mode (i.e., competitive, uncompetitive, or mixed), and association and dissociation kinetics of enzyme inhibitors. Further, we provide examples of ITC applied to complex, heterogeneous mixtures, such as insoluble substrates and live cells. These studies exemplify the wide range of problems where ITC can provide answers, and illustrate the versatility of the technique and potential for future development and applications.

Keywords: enzyme catalysis, inhibition, activation, allostery, kinetics, ITC

OPEN ACCESS

Edited by:
Pemra Doruker,
University of Pittsburgh, United States

Reviewed by:
Lee D. Hansen,
Brigham Young University,
United States
Igor N. Berezovsky,
*Bioinformatics Institute (A*STAR),*
Singapore

***Correspondence:**
Anthony K. Mittermaier
anthony.mittermaier@mcgill.ca

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *15 July 2020*
Accepted: *11 September 2020*
Published: *19 October 2020*

Citation:
Wang Y, Wang G, Moitessier N and Mittermaier AK (2020) Enzyme Kinetics by Isothermal Titration Calorimetry: Allostery, Inhibition, and Dynamics. Front. Mol. Biosci. 7:583826. doi: 10.3389/fmolb.2020.583826

INTRODUCTION

Enzymes are catalytic proteins that are ubiquitous in living systems and play central roles in virtually all cellular processes, such as metabolism, active transport, sensing, regulation, communication, and signal transduction and integration (Hunter, 1995; Capaldi and Aggeler, 2002; Benhar et al., 2009; Reyes-Turcu et al., 2009). Consequently enzymes constitute approximately 44% of all validated drug targets, including human enzymes whose dysregulation is linked to disease, and foreign enzymes expressed by pathogens (Zheng et al., 2006). In addition, enzymes are the most efficient catalysts known and have many industrial and medical applications (Choi et al., 2015). For example, hydrolases break polysaccharides down into their component sugars, with applications to food processing, pulp and paper, and biofuel industries (Guzman-Maldonado and Paredes-Lopez, 1995; Sun and Cheng, 2002; Kuhad et al., 2011). Their high selectivity and

biocompatibility have also made enzymes useful as therapeutics, for instance in the treatment of phytobezoars (Kramer and Pochapin, 2012).

In general, enzymes show saturation kinetics, which can be rationalized according to the Michaelis–Menten/Briggs–Haldane (MM/BH) model shown in the scheme below

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P \quad (1)$$

where an enzyme molecule (E) binds a substrate (S) with association and dissociation rate constants k_1 and k_{-1}, respectively, to form the Michaelis complex (ES). The enzyme then acts on the substrate to produce the product (P) with a rate constant k_{cat}. This kinetic scheme gives rise to the familiar MM/BH equation where the enzyme velocity, v_0, has a saturable dependence on the substrate concentration:

$$v_0 = \frac{d[P]}{dt} = -\frac{d[S]}{dt} = \frac{V_{max}[S]}{K_m + [S]} \quad (2)$$

V_{max} is the maximum rate of catalysis in the theoretical presence of an infinite quantity of substrate and K_m is the concentration of substrate required to achieve half-maximal velocity, as illustrated in **Figure 1**. In terms of the rate constants in Scheme 1,

$$V_{max} = k_{cat}[E] \quad (3)$$

and

$$K_m = \frac{k_{-1} + k_{cat}}{k_1} \quad (4)$$

The relationship between enzyme velocity and substrate concentration can be linearized according to the double-reciprocal or Lineweaver–Burk plot, in which ${v_0}^{-1}$ is plotted as a function of $[S]^{-1}$, shown below:

$$\frac{1}{v_0} = \frac{K_m}{V_{max}[S]} + \frac{1}{V_{max}} \quad (5)$$

The slope of the resulting straight line is K_m/V_{max}, the x-intercept is $-{K_m}^{-1}$ and the y-intercept is ${V_{max}}^{-1}$. The parameters K_m and k_{cat} provide simple metrics of an enzyme's behavior and quantify how activity changes in response to changing solution conditions, addition of inhibitors or activators, changes in the amino acid sequence of the enzyme, chemical modification of the substrate, or exchanging one cofactor for another, among other factors. Thus, methods for measuring K_m and k_{cat} are among the foundational techniques of molecular biosciences.

Most enzyme assays measure the concentrations of substrate and/or product as a function of time. The rates of disappearance and/or appearance give the enzyme velocity, which can be fitted according to Equations 2 or 5. Note that care must be taken in the choice of enzyme and substrate concentrations in order to ensure that both K_m and k_{cat} can be robustly extracted from the data (Stroberg and Schnell, 2016). These experiments can be classified in two types: continuous (or real-time) and discontinuous assays (Harris and Keshwani, 2009). In a continuous assay, the concentrations of substrates or products are measured in the reaction mixture at the same time as catalysis proceeds. For the most part, they employ spectroscopies, such as fluorimetry, UV/vis absorption, or nuclear magnetic resonance, and rely on substrates and products having different spectroscopic signatures (Easterby, 1973; Seethala and Menzel, 1997; Reetz, 2001; Shu and Frieden, 2005). While this is sometimes true in the native reaction, in many cases continuous assays require experimental modifications. Substrates can be chemically altered so that they change color or fluoresce when converted to products (Reetz, 2001). While convenient, this approach has the drawback that a customized substrate must be produced for each enzyme of interest, and non-native chromogenic or fluorogenic substrates do not necessarily have the same reaction kinetics as the natural substrate. Alternatively, in coupled enzyme assays, the reaction mixture includes secondary enzymes that accept the product of the first enzymatic reaction as a substrate and produce downstream spectroscopic changes, such as the interconversion of NAD^+ and NADH that have very different extinction coefficients for light at 340 nm (Easterby, 1973; McKay and Wright, 1995). This approach allows native substrates to be used, but the assay places limitations on the composition of the reaction mixtures, for example product inhibition or activation studies are impossible (McKay and Wright, 1995) and accurate results depend on choosing appropriate concentrations of the coupled enzymes and secondary substrates. When it is not possible to monitor substrate or product concentrations in real time, discontinuous enzyme assays must be used. In these experiments, the reaction is quenched at various time points after initiation and the substrates and products are separated by an ancillary technique, such as liquid chromatography, gel electrophoresis, centrifugation, or mass spectrometry (Reetz et al., 2004; Hooff et al., 2012; Tauran et al., 2014) and quantified, for instance spectroscopically, radiometrically, or by an immunosorbent assay (Butler, 2000; Kerner and Hoppel, 2002; Hastie et al., 2006). These additional steps add time, expense, and uncertainty to the characterization process.

Isothermal titration calorimetry (ITC) is well known as a powerful tool for studying host/guest binding interactions, but has recently gained in popularity as a general and versatile kinetic assay (Todd and Gomez, 2001; Di Trani et al., 2018a,b). ITC has the advantage of directly measuring the heat flow produced by catalysis in real time (Todd and Gomez, 2001; Di Trani et al., 2017). Since most chemical reactions are either exothermic or endothermic, ITC can be applied to study virtually any enzymatic reaction, without the need for customized reporter molecules, additional coupled enzymes, or post-reaction separation. Furthermore, kinetic ITC experiments can be performed with conventional dilute enzymatic reaction mixtures, even with opaque samples, and require far less enzyme than ITC binding studies (Oezen and Serpersu, 2004). The study of enzyme kinetics has been briefly described in several surveys of the ITC field (Freyer and Lewis, 2008; Liang, 2008; Ghai et al., 2012; Atri et al., 2015) and has been the focus of more technically detailed reviews (Hansen et al., 2016; Mazzei et al., 2016). Here, we discuss how ITC can be applied to a broad array of problems in enzyme biochemistry, including understanding inhibition and allosteric modulation and studying heterogeneous reaction mixtures, from the perspective of our own work in

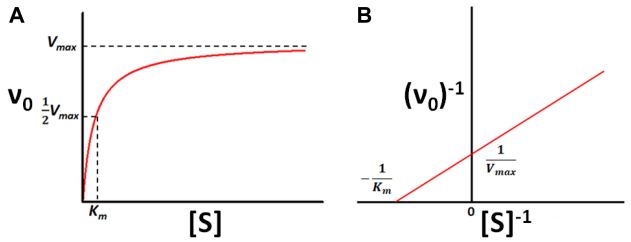


FIGURE 1 | Michaelis-Menten kinetics. **(A)** Velocity (v_0) of a typical enzyme-catalyzed reaction versus substrate concentration ([*S*]). **(B)** Double-Reciprocal (Lineweaver–Burk) linearized plot ($1/v_0$ vs. $1/[S]$) of the rates in **(A)**.

the field. We have tried to choose examples that illustrate how ITC studies can go beyond measuring the parameters usually associated with the term "enzyme kinetics" such as K_m, k_{cat}, K_i, etc. and extend to observing additional dynamic phenomena like inhibitor association and release, substrates slowly entering the bacterial periplasm, or rearrangements of crystalline chitosan, as described below.

ENZYME KINETICS BY ISOTHERMAL TITRATION CALORIMETRY

ITC Instrumentation

Isothermal titration calorimetry instruments measure in real time the thermal power that results when one solution (in a syringe) is titrated into another (in a sample cell), as illustrated in **Figure 2**. A pair of cells, typically coin-shaped or cylindrical with volumes on the order of 200–1,400 μL, are termed the sample and reference cells and contain the analyte solution and reference buffer (or pure water) respectively (Malvern, 2016; TA, 2019). The cells are housed inside a thermostated adiabatic jacket, that is maintained at a temperature slightly below the user-specified value for the cells. Electric resistive heaters, termed the feedback and reference heaters are located on the outer surfaces of the sample and reference cells, respectively, and must supply a constant flow of heat to maintain the cell temperatures at their set point. A Seebeck device sandwiched between the two cells detects any differences in temperature (ΔT) and modulates the power supplied to the feedback heater in order to keep the temperatures of the two cells identical. An automated injection syringe protrudes into the sample cell, which is stirred either by rotation of the paddle-shaped syringe, or by the action of a separate propeller, depending on the make and model of the instrument. A series of injections (typically between 1 and 20 μL) is made into the sample cell. If the reaction between the injectant and analyte is exothermic, there will be a concomitant drop in the power supplied by the feedback heater to maintain a constant temperature. Conversely, if the reaction is endothermic, there will be an increase in feedback power. Once the reaction is complete or the rate becomes negligible, and no further heat is produced or absorbed in the sample cell, the feedback power returns to baseline. The raw output of an ITC instrument is the feedback power measured as a function of time (typically at 1 s intervals). When characterizing binding or reaction thermodynamics, the deflection of the ITC signal from baseline is integrated over the entire injection, and is used to extract enthalpy differences between the unreacted and reacted states (i.e., free vs. bound or substrates vs. products). When characterizing kinetics, the instantaneous output power is interpreted in terms of the reaction velocity, since the rate of heat production or absorption in sample cell is directly proportional to the rate of the reaction. This is slightly complicated by the fact that the ITC signal lags behind heat events in the cell, however, there are several approaches to overcoming this issue, as discussed in later sections. Furthermore, it should be noted that obtaining accurate reaction rates requires accurate heat rates, so it is important to calibrate the calorimetric response (Demarse et al., 2011).

ITC Kinetics Methods

The instantaneous rate of heat production in the ITC sample cell, dQ/dt, is directly proportional to the reaction velocity ($v_0 = d[P]/dt$) and the enthalpy change of the reaction catalyzed ($\Delta_r H = H_{product} - H_{substrate}$), according to

$$\frac{dQ}{dt} = V_{cell}\Delta_r H\frac{d[P]}{dt} \tag{6}$$

where V_{cell} is the volume of the sample cell. Thus with ITC-derived dQ/dt values obtained as a function of time, it is straightforward to precisely calculate enzyme velocity at any point in the experiment, provided $\Delta_r H$ and V_{cell} are known. This is obtained from the integrated area of an ITC peak obtained by injecting a known amount of substrate into a sample cell containing sufficient enzyme to rapidly convert it entirely to product,

$$\Delta_r H = \frac{\int_{t=0}^{\infty}\frac{dQ}{dt}dt}{n_S} \tag{7}$$

where n_S is the number of moles of substrate injected. In their seminal 2001 paper, Todd and Gomez describe two main approaches for designing ITC experiments that rapidly measure v_0 as a function of substrate concentration, allowing the enzyme kinetic parameters to be extracted by fits to Equations 2 or 5. They referred to these as "Pseudo-first Order" and "Continuous" assays, although these terms have been largely replaced with "multiple injection" and "single injection" and we will use the latter terms here. A broad variety of ITC enzyme kinetics experiments have been developed in subsequent years, however, most build on one or the other approach, so it is worthwhile to describe them in some detail, as foundational to the field. In both types of experiment, the reaction is initiated one or more times by mixing enzyme and substrate solutions via injection(s) from the syringe into the sample cell. However, the two methods differ in the concentrations of enzyme and substrate used, the appearance and information content of the data, and the analysis.

Multiple Injection Assays

In a multiple injection ITC enzyme kinetic assay, the enzyme concentration is chosen to be sufficiently low so that substrate

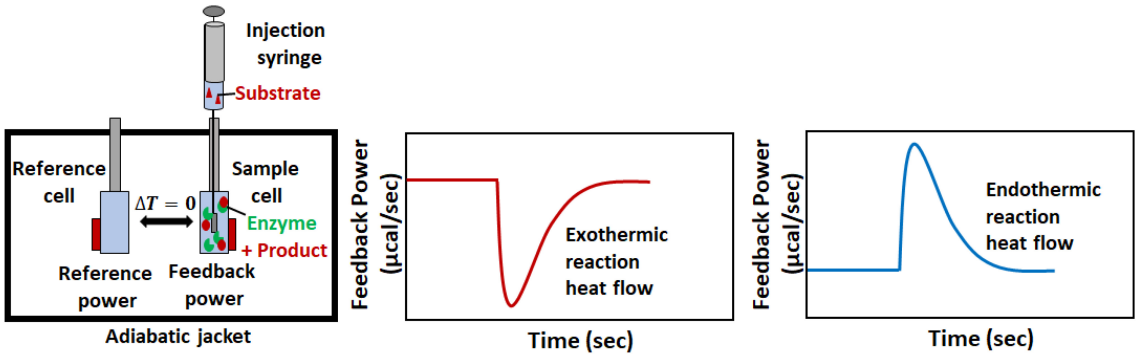


FIGURE 2 | A typical ITC enzyme kinetics experiment. The reaction is initiated when substrate in the syringe is injected into the sample cell containing enzyme. If the reaction is exothermic (endothermic), less (more) feedback power must be supplied to the sample cell to keep it at the same temperature as the reference cell. The instantaneous value of the feedback power is the ITC output.

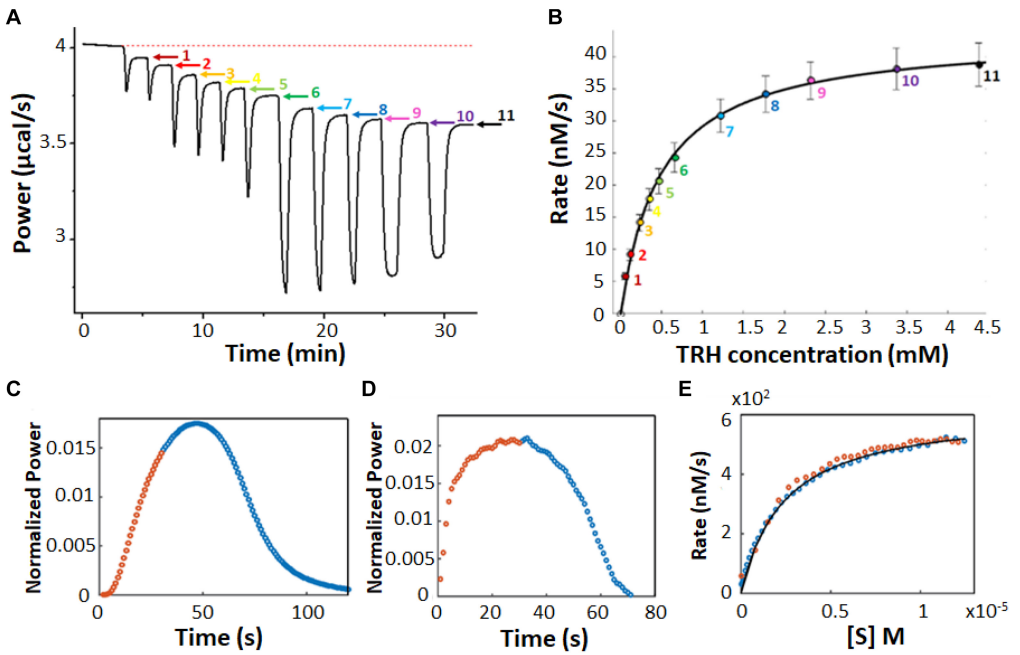


FIGURE 3 | Multiple injection and single injection ITC enzyme kinetic data. **(A)** Multiple injection assay of prolyl oligopeptidase in the sample cell and one of its substrates, thyrotropin releasing hormone, in the syringe (Di Trani et al., 2017). The downward spikes correspond to dilution artifacts from each injection (3, 3, 6, 6, 6, 10, 30, 30, 30, 60, and 60 µL). Larger injections produce larger spikes. The displacement, following each injection, of the horizontal baseline relative to the initial baseline (red dotted line) is proportional to the enzyme velocity. **(B)** MM/BH plot calculated from the data in **(A)**. Error bars correspond to the standard deviations of three repeat experiments. **(C)** Single injection assay with trypsin in the sample cell and one of its substrates, benzoyl-L-arginine ethyl ester, in the syringe (Di Trani et al., 2018b). Data collected during (after) the 30 s injection are plotted in orange (blue). **(D)** Deconvolution of the data in **(C)** to remove the effect of the delayed instrument response according to Equation 10 and an empirical response function. Note that during the first 30 s, the substrate is injected into the sample cell faster than it is consumed and its concentration gradually increases, while the reaction velocity asymptotically approaches the maximum V_{max} value, in accordance with the MM/BH Equation (orange circles). After the injection ends, the substrate continues to be consumed and its concentration gradually drops, while the reaction velocity decreases to zero once more (blue circles). **(E)** MM/BH plot generated from the data in **(D)** using Equation 8. The rate versus [*S*] values are superimposable for the injection (orange, increasing [*S*]) and post-injection (blue, decreasing [*S*]) halves of the experiment, providing cross-validation for the data.

depletion during the experiment is negligible but high enough to provide good signal (Todd and Gomez, 2001). As a result, the instantaneous heat (dQ/dt) and ITC signal are ideally constant (horizontal) between substrate injections and resemble a series of steps, one per injection (**Figure 3A**). The displacement of each step relative to the initial baseline is directly proportional to v_0, according to Equation 6. Exothermic and endothermic reactions give descending and ascending steps, respectively, if the raw feedback power is plotted as a function of time. The injections are designed such that early steps have $[S] << K_m$ and the final injections have nearly saturated the enzyme with $[S] >> K_m$. The concentration of substrate present in the sample cell after each

injection is known from the concentration of substrate in syringe and volumes of all injections, while the reaction velocity can be read directly from the vertical position of each step, tracing out a complete Michaelis Menten curve (**Figure 3B**). In practice, we find that the condition of negligible substrate consumption is met when $[E] \leq (10^{-4}\ \mathrm{s}) \times K_\mathrm{m}/k_\mathrm{cat}$. Enzyme concentrations that are too high will give steps that slope toward the initial baseline, and will lead to overestimates in the amount of substrate present at each step. Enzyme concentrations that are too low will lead to disappearingly small steps that are obscured by instrument noise.

There are several potential advantages to multiple injection assays compared to single injection ones. Firstly, they can accommodate substantially lower enzyme concentrations. For example in **Figures 3A,B**, saturation is reached at about 4.5 mM substrate with a V_max of 40 nM s^{-1}. At that rate, it would take more than 10^5 s or 28 h for the enzyme to convert a sufficient quantity of S to P to complete a single injection assay (see below), which is too long for practical purposes. It should be noted that $\Delta_r H$ must be determined in a separate measurement for multiple injection assays, while it is obtained directly from single injection data, thus comparable amounts of enzymes can be consumed when all the necessary experiments are factored in. Secondly, the readout portions of the experiment, i.e., the approximately horizontal signals, are easy to distinguish from injection artifacts, which themselves tend to be smaller since less substrate is added in each injection. Secondly, product accumulation is also less than for single injection assays. In a single injection assay, the amount of product present near the end of an ITC peak is necessarily several-fold greater than the K_m, since the enzyme is initially saturated with substrate. In contrast, much less substrate is converted to product during a multiple-injection experiment, ideally less than 5% (Hansen et al., 2016). Thus much less product is produced during a multiple injection assay compared to a single injection one. This is advantageous when strong product inhibition is present (Wang et al., 2019), although conversely, if product inhibition is of interest yet relatively weak, single injection assays would be the more sensitive option. Furthermore, since the data are drawn from post-injection periods where the enzyme velocities have stabilized to constant values, the timescale of the instrument response to changing heat flow can be ignored, simplifying the analysis (as described below). The main disadvantages are that the determination of a single pair of k_cat and K_m values requires a complete series of injections, making this technique relatively slow, and that the total amount of heat generated is much less, making it more susceptible to instrument noise.

Single Injection Assays

In a single injection ITC kinetic assay, the amount of enzyme is typically chosen to be large enough so that the injected substrate can be fully converted to product on the timescale of minutes or tens of minutes. The concentration of substrate is chosen so that the injection appreciably saturates the enzyme, i.e., the concentration of substrate in the sample cell immediately after the injection is several-fold higher than the K_m (Transtrum et al., 2015; Di Trani et al., 2018b). Single injection assays can be initiated either by injecting substrate (syringe) into enzyme (cell) or enzyme (syringe) into substrate (cell). Either case, the ITC feedback power exhibits a large deflection immediately after the injection, decreasing for exothermic reactions and increasing for endothermic ones due to the heat released or absorbed by catalysis. Large heat flows continue as long as the enzyme remains saturated with substrate. The signal gradually returns to the pre-injection baseline as the substrate is consumed as shown in **Figure 3C**. It should be noted that single injection assays are usually performed with substantially more dilute enzyme, leading to peaks that are much broader, on the order of 20–60 min, in contrast to the 100 s shown here. Our group has been developing ways to collect and analyze data for rapid single injection experiments, such as those shown in **Figure 3C**, offering over 10-fold reductions in measurement time and opening the door to new types of experiment (Di Trani et al., 2018b; Wang et al., 2019). Single-injection ITC data can be fitted directly by numerically integrating Equation 2 to give $[S](t)$ and $v_0(t)$, and calculating dQ/dt as a function of time according to Equation 6, or using non-linear least squares optimization to find the values of K_m and k_cat that best reproduce the experimental values. This approach has the advantage that the baseline and instrument response time (see below) can be fitted along with the MM/BH parameters (Transtrum et al., 2015; Di Trani et al., 2017). Alternatively, the concentration of substrate present at any time, t, during the heat spike can be calculated by recognizing that the fraction of total substrate remaining at t is equal to the ratio of the heat generated after time t relative to the total amount of heat generated during the heat spike:

$$[S]\,(t) = \frac{\int_t^\infty \frac{\mathrm{dQ}}{\mathrm{dt}}\mathrm{dt}}{\int_0^\infty \frac{\mathrm{dQ}}{\mathrm{dt}}\mathrm{dt}}\,[S]\,(t=0) \tag{8}$$

Together, the v_0 and $[S]$ values trace out a complete Michaelis–Menten curve. We find that substrate is consumed sufficiently rapidly for this technique to be applied when $[E] \geq (10^{-2}\ \mathrm{s}) \times K_\mathrm{m}/k_\mathrm{cat}$. When the concentration of enzyme is too low, the heat spike persists for such a long time (several hours or more) that the return to baseline is difficult to distinguish. However, the enzyme must be at a low enough concentration so that the return to baseline takes at least seconds to tens of seconds. More rapid reactions start to become obscured by the response function of the instrument (as described below) (Di Trani et al., 2017).

The defining feature of this approach is that a full enzyme kinetic characterization is achieved in a single injection. Thus, with substrate in the syringe, it is straightforward to perform many single injection measurements within the same ITC experiment, simply by programming several injections (as many as 10 or 20) spaced at appropriate intervals (Cai et al., 2001; Ertan et al., 2012; Pedroso et al., 2014; Siddiqui et al., 2014; Maximova and Trylska, 2015; Di Trani et al., 2018b; Mason et al., 2018; Abis et al., 2019; Maximova et al., 2019; Wang et al., 2019). For the sake of clarity, we will refer to these as recurrent single injection experiments, to distinguish them from the very different approach termed multiple injection experiments (see section "Multiple Injection Assays," above). At the simplest level, recurrent single injection experiments

provide repeat measurements of enzyme kinetic parameters and improve the effective signal to noise ratio, although they do not replace true replicate experiments for estimating parameter uncertainties. Catalytic activity is repeatedly characterized over a period of time, giving information on the stability of the enzyme (Mason et al., 2018). These experiments also provide a sensitive measure of product inhibition (or activation), since the product accumulates in the sample cell with each injection (Cai et al., 2001; Wang et al., 2019). The recurrent single injection approach can be adapted to rapidly characterize other types of inhibition as well, as described below (Di Trani et al., 2018b).

Alternatively, single-injection assays can be performed with enzyme in the syringe. This variation is preferable for substrates that are poorly soluble, or those that form suspensions rather than solutions, since they can remain at working (diluted) concentration in the sample cell with constant stirring throughout the experiment (Lonhienne et al., 2001; Ali et al., 2013a,b; Commin et al., 2013; Pedroso et al., 2014; Lehoczki et al., 2016; Kaeswurm et al., 2019). Similarly, if very high substrate concentrations (100 s of mM) are needed (e.g., for enzymes with very large K_m values), it can be unfeasible to inject sufficient amounts of substrate without generating large injection heat artifacts, related to the large dilutions. Instead, the concentrated substrate can be equilibrated in the sample cell and small injections of dilute enzyme can be used to initiate the reaction (Lonhienne and Winzor, 2002; Lonhienne et al., 2003). Lastly, the barrel of the injection syringe lies exterior to ITC insulated jacket. If experiments are being performed at temperatures approaching the enzyme melting point, placing the enzyme in the syringe (which is at ambient temperature) allows it to spend as little time as possible at high and destabilizing temperatures (Ali et al., 2013a,b, 2015). It is worth noting that with enzyme in the syringe, recurrent single injection assays, as described above, are not possible, since the maximum concentration of substrate is necessarily present at the beginning of the measurement and cannot be replenished once conversion to product is complete. Another consideration is that small amounts of material can leak from the tip of the injection syringe during the long initial equilibration step, as well as between injections (although these delays are shorter). While this is true for both substrate- and enzyme-injection setups, the leakage is potentially far more serious with enzyme in the syringe, as this can act on the substrate in sample cell throughout the equilibration period, consuming much or all of it before the experiment has begun. In contrast, leakage of a few μL of substrate from the tip of the syringe does not dramatically imperil the procedure. Consequently, it is recommended to employ a buffer "plug" when injecting enzyme, a few μL of buffer that is drawn up into the needle after loading the syringe with an enzyme solution (Malvern, 2010, Malvern, 2014).

Rapid Enzyme Kinetics Measured by ITC

In many cases, the ITC signal can be considered nearly equal (and technically opposite) to the instantaneous rate of heat generation in the sample cell (i.e., ≈−dQ/dT). This approximation holds when the relevant portions of the heat signal vary slowly with time, such as in multiple injection assays and in cases where the peaks for single injection assays are broad (tens of minutes). For short reactions with rapidly varying heat signals, the situation becomes substantially more complicated. There are several physical processes that must occur before the heat generated by enzymatic catalysis is detected in the ITC output (Todd and Gomez, 2001; Burnouf et al., 2012). These include a heat transfer delay, which is the length of time necessary for the solid phase thermocouple to detect the small change in sample cell temperature (Wiseman et al., 1989; Freire et al., 1990; García-Fuentes et al., 1998; Ebrahimi et al., 2015) and the electronic response that alters the power supplied to the feedback heater, driving the temperature gradient between the cells back to zero (Wiseman et al., 1989). These steps are typically described collectively as an instrument response function, $f(t)$ which can be thought of as the instrument signal that would result from an instantaneous burst of heat being released in the sample cell. If the release of heat in the sample cell is described by the time-dependent function $h(t)$, then the instrument output is given by

$$g(t) = f(t) \otimes h(t) = \int_0^t f(\tau)\, g(t-\tau)\, d\tau \tag{9}$$

where $\otimes$ indicates the convolution. The finite instrument response has the effect of spreading out the observed signal compared to the actual heat profile, such that peaks begin more gradually and die away more slowly. The instrument response function is often assumed to have a simple exponential shape (López-Mayorga et al., 1987; García-Fuentes et al., 1998; Velázquez-Campoy et al., 1999; Burnouf et al., 2012; Vander Meulen and Butcher, 2012) $f(t) \propto exp\left\{-\frac{t}{\tau}\right\}$, where τ is referred to as the response time and is typically on the order of 5–15 s (Burnouf et al., 2012). Accounting for the instrument response can be done in one of two ways. In the first, non-linear least squares fitting can be used to find the enzyme kinetic parameters that generate an instantaneous heat function, $h(t)$, which when convoluted with the assumed instrument response function, $f(t)$, best reproduces the ITC peak shape, $g(t)$. In the second, one can use the Tian equation and the assumed value of τ to mathematically remove the spreading effect of the instrument response (Backman et al., 1994).

We have recently shown that the assumption of a simple exponential response function is incompatible with experimental ITC peak shapes, and that $f(t)$ is a more complicated function of time. We found that the response function can instead be equated to the signal obtained from very short (0.1 s) injections of a model host/guest system, such as EDTA injected with Ca^{2+}. We termed this approach an empirical response model (ERM), and it reproduces ITC peaks quantitatively, producing sub-second time resolution (Di Trani et al., 2017). The empirical response model can be used for direct fitting to raw ITC data, according to Equation 9, or can be used to deconvolute the instantaneous heat function from the instrument response (Di Trani et al., 2017, 2018b). This second approach relies on the convolution theorem which states that, given Equation 9, then

$$\mathcal{F}\left(g(t)\right) = \mathcal{F}\left(f(t)\right) \cdot \mathcal{F}\left(h(t)\right) \tag{10}$$

where $\mathcal{F}$ indicates the Fourier transform. The deconvoluted instantaneous heat function is then given by $h(t) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(g(t))}{\mathcal{F}(f(t))}\right)$, as exemplified in **Figures 3C–E**. It must be emphasized that the instrument response $[f(t)]$ varies with the manufacturer and model, as well as the temperature, solution viscosity, and stirring speed, among other factors, and must be measured using very short injections (e.g., Ca^{2+}/EDTA) performed under conditions as close to those of the experiment of interest as possible (Di Trani et al., 2017). This approach is only really necessary when measuring reactions that take place on the same timescale as the instrument response (i.e., roughly less than 20–30 s). For slow reactions that take tens of minutes or more to complete, the instrument response can be largely ignored, while for intermediate timescale reactions, the approximation of a single instrument response time, τ, is adequate.

An alternative approach, termed initial rate calorimetry (IrCal), avoids the issue of modeling the instrument response function altogether (Honarmand Ebrahimi et al., 2015). Honarmand Ebrahimi et al. (2015) found empirically that the initial slope of the ITC signal is proportional to the peak velocity of the enzyme after the injection. The constant of proportionality can be determined by calibration experiments. A series of substrate injections of different sizes is made, and the initial slopes of the injections reveal how the v_0 varies with $[S]$. One drawback of this approach is that each injection gives only a single v_0 value, in contrast to a typical ITC single injection assay, which yields tens to hundreds of v_0 values at different $[S]$.

ITC Enzyme Kinetics Applications

Overview

We have performed a comprehensive search of the scientific literature and identified 73 publications between 2001 and 2019 reporting ITC-derived kinetic data on 59 different enzymes including hydrolases, transferases, oxidoreductases, lyases, ligases, and a protein folding chaperonin, listed in **Supplementary Table 1**. The authors explained their choice of ITC with a variety of reasons, including that ITC can represent the only continuous assay available, that it can exploit the native substrate where alternative continuous assays require chemically-modified chromogenic or fluorogenic substrates, that ITC avoids potential artifacts associated with coupled enzyme assays, and that ITC allows continuous assays to be performed on heterogeneous and spectroscopically opaque mixtures. multiple injection-type ITC experiments were used for 35 enzymes, single injection-type ITC experiments were used for 27 enzymes, and enzyme-injection assays were used for 8 enzymes. Several of these publications focused on the development of new ITC kinetics approaches, such as IrCal and ERM above, and others are described below. Many of these studies focused on characterizing homogeneous enzymes exhibiting classical MM/BH kinetics. However, many others described more complex systems, such as enzymes with cooperative kinetics, those interacting with allosteric effectors or inhibitors, and those in heterogeneous media, such as insoluble hydrated polymers or even living cells. We describe some interesting examples from our own work and the work of others below.

Allostery and Cooperativity

Allostery is a key feature of biological systems in which covalent modification or ligand binding at one site influences the activity at distant sites in a macromolecule or macromolecular assembly. Allosteric regulation plays a central role in metabolism and cell signaling (Guarnera and Berezovsky, 2019) and has been identified as a source of new drug targets (Monod et al., 1965; Koshland et al., 1966; Perutz, 1989; DeDecker, 2000; Fenton, 2008; Guarnera and Berezovsky, 2020); thus, detailed descriptions of allostery have far-reaching implications (Zhang et al., 2020). For example, the downstream products of a biosynthetic pathway can down-regulate the activity of the enzyme catalyzing the first committed step, maintaining balance between different branches of core metabolism through the process of feedback inhibition (Spencer and Raffa, 2004; Gerhart, 2014). Alternatively, enzymes may require allosteric activators in order to function, providing an extra layer of control (Willemoës and Sigurskjold, 2002; Lunn et al., 2008). In the special case that the substrate itself acts as an allosteric effector, enzyme kinetics necessarily deviate from the classical MM/BH model. This can often be accounted for mathematically with a Hill coefficient of cooperativity, n, such that the enzyme velocity is given by the expression

$$v_0 = \frac{V_{\max}[S]^n}{K_{\mathrm{m}}^{\mathrm{n}} + [S]^n} \tag{11}$$

Values of $n > 1$ indicate positive cooperativity, such that substrate binding makes an enzyme more active toward additional substrates, and give characteristically sigmoidal v_0 vs. $[S]$ plots. In a simple interpretation, an enzyme with a given Hill coefficient, n, either binds exactly n molecules of substrate or none at all. When binding a molecule of substrate at an allosteric site reduces enzyme activity toward additional substrates (substrate inhibition), enzyme velocity can often be described by the expression

$$\frac{d[P]}{\mathrm{dt}} = \frac{V_{\max}[S] + V'_{\max}\left(\frac{[S]^2}{K'_{\mathrm{i}}}\right)}{K_{\mathrm{m}} + [S]\left(1 + \frac{K_{\mathrm{m}}}{K_{\mathrm{i}}}\right) + \left(\frac{[S]^2}{K'_{\mathrm{i}}}\right)} \tag{12}$$

where $V_{\max}$ is the maximum velocity of the reaction when the allosteric site is empty, $V'_{\max}$ is the maximum velocity when the allosteric site is filled, and K_{i} and K'_{i} are the equilibrium dissociation constants for substrate binding at the allosteric site when the active site is empty and filled, respectively (Jeoh et al., 2005).

Isothermal titration calorimetry represents a powerful tool for characterizing complex enzyme allosteric interactions. For instance, ITC was used to measure the kinetics of pyruvate kinase (PK) (Lonhienne and Winzor, 2002; Lonhienne et al., 2003) which catalyzes the transfer of a phosphate from phosphoenolpyruvate to ADP as part of the last step of glycolysis. Allosteric binding of the amino acid phenylalanine (Phe) shifts PK to an inactive form, and is believed to be related to cellular damage in the genetic disease phenylketonuria (Hörster et al., 2006; van Spronsen et al., 2009). Lonhienne and Winzor (2002), Lonhienne et al. (2003) were interested in how the presence

of osmolytes affected the active/inactive transition. PK has traditionally been studied using a coupled enzyme assay, which is suboptimal for studying the effects of high concentrations of osmolytes, since it is challenging to distinguish effects on PK from effects on the secondary enzymes of the coupled assay. ITC avoids these issues, since enzyme activity is detected directly. They performed enzyme injection assays (**Figure 4A**), where the displacement in the baseline after enzyme is injected is proportional to the velocity of the reaction. While this approach avoids large injection artifacts, it is somewhat time consuming as separate experiment must be performed for each data point in **Figure 4B**. They obtained standard MM/BH curves in the absence of Phe. In the presence of 6 mM Phe, the curve shifts to right, indicating a lower substrate affinity, and develops sigmoidal character, a hallmark of positive cooperativity. Interestingly, addition of the osmolyte proline shifted the curve back to the original location, corresponding to a return of the inactive state to the active state.

In another example, Rohatgi et al. (2015) used ITC to fully characterize the complex kinetic mechanism of gluconokinase, which transfers a phosphate from ATP to the common nutrient gluconate. They used multiple injection enzyme assays, injecting gluconate into enzyme at a constant ATP concentration. The traces clearly show declining activity at higher substrate concentrations indicative of substrate inhibition (**Figures 4C,D**). Interestingly the shapes of the plots varied as a function of ATP concentration in a way that was consistent with substrate inhibition occurring via the formation of an enzyme·ADP·gluconate ternary complex.

Our lab recently used ITC to characterize prolyl-oligopeptidase (POP) a validated drug target for multiple myeloma (Di Trani et al., 2018a) allowing study of the native peptide substrate, rather than the chemically-modified colorigenic substrate analog that is typically used. We performed ITC single injection assays and found that at lower enzyme concentrations, data were well fit by the standard MM/BH equation, while at higher concentrations, cooperativity became more apparent, with $n > 2$ at an enzyme concentration of 2 μM (**Figures 4E,F**). At the high enzyme concentrations where this behavior becomes apparent, the reactions go to completion in 10 s or less, underlining the potential of ITC to characterize rapid reaction kinetics.

More exotic ITC thermograms were obtained for the versatile peroxidase (VP) from *Bjerkandera adusta*, which has potential applications in the degradation of the industrial and agricultural materials (Abdel-Hamid et al., 2013; Wang et al., 2013). VP can employ both lignin peroxidase and manganese peroxidase mechanisms in the degradation of humic materials (Siddiqui et al., 2014). ITC single injection assays (fulvic acid injected into VP) performed under conditions where both mechanisms are active gave biphasic heat spikes where each phase can be attributed to one of the mechanisms and each showed cooperative kinetics (**Figures 4G,H**). The data were well fit by the modified Hill Equation

$$v = \frac{V_{\max 1}[S]^{n1}}{K_{\mathrm{m1}}^{\mathrm{n1}} + [S]^{n1}} + \frac{V_{\max 2}[S]^{n2}}{K_{\mathrm{m2}}^{\mathrm{n2}} + [S]^{n2}} \quad (13)$$

with $n1 = 1.7$ for the 1st phase and $n2 = 10$ for the 2nd phase. Interestingly, the second injection gave broader peaks than the first, indicative of product inhibition, as described in more detail below.

Enzyme Inhibitors

Quantitative information on inhibitor binding is critical for developing drugs (Su and Xu, 2018) and understanding how enzymes function in living systems (Hulme and Trevethick, 2010). In fact ITC is primarily used to measure these types of host–guest interactions and a quick literature search for "ITC and inhibitor" yields more than 1,300 articles (using Clarivate Analytics Web of Science). In a traditional ITC experiment, the enzyme and inhibitor are placed in the sample cell and injection syringe, respectively, and a series of injections are made, while the instrument records the heat released or absorbed by the binding process itself (Su and Xu, 2018). This gives a wealth of thermodynamic information on the interaction, since the Gibbs energy and enthalpy of binding are detected separately, as the shape and overall magnitude of the saturation isotherm, respectively (Sigurskjold, 2000; Tellinghuisen, 2008; Keller et al., 2012; Su and Xu, 2018). Under favorable circumstances, the kinetics of binding can be measured as well (Burnouf et al., 2012; Piñeiro et al., 2019).

However, there are some drawbacks to this approach. Firstly, traditional ITC experiments require substantially more material than many other techniques used to measure binding, such as fluorescence or surface plasmon resonance. The recommended concentration of enzyme in the sample cell is roughly 5–500 times the inhibitor dissociation constant, K_i, [i.e., Wiseman "c" values of 5–500 (Wiseman et al., 1989)] often leading to requirements for protein on the micromolar to tens of micromolar scale (Malvern, 2010, Malvern, 2014). On the flip side, enzyme/inhibitor interactions that are too tight can be challenging since the enzyme concentration should not be more than about 1,000-fold greater than K_i, and very low enzyme concentrations lead to vanishingly small heat signals. This can be overcome with competition assays (Velazquez-Campoy and Freire, 2006), but the procedure is far more complicated. Finally, a traditional ITC binding experiment is not suited to characterizing all modes of inhibition. Competitive inhibitors bind exclusively to the free enzyme (E) and are suitable for traditional ITC binding experiments. In contrast, uncompetitive inhibitors bind exclusively to the enzyme/substrate Michaelis complex (ES), and in principle would not show an interaction at all in a traditional experiment. Mixed inhibitors bind to both E and ES, but the results of a traditional experiment would not reflect the true inhibition properties of the compound. Furthermore, an ITC binding experiment alone does not contain sufficient information to identify the inhibition mode.

These drawbacks can be overcome with ITC-based enzyme kinetic experiments. Firstly ITC kinetics experiments require far less enzyme than binding experiments. In a binding experiment, a single molecule of enzyme generates heat only once, when it forms a complex with the inhibitor. Whereas in a kinetics experiment, a single molecule of enzyme produces heat continuously as it undergoes multiple turnover. This allows

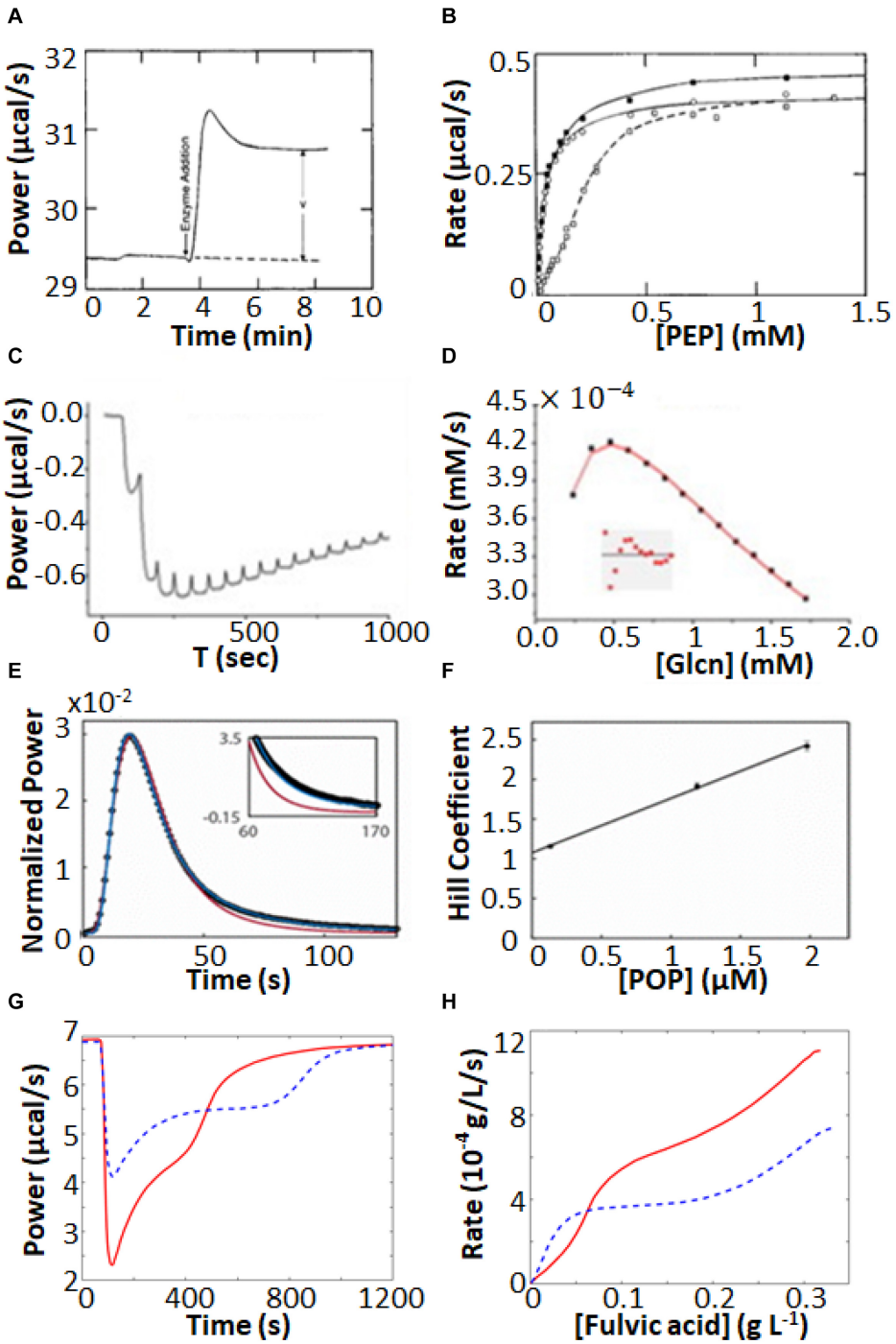


FIGURE 4 | Non-MM/BH enzyme kinetics observed by ITC. **(A)** Single injection experiments with pyruvate kinase in the syringe and phosphoenolpyruvate and ADP in the sample cell (Lonhienne and Winzor, 2002). The displacement of the horizontal baseline is proportional to the velocity of the enzyme. **(B)** Baseline displacements (ΔP) obtained at different [PEP] (O), in the presence of phenylalanine as an allosteric effector (□), and in the presence of phenylalanine and proline as a molecular crowding agent (■). **(C)** Multiple injection assay with gluconokinase and ATP in the sample cell and gluconate in the syringe (Rohatgi et al., 2015). **(D)** Enzyme velocities from **(C)**, fitted to a variant of Equation 12 that accounts for formation of the non-productive E·ADP·gluconate ternary complex. **(E)** Single injection assay with substrate (thyrotropin releasing hormone) in the syringe and prolyl oligopeptidase (POP) in the sample cell (Di Trani et al., 2017). Points are experimental data, red and blue curves are the best fits with classical MM/BH model, and cooperative model (Equation 11) with $n = 2.4$, respectively. **(F)** Dependence of the extracted Hill coefficient on POP concentration (0.125, 1.2, and 2 μM). **(G)** Single injection assay with versatile peroxidase in the sample cell and fulvic acid in the syringe, exhibiting biphasic cooperative kinetics (Siddiqui et al., 2014). **(H)** Reaction rates as s function of substrate concentration calculated from **(G)**. In **(G,H)**, data were extracted from the original reference using Graph Grabber v2.0.2 (Quintessa) and plotted using MATLAB (MathWorks); red solid curves indicate the first injection and blue dashed curves indicate the second injection.

ITC enzyme kinetics experiments to routinely be performed with sub-nM protein concentrations, which is outside the typical concentration range of ITC binding experiments. Furthermore ITC kinetics experiments are suitable for all modes of inhibition and can be performed in such a way that the mode and associated parameters are clearly evident. Finally, as detailed below, ITC differs fundamentally from other enzyme assays in that it detects the instantaneous velocity directly, while other methods measure concentrations of substrates, products, or reporters as a function of time and extract enzyme velocity indirectly. This makes ITC uniquely sensitive to how enzyme velocity changes with time, for instance as inhibitors exert their influence. Thus ITC has great potential for measurement of inhibitor association and dissociation rates.

A quantitative analysis of enzyme inhibition typically involves determination of the mode (competitive, uncompetitive, or mixed) and the inhibitor dissociation constant K_i. For mixed-mode inhibitors, there are separate K_i values for binding to E and to ES. Apparent $K_m{}^{app}$ and $k_{cat}{}^{app}$ values are measured at different concentrations of inhibitor [I] and analyzed collectively to extract the inhibition parameters. For a competitive inhibitor

$$k_{cat}^{app} = k_{cat}; \quad K_m^{app} = K_m\left(1 + \frac{[I]}{K_i}\right) \tag{14}$$

and a double-reciprocal plot of $1/v_0$ vs. $1/[S]$ obtained at different $[I]$ gives a series of lines that intersect at the y-axis. For an uncompetitive inhibitor

$$k_{cat}^{app} = \frac{k_{cat}}{1 + \frac{[I]}{K_i'}}; \quad K_m^{app} = \frac{K_m}{1 + \frac{[I]}{K_i'}} \tag{15}$$

where K'_i is the dissociation constant for the inhibitor and ES complex and a double-reciprocal plot gives a series of parallel lines. For mixed inhibitors

$$k_{cat}^{app} = \frac{k_{cat}}{1 + \frac{[I]}{K_i'}}; \quad K_m^{app} = K_m\frac{\left(1 + \frac{[I]}{K_i}\right)}{\left(1 + \frac{[I]}{K_i'}\right)} \tag{16}$$

and a double-reciprocal plot gives a series of lines that intersect elsewhere than the y-axis. In the case that $K_i = K'_i$, the inhibitor is said to be non-competitive and the lines intersect at the x-axis.

Characterization of enzyme inhibition can largely be accomplished with the experiments described in Section "ITC Kinetics Methods." For example, ITC was used to characterize inhibitors of pancreatic α-amylase, which hydrolyses starches into monosaccharides in the gut (Kaeswurm et al., 2019). It has been proposed that a variety of polyphenol plant metabolites inhibit α-amylase, slowing glucose absorption by the intestine, and reducing spikes in insulin levels with implications for the management of diabetes (Hanhineva et al., 2010). The authors tested a panel of naturally occurring polyphenols, together with the known potent α-amylase inhibitor acarbose. They used a single injection assay where α-amylase was injected into 1 mM trisaccharide substrate, with or without 100 μM of each phenolic inhibitor. The enzyme injection technique allowed them to initiate the reaction with very small (1 μL) additions of dilute reagent (1 μM enzyme), thereby almost entirely avoiding injection artifacts. Each experiment consists of a single peak, corresponding to the complete conversion of substrate to product over the course of about 2.5 h. The resulting ITC isotherms are shown in **Figure 5A** and MM/BH plots calculated from these data are shown in **Figure 5B**. Interestingly, all of the polyphenols affected both the k_{cat} of the enzyme (indicated by the height of the asymptote in **Figure 5B**) and the K_m, indicating mixed modes of inhibition, although the curves were not fitted quantitatively in this study. Mixed inhibition was also observed using experiments with a colorimetric assay detected by UV/vis spectroscopy. Note that while these assays were performed injecting enzyme into substrate pre-incubated with inhibitor, similar assays can also be performed by injecting substrate into enzyme pre-incubated with inhibitors (Catucci et al., 2019).

When ITC inhibition experiments are performed with the inhibitor loaded in the sample cell prior to data collection, as in the examples above, then the procedure must be repeated several times in order to accurately measure inhibition parameters. This demands a considerable investment of time, since the cleaning, loading, equilibration, and data collection must be performed separately for each inhibitor concentration. Our lab has developed a procedure for considerably shortening this timeline, allowing much higher throughput of samples (Di Trani et al., 2018b). This approach is a variation of a standard recurrent single injection assay, with the modification that the syringe contains both substrate and inhibitor. A series of injections is made with each ITC peak giving a k_{cat} and K_m pair. The inhibitor accumulates with each injection, such that the activity of the enzyme decreases in each successive peak, and provides a thorough sampling of different inhibitor concentrations. Typical data are shown in **Figure 5C**, with each peak clearly broader than the preceding one, reflecting slower catalysis by the enzyme (trypsin inhibited with benzamidine in this case). Each peak is complete within 5–10 min and a total data set can be collected in under an hour. The peaks were analyzed by fitting directly to the raw ITC data. The extracted $k_{cat}{}^{app}$ was very similar for each peak, while the $K_m{}^{app}$ increased linearly with increasing peak number and inhibitor concentration (**Figure 5D**), as expected for a competitive inhibitor (Equation 14). The y-intercept of the line is K_m and the slope is K_m/K_i. Alternatively, the peaks can be deconvoluted using the empirical response function (see Rapid Enzyme Kinetics Measured by ITC, above) and converted to double-reciprocal plots in a model-free manner (**Figure 5E**). The data give a series of straight lines that intersect at the y-axis, consistent with the competitive inhibition mechanism of benzamidine. This approach gives full kinetic characterization for as many as ten different inhibitor concentrations in a single experiment, providing assessment of inhibition mode and strength. Note that direct fitting of ITC peaks is more straightforward and is generally preferred to deconvolution-based approaches. The double-reciprocal plot is primarily useful for illustrating the extent to which data follow the MM/BH equation and obey a given pattern of inhibition, as it is easier to spot deviations from linearity and identify intersections than it is to judge the shapes of ITC peaks by eye. More recently, this method was applied to the study of human soluble epoxide hydrolase, which is involved in cardiovascular homeostasis, hypertension, nociception, and insulin sensitivity through the

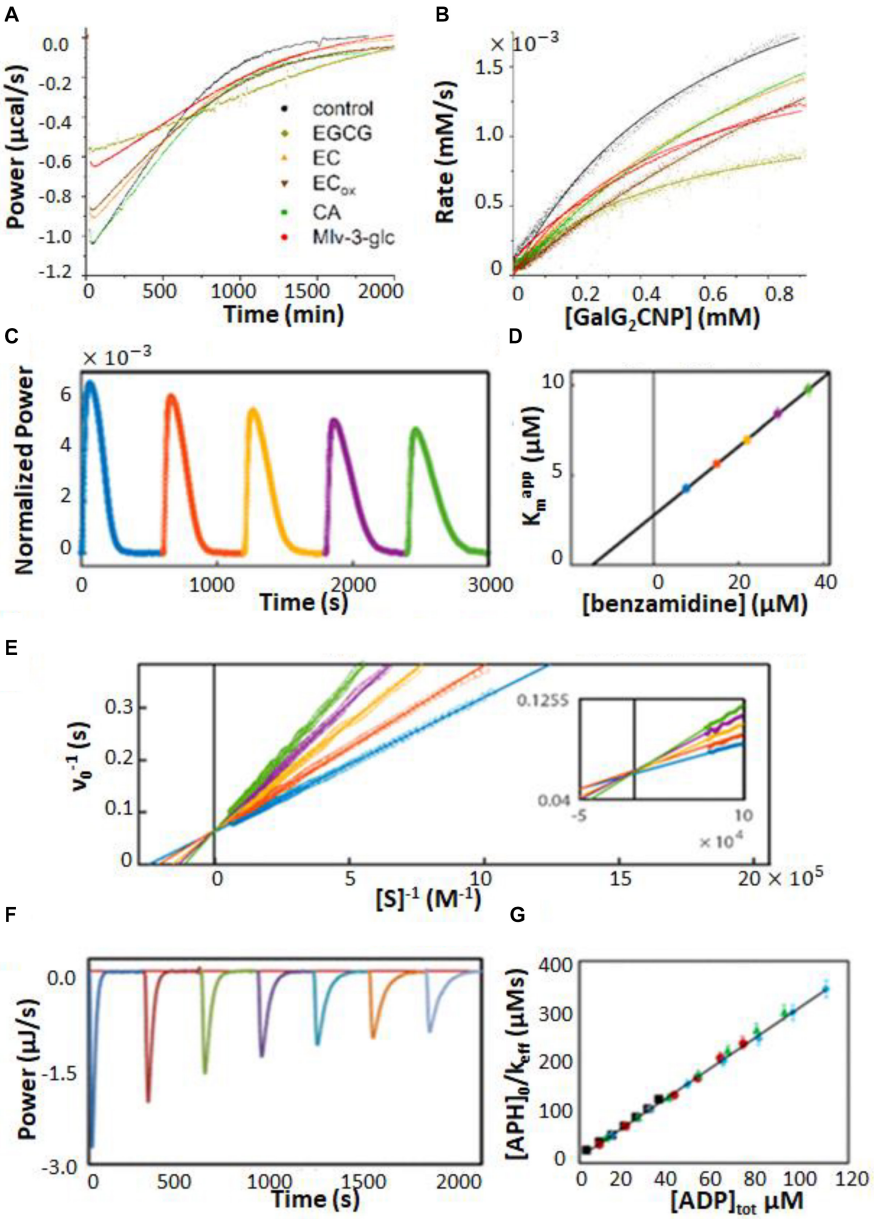


FIGURE 5 | Enzyme inhibition characterized by ITC single injection-type assays. **(A)** Inverse injection assay with α-amylase in the syringe and the substrate 2-chloro-4-nitrophenyl-maltoside ($GalG_2CNP$) in the syringe together with a variety of inhibitors: ACA (acarbose), CA (chlorogenic acid), EC (epicatechin), ECox (oxidized epicatechin), EGCG (epigallocatechin gallate), Mlv-3-glc (malvidin-3-glucoside) (Hanhineva et al., 2010). **(B)** MM/BH curves calculated from the curves in **(A)**. **(C)** Single injection assay with substrate (benzoyl-L-arginine ethyl ester) and inhibitor (benzamidine) in the syringe and trypsin in the sample cell (Di Trani et al., 2018b). **(D)** K_m^{app} values extracted from direct fits to each of the injections (different colors) in **(C)**. **(E)** Data from **(C)**, deconvoluted using the empirical response model (Equation 10), converted to ν_0 and [*S*] and presented as a double-reciprocal plot. **(F)** Single injection assay with substrate (ATP) in the syringe and aminoglycoside-3′-phosphotransferase IIIa (APH) and kanamycin A in the sample cell (Wang et al., 2019). Under these dilute conditions [ATP] $<< K_m$, ITC peaks decay exponentially with rate constant $k_{eff} = k_{cat}/K_m$. k_{eff} decreases with each injection due to product inhibition by ADP. **(G)** Plot of [APH]/k_{eff} as a function of total accumulated ADP concentration.

metabolism of a variety of epoxy-fatty acids (Abis et al., 2019). There is interest in inhibiting this enzyme in order to raise physiological levels of epoxy-fatty acids, which have been shown to have beneficial biological activities, and to reduce the levels of the reaction products, dihydroxy fatty acids, many of which are cytotoxic (Wagner et al., 2017). Previous assays of this enzyme relied on liquid chromatography/tandem mass spectrometry which is a discontinuous method with non-negligible liquid handling steps. The authors showed that single injection ITC assays can clearly distinguish the different kinetics of different epoxy fatty acids providing a simpler and more robust route to characterization. As well, they validated the multiple inhibitor

injection method for this system using a previously identified antagonist, setting the stage for rapid ITC-based screening for inhibitors of this enzyme.

A similar type of situation occurs when the enzyme is inhibited by the reaction product. In this case, the product of the reaction accumulates after each injection, leading to progressively slower catalysis. In fact, slowing catalysis with subsequent injections in a single injection ITC experiment is a hallmark of product inhibition (Cai et al., 2001). A challenge can arise when inhibition is strong, since the product generated by a single injection can be sufficient to essentially abrogate activity of the enzyme in subsequent injections. This was the case in our studies of several bacterial kinases and their inhibition by the reaction product ADP (Wang et al., 2019). Interestingly, ADP inhibition of kinases cannot be studied using the standard coupled enzyme assay for kinase activity, since ADP is reconverted to ATP by the secondary enzymes (McKay and Wright, 1995). In order to avoid excessive inhibition of the enzymes, we injected the substrate ATP at a concentration well below the K_m. Under these conditions, substrate consumption follows simple first-order kinetics with rate constant, $k_{eff} = k_{cat}[E]/K_m^{app}$, where K_m^{app} increases with each injection as product (i.e., inhibitor) accumulates. Data for dilute ATP injected into aminoglycoside-3′-phosphotransferase IIIa are shown in **Figure 5F**, clearly exhibiting broadening of successive peaks. A plot of $[E]/k_{eff}$ vs. $[I]$ is linear (**Figure 5G**) with a slope of $\frac{K_m}{k_{cat}} \cdot \frac{1}{K_i}$ and a y-intercept of $\frac{K_m}{k_{cat}}$; the ratio of the two gives the inhibition constant, K_i. Using this approach, we found that the K_i for ADP was comparable or lower than the K_m for ATP, for all three kinases studied, implying that inhibition by ADP influences kinase activity *in vivo*. We also identified a more complex mechanism for a dimeric pantothenate kinase, wherein ADP is activating at low concentrations and becomes inhibitory at higher concentrations, consistent with allosteric communication between the two active sites (Wang et al., 2019).

As discussed in Section "Multiple Injection Assays," there are some advantages associated with multiple injection ITC enzyme assays, and this holds true for inhibitor characterization as well. For instance, the urease enzyme acts on urea to produce bicarbonate and two equivalents of ammonia. Multiple injection ITC assays produce far less reaction product than do single injection ones; in this case it helps to minimize the production of ammonia which is alkaline, volatile, and corrosive. Urease inhibitors have potential antimicrobial and agricultural applications (Kosikowska and Berlicki, 2011; Upadhyay, 2012). Benini et al. (2014) used multiple injection ITC assays to characterize inhibition of urease by fluoride ions. Typical raw ITC data are shown in **Figure 6A**, while the corresponding MM/BH curves for urease activity in the presence of 0, 400, and 800 μM NaF are shown in **Figure 6B**, revealing a mixed mode of inhibition. Interestingly, the competitive and uncompetitive components (K_i and K'_i in Equation 16) showed different pH dependencies, and the authors could link the two different inhibition modes to two different locations of fluoride ion binding in the active site, as determined by X-ray crystallography.

Our lab has recently designed a pair of experiments which build on the multiple injection ITC experiment to give additional information on inhibitor association and dissociation rates (Di Trani et al., 2018a). In the association experiment (**Figure 6C**), the syringe contains the inhibitor and the sample cell contains dilute enzyme and sufficient substrate to maintain an essentially constant concentration throughout the experiment. The rate of catalysis is initially constant, giving a horizontal line. A series of injections is made, in this case of reversible covalent inhibitors targeting prolyl oligopeptidase in the sample cell. In each case, the enzyme was increasingly inhibited and the power values shifted upward, since the rate of (exothermic) catalysis was reduced after each injection. As highlighted in **Figure 6D**, this shift occurred gradually over tens to hundreds of seconds, which corresponds to the time required for the inhibitor to bind in the active site. Furthermore, the upward shift of the ITC signal became smaller for each subsequent injection, as the enzyme became increasingly saturated with inhibitor. The decrease in the sizes of the steps is related to the inhibition constant, K_i, and the values of k_{on} and K_i can be fitted numerically to the data. The disassociation rate can then be calculated as $k_{off} = k_{on} \times K_i$.

In the dissociation experiments, the sample cell contains only the substrate and the syringe contains enzyme saturated with an inhibitor (prolyl oligopeptidase and a reversible covalent inhibitor), which is added to the cell in a series of injections (**Figure 6E**). Immediately following each injection there was no change in the rate of catalysis in the sample cell as the added enzyme was fully inhibited. However, the large dilution (>20-fold) experienced by the injectant led to a net dissociation of the inhibitor and a gradual downward shift of the ITC signal as the freshly released enzyme began to act on the substrate (**Figure 6F**). The downward shift of the ITC signal became smaller for each subsequent injection, as the inhibitor accumulated in the sample cell and the net dissociation of each injection diminished. The decrease in the sizes of the steps is governed by the value of K_i. Data for the series of injections can be fitted simultaneously to yield k_{off} and K_i (**Figure 6F**). The association rate can then be calculated as $k_{on} = k_{off}/K_i$. Note that concentrations of enzyme are so low in these experiments (≈10 nM) that ITC detects only heats of catalysis, while heats of inhibitor/enzyme binding can be safely ignored. This method exploits the fact that ITC measures enzyme velocity directly. A traditional concentration-based enzyme assay would detect the gradual decreases and increases in enzyme velocity vividly illustrated in **Figures 6D,F** as slight curvature in the product buildup curve, making quantitative analysis far more difficult (Di Trani et al., 2018a).

Heterogeneous Mixtures

A unique aspect of ITC enzyme kinetic assays is their general ability to provide real-time measurements on opaque systems that are unsuitable for typical bulk spectroscopic techniques. One example of this is ITC enzyme kinetics experiments performed on suspensions of living cells (Wang et al., 2017, 2018; Zhang et al., 2018; Lv et al., 2019). Comparing the behavior of enzymes *in vitro* and *in situ* is critical for understanding how they work in living

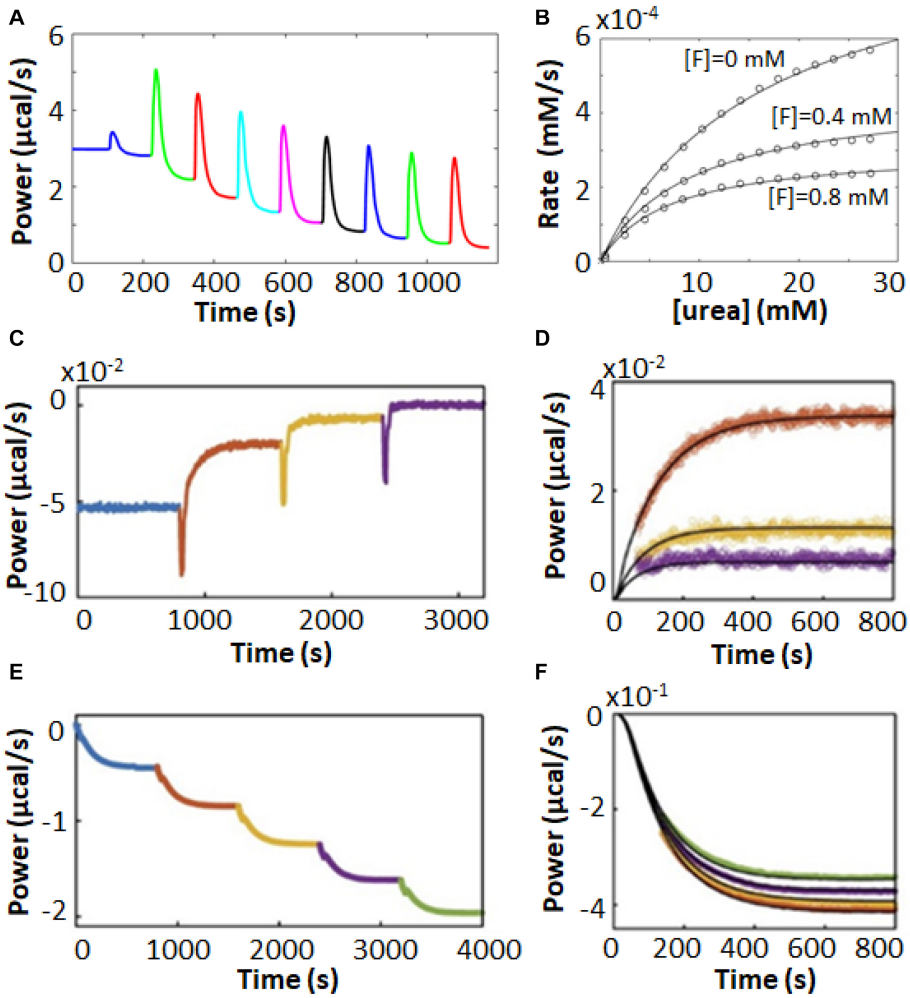


FIGURE 6 | Enzyme inhibition characterized by ITC Pseudo-First-Order-type assays. **(A)** Multiple injection-type ITC assays with urea in the syringe and urease in the sample cell (Benini et al., 2014). **(B)** MM/BH plots from data similar to **(A)** with fluoride ion concentrations of 0, 0.4, and 0.8 mM, fitted to a mixed inhibition model (Equation 16). In **(A,B)**, data were extracted from the original reference using Graph Grabber v2.0.2 (Quintessa) and plotted using MATLAB (MathWorks). **(C)** Inhibitor association kinetics experiment with prolyl oligopeptidase (POP) and substrate (thyrotropin releasing hormone, TRH) in the sample cell and reversible covalent inhibitor in the syringe (Di Trani et al., 2018a). **(D)** Overlay of injections 1–3 from **(C)** (colored points) with best global fits to a kinetic model of association (black curves). **(E)** Inhibitor dissociation kinetics experiment with TRH in the sample cell and POP and a reversible covalent inhibitor in the syringe (Di Trani et al., 2018a). **(F)** Overlay of injections 2–5 from **(E)** (colored points) with fit best global fits to a kinetic model of dissociation (black curves).

systems and can reveal how enzyme kinetics are tied to additional layers of biological dynamics. Furthermore, studying enzymes in the intact organism avoids the question of whether activity has been compromised by extraction and circumvents the need for purification steps at all. For instance, Zhang et al. (2018) used ITC to study the metallo-β-lactamase NDM-1 in living cultures of *Escherichia coli*. NDM-1 cleaves carbapenems, providing bacterial resistance to these "last resort" β-lactam antibiotics. Development of NDM-1 inhibitors has the potential to resensitize resistant bacteria and offers an avenue for treating these kinds of serious drug-resistant infections (Zhang et al., 2018). NDM-1 is located in the periplasm of Gram-negative bacteria, anchored to the inner leaflet of the outer membrane by a lipidated cysteine residue (Palzkill, 2013). Thus the natural environment of NDM-1 is not well recapitulated by purified enzymes in solution. The authors used cefazolin as a test substrate, injecting it into either purified NDM-1 (**Figure 7A**) or live *E. coli* bacteria expressing the enzyme (**Figure 7B**). Very similar single injection-type heat signals were obtained in both cases, with all injected substrate being consumed within 100–200 s in the case of purified protein and 400–600 s in the case of live cells. Cefazolin injected in live bacterial cultures not expressing NDM-1 gave negligible heat signals. The amount of NDM-1 present in the live cells was not determined and differences in enzyme concentration could have contributed to differences in kinetic profiles in **Figures 7A,B**. However, the authors also made the intriguing suggestion that the slower kinetics in live cells could be due, at least in part, to the time lag of cefazolin entering the periplasm and hydrolyzed products leaving. The results presented in their study suggest that ITC could be a powerful tool for studying these processes

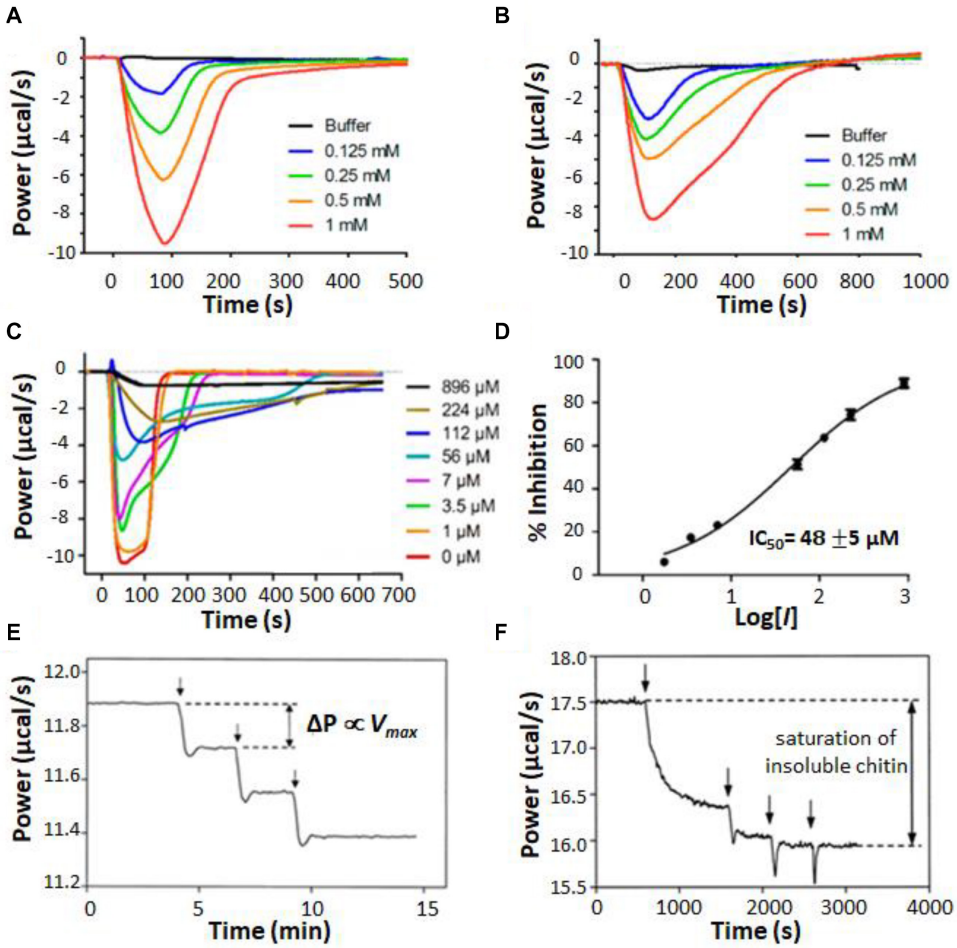


FIGURE 7 | Isothermal titration calorimetry characterization of heterogeneous mixtures. **(A)** Single injection assays with substrate (cefazolin) in the syringe and purified NDM-1 in the sample cell (Zhang et al., 2018). **(B)** Single injection assays with cefazolin in the syringe and a suspension of live *E. coli* bacteria expressing NDM-1 in the sample cell. **(C)** Experiments in **(B)** repeated with various concentrations of an inhibitor (D-captopril) added to the *E. coli* suspension. **(D)** IC_{50} calculation, taking the magnitude of each peak in **(C)** as proportional to enzyme activity. Single injection assays with chitinase in the injection syringe and **(E)** soluble chitin fragments or **(F)** insoluble chitin in the sample cell (Lonhienne et al., 2001). Vertical arrows indicate timings of injections.

in future work. The authors went on to test a live-cell screen for NDM-1 inhibitors, using a panel of previously reported compounds. Data are shown in **Figure 7C** for a similar assay to **Figure 7B**, with cefazolin injected in live cells pre-incubated with different concentrations of known inhibitor, D-captopril; as the concentration of inhibitor was raised, the ITC heat peaks became increasingly broad, indicative of slower catalysis and enzyme inhibition. Taking the absolute amplitudes of the peaks as a measure of enzyme activity, the set of ITC data gave an IC_{50} of about 50 µM (**Figure 7D**), in good agreement with previous measurements on purified protein (Zhang et al., 2018). The authors repeated the experiments on four different clinical strains of pathogenic bacteria. Interestingly, different strains showed different levels of activity, perhaps reflecting different levels of NDM-1 expression, different accessibility of the periplasmic space, or both. This study paves the way for using ITC both as a rapid screen for inhibitors of antibiotic resistance genes, and also as a tool for probing the resistance profiles of clinical isolates.

Other examples of opaque reaction mixtures are those involving insoluble substrates (Lonhienne et al., 2001; Murphy et al., 2010a,b, 2012, 2013) or enzymes immobilized on insoluble matrices (Henzler et al., 2008; Henao-Escobar et al., 2014; Ali et al., 2015; Mason et al., 2018) where the components are combined as a suspension or slurry. These mixtures are of great industrial importance, in part because the insoluble carbohydrate polymers cellulose and chitin are the two most abundant organic compounds on earth, present in large quantities in vascular plants and arthropod exoskeletons, respectively (Berlemont et al., 2016). Cellulose and chitin modifying enzymes have many potential applications in biofuel production, chemical upcycling, agriculture, and textile production (Turner et al., 2007). Lonhienne et al. (2001) used ITC to characterize the activity of psychrophilic bacterial chitinases, which hydrolyse glycosidic bonds in chitin. They studied both an insoluble suspension of powdered chitin and a sample in which the chitin had been cleaved by acid hydrolysis into soluble fragments of

40–100 kDa. When a series of injections of dilute chitinase were made into a sample of soluble chitin fragments, the heat flow increased in a series of steps of equal size (**Figure 7E**), as expected, since for normal MM/BH kinetics, the total reaction velocity is proportional to the concentration of enzyme. In contrast, when chitinase was injected into the chitin suspension, the first injection produced a large increase in heat flow, the step associated with the second injection was much smaller, and subsequent injections did not increase the rate of catalysis at all (**Figure 7F**). The authors attributed this to saturation of the chitin substrate. The bulk of the material forms an interconnected network buried in the colloidal particle and is therefore inaccessible to enzymatic attack. Once the surfaces of the particles are covered by enzyme molecules, the addition of further enzyme does not increase the rate of hydrolysis. An additional interesting feature of the ITC data is the slow and gradual increase in enzyme velocity after the first injection. The steady-state rate of catalysis was not reached until roughly 15 min after the enzyme was added. The authors attribute this lag to the opening of crystalline regions on the colloid surface to expose new fibril ends that are susceptible to the enzyme. This study vividly illustrates how ITC data can give information not only on enzyme kinetics, but also on the dynamic processes to which the enzyme activity is linked.

DISCUSSION

The methods and examples discussed above illustrate the power and potential of ITC as a universal enzyme assay. ITC offers real-time monitoring of enzymatic reactions in cases where other types of continuous assays are unavailable. This is exemplified by human soluble epoxide hydrolase (Abis et al., 2019) discussed above, where previous work had relied on a combination of quenching the reaction at various time points and analyzing the composition by liquid chromatography and tandem mass spectrometry (Morisseau et al., 2010). The ability to employ natural substrates is another large asset for ITC. This is particularly true when the MM/BH parameters obtained for chemically modified colorigenic or fluorogenic substrate analogs do not match those obtained for the native substrate by ITC. For example, glycosidase activity can be measured spectrophotometrically with synthetic substrates, such as maltooligomer derivatives with chromogenic chloronitrophenyl (CNP) groups attached (Klaus et al., 1999; Morishita et al., 2000; Ramasubbu et al., 2005). Separate studies on α-amylase and glycogen phosphorylase found the K_m values of the fluorogenic substrate analogs to be substantially lower than those of the native substrates determined by ITC, possibly due to interactions of the chromophore with the active site of the enzyme (Lehoczki et al., 2016; Szabó et al., 2019). Thus ITC represents a simple way to accurately characterize how enzymes interact with their biologically relevant molecular partners. ITC also offers advantages for enzymes where the standard assay involves indirect readout with a coupled-enzyme system. This is particularly true when adding co-solutes or inhibitors that affect enzymatic activity since the secondary enzymes can be affected as well as the enzyme of interest, as discussed above for pyruvate kinase (Lonhienne and Winzor, 2002; Lonhienne et al., 2003). In addition, testing spectroscopically active inhibitors or other effector molecules can become a challenge when using spectrophotometric assays, i.e., with chromogenic or fluorogenic probes, or with coupled assays. In contrast, deeply-colored inhibitors are fully compatible with ITC inhibition assays (Zebisch et al., 2012). Furthermore ITC's ability to characterize opaque samples further extends the reach of this technique beyond spectroscopically-accessible systems. The examples described above involving suspensions of live cells (Wang et al., 2017, 2018; Zhang et al., 2018; Lv et al., 2019) and insoluble substrates (Lonhienne et al., 2001; Murphy et al., 2010a,b, 2012, 2013), illustrate how the surrounding milieu can influence enzyme activity and how ITC can be a probe of these more complex dynamics.

Over the past 20 years, the number of publications using ITC to measure enzyme kinetics has been growing at an ever-accelerating rate (**Supplementary Figure 1**). The advantages of ITC described above are becoming increasingly recognized, and we expect that this trend will continue as the technique becomes more visible and widely known. Most of the studies to date have employed the two main types of experiment described in the original paper by Todd and Gomez (2001), i.e., multiple injection and single injection assays. However, we believe that the full potential of ITC as an enzyme kinetic technique is only starting to be explored and that the development of innovative methods and novel capabilities will help to drive the further growth of the user community. Our group (Di Trani et al., 2017, 2018a,b; Wang et al., 2019) and others (Honarmand Ebrahimi et al., 2015; Transtrum et al., 2015) have recently reported methodological advances that push limits of the technique. One area of interest is the development of strategies for quantitatively addressing the finite response time of ITC instruments (Honarmand Ebrahimi et al., 2015; Transtrum et al., 2015; Di Trani et al., 2017). For example, our Empirical Response Model, described above, has performed well under a variety of conditions, and allowed us to, for example, characterize very rapid reactions and clearly identify non-MM/BH kinetics from single ITC peaks (Di Trani et al., 2017). In developing this model, we found that the kinetics of post-injection mixing can be largely ignored under many conditions. However, this approximation should likely be revisited and is an area ripe for advancement, perhaps by incorporating some ideas used in analyzing binding kinetics by ITC (Burnouf et al., 2012). More generally, there are an enormous number of ways to generate different ITC experiments. As an illustration, for a simple ternary system of enzyme/substrate/inhibitor, one can imagine that each of the components can be loaded in either the syringe or the sample cell, and can be either dilute ($C << K_m$, K_i) or concentrated ($C >> K_m$, K_i). This, in principle, gives 64 distinct arrangements, only a few of which have been investigated to date. There are likely scenarios in which many of these hypothetical experiments would yield uniquely useful data. When one considers the number of multi-substrate enzymes (Wang et al., 2019) and multiple enzyme systems with shared substrates or products (Murphy et al., 2010a), as well as the effects of

allostery and substrate or product inhibition, the complexity of the experiment-space and versatility of ITC starts to become apparent. Finally, the ability of ITC to extract meaningful kinetic data from systems as complicated as biopolymer suspensions (Henzler et al., 2008; Ali et al., 2015) or even living cells (Wang et al., 2017, 2018; Zhang et al., 2018; Lv et al., 2019) holds great promise for understanding enzyme behavior *in situ* and *in vivo*. One could imagine expanding this approach to a multitude of other complex and heterogeneous media, such as purified cellular components, homogenized tissue or soil samples, and nanostructured materials, to name a few. It is our belief that between advancements in experimental design and analysis and sample selection and preparation, the full potential of ITC to study enzyme kinetics will become evident in the coming years.

AUTHOR CONTRIBUTIONS

YW, GW, NM, and AM wrote the article. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant 327028-09.

ACKNOWLEDGMENTS

AM is a member of the Quebec Network for Research on Protein Function, Engineering, and Applications (PROTEO) and the McGill Centre for Structural Biology. The authors would like to thank Prof. Lee D. Hansen for useful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.583826/full#supplementary-material

REFERENCES

Abdel-Hamid, A. M., Solbiati, J. O., and Cann, I. K. O. (2013). "Insights into lignin degradation and its potential industrial applications," in *Advances in Applied Microbiology*, Vol. 82, eds S. Sariaslani and G. M. Gadd (Cambridge, MA: Academic Press), 1–28. doi: 10.1016/b978-0-12-407679-2.00001-6

Abis, G., Pacheco-Gómez, R., Bui, T. T. T., and Conte, M. R. (2019). Isothermal titration calorimetry enables rapid characterization of enzyme kinetics and inhibition for the human soluble epoxide hydrolase. *Anal. Chem.* 91, 14865–14872. doi: 10.1021/acs.analchem.9b01847

Ali, G., Dulong, V., Gasmi, S. N., Rihouey, C., Picton, L., and Le Cerf, D. (2015). Covalent immobilization of pullulanase on alginate and study of its hydrolysis of pullulan. *Biotechnol. Progr.* 31, 883–889. doi: 10.1002/btpr.2093

Ali, G., Rihouey, C., Larreta-Garde, V., Le Cerf, D., and Picton, L. (2013a). Molecular size characterization and kinetics studies on hydrolysis of pullulan by pullulanase in an entangled alginate medium. *Biomacromolecules* 14, 2234–2241. doi: 10.1021/bm400371r

Ali, G., Rihouey, C., Le Cerf, D., and Picton, L. (2013b). Effect of carboxymethyl groups on degradation of modified pullulan by pullulanase from *Klebsiella pneumoniae*. *Carbohydr. Polym.* 93, 109–115. doi: 10.1016/j.carbpol.2012.07.039

Atri, M. S., Saboury, A. A., and Ahmad, F. (2015). Biological applications of isothermal titration calorimetry. *Phys. Chem. Res.* 3, 319–330.

Backman, P., Bastos, M., Hallen, D., Lonnbro, P., and Wadso, I. (1994). Heat conduction calorimeters: time constants, sensitivity and fast titration experiments. *J. Biochem. Biophys. Methods* 28, 85–100. doi: 10.1016/0165-022x(94)90023-x

Benhar, M., Forrester, M. T., and Stamler, J. S. (2009). Protein denitrosylation: Enzymatic mechanisms and cellular functions. *Nat. Rev. Mol. Cell Biol.* 10, 721–732. doi: 10.1038/nrm2764

Benini, S., Cianci, M., Mazzei, L., and Ciurli, S. (2014). Fluoride inhibition of Sporosarcina pasteurii urease: structure and thermodynamics. *J. Biol. Inorg. Chem.* 19, 1243–1261. doi: 10.1007/s00775-014-1182-x

Berlemont, R., Martiny, A. C., and Martiny, A. C. (2016). Glycoside hydrolases across environmental microbial communities. *PLoS Comput. Biol.* 12:e1005300. doi: 10.1371/journal.pcbi.1005300

Burnouf, D., Ennifar, E., Guedich, S., Puffer, B., Hoffmann, G., Bec, G., et al. (2012). kinITC: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry. *J. Am. Chem. Soc.* 134, 559–565. doi: 10.1021/ja209057d

Butler, J. E. (2000). Enzyme-Linked immunosorbent assay. *J. Immun.* 21, 165–209.

Cai, L., Cao, A., and Lai, L. (2001). An isothermal titration calorimetric method to determine the kinetic parameters of enzyme catalytic reaction by employing the product inhibition as probe. *Anal. Biochem.* 299, 19–23. doi: 10.1006/abio.2001.5397

Capaldi, R. A., and Aggeler, R. (2002). Mechanism of the F1F0-type ATP synthase, a biological rotary motor. *Trends Biochem. Sci.* 27, 154–160. doi: 10.1016/s0968-0004(01)02051-5

Catucci, G., Sadeghi, S. J., and Gilardi, G. (2019). A direct time-based ITC approach for substrate turnover measurements demonstrated on human FMO3. *Chem. Commun.* 55, 6217–6220. doi: 10.1039/c9cc01356c

Choi, J.-M., Han, S.-S., and Kim, H.-S. (2015). Industrial applications of enzyme biocatalysis: current status and future aspects. *Biotechnol. Adv.* 33, 1443–1454. doi: 10.1016/j.biotechadv.2015.02.014

Commin, C., Aumont-Nicaise, M., Claisse, G., Feller, G., and Da Lage, J.-L. (2013). Enzymatic characterization of recombinant α-amylase in the *Drosophila melanogaster* species subgroup: is there an effect of specialization on digestive enzyme? *Genes Genet. Syst.* 88, 251–259. doi: 10.1266/ggs.88.251

DeDecker, B. S. (2000). Allosteric drugs: thinking outside the active-site box. *Chem. Biol.* 7, R103–R107.

Demarse, N. A., Quinn, C. F., Eggett, D. L., Russell, D. J., and Hansen, L. D. (2011). Calibration of nanowatt isothermal titration calorimeters with overflow reaction vessels. *Anal. Biochem.* 417, 247–255. doi: 10.1016/j.ab.2011.06.014

Di Trani, J. M., De Cesco, S., O'Leary, R., Plescia, J., do Nascimento, C. J., Moitessier, N., et al. (2018a). Rapid measurement of inhibitor binding kinetics by isothermal titration calorimetry. *Nat. Commun.* 9:893.

Di Trani, J. M., Moitessier, N., and Mittermaier, A. K. (2017). Measuring rapid time-scale reaction kinetics using isothermal titration calorimetry. *Anal. Chem.* 89, 7022–7030. doi: 10.1021/acs.analchem.7b00693

Di Trani, J. M., Moitessier, N., and Mittermaier, A. K. (2018b). Complete kinetic characterization of enzyme inhibition in a single isothermal titration calorimetric experiment. *Anal. Chem.* 90, 8430–8435. doi: 10.1021/acs.analchem.8b00993

Easterby, J. S. (1973). Coupled enzyme assays: a general expression for the transient. *Biochim. Biophys. Acta* 293, 552–558. doi: 10.1016/0005-2744(73)90362-8

Ebrahimi, K., Hagedoorn, P.-L., Jacobs, D., and Hagen, W. R. (2015). Accurate label-free reaction kinetics determination using initial rate heat measurements. *Sci. Rep.* 5:16380.

Ertan, H., Siddiqui, K. S., Muenchhoff, J., Charlton, T., and Cavicchioli, R. (2012). Kinetic and thermodynamic characterization of the functional properties of a hybrid versatile peroxidase using isothermal titration calorimetry: Insight into manganese peroxidase activation and lignin peroxidase inhibition. *Biochimie* 94, 1221–1231. doi: 10.1016/j.biochi.2012.02.012

Fenton, A. W. (2008). Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem. Sci.* 33, 420–425. doi: 10.1016/j.tibs.2008.05.009

Freire, E., Mayorga, O. L., and Straume, M. (1990). Isothermal titration calorimetry. *Anal. Chem.* 62, 950A–959A.

Freyer, M. W., and Lewis, E. A. (2008). Methods in cell biology. *Acad. Press.* 84, 79–113.

García-Fuentes, L., Barón, C., and Mayorga, O. L. (1998). Influence of dynamic power compensation in an isothermal titration microcalorimeter. *Anal. Chem.* 70, 4615–4623. doi: 10.1021/ac980203u

Gerhart, J. (2014). From feedback inhibition to allostery: the enduring example of aspartate transcarbamoylase. *FEBS J.* 281, 612–620. doi: 10.1111/febs.12483

Ghai, R., Falconer, R. J., and Collins, B. M. (2012). Applications of isothermal titration calorimetry in pure and applied research-survey of the literature from 2010. *J. Mol. Recognit.* 25, 32–52. doi: 10.1002/jmr.1167

Guarnera, E., and Berezovsky, I. N. (2019). On the perturbation nature of allostery: sites, mutations, and signal modulation. *Curr. Opin. Struct. Biol.* 56, 18–27. doi: 10.1016/j.sbi.2018.10.008

Guarnera, E., and Berezovsky, I. N. (2020). Allosteric drugs and mutations: chances, challenges, and necessity. *Curr. Opin. Struct. Biol.* 62, 149–157. doi: 10.1016/j.sbi.2020.01.010

Guzman-Maldonado, H., and Paredes-Lopez, O. (1995). Amylolytic enzymes and products derived from starch: a review. *Crit. Rev. Food Sci. Nutr.* 35, 373–403. doi: 10.1080/10408399509527706

Hanhineva, K., Torronen, R., Bondia-Pons, I., Pekkinen, J., Kolehmainen, M., Mykkanen, H., et al. (2010). Impact of dietary polyphenols on carbohydrate metabolism. *Int. J. Mol. Sci.* 11, 1365–1402. doi: 10.3390/ijms11041365

Hansen, L. D., Transtrum, M. K., Quinn, C., and Demarse, N. (2016). Enzyme-catalyzed and binding reaction kinetics determined by titration calorimetry. *Biochim. Biophys. Acta Gen. Subj.* 1860, 957–966. doi: 10.1016/j.bbagen.2015.12.018

Harris, T. K., and Keshwani, M. M. (2009). *Methods Enzymology*, Vol. 463. Cambridge, MA: Academic Press, 57–71.

Hastie, C. J., McLauchlan, H. J., and Cohen, P. (2006). Assay of protein kinases using radiolabeled ATP: a protocol. *Nat. Protoc.* 1, 968–971. doi: 10.1038/nprot.2006.149

Henao-Escobar, W., Domínguez-Renedo, O., Alonso-Lomillo, M. A., Cascalheira, J., Dias-Cabral, A., and Arcos-Martínez, M. (2014). Characterization of a disposable electrochemical biosensor based on putrescine oxidase from micrococcus rubens for the determination of putrescine. *Electroanalysis* 27, 368–377. doi: 10.1002/elan.201400387

Henzler, K., Haupt, B., and Ballauff, M. (2008). Enzymatic activity of immobilized enzyme determined by isothermal titration calorimetry. *Anal. Biochem.* 378, 184–189. doi: 10.1016/j.ab.2008.04.011

Honarmand Ebrahimi, K., Hagedoorn, P.-L., Jacobs, D., and Hagen, W. R. (2015). Accurate label-free reaction kinetics determination using initial rate heat measurements. *Sci. Rep.* 5:16380.

Hooff, G. P., van Kampen, J. J. A., Meesters, R. J. W., van Belkum, A., Goessens, W. H. F., and Luider, T. M. (2012). Characterization of β-Lactamase enzyme activity in bacterial lysates using MALDI-mass spectrometry. *J. Prot. Res.* 11, 79–84. doi: 10.1021/pr200858r

Hörster, F., Schwab, M. A., Sauer, S. W., Pietz, J., Hoffmann, G. F., Okun, J. G., et al. (2006). Phenylalanine reduces synaptic density in mixed cortical cultures from mice. *Pediatr. Res.* 59, 544–548. doi: 10.1203/01.pdr.0000203091.45988.8d

Hulme, E. C., and Trevethick, M. A. (2010). Ligand binding assays at equilibrium: validation and interpretation. *Br. J. Pharmacol.* 161, 1219–1237. doi: 10.1111/j.1476-5381.2009.00604.x

Hunter, T. (1995). Protein kinases and phosphatases: the Yin and Yang of protein phosphorylation and signaling. *Cell* 80, 225–236. doi: 10.1016/0092-8674(95)90405-0

Jeoh, T., Baker, J. O., Ali, M. K., Himmel, M. E., and Adney, W. S. (2005). β-d-Glucosidase reaction kinetics from isothermal titration microcalorimetry. *Anal. Biochem.* 347, 244–253. doi: 10.1016/j.ab.2005.09.031

Kaeswurm, J. A. H., Claasen, B., Fischer, M.-P., and Buchweitz, M. (2019). Interaction of structurally diverse phenolic compounds with porcine pancreatic α-Amylase. *J. Agric. Food Chem.* 67, 11108–11118. doi: 10.1021/acs.jafc.9b04798

Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A., and Schuck, P. (2012). High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.* 84, 5066–5073. doi: 10.1021/ac3007522

Kerner, J., and Hoppel, C. L. (2002). Radiochemical malonyl-CoA decarboxylase assay: Activity and subcellular distribution in heart and skeletal muscle. *Anal. Biochem.* 306, 283–289. doi: 10.1006/abio.2002.5696

Klaus, L., Barbara, G., and Florian, R. (1999). Evaluation of a direct α-amylase assay using 2-Chloro-4-nitrophenyl-α-D-maltotrioside. *Clin. Chem. Lab. Med.* 37, 1053–1062.

Koshland, , D. E. Jr, Nemethy, G., and Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5, 365–385. doi: 10.1021/bi00865a047

Kosikowska, P., and Berlicki, L. (2011). Urease inhibitors as potential drugs for gastric and urinary tract infections: a patent review. *Exp. Opin. Ther. Pat.* 21, 945–957. doi: 10.1517/13543776.2011.574615

Kramer, S. J., and Pochapin, M. B. (2012). Gastric phytobezoar dissolution with ingestion of diet coke and cellulase. *Gastroenterol. Hepatol.* 8, 770–772.

Kuhad, R. C., Gupta, R., and Singh, A. (2011). Microbial cellulases and their industrial applications. *Enzyme Res.* 280696:280610.

Lehoczki, G., Szabó, K., Takács, I., Kandra, L., and Gyémánt, G. (2016). Simple ITC method for activity and inhibition studies on human salivary α-amylase. *J. Enzyme Inhibit. Med. Chem.* 31, 1648–1653. doi: 10.3109/14756366.2016.1161619

Liang, Y. (2008). Applications of isothermal titration calorimetry in protein science. *Acta Biochim. Biophys. Sin.* 40, 565–576. doi: 10.1111/j.1745-7270.2008.00437.x

Lonhienne, T., Baise, E., Feller, G., Bouriotis, V., and Gerday, C. (2001). Enzyme activity determination on macromolecular substrates by isothermal titration calorimetry: application to mesophilic and psychrophilic chitinases. *Biochim. Biophys. Acta* 1545, 349–356. doi: 10.1016/s0167-4838(00)00296-x

Lonhienne, T. G. A., Reilly, P. E. B., and Winzor, D. J. (2003). Further evidence for the reliance of catalysis by rabbit muscle pyruvate kinase upon isomerization of the ternary complex between enzyme and products. *Biophys. Chem.* 104, 189–198. doi: 10.1016/s0301-4622(02)00366-6

Lonhienne, T. G. A., and Winzor, D. J. (2002). Calorimetric demonstration of the potential of molecular crowding to emulate the effect of an allosteric activator on pyruvate kinase kinetics. *Biochemistry* 41, 6897–6901. doi: 10.1021/bi020064h

López-Mayorga, O., Mateo, P. L., and Cortijo, M. (1987). The use of different input signals for dynamic characterisation in isothermal microcalorimetry. *Sci. Instr.* 20:265. doi: 10.1088/0022-3735/20/3/006

Lunn, F. A., MacDonnell, J. E., and Bearne, S. L. (2008). Structural requirements for the activation of *Escherichia coli* CTP synthase by the allosteric effector GTP are stringent, but requirements for inhibition are lax. *J. Biol. Chem.* 283, 2010–2020. doi: 10.1074/jbc.m707803200

Lv, M., Zhang, Y.-J., Zhou, F., Ge, Y., Zhao, M.-H., Liu, Y., et al. (2019). Real-time monitoring of D-Ala-D-Ala dipeptidase activity of VanX in living bacteria by isothermal titration calorimetry. *Anal. Biochem.* 578, 29–35. doi: 10.1016/j.ab.2019.05.002

Malvern (2010). *VP-ITC Microcalorimeter User's Manual*. Cambridge: Malvern.

Malvern (2014). *ITC-200 Microcalorimeter User's Manual*. Cambridge: Malvern.

Malvern (2016). *Microcal itc Systems:Understanding Biomolecular Interactions*. Cambridge: Malvern.

Mason, M., Scampicchio, M., Quinn, C. F., Transtrum, M. K., Baker, N., Hansen, L. D., et al. (2018). Calorimetric methods for measuring stability and reusability of membrane immobilized enzymes. *J. Food Sci.* 83, 326–331. doi: 10.1111/1750-3841.14023

Maximova, K., and Trylska, J. (2015). Kinetics of trypsin-catalyzed hydrolysis determined by isothermal titration calorimetry. *Anal. Biochem.* 486, 24–34. doi: 10.1016/j.ab.2015.06.027

Maximova, K., Wojtczak, J., and Trylska, J. (2019). Enzyme kinetics in crowded solutions from isothermal titration calorimetry. *Anal. Biochem.* 567, 96–105. doi: 10.1016/j.ab.2018.11.006

Mazzei, L., Ciurli, S., and Zambelli, B. (2016). Methods enzymol. *Elsevier* 567, 215–236. doi: 10.1016/s0962-8924(00)89005-4

McKay, G. A., and Wright, G. D. (1995). Kinetic mechanism of aminoglycoside phosphotransferase type IIIa. Evidence for a Theorell-Chance mechanism. *J. Biol. Chem.* 270, 24686–24692. doi: 10.1074/jbc.270.42.24686

Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12, 88–118. doi: 10.1016/s0022-2836(65)80285-6

Morishita, Y., Iinuma, Y., Nakashima, N., Majima, K., Mizuguchi, K., and Kawamura, Y. (2000). Total and pancreatic amylase measured with 2-chloro-4-nitrophenyl-4-O-β-D-galactopyranosylmaltoside. *Clin. Chem.* 46, 928–933. doi: 10.1093/clinchem/46.7.928

Morisseau, C., Inceoglu, B., Schmelzer, K., Tsai, H. J., Jinks, S. L., Hegedus, C. M., et al. (2010). Naturally occurring monoepoxides of eicosapentaenoic acid and docosahexaenoic acid are bioactive antihyperalgesic lipids. *J. Lipid Res.* 51, 3481–3490. doi: 10.1194/jlr.m006007

Murphy, L., Baumann, M. J., Borch, K., Sweeney, M., and Westh, P. (2010a). An enzymatic signal amplification system for calorimetric studies of cellobiohydrolases. *Anal. Biochem.* 404, 140–148. doi: 10.1016/j.ab.2010.04.020

Murphy, L., Bohlin, C., Baumann, M. J., Olsen, S. N., Sørensen, T. H., Anderson, L., et al. (2013). Product inhibition of five Hypocrea jecorina cellulases. *Enzyme Microb. Technol.* 52, 163–169. doi: 10.1016/j.enzmictec.2013.01.002

Murphy, L., Borch, K., McFarland, K. C., Bohlin, C., and Westh, P. (2010b). A calorimetric assay for enzymatic saccharification of biomass. *Enzyme Microb. Technol.* 46, 141–146. doi: 10.1016/j.enzmictec.2009.09.009

Murphy, L., Cruys-Bagger, N., Damgaard, H. D., Baumann, M. J., Olsen, S. N., Borch, K., et al. (2012). Origin of initial burst in activity for Trichoderma reesei endo-glucanases hydrolyzing insoluble cellulose. *J. Biol. Chem.* 287, 1252–1260. doi: 10.1074/jbc.m111.276485

Oezen, C., and Serpersu, E. H. (2004). Thermodynamics of aminoglycoside binding to Aminoglycoside-3'-phosphotransferase IIIa studied by isothermal titration calorimetry. *Biochemistry* 43, 14667–14675. doi: 10.1021/bi0487286

Palzkill, T. (2013). Metallo-β-lactamase structure and function. *Ann. N.Y. Acad. Sci.* 1277, 91–104. doi: 10.1111/j.1749-6632.2012.06796.x

Pedroso, M. M., Ely, F., Lonhienne, T., Gahan, L. R., Ollis, D. L., Guddat, L. W., et al. (2014). Determination of the catalytic activity of binuclear metallohydrolases using isothermal titration calorimetry. *Eur. J. Biochem.* 19, 389–398. doi: 10.1007/s00775-013-1079-0

Perutz, M. F. (1989). Mechanisms of cooperativity and allosteric regulation in proteins. *Quart. Rev. Biophys.* 22, 139–237. doi: 10.1017/s0033583500003826

Piñeiro, Á, Muñoz, E., Sabín, J., Costas, M., Bastos, M., Velázquez-Campoy, A., et al. (2019). AFFINImeter: a software to analyze molecular recognition processes from experimental data. *Anal. Biochem.* 577, 117–134. doi: 10.1016/j.ab.2019.02.031

Ramasubbu, N., Ragunath, C., Sundar, K., Mishra, P. J., Gyemant, G., and Kandra, L. (2005). Structure-function relationships in human salivary α-amylase: role of aromatic residues. *Biologia* 60, 47–56.

Reetz, M. T. (2001). Combinatorial and evolution-based methods in the creation of enantioselective catalysts. *Angew. Chem., Int. Ed.* 40, 284–310. doi: 10.1002/1521-3773(20010119)40:2<284::aid-anie284>3.0.co;2-n

Reetz, M. T., Daligault, F., Brunner, B., Hinrichs, H., and Deege, A. (2004). Directed evolution of cyclohexanone monooxygenases: enantioselective biocatalysts for the oxidation of prochiral thioethers. *Angew. Chem. Int. Ed. Engl.* 43, 4078–4081. doi: 10.1002/anie.200460311

Reyes-Turcu, F. E., Ventii, K. H., and Wilkinson, K. D. (2009). Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu. Rev. Biochem.* 78, 363–397. doi: 10.1146/annurev.biochem.78.082307.091526

Rohatgi, N., Gudmundsson, S., and Rolfsson, O. (2015). Kinetic analysis of gluconate phosphorylation by human gluconokinase using isothermal titration calorimetry. *FEBS Lett.* 589, 3548–3555. doi: 10.1016/j.febslet.2015.10.024

Seethala, R., and Menzel, R. (1997). A homogeneous, fluorescence polarization assay for Src-family tyrosine kinases. *Anal. Biochem.* 253, 210–218. doi: 10.1006/abio.1997.2365

Shu, Q., and Frieden, C. (2005). Relation of enzyme activity to local/global stability of murine Adenosine Deaminase: 19F NMR Studies. *J. Mol. Biol.* 345, 599–610. doi: 10.1016/j.jmb.2004.10.057

Siddiqui, K. S., Ertan, H., Charlton, T., Poljak, A., Daud Khaled, A. K., Yang, X., et al. (2014). Versatile peroxidase degradation of humic substances: use of isothermal titration calorimetry to assess kinetics, and applications to industrial wastes. *J. Biotechnol.* 178, 1–11. doi: 10.1016/j.jbiotec.2014.03.002

Sigurskjold, B. W. (2000). Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal. Biochem.* 277, 260–266. doi: 10.1006/abio.1999.4402

Spencer, S. D., and Raffa, R. B. (2004). Isothermal titration calorimetric study of RNase-A kinetics (cCMP → 3′-CMP) involving end-product inhibition. *Pharm. Res.* 21, 1642–1647. doi: 10.1023/b:pham.0000041460.78128.0f

Stroberg, W., and Schnell, S. (2016). On the estimation errors of KM and V from time-course experiments using the Michaelis–Menten equation. *Biophys. Chem.* 219, 17–27. doi: 10.1016/j.bpc.2016.09.004

Su, H., and Xu, Y. (2018). Application of ITC-based characterization of thermodynamic and kinetic association of ligands with proteins in drug design. *Front. Pharmacol.* 9:1133. doi: 10.3389/fphar.2018.01133

Sun, Y., and Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour. Technol.* 83, 1–11. doi: 10.1016/s0960-8524(01)00212-7

Szabó, K., Kandra, L., and Gyémánt, G. (2019). Studies on the reversible enzyme reaction of rabbit muscle glycogen phosphorylase b using isothermal titration calorimetry. *Carbohydr. Res.* 477, 58–65. doi: 10.1016/j.carres.2019.03.014

TA (2019). *Microcalorimetry: Itc & Dsc*. New Castle, DE: TA Instruments.

Tauran, Y., Anjard, C., Kim, B., Rhimi, M., and Coleman, A. W. (2014). Large negatively charged organic host molecules as inhibitors of endonuclease enzymes. *Chem. Commun.* 50, 11404–11406. doi: 10.1039/c4cc04805a

Tellinghuisen, J. (2008). Isothermal titration calorimetry at very low c. *Anal. Biochem.* 373, 395–397. doi: 10.1016/j.ab.2007.08.039

Todd, M. J., and Gomez, J. (2001). Enzyme kinetics determined using calorimetry: a general assay for enzyme activity? *Anal. Biochem.* 296, 179–187. doi: 10.1006/abio.2001.5218

Transtrum, M. K., Hansen, L. D., and Quinn, C. (2015). Enzyme kinetics determined by single-injection isothermal titration calorimetry. *Methods* 76, 194–200. doi: 10.1016/j.ymeth.2014.12.003

Turner, P., Mamo, G., and Karlsson, E. N. (2007). Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microb. Cell Fact.* 6:9. doi: 10.1186/1475-2859-6-9

Upadhyay, L. S. B. (2012). Urease inhibitors: a review. *Indian J. Biotechnol.* 11, 381–388.

van Spronsen, F. J., Hoeksma, M., and Reijngoud, D.-J. (2009). Brain dysfunction in phenylketonuria: Is phenylalanine toxicity the only possible cause? *J. Inherit. Metab. Dis.* 32:46. doi: 10.1007/s10545-008-0946-2

Vander Meulen, K. A., and Butcher, S. E. (2012). Characterization of the kinetic and thermodynamic landscape of RNA folding using a novel application of isothermal titration calorimetry. *Nucl. Acids Res.* 40, 2140–2151. doi: 10.1093/nar/gkr894

Velazquez-Campoy, A., and Freire, E. (2006). Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat. Protoc.* 1, 186–191. doi: 10.1038/nprot.2006.28

Velázquez-Campoy, A., López-Mayorga, O., and Cabrerizo-Vílchez, M. (1999). Determination of the rigorous transfer function of an isothermal titration microcalorimeter with peltier compensation. *J. Ther. Anal. Calorim.* 57, 343–359.

Wagner, K. M., McReynolds, C. B., Schmidt, W. K., and Hammock, B. D. (2017). Soluble epoxide hydrolase as a therapeutic target for pain, inflammatory and neurodegenerative diseases. *Pharmacol. Ther.* 180, 62–76. doi: 10.1016/j.pharmthera.2017.06.006

Wang, F.-Q., Xie, H., Chen, W., Wang, E.-T., Du, F.-G., and Song, A.-D. (2013). Biological pretreatment of corn stover with ligninolytic enzyme for high efficient enzymatic hydrolysis. *Bioresour. Technol.* 144, 572–578. doi: 10.1016/j.biortech.2013.07.012

Wang, Q., He, Y., Lu, R., Wang, W.-M., Yang, K.-W., Hai, M. F., et al. (2018). Thermokinetic profile of NDM-1 and its inhibition by small carboxylic acids. *Biosci. Rep.* 38:BSR20180244.

Wang, W.-J., Wang, Q., Zhang, Y., Lu, R., Zhang, Y.-L., Yang, K.-W., et al. (2017). Characterization of β-lactamase activity using isothermal titration calorimetry. *Biochim. Biophys. Acta* 1861, 2031–2038.

Wang, Y., Guan, J., Di Trani, J. M., Auclair, K., and Mittermaier, A. K. (2019). Inhibition and activation of kinases by reaction products: a reporter-free assay. *Anal. Chem.* 91, 11803–11811. doi: 10.1021/acs.analchem.9b02456

Willemoës, M., and Sigurskjold, B. W. (2002). Steady-state kinetics of the glutaminase reaction of CTP synthase from Lactococcus lactis. *Eur. J. Biochem.* 269, 4772–4779. doi: 10.1046/j.1432-1033.2002.03175.x

Wiseman, T., Williston, S., Brandts, J. F., and Lin, L.-N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* 179, 131–137. doi: 10.1016/0003-2697(89)90213-3

Zebisch, M., Krauss, M., Schäfer, P., and Sträter, N. (2012). Crystallographic evidence for a domain motion in rat nucleoside triphosphate diphosphohydrolase (NTPDase) 1. *J. Mol. Biol.* 415, 288–306. doi: 10.1016/j.jmb.2011.10.050

Zhang, Y., Doruker, P., Kaynak, B., Zhang, S., Krieger, J., Li, H., et al. (2020). Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Curr. Opin. Struct. Biol.* 62, 14–21. doi: 10.1016/j.sbi.2019.11.002

Zhang, Y.-J., Wang, W.-M., Oelschlaeger, P., Chen, C., Lei, J.-E., Lv, M., et al. (2018). Real-time monitoring of NDM-1 activity in live bacterial cells by isothermal titration calorimetry: a new approach to measure inhibition of antibiotic-resistant bacteria. *ACS Infect. Dis.* 4, 1671–1678. doi: 10.1021/acsinfecdis.8b00147

Zheng, C. J., Han, L. Y., Yap, C. W., Ji, Z. L., Cao, Z. W., and Chen, Y. Z. (2006). Therapeutic targets: progress of their exploration and investigation of their characteristics. *Pharmacol. Rev.* 58, 259–279. doi: 10.1124/pr.58.2.4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 05 November 2020
doi: 10.3389/fmolb.2020.586970

Is Crocin a Potential Anti-tumor Candidate Targeting Microtubules? Computational Insights From Molecular Docking and Dynamics Simulations

Ze Wang[1†], Juan Ren[1†], Nengzhi Jin[2], Xingyi Liu[3] and Xiaofei Li[1*]*

[1] Department of Pharmaceutical Sciences, Zunyi Medical University at Zhuhai Campus, Zhuhai, China, [2] Gansu Computing Center, Lanzhou, China, [3] Center for Systems Biology, Department of Bioinformatics, School of Biology and Basic Medical Sciences, Soochow University, Suzhou, China

OPEN ACCESS

Edited by:
Hongchun Li,
Shenzhen Institutes of Advanced Technology (CAS), China

Reviewed by:
Zhiwei Feng,
University of Pittsburgh, United States
Khurshid Ahmad,
Yeungnam University, South Korea

***Correspondence:**
Ze Wang
wz@zmu.edu.cn;
felixwang08@hotmail.com
Xiaofei Li
lixiaofei@zmu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *24 July 2020*
Accepted: *13 October 2020*
Published: *05 November 2020*

Citation:
Wang Z, Ren J, Jin N, Liu X and Li X (2020) Is Crocin a Potential Anti-tumor Candidate Targeting Microtubules? Computational Insights From Molecular Docking and Dynamics Simulations. Front. Mol. Biosci. 7:586970. doi: 10.3389/fmolb.2020.586970

Although it is known crocin, a hydrophilic compound from the herbal plant *Crocus sativus* L., has promising antitumor activity, the detailed mechanism of its antitumor activity was not well understood. Recent experiments suggested tubulin as the primary target for the antitumor activity of crocin. However, due to a lack of crystal structure of tubulin bound with crocin, the exact binding mode and interaction between crocin and tubulin remains exclusive. In the present work, a computational study by integrating multiple conformation docking, molecular dynamics simulation as well as residue interaction network analysis was performed to investigate the molecular mechanism of crocin-tubulin interaction. By comparing the docking score, the most likely binding mode CRO_E1 were identified from 20 different binding modes of crocin in the vinca binding pockets. Further molecular dynamics simulation of CRO_E1 complex showed the binding of crocin is more stable than the inhibitor soblidotin and vinblastine. During the simulation course, an excessive number of hydrogen bonds were observed for the ligand crocin. The binding free energy of crocin-tubulin complex was calculated as -79.25 ± 7.24 kcal/mol, which is almost twice of the ligand soblidotin and vinblastine. By using energy decomposition, hot residues for CRO_E1 were identified as Gln^{11}, Gln^{15}, Thr^{72}, Ser^{75}, Pro^{173}-Lys^{174}-Val^{175}-Ser^{176}-Asp^{177}, Tyr^{222}, and Asn^{226} in the β-chain, and Asp^{245}, Ala^{247}-Leu^{248}, Val^{250}, Asn^{329}, and Ile^{332} in the α-chain. Residue interaction network analysis also showed the importance of these hot residues in the interaction network of crocin-tubulin complex. In addition, a common residue motif Val^{175}-Xxx^{176}-Asp^{177} was discovered for all three bindings, suggesting its importance in future drug design. The study could provide valuable insights into the interaction between crocin and tubulin, and give suggestive clues for further experimental studies.

Keywords: tubulin, anti-tumor activity, molecular docking, molecular dynamics simulation, binding free energy, residue interaction network

INTRODUCTION

Exploiting drug candidates from traditional Chinese medicine is of great interests in drug discovery. Saffron is the dried stigma of *Crocus sativus* L., which is a species of the Iridaceae family widely cultivated in China, Iran, India, Italy, Israel, Spain, and Turkey (Bathaie and Mousavi, 2010; Alavizadeh and Hosseinzadeh, 2014). Since ancient times, saffron is used as a dietary ingredient as well as medicinal herb in the treatment of various diseases (Bathaie and Mousavi, 2010). Crocin (CRO) is a hydrophilic carotenoid that are separated from saffron (Alavizadeh and Hosseinzadeh, 2014). As one of the main characteristic ingredients, CRO and its derivatives account for nearly 10% of total compounds in saffron (Pfander and Wittwer, 1975; Tsimidou and Tsatsaroni, 1993). Chemically, CRO is a di-glycosyl polyene ester of crocetin containing a 20-carbon carotenoid backbone and two D-gentiobioses as carbohydrate moieties (Alavizadeh and Hosseinzadeh, 2014). Experiments have shown that CRO has wide pharmacological effects including antioxidant, neuroprotective, antidepressant and antiproliferative (Alavizadeh and Hosseinzadeh, 2014). More importantly, the good hydrophilic property of CRO made it an attractive candidate in drug development.

Pharmacological studies showed that CRO exhibits promising antitumor activities (Bolhassani et al., 2014; Hoshyar and Mollaei, 2017). Several mechanisms were proposed to understand the antitumor activity of CRO, including inhibition of DNA and RNA synthesis (Abdullaev et al., 2003), interaction with topoisomerases (Bajbouj et al., 2012), induction of apoptosis (Sun et al., 2013; Amin et al., 2015), and so on. However, one of the drawbacks of these mechanisms is the lack of clarifying the primary target protein of CRO. Recently, biochemical as well as proteomic approaches suggested microtubules as the primary target of CRO (Hosseinzadeh et al., 2013; Hire et al., 2017; Sawant et al., 2019). Microtubule is a dynamic biopolymer composed of tubulin, which is a heterodimer composed of β and α subunit (Dumontet and Jordan, 2010). Microtubule dynamics, i.e., the assembly or disassembly of tubulin, plays essential roles in cell cycle (Dumontet and Jordan, 2010). The interference of microtubule dynamics could induce mitotic arrest and cell apoptosis. Due to the reason, tubulin is a target for a number of antitumor drugs including vinblastine (VBL), paclitaxel and colchicine (Dumontet and Jordan, 2010). It was found that CRO could competitively bind with tubulin at VBL site, disrupting microtubule dynamics and inhibiting cell proliferation (Hire et al., 2017; Sawant et al., 2019).

However, due to a lack of crystal structure of tubulin bound with CRO, the binding mode and detailed molecular interaction between tubulin and CRO is still unknown. In this work, we investigated the interaction between tubulin and CRO through computational approaches. The possible binding modes of CRO were explored through multiple conformation docking strategy. Then, molecular dynamics simulation was performed to fully consider the flexibility of tubulin and CRO. Molecular mechanics/generalized born surface area (MM/GBSA) method was applied to obtain a detailed energy contribution from key contact residues. Additionally, the underlying characteristics of key residues were analyzed from residue interaction network. Our study could provide valuable insights into the interaction between CRO and tubulin at molecular level, and give suggestive clues for further experimental studies.

MATERIALS AND METHODS

Structure Preparation

The structure of tubulin having different vinca binding pockets were obtained from the Research Collaboration for Structural Bioinformatics protein database, including 1Z2B (bound with VBL), 3DU7 (bound with phomopsin A), 3E22 (bound with soblidotin, SBD) and 5NJH (bound with triazolopyrimidine). Molecular Operating Environment (MOE, 2019) software was used for structural preparation. Each structural data was cleaned by removing all unnecessary subunits and small molecules, leaving a ligand molecule, β and α-subunit. Missing amino acid residues and hydrogen atoms were added by QuickPrep in MOE. Energy minimization was performed by using Amber10 force field, with 0.1 RMS kcal/mol/A^2 as a gradient.

Multiple Conformation Docking With MOE

After the preparation of tubulin dimer with different vinca pocket conformations, multiple conformation docking was performed with MOE. In the multiple conformation docking strategy, an ensemble of different pocket conformations was used instead of a specific pocket conformation. Multiple conformation docking is different from traditional docking protocol, allowing the investigation and comparison of conformational variations of binding pockets. In order to compare the binding mode, all prepared tubulin structures were superimposed with reference to 1Z2B. The conformational difference of the binding pockets was measured by an MOE SVL script.

Retrieved from PubChem, the ligand structure of CRO (**Figure 1D**) was imported in MOE and docked into the vinca binding site of each prepared conformation. The vinca binding site was defined as residues within 4.5 Å to the ligand of each prepared tubulin structure. For ligand docking with each receptor conformation, a set of 30 ligand conformations was produced to account for ligand flexibility. Docking structures were then refined by Amber10 force field and finally five poses were generated and ranked according to GBVI/WSA ΔG scoring method. The scoring function is defined as following:

$$\Delta G_{bind} \approx \gamma \left[\frac{2}{3} (\Delta E_{ele} + \Delta E_{sol}) + \Delta E_{vdW} + \delta \Delta SA \right] + K \quad (1)$$

where ΔE_{ele}, ΔE_{sol} and ΔE_{vdW} are the electrostatic, solvation, and van der Waals terms, respectively, ΔSA is exposed SA, K is the average entropy change, γ and δ are two parameters obtained by training. The GBVI/WSA ΔG scoring function was trained by the forcefield MMFF94x and AMBER99 to estimate the binding free energy between a ligand conformer and a binding pocket

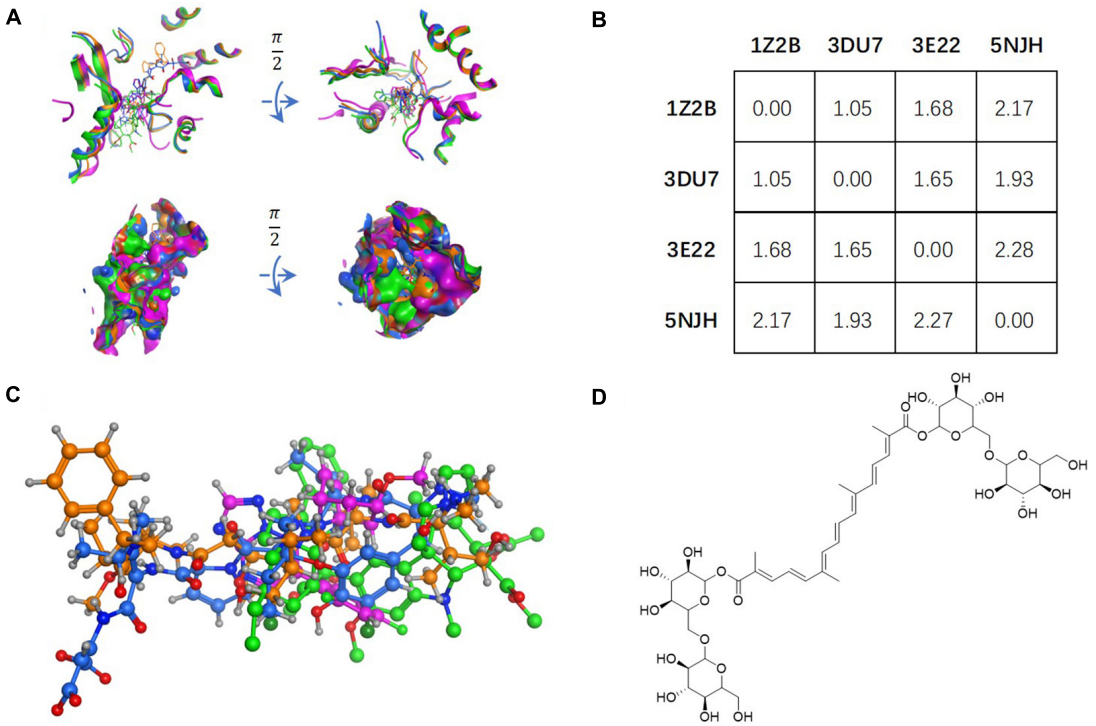


FIGURE 1 | Tubulin bound with four different ligands (1Z2B, 3DU7, 3E22, and 5NJH, see the "Materials and Methods" part for more information) were compared by: **(A)** superposition of four binding pockets; **(B)** calculating RMSD matrix in angstrom; **(C)** superposition of four ligands. The chemical structure of CRO was shown in panel **(D)**.

(Naïm et al., 2007). The CRO poses at each corresponding tubulin were filtered to remain the one with the highest ΔG score. After that, the CRO-tubulin complex was exported for further molecular dynamics simulations.

Molecular Dynamics Simulations

After the docking step, a set of CRO binding modes were obtained for different pocket conformations of tubulin. Each binding complex was further analyzed by molecular dynamics (MD) simulation. The AMBER18 program was used to perform MD simulation. The ff14SB force field parameters were assigned to the prepared tubulin structure. For ligand molecules, the force field parameters described by General Force Field (GAFF) (Wang et al., 2004) were generated using the Antechamber program in AMBER18. The RESP charge fitting technique (Bayly et al., 1993; Cieplak et al., 1995; Fox and Kollman, 1998) was applied to calculate partial charges of ligands. The ligand and tubulin structure were then combined by using the LEaP program. A rectangular periodic box of water molecules was generated by using TIP3P water model (Jorgensen et al., 1983), extending at least 10 Å in each direction. The whole system was neutralized with sodium ions as counterions.

Three steps of minimization were performed in prior to MD simulation. In the first stage, only the positions of water molecules were optimized by fixing ligand-tubulin complex with a restraint force constant of 10.0 kcal/mol/Å^2. In the second stage, the restrains on the complex were partially released by only fixing C_α, N, O with a restraint constant of 5.0 kcal/mol/Å^2. In the third stage, the entire system in solvated box was minimized by releasing all restraints. Each minimization steps contained 10,000 cycles including the first 1,000 cycles of the steepest descent algorithm and the remaining 9,000 cycles of conjugate gradient method. The minimized structure was used as starting input for MD simulation. The temperature of system was gradually raised from 0 to 300 K in 200 ps canonical ensemble (fixed N, V, and T) heating process by applying the Langevin dynamics with a collision frequency of 2.0. The system was equilibrated by 300 ps NPT equilibration (fixed N, P, and T) at 1.0 bar and 300 K, with all residues restrained by a force constant of 1.0 kcal/mol/Å^2. Finally, the position restraints were released, and a production phase of 90 ns was performed under the same conditions as in NPT equilibration. Coordinates were saved for every 10 ps. In all of the MD simulations, 2.0 fs was used as time step and 8.0 Å was used as short-range cutoff value for non-bonded interactions. The long-range electrostatic interactions were calculated through the particle-mesh Ewald (PME) method (Darden et al., 1993). Bond restraints including hydrogen atoms were realized by applying SHAKE algorithm (Ryckaert et al., 1977). MD trajectories were processed and analyzed by evaluating RMSD value of the tubulin and ligands. The RMSF and the hydrogen bond analysis were performed by cpptraj tool in AMBER18. The same protocol was applied for all simulation processes of different conformation of binding pockets and ligands.

MM/GBSA Binding Energy Calculation

For each ligand-tubulin complex, the MD trajectory was used to estimate the binding energy (ΔG_{total}) between ligand and tubulin, which is the sum of van der Waals, electrostatic, polar and non-polar solvent energies. To effectively calculate the binding energy, MM/GBSA method (Wang et al., 2017, 2019) was applied to the following thermodynamic relation:

$$\Delta G_{bind,sol} = \Delta G_{bind,vac} + \Delta G_{com,sol} - (\Delta G_{lig,sol} + \Delta G_{rec,sol}) \quad (2)$$

where $\Delta G_{bind,\ sol}$ and $\Delta G_{bind,vac}$ are the binding energies in solvent condition and vacuum condition, respectively, and $\Delta G_{com,sol}$, $\Delta G_{lig,sol}$ and $\Delta G_{rec,sol}$ are the solvation free energies of complex, ligand, and receptor, respectively. The solvation free energy can be attributed to an electrostatic and a non-electrostatic contribution through the equation:

$$\Delta G_{sol} = G_{ele}|_{\epsilon=1}^{\epsilon=80} + \Delta G_{nonele} \quad (3)$$

The electrostatic contribution can be solved by the linearized GB method, while the non-electrostatic contribution can be estimated by an empirical SA term. In this study, we used the solute dielectric constant of 1, the solvent dielectric constant of 80, and water probe radius of 1.4 Å. ΔG_{vac} is determined by calculating non-bonded interaction energy (ΔE_{MM}) between ligand and receptor and entropy change (ΔS_{NMA}) during ligand binding:

$$\Delta G_{vac} = \Delta E_{MM} - T \bullet \Delta S_{NMA} \quad (4)$$

In case of different ligands binding to the same protein, the entropy contribution can be neglected if only the hotspot residues and interaction features rather than the absolute Gibbs free energy were to be evaluated. For this reason, we collected multiple snapshots from MD trajectory for the MM/GBSA calculation at 100 ps intervals. The binding energies between different conformations of binding pocket of tubulin and ligands were obtained and compared for further analysis. In addition, to achieve a detailed picture of the interaction between ligand and tubulin, MM/GBSA method was applied to decompose the interaction energy at a per-residue basis without considering entropy contributions.

Residue Network Calculation

The web server RING-2.0 (Piovesan et al., 2016) was used to build the residue interaction network by using protein and protein-ligand structures. RING-2.0 algorithm could derive a network through two steps, i.e., identifying node-node pair by measuring physical distance and recognizing the interaction type of each pair (Piovesan et al., 2016). In the computation, we have considered all atoms of each residue for distance measurement and display only one interaction per interaction type for simplicity reason. Then, the derived networks were imported into Cytoscape (Shannon et al., 2003) for topological analysis. In the network graph, residues and interactions between residues were represented as nodes and edges between nodes, respectively. The degree, betweenness and closeness centrality was computed by using NetworkAnalyzer (Assenov et al., 2007), which are key values measuring the importance or centrality of a node in the network.

RESULTS AND DISCUSSION

Multiple Conformation Docking

The vinca binding pocket of tubulin dimer has different conformations while bound to different inhibitors. As screened from the Protein Data Bank, at least four entities were found to represent tubulin bound to structurally different inhibitors at vinca binding pocket. The PDB structures include tubulin-VBL complex (PDB ID: 1Z2B), tubulin-phomopsin A complex (PDB ID: 3DU7), tubulin-SBD complex (PDB ID: 3E22) and tubulin-triazolopyrimidine complex (PDB ID: 5NJH) (**Figure 1C**). According to induced-fit theory, the shape of the binding cavity will change according to ligand geometries. Comparison of these binding pockets indicated a great deal of structural variety upon binding of structurally diverse ligands (**Figure 1A**). The RMSD matrix showed the structural differences among four binding pockets. Despite the similarity between the binding pockets of 1Z2B and 3DU7, the binding cavities varies significantly (with RMSDs > 1.5 Å) (**Figures 1A,B**).

Ligand geometry could significantly change the conformation of the same binding pocket. Since the binding mode of CRO is largely unknown, the exploration of docking by different conformations of binding pockets allows to probe the binding mode and interaction between CRO and tubulin. In the study, CRO was docked into four different conformations of the binding pocket of tubulin by using MOE software. The selection of tubulin structures (PDB ID: 1Z2B, 3DU7, 3E22, and 5NJH) from the Protein Data Bank helps to investigate and compare different binding modes of CRO.

Figure 2A showed the docking matrix of possible binding modes of CRO by multiple conformation docking method. Each row represents the conformation of the binding pocket of tubulin, where Z, D, E, and N stands for 1Z2B, 3DU7, 3E22, and 5NJH, respectively. By MOE docking, the first five top ranked conformers of CRO were listed for each binding pocket. The binding matrix therefore has collected a total number of 20 different binding modes of CRO (**Figure 2A**). The RMSD values of the screened conformers of CRO were calculated and listed in **Figure 2B**. Ranging from 7.07 to 12.26 Å, the RMSD matrix indicated that a significantly diversity of the ligand geometry was obtained from the multiple conformation docking method. The conformers of CRO with the highest score in each binding pocket were shown in **Figure 3**. As shown in the figure, the geometry and orientation of CRO differs significantly in the four binding pockets. In fact, the conformers of VBL and their locations in four pockets also varies in different binding pockets (**Figure 2C**). This indicated the flexibility of ligand in binding with a specified pocket conformation, and also rationalized the necessity for performing multiple conformation docking.

The detailed interactions between tubulin and ligand for different binding modes were analyzed, and residues involving the binding interaction were plotted in **Supplementary**

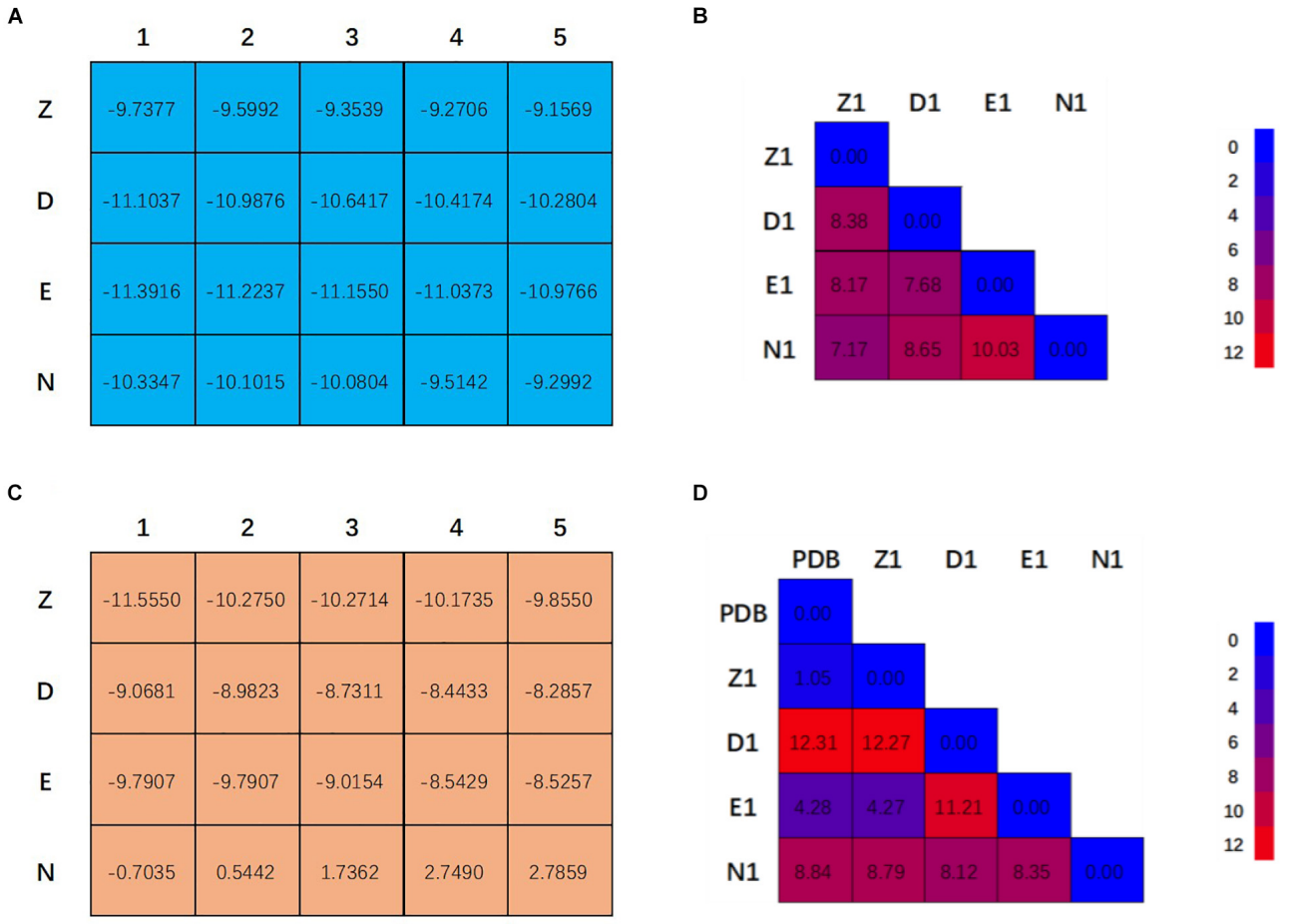


FIGURE 2 | Twenty binding modes for CRO **(A)** and VBL **(C)** discovered by multiple conformation docking strategy. The RMSD matrix of CRO **(B)** and VBL **(D)** was created to compare the different ligand conformations. For VBL, PDB structure was used to demonstrate the possibility of screening correct ligand conformation by using multiple conformation docking strategy. All units are in Å.

Figure S1. The interacting residues were highlighted in the protein sequence as shown in **Supplementary Figure S2** (for ligand CRO) and S3 (for ligand VBL). The 2D map is a projection of 3D structure in **Figure 3**, which provides a clear representation of the binding interaction in 3D structure. Interestingly, although the ligand pose varies significantly (**Figures 2B,D**), some common modes were observed for the protein residues involving the binding interaction. For the ligand CRO, the common modes shared Gln^{15}, Val^{175}-Ser^{176}, Tyr^{208}, Pro^{220}-Thr^{221}-Tyr^{222} in the β-chain, and Leu^{248}, Pro^{325}, Val^{328}-Asn^{329}, Ile^{332}, Phe^{351}, Val^{353}, Ile^{355} in the α-chain. Similar patterns were observed in the binding mode of the ligand VBL, including Val^{175}-Ser^{176}, Pro^{220}-Thr^{221}-Tyr^{222} in the β-chain, and Leu^{248}, Pro^{325}, Val^{328}-Asn^{329}, Phe^{351}, Val^{353}, Ile^{355} in the α-chain.

As can be seen in **Figure 2A**, E1 for CRO is the most favorable binding mode from the perspective of binding energy. Actually, the GBVI/WSA ΔG scores for the first three modes in 3E22 binding pocket are higher than other investigated binding pockets, indicating 3E22 is the most likely conformation for the binding pocket of CRO. In comparison, the most favorable binding mode for VBL is Z1. Since the crystal structure of tubulin bound to VBL has been solved, we compared the predicted binding mode Z1 with its crystal structure. As shown in **Figure 2D**, the RMSD value between Z1 and its PDB structure is 1.08 Å, meaning the computed binding mode is highly similar to its crystal structure. This suggests our method of multiple conformation docking is useful in finding the correct binding mode. For this reason, we will use the binding mode E1 for CRO as the starting structure for further investigation.

Ideally, the screening of the correct binding modes was achieved through calculating of some physical quantities, such as binding energy, by averaging over an infinite conformational space of both ligand and binding pocket. According to the ergodic hypothesis, this is equivalent to performing time average from zero to infinity (Cramer, 2004). In molecular dynamics simulation, a finite period of time (typically in nanosecond scale)

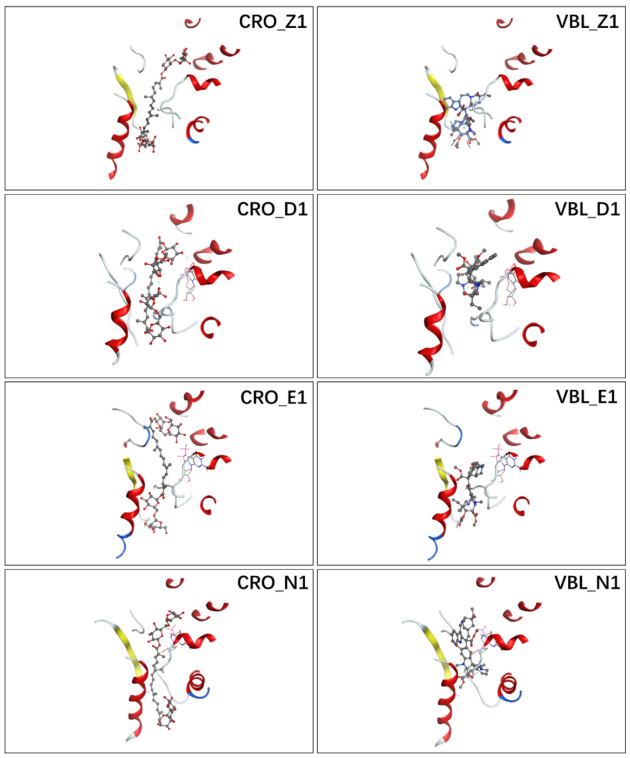


FIGURE 3 | Ligand conformations with the highest docking scores of CRO **(left)** and VBL **(right)** in the binding pocket.

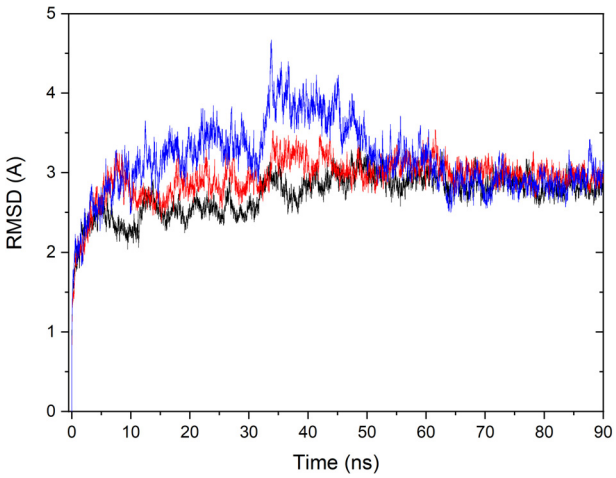


FIGURE 4 | Monitoring of RMSD change over the MD simulation course for the tubulin bound with ligand CRO (black), VBL (blue), and SBD (red). The RMSD value of C_α of each MD trajectory was calculated and plotted against simulation time.

was engaged to focus on the most representative microstates of an ensemble. Therefore, it is necessary to enumerate representative microstates of the CRO-tubulin complex.

Molecular Dynamics Simulations

Based on the constructed structure of CRO-tubulin complex identified in the multiple conformation docking step, MD simulations were performed to further achieve rationalized and stable complex. The stability of tubulin and CRO in the binding site were assessed by the root-mean-square derivation (RMSD) values of C_α atoms with respect to the initial conformation during the MD simulation period, as shown in **Figure 4**. Since SBD is the ligand molecule in crystal structure of 3E22, MD trajectories of SBD and VBL were obtained and RMSD values of C_α atoms were plotted accordingly for comparison (**Figure 4**). Significant fluctuations in RMSD plots were observed in the first 60 ns for all three ligands CRO, SBD, and VBL, indicating protein domain movements upon ligand binding. Then the three RMSD curves achieved stable plateaus for the last 30 ns. In the stable stage, the RMSD values kept at around 2.8 Å with respect to the initial protein conformation. However, in the first 60 ns the RMSD fluctuations of CRO is significantly smaller than VBL and SBD. This means a slighter conformational change of tubulin upon CRO binding as compared to VBL and SBD. It is likely the ligand CRO is better accommodated in the protein than VBL and SBD.

To further investigate the flexible protein segments attributing the RMSD fluctuations, the root-mean-square fluctuation (RMSF) values of tubulin upon binding of each ligand were calculated based on the all-atom MD trajectories (**Figure 5**). It could be discovered that the average fluctuations of CRO binding is smaller SBD and VBL. The RMSF curves of SBD and VBL are highly similar, but are significantly distinct from CRO. This suggests a different binding mode of CRO from the traditional inhibitors SBD and VBL. Furthermore, a lower average RMSF value throughout tubulin indicate the CRO binding mode is more favorable than SBD and VBL.

Hydrogen Bond Analysis

To primarily investigate the binding affinity between the ligands and tubulin, we performed hydrogen bond analysis along the 90 ns MD trajectories of each ligands. The results were presented in **Figure 6**. The frequencies of hydrogen bonding between tubulin and the ligand CRO, SBD, and VBL were plotted versus snapshots extracted from MD trajectories. As demonstrated in **Figure 6**, the average frequency of hydrogen bonding of CRO was around 6, which is larger than the average frequency of SBD and VBL. Although the strength of each hydrogen bond was not considered yet, but it is highly likely that the formation of excessive amounts of hydrogen bond between CRO and tubulin will lead to a much more stable binding mode than SBD and VBL. In the next part, the binding energy of each ligand will be further analyzed by MM/GBSA methodology.

MM/GBSA Binding Energy Calculation

To estimate the binding energy of ligands and tubulin, MM/GBSA method was performed to calculate energy contributions (Wang et al., 2017, 2019). The three methodologies of MM, GB, and SA were utilized to compute energy

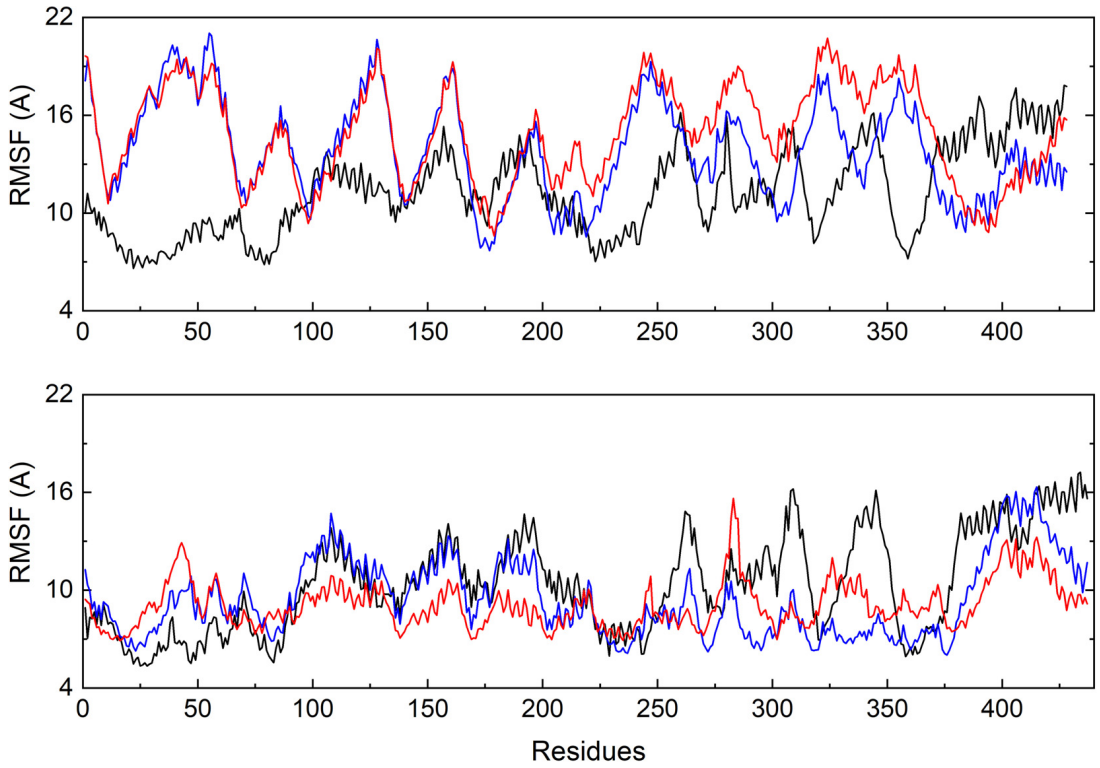


FIGURE 5 | Comparison of the backbone RMSF values of tubulin bound with different ligands CRO (black), SBD (blue) and VBL (red). The β **(top)** and α **(below)** chain were plotted separately.

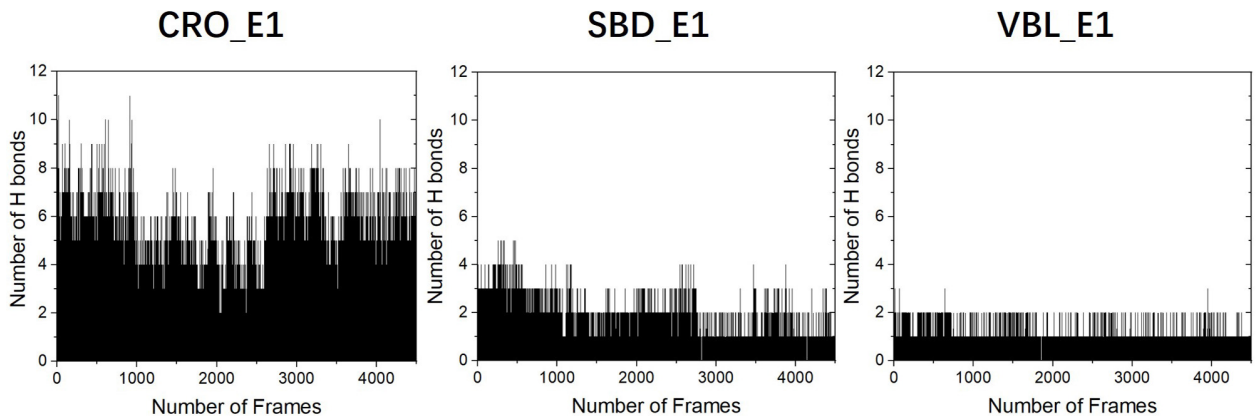


FIGURE 6 | The analysis of hydrogen bonds between tubulin and different ligands CRO **(left)**, SBD **(middle)**, and VBL **(right)**. The density of frame was 50 frames/ns.

contributions from van der Waals (vdw), electrostatic (ele), polar and non-polar surface solvation interactions (Wang et al., 2017, 2019). According to the all-atom MD trajectories shown in **Figure 4**, the last 20 ns frames were all considered to perform MM/GBSA for all three ligands. A total number of 2,000 frames were extracted for the computation to obtain reliable binding free energies. It should be pointed out that a complete estimation of binding free energy includes the calculation of entropy contribution. However, since we are interested in elucidating the dominate factors in different binding modes rather than computing the exact value of free energy, therefore the computationally expensive entropy calculations were neglected in this part.

The computed results of MM/GBSA and corresponding energy components terms for the three ligands were listed in **Table 1**. The methodology of MM/GBSA allows detailed decomposition of the free energy into different interaction contributions, which is convenient for the analysis of each term

TABLE 1 | MM/GBSA binding energy between the protein and different ligand CRO, SBD and VBL.

Mode	Contribution (kcal/mol)				ΔG_{total} (kcal/mol)
	vdw	ele	Polar	Non-polar	
CRO_E1	−89.64	−115.30	138.94	−13.25	−79.25 ± 7.24
SBD_E1	−56.84	−381.00	404.62	−7.72	−40.94 ± 3.71
VBL_E1	−66.07	−39.94	71.87	−8.32	−42.49 ± 2.95

separately. As shown in **Table 1**, the polar solvation energies are the only unfavorable terms for all three ligands. And the remaining terms of the van der Waals, the electrostatic and the non-polar solvation interactions have attributed a total energy of −79.25 ± 7.24, −40.94 ± 3.71, and −42.49 ± 2.95 kcal/mol for the ligand CRO, SBD, and VBL, respectively. This means the binding of three ligands are thermodynamically favorable, which is accordance with experimental observations that all three ligands are good inhibitors for tubulin. On the other hand, the binding free energy of CRO is almost twice of traditional inhibitors SBD and VBL, suggesting its potential high inhibition efficiency toward tubulin.

Key Residues Analysis

The energy contribution of each residue-ligand pair was decomposed to obtain a detailed energy analysis on the interaction between tubulin and the ligands. By using this quantitative analysis, it is helpful to investigate and identify key residues as hot spots involving in the binding interaction. The decomposed binding free energies of each residue-ligand pair

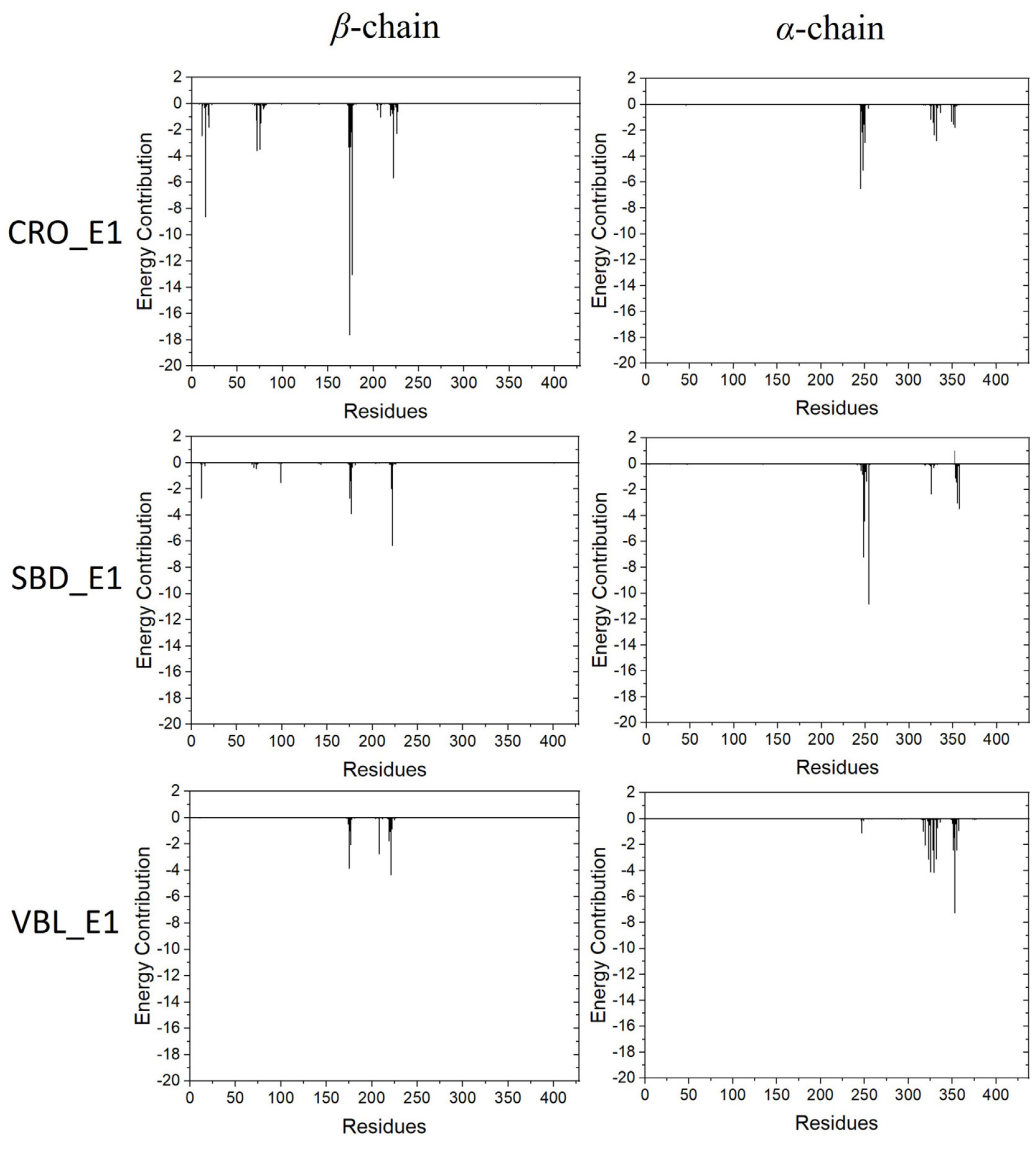


FIGURE 7 | Energy contributions of each residue-ligand pair for the ligand CRO **(top)**, SBD **(middle)** and VBL **(bottom)**.

were plotted versus the position number of each amino acid in **Figure 7**. The peaks in the figure showed the energy contributions of each residue.

Residues with an absolute energy contribution larger than 2 kcal/mol were identified as hot residues. For the CRO binding, 11 (Gln^{11}, Gln^{15}, Thr^{72}, Ser^{75}, Pro^{173}-Lys^{174}-Val^{175}-Ser^{176}-Asp^{177}, Tyr^{222}, and Asn^{226}) and 6 (Asp^{245}, Ala^{247}-Leu^{248}, Val^{250}, Asn^{329}, and Ile^{332}) hot residues were identified in the β and α chain, respectively (**Figure 8**). In comparison, 4 (in β chain) and 6 (in α chain) hot residues were identified for SBD, and 4 (in β chain) and 9 (in α chain) were identified for VBL (**Figure 8**). Clearly, the total energy contributions of hot residues in the binding of CRO is larger than SBD and VBL. This is in line with energy analysis by MM/GBSA method, indicating a strong interaction between CRO and tubulin. In addition, the hot residues in the beta chain involving in the binding of CRO were significantly different from SBD and VBL, suggesting a distinct binding mode of CRO. An interesting binding motif of Val^{175}-Xxx^{176}-Asp^{177} in the beta chain was discovered to the common element involving the binding of different ligands CRO, SBD and VBL. This peptide motif may serve as a critical site for further development of tubulin inhibitor.

In order to compare the CRO-tubulin complex structure before and after molecular dynamics simulation, a snapshot at 80 ns in the stable plateau of MD trajectory of CRO_E1 was extracted as a representative structure of the post-MD structure. The 2D interaction map and pocket residues were shown in **Supplementary Figure S4**. It should be noted that the 2D interaction map in **Supplementary Figure S4** is different from the hot residue map in **Figure 8**. The hot residue map considers the average energy contribution (>2 kcal/mol) throughout the MD simulation period, while the 2D interaction map identifies important pocket residues from a static structure. The comparison of 2D interaction map between pre-MD (**Supplementary Figure S2**) and post-MD (**Supplementary Figure S4**) showed the common residues were reserved, including Gln^{15}, Val^{175}-Ser^{176}, Pro^{220}-Thr^{221}-Tyr^{222} in the β-chain, and Leu^{248}, Pro^{325}, Val^{328}-Asn^{329}, Ile^{332}, Phe^{351}, Val^{353} in the α-chain, which indicates the interacting residues in the binding pocket were conserved features for CRO.

Community Network Between CRO and Tubulin

Community network analysis of protein, also named as residue interaction network (RIN) analysis, is a valuable method in deciphering the topology and dynamics of protein structure (Shcherbinin et al., 2019). By modeling a protein structure as residue nodes and interaction edges, the RIN approach allows to uncover key characteristics of the protein as well as rationalize drug design by topologically measuring the binding interactions (Hu et al., 2014; Liang et al., 2018, 2019, 2020). To better understand the interaction between CRO and tubulin, the RIN of tubulin and its bound state with CRO were constructed accordingly (**Figure 9**). As shown in **Figure 9**, residues involving in more than one interaction with the remaining residues or ligand (blue node) were represented as nodes. Residues in the β and α chain of tubulin were colored in pink and green, respectively. In the RIN, the non-covalent interactions including hydrogen bonding (purple), ionic interaction (blue), van der Waals interaction (yellow), and π–π stacking (orange) were represented as undirected edges between nodes. For simplicity, only one interaction per interaction type was plotted in the network. The discussion here was based on the interaction simplified network, but it should be pointed out that the

FIGURE 8 | Key residues distribution in β and α chain for ligand CRO **(top)**, SBD **(middle)** and VBL **(bottom)** by MM/GBSA energy decomposition method.

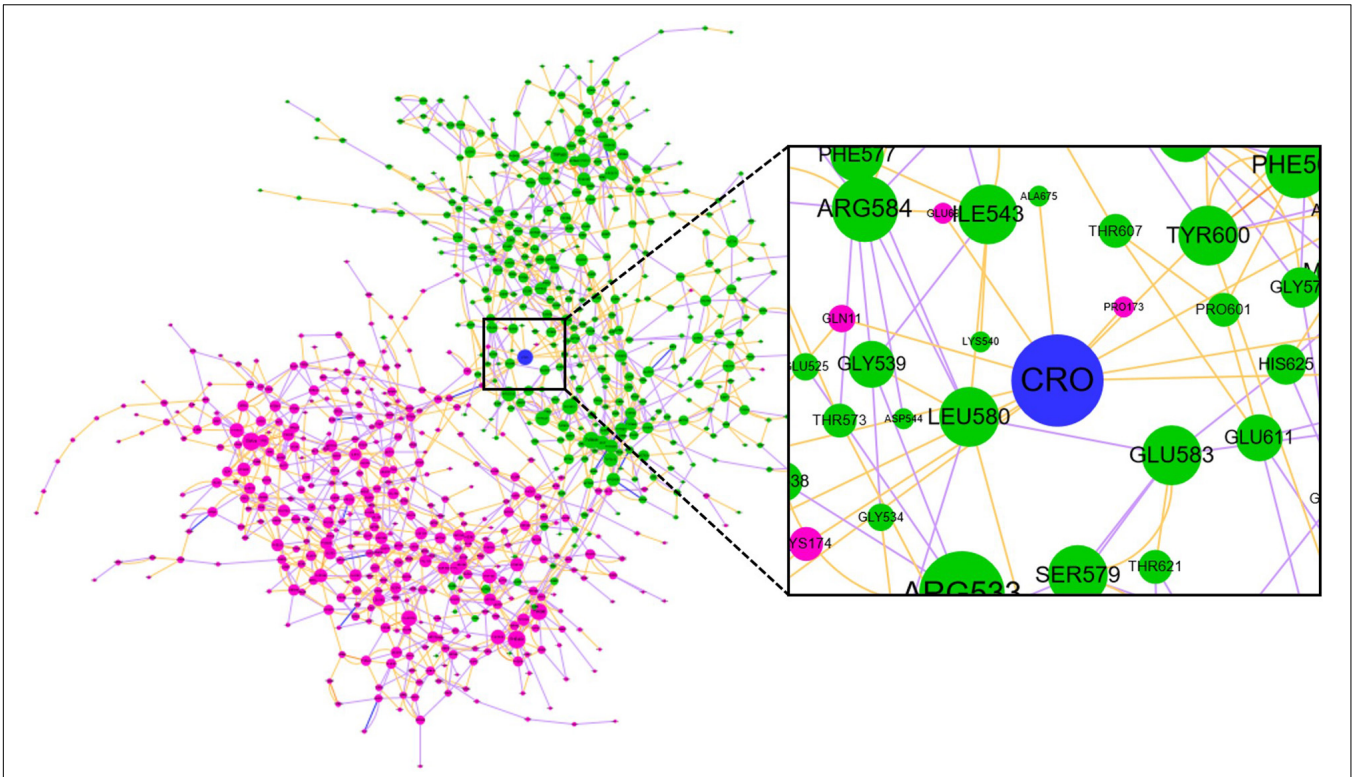


FIGURE 9 | The residue interaction network of tubulin-CRO complex. Each node represents a residue in β chain (pink), α chain (green) or ligand (blue). The size of each node was linearly correlated with the degree of the node. The label of each node was numbered sequentially by taking the two chains as one sequence, which means the number 1–428 and 429–865 represents β and α chain, respectively. Each edge represents the interaction between nodes, including hydrogen bonding (purple), ionic interaction (blue), van der Waals interaction (yellow), and π–π stacking (orange).

TABLE 2 | Comparison of the degree, betweenness and closeness centrality of the key residues of tubulin and tubulin-CRO complex.

Chains	Residues	Tubulin-CRO complex			Tubulin		
		Betweenness	Closeness	Degree	Betweenness	Closeness	Degree
B	GLN11	0.0000	0.1307	2	0.0000	0.0775	1
B	THR72	0.0449	0.1331	5	0.0053	0.0839	4
B	PRO173	0.0000	0.1290	1	NA	NA	0
B	LYS174	0.0080	0.1293	3	0.0000	0.0841	2
B	VAL175	0.0945	0.1413	4	0.0013	0.1014	3
B	TYR222	0.1619	0.1417	5	0.0006	0.0968	4
C	ALA675	0.0000	0.1290	1	NA	NA	0
C	LEU676	0.0109	0.1302	3	0.0009	0.0881	2
C	VAL678	0.1983	0.1425	3	0.0039	0.1011	2
C	ASN757	0.0381	0.1339	4	0.0030	0.1011	3

1) NA means the node was neglected by Cytoscape since its degree was deduced to 0; 2) the label of each node was numbered sequentially by taking the two chains as one sequence, which means the number 1–428 and 429–865 represents β and α chain, respectively.

discussion could be extended to an advanced network with multiple edges between nodes.

In network graph theory, the degree, betweenness and closeness centrality are characteristic values for measuring the importance of a node in a network (Shcherbinin et al., 2019). To investigate the ligand-binding induced change of the key residues as discovered in molecular dynamics simulation, we have computed the degree, betweenness and closeness centrality of the ligand and key residues. A comparison of the degree, betweenness and closeness centrality of the key residues between tubulin and tubulin bound with CRO were listed in **Table 2**. It could be found that the degree, betweenness and closeness centrality of each node was increased after the binding of the ligand CRO. Actually, the betweenness and closeness centrality of CRO (0.3199 and 0.1481, respectively) were ranked the highest in the network, indicating its vital importance in the interaction with tubulin. Therefore,

the key residues were deeply connected with other parts of the network through the interaction with CRO. In other words, their importance or centrality in the network was increased after binding with the ligand, supporting the conclusion from the previous MD analysis.

CONCLUSION

Currently, the crystal structure of tubulin bound with CRO is still lacking, which hinders our understanding of the interaction between CRO and tubulin. In this paper, we have screened the most likely binding mode CRO_E1 of CRO in the vinca binding pocket of tubulin based on multiple conformation docking strategy. Furthermore, molecular dynamics simulation method was involved to investigate the mechanism of interaction of CRO_E1. The results showed the excessive number of hydrogen bonds of CRO_E1 plays an important role in the CRO-tubulin binding. Energic analysis showed the binding free energy of CRO is almost as twice as the inhibitor soblidotin and VBL, suggesting a favored binding of CRO in the vinca binding pocket of tubulin. Hot residues were analyzed by energy decomposition, and were shown to be in accordance with their topological characteristics in the interaction network. Although hot residues involving the binding were different, a common residue motif Val^{175}-Xxx^{176}-Asp^{177} was identified for the three ligands, suggesting its importance in future drug design. The results in this paper provide new insights into structural basis of the interaction between CRO and tubulin, which is valuable for future drug design and development targeting tubulin.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

ZW conceptualized the methodology. ZW, JR, and NJ performed the MD simulations. ZW and XLn performed residue interaction network analysis. ZW and XLa analyzed the data. ZW and JR wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Innovation Talent Team of Guizhou Science and Technology Department (qiankehe platform talents [2020] 5007), the Scientific Research Foundation for the Returned Overseas Chinese Scholars of Guizhou Province (2018) 0013, Guizhou Provincial Natural Science Foundation (QKHJ [2020] 1Y045), and Zunyi Science and Technology Project (2018) 21.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.586970/full#supplementary-material

REFERENCES

Abdullaev, F. I., Riverón-Negrete, L., Caballero-Ortega, H., Manuel Hernández, J., Pérez-López, I., Pereda-Miranda, R., et al. (2003). Use of in vitro assays to assess the potential antigenotoxic and cytotoxic effects of saffron (crocus sativus l.). *Toxicol. In Vitro* 17, 731–736. doi: 10.1016/s0887-2333(03)00098-5

Alavizadeh, S. H., and Hosseinzadeh, H. (2014). Bioactivity assessment and toxicity of crocin: a comprehensive review. *Food Chem. Toxicol.* 64, 65–80. doi: 10.1016/j.fct.2013.11.016

Amin, A., Bajbouj, K., Koch, A., Gandesiri, M., and Schneider-Stock, R. (2015). Defective autophagosome formation in p53-null colorectal cancer reinforces crocin-induced apoptosis. *Int. J. Mol. Sci.* 16, 1544–1561. doi: 10.3390/ijms16011544

Assenov, Y., Ramírez, F., Schelhorn, S.-E., Lengauer, T., and Albrecht, M. (2007). Computing topological parameters of biological networks. *Bioinformatics* 24, 282–284. doi: 10.1093/bioinformatics/btm554

Bajbouj, K., Schulze-Luehrmann, J., Diermeier, S., Amin, A., and Schneider-Stock, R. (2012). The anticancer effect of saffron in two p53 isogenic colorectal cancer cell lines. *BMC Comp. Altern. Med.* 12:69. doi: 10.1186/1472-6882-12-69

Bathaie, S. Z., and Mousavi, S. Z. (2010). New applications and mechanisms of action of saffron and its important ingredients. *Crit. Rev. Food Sci. Nutr.* 50, 761–786. doi: 10.1080/10408390902773003

Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *J. Phys. Chem.* 97, 10269–10280. doi: 10.1021/j100142a004

Bolhassani, A., Khavari, A., and Bathaie, S. Z. (2014). Saffron and natural carotenoids: biochemical activities and anti-tumor effects. *Biochim. Biophys. ActaRev. Cancer* 1845, 20–30. doi: 10.1016/j.bbcan.2013.11.001

Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995). Application of the multimolecule and multiconformational resp methodology to biopolymers: charge derivation for DNA, rna, and proteins. *J. Comp. Chem.* 16, 1357–1377. doi: 10.1002/jcc.540161106

Cramer, C. J. (2004). *Essentials of Computational Chemistry: Theories and Models*. Hoboken, NJ: John Wiley & Sons Ltd.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: an n·log(n) method for ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397

Dumontet, C., and Jordan, M. A. (2010). Microtubule-binding agents: a dynamic field of cancer therapeutics. *Nat. Rev. Drug Discov.* 9, 790–803. doi: 10.1038/nrd3253

Fox, T., and Kollman, P. A. (1998). Application of the resp methodology in the parametrization of organic solvents. *J. Phys. Chem. B* 102, 8070–8079. doi: 10.1021/jp9717655

Hire, R. R., Srivastava, S., Davis, M. B., Kumar Konreddy, A., and Panda, D. (2017). Antiproliferative activity of crocin involves targeting of microtubules in breast cancer cells. *Sci. Rep.* 7:44984. doi: 10.1038/srep44984

Hoshyar, R., and Mollaei, H. (2017). A comprehensive review on anticancer mechanisms of the main carotenoid of saffron, crocin. *J. Pharm. Pharmacol.* 69, 1419–1427. doi: 10.1111/jphp.12776

Hosseinzadeh, H., Mehri, S., Heshmati, A. A., Ramezani, M., Sahebkar, A. H., and Abnous, K. (2013). Proteomic screening of molecular targets of crocin. *DARU J. Pharmaceut. Sci.* 22:5. doi: 10.1186/2008-2231-22-5

Hu, G., Yan, W., Zhou, J., and Shen, B. (2014). Residue interaction network analysis of dronpa and a DNA clamp. *J. Theoretical Biol.* 348, 55–64. doi: 10.1016/j.jtbi.2014.01.023

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Liang, Z., Hu, J., Yan, W., Jiang, H., Hu, G., and Luo, C. (2018). Deciphering the role of dimer interface in intrinsic dynamics and allosteric pathways underlying the functional transformation of dnmt3a. *Biochim. Biophys. Acta Gen. Subj.* 1862, 1667–1679. doi: 10.1016/j.bbagen.2018.04.015

Liang, Z., Verkhivker, G. M., and Hu, G. (2019). Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *Brief. Bioinformat.* 21, 815–835. doi: 10.1093/bib/bbz029

Liang, Z., Zhu, Y., Long, J., Ye, F., and Hu, G. (2020). Both intra and inter-domain interactions define the intrinsic dynamics and allosteric mechanism in dnmt1s. *Comput. Struct. Biotechnol. J.* 18, 749–764. doi: 10.1016/j.csbj.2020.03.016

Naïm, M., Bhat, S., Rankin, K. N., Dennis, S., Chowdhury, S. F., Siddiqi, I., et al. (2007). Solvated interaction energy (sie) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. *J. Chem. Inform. Model.* 47, 122–133. doi: 10.1021/ci600406v

Pfander, H., and Wittwer, F. (1975). Carotinoid-glycoside. 3. Mitteilung. Untersuchungen zur carotinoidzusammensetzung im safran. *Helvet. Chim. Acta* 58, 2233–2236. doi: 10.1002/hlca.19750580736

Piovesan, D., Minervini, G., and Tosatto Silvio, C. E. (2016). The ring 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* 44, W367–W374. doi: 10.1093/nar/gkw315

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5

Sawant, A. V., Srivastava, S., Prassanawar, S. S., Bhattacharyya, B., and Panda, D. (2019). Crocin, a carotenoid, suppresses spindle microtubule dynamics and activates the mitotic checkpoint by binding to tubulin. *Biochem. Pharmacol.* 163, 32–45. doi: 10.1016/j.bcp.2019.01.023

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shcherbinin, D., Veselovsky, A., and Mohan, C. G. (eds) (2019). "Analysis of protein structures using residue interaction networks," in *Structural Bioinformatics: Applications in Preclinical Drug Discovery Process*, (Cham: Springer International Publishing).

Sun, Y., Xu, H.-J., Zhao, Y.-X., Wang, L.-Z., Sun, L.-R., Wang, Z., et al. (2013). Crocin exhibits antitumor effects on human leukemia hl-60 cells in vitro and in vivo. *Evidence Based Comp. Altern. Med.* 2013:690164. doi: 10.1155/2013/690164

Tsimidou, M., and Tsatsaroni, E. (1993). Stability of saffron pigments in aqueous extracts. *J. Food Sci.* 58, 1073–1075. doi: 10.1111/j.1365-2621.1993.tb06116.x

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019). End-point binding free energy calculation with mm/pbsa and mm/gbsa: strategies and applications in drug design. *Chem. Rev.* 119, 9478–9508. doi: 10.1021/acs.chemrev.9b00055

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035

Wang, P., Fu, T., Zhang, X., Yang, F., Zheng, G., Xue, W., et al. (2017). Differentiating physicochemical properties between ndris and snris clinically important for the treatment of adhd. *Biochim. Biophys. Acta Gen. Subj.* 1861, 2766–2777. doi: 10.1016/j.bbagen.2017.07.022

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

BRIEF RESEARCH REPORT
published: 19 November 2020
doi: 10.3389/fmolb.2020.582702

ANCA: A Web Server for Amino Acid Networks Construction and Analysis

Wenying Yan[1†], Chunjiang Yu[2†], Jiajia Chen[3], Jianhong Zhou[4] and Bairong Shen[5*]*

[1] Center for Systems Biology, School of Biology and Basic Medical Sciences, Soochow University, Suzhou, China, [2] School of Biotechnology, Suzhou Industrial Park Institute of Services Outsourcing, Suzhou, China, [3] School of Chemistry, Biology and Materials Engineering, Suzhou University of Science and Technology, Suzhou, China, [4] Public Library of Science, San Francisco, CA, United States, [5] Institutes for Systems Genetics, West China Hospital, Sichuan University, Chengdu, China

Amino acid network (AAN) models empower us to gain insights into protein structures and functions by describing a protein 3D structure as a graph, where nodes represent residues and edges as amino acid interactions. Here, we present the ANCA, an interactive Web server for Amino Acids Network Construction and Analysis based on a single structure or a set of structures from the Protein Data Bank. The main purpose of ANCA is to provide a portal for three types of an environment-dependent residue contact energy (ERCE)-based network model, including amino acid contact energy network (AACEN), node-weighted amino acid contact energy network (NACEN), and edge-weighted amino acid contact energy network (EACEN). For comparison, the C-alpha distance-based network model is also included, which can be extended to protein–DNA/RNA complexes. Then, the analyses of different types of AANs were performed and compared from node, edge, and network levels. The network and corresponding structure can be visualized directly in the browser. The ANCA enables researchers to investigate diverse concerns in the framework of AAN, such as the interpretation of allosteric regulation and functional residues. The ANCA portal, together with an extensive help, is available at http://sysbio.suda.edu.cn/anca/.

Keywords: Amino acids network, ANCA portal, network analysis, protein structure, allosteric regulation, functional residues

OPEN ACCESS

Edited by:
Hongchun Li,
Shenzhen Institutes of Advanced Technology (CAS), China

Reviewed by:
Jian Huang,
University of Electronic Science and Technology of China, China
Dmitry Suplatov,
Lomonosov Moscow State University, Russia

***Correspondence:**
Wenying Yan
wyyan@suda.edu.cn
Bairong Shen
bairong.shen@scu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *13 July 2020*
Accepted: *19 October 2020*
Published: *19 November 2020*

Citation:
Yan W, Yu C, Chen J, Zhou J and Shen B (2020) ANCA: A Web Server for Amino Acid Networks Construction and Analysis. Front. Mol. Biosci. 7:582702. doi: 10.3389/fmolb.2020.582702

INTRODUCTION

With the increasing number of high-resolution 3D structures of biomolecules, including proteins, protein–DNA complexes, and protein–RNA complexes, the development of rapid and efficient methods to perform large-scale analysis for them is needed. A variety of structure-based computational tools and methods is developed to satisfy the new challenges (Romero-Rivera et al., 2016; Liu et al., 2019; Sequeiros-Borja et al., 2020), such as consensus-based, machine learning-based, molecular dynamics (MD) simulation-based, quantum-mechanic simulation-based methods, and so on. The network concepts and methods have been widely used in numerous problems in different fields of biological science including the study of protein structures and functions (Hu et al., 2017). Amino acid network (AAN) models, which are undirected networks

consisting of amino acid residues and their interactions, have opened numerous opportunities to reveal new insights in understanding the function of biomolecules from large-scale 3D structure data. Compared with traditional structure-based methods, studying a biomolecule from a network perspective not only gives a systems-level understanding of the biomolecule structure through topological information and global connectivity (Yan et al., 2014; Zhou et al., 2014) but also provides an efficient way for characterization of each individual amino acid within the complex interaction network, such as protein–protein interfaces (Di Paola et al., 2015), catalytic residues (Zhou et al., 2016), and allosteric regulation (Di Paola and Giuliani, 2015; Yan et al., 2018).

Nowadays, several Web tools that construct different types of AANs have facilitated progress in this area of research. RING2.0 constructs an AAN based on the physicochemical interactions between the residues, which include covalent and non-covalent interactions (Piovesan et al., 2016). Protein contact atlas focuses on the non-covalent interactions within structures and shows them at different scales ranging from atomic level to the entire macromolecule level (Kayikci et al., 2018). webPSN investigates structural communication in macromolecules by constructing static and dynamic AANs (Felline et al., 2020). Furthermore, AAN-based Web servers such as MDN (Ribeiro and Ortiz, 2015), NAPS (Chakrabarty and Parekh, 2016), and RIP-MD (Contreras-Riquelme et al., 2018) provide tools for quantifying protein dynamic based on MD simulation trajectories. More tools and Web servers for network can be found in a recent review (Liang et al., 2020). However, many of the AAN models only considered amino acid interactions on a geometric level but not on the chemical properties of the proteins. An alternative strategy for the simulation of the interactions is using the energy between residues. We proposed an amino acid contact energy network (AACEN) based on a coarse-grained contact energy called environment-dependent residue contact energy (ERCE; Yan et al., 2014), which takes into account the type of secondary structure for each residue and is more efficient and easier for characterizing the energy between residues (Zhang and Kim, 2000; Shen and Vihinen, 2003). Moreover, another inadequacy of most AAN models is the disregard for heterogeneity of residues and treating all nodes as the same in the network. To address this, we improved our AACEN model by assigning residue properties as node weights and named it as node-weighted amino acid contact energy network (NACEN; Yan et al., 2018).

In this paper, we developed a Web server called ANCA (Amino Acids Network Construction and Analysis) for construction and analysis of our previously proposed ERCE-based network models AACEN and NACEN. To refine our ERCE-based models, we also added the edge-weighted amino acid contact energy network (EACEN) model in our Web server using the ERCE as link weights. Moreover, a C-alpha distance-based network (C-alpha) model was also included in our ANCA for two purposes. Firstly, the C-alpha model can be used as the comparison network for our ERCE-based models for proteins or protein complexes. Secondly, ANCA also provides the construction and analysis for single and multiple protein–DNA/RNA complexes based on the C-alpha model. The organization of our portal ANCA was shown in **Figure 1**.

METHODS AND IMPLEMENTATION

The ANCA Web server is comprised of two core modules entitled "single structure" and "multiple structures." The single-structure module provides the construction and analysis for one structure with any one of the AAN models at a time, while in the multiple-structures module, the structures can be analyzed in batches using any of the four types of AAN models. The former module is more suitable for carrying out a detailed analysis for one structure (either a PDB code or a PDB file). The latter module can be used for comparison analysis of different structures. Both modules support four types of AAN construction, analysis, and visualization. Moreover, ANCA can provide the option of distance-based AAN construction for protein–DNA/RNA complexes.

Amino Acid Network

Amino Acid Contact Energy Network

As defined in our previous studies (Yan et al., 2014; Zhou et al., 2014), an amino acid residue in the protein or protein complex is denoted as a node and a link is set to two nodes if the ERCE (Zhang and Kim, 2000; Shen and Vihinen, 2003) between them is less than 0. ERCE is an improvement of Miyazawa–Jernigan's model by an extension of residue alphabet from 20 to 60, which considers the 20 amino acids in three secondary structural states. ERCE e_{ij} between residues i and j was defined as in our previous studies (Yan et al., 2014), and then according to the e_{ij}, the element in the adjacent matrix AM of AACEN was set to 1 if e_{ij} was less than 0, otherwise the element was set to 0 (Yan et al., 2018).

Node-Weighted Amino Acid Contact Energy Network

Based on AACEN, we have developed a NACEN module to characterize and predict functional residues. In this network representation, links between residues were defined the same as in AACEN, and the properties of residues, including relative solvent accessibility (SAS), mass, hydrophobicity, polarity, or user-self defined node weights (Yan et al., 2018).

Edge-Weighted Amino Acid Contact Energy Network

In EACEN, the links between residues were weighted by ERCE and the adjacent matrix AM of EACEN was defined as:

$$\mathrm{AM}_{ij} = \begin{cases} 0, & e_{ij} \geq 0 \\ w_{ij}, & e_{ij} < 0 \end{cases} \tag{1}$$

Where w_{ij} is the normalization of the contact energy e_{ij} between i and j:

$$\mathrm{w}_{ij} = \begin{cases} 0.0001, & if \ \left|e_{ij}\right| = \left|e_{ij}\right|_{min} \\ \frac{\left|e_{ij}\right| - \left|e_{ij}\right|_{min}}{\left|e_{ij}\right|_{max} - \left|e_{ij}\right|_{min}}, & if \ \left|e_{ij}\right| \neq \left|e_{ij}\right|_{min} \end{cases} \tag{2}$$

C-Alpha Distance-Based Network

ANCA can construct the network for protein–protein, protein–DNA, or protein–RNA complex based on the distance between represented atoms. For a protein–protein complex, the link between two residue nodes in the network was established if the distance between C-alphas of the

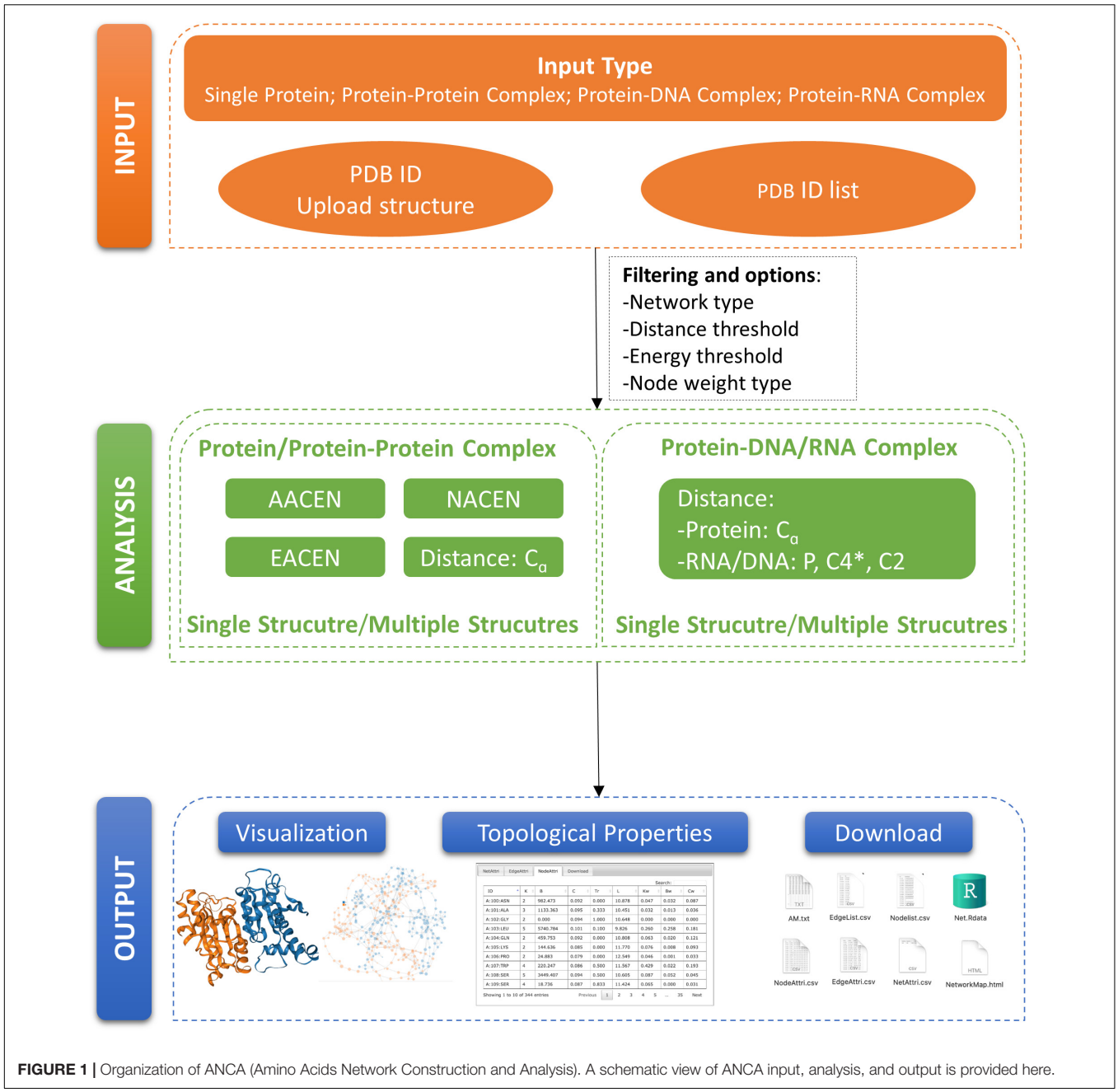


FIGURE 1 | Organization of ANCA (Amino Acids Network Construction and Analysis). A schematic view of ANCA input, analysis, and output is provided here.

residues is lower than a threshold (Di Paola et al., 2013). For a protein–DNA complex or protein–RNA complex, we use one node to represent one amino acid, and three nodes of P, C4*(sugar group), and C2 (base group) atoms to represent each nucleotide of the DNA or RNA (Delarue and Sanejouand, 2002).

As mentioned above, our portal provides the above four AAN models, including two unweighted AANs (C-alpha and AACEN) and two weighted AANs (NACEN and EACEN). The C-alpha model can be used not only for protein and protein–protein complex but also for protein–DNA/RNA complex. But since the network is constructed just based on the distance between C-alpha atoms, it is a relatively coarse model. AACEN is an ERCE-based network model that can be used just for protein or protein complex but provide more detailed information by considering the local environment of the residues (Zhang and Kim, 2000), and it has been used to compare protein structures and evolution (Yan et al., 2014, 2016). NACEN and EACEN are also ERCE based. The difference is the former one also employs the characters of residues as node weights, so it is more suitable to explore residue function (Yan et al., 2018), while the latter one assigns the ERCE between residues as the link weights that provide more detailed information on the links between residues than the unweighted model, so it can be

helpful for studying the communication between residues, such as allosteric regulation.

Analysis and Visualization of the Amino Acid Network

ANCA can be used for the network analysis of proteins, protein–protein complexes, and protein–DNA/RNA complexes from the node level, edge level, and network level. The detailed definition of the parameters was listed in http://sysbio.suda.edu.cn/anca/.

At node level, the topological parameters of nodes are calculated, including degree, betweenness, closeness, transitivity, and average shortest path length (L_{net}). Moreover, for NACEN, we also calculated the weighted degree, betweenness, and closeness centralities based on the node weights. Their definitions were in our previous work (Yan et al., 2018). At the edge level, edge betweenness centrality is calculated for each edge to evaluate the importance of the edge. Moreover, the long-range link, which is related to protein secondary structure density and residue evolution rate (Yan et al., 2014), is also labeled. At the network level, the node number (n), edge number, L_{net}, density, and diameter of the network are calculated.

The ANCA provides two types of visualization for protein molecule 3D structure and AAN. ANCA uses NGL Viewer (Rose and Hildebrand, 2015) to display the protein molecule 3D structure using the NGL JavaScript library. The visualization for the AAN is implemented using R package networkD3 (Allaire et al., 2017). In the AAN view, the lines represent the edges and the dots represent the amino acid. When the mouse pointer hovers over the dot, the amino acid name will be shown beside the dot. Users can use the mouse to manipulate the graph, such as scroll mouse wheel to zoom in or out of the graph, move the mouse by pressing the left button to rotate or drag the graph, and so on. The color of the protein molecule 3D structure and the AAN can consist with the chain name.

Implementation

The ANCA portal can be accessed by modern popular Web browsers, including Chrome, Internet Explorer, Safari, and Firefox, without installing any specialized software or browser plug-ins. The Apache[1] was used as the Web server, which is a secure, efficient, and extensible open-source HTTP server. The application was realized using three-tiered architecture. In the view tier, the front-end program was developed using PHP[2], the user interface interaction was realized using jQuery[3], and the advanced interaction control DataTable[4] was adopted to represent the result data. In the controller tier, we used C# and.NET Framework 4.0[5] to implement the logic process program, and the R program was used to construct and analyze the AANs. In the model tier, MySQL[6] was used to store execution-related information.

[1]http://www.apache.org/

[2]https://www.php.net/

[3]https://jquery.com

[4]https://www.datatables.net/

[5]https://docs.microsoft.com/en-us/dotnet/csharp/

[6]https://www.mysql.com/

WORKFLOW OF ANCA

Step 1: Module Selection

The step-by-step workflow of ANCA is shown in **Figure 2**. The first step is module selection for single structure or multiple structures (step 1 in **Figure 2**). Both modules support the four types of AAN construction and analysis including AACEN, NACEN, EACEN, and $C_{upalpha}$ distance-based AAN. In the single-structure module, one of the AAN is constructed. The results page shows the topological properties of the AAN and the visualization of structure and network. While in multiple-structures module, any type from the four AAN models can be constructed and analyzed for each structure, and the results page will demonstrate summary information for all the AANs.

Step 2: Input Data Upload and Parameter Setting

This step contains three procedures (**Figure 2**). First, PDB ID or file in PDB format of the structure should be filled in or uploaded (A in **Figure 2**). The input file should have a.pdb extension. The second procedure (B in **Figure 2**) is the selection of AAN type, i.e., AACEN, NACEN, EACEN, and C-alpha distance-based AAN. Then, the parameters of the corresponding AAN type should be specified. For AACEN and NACEN, the threshold of energy and distance between residues should be set. Besides these two parameters, users should also specify the node weights of residues either by selecting the default properties of residues, including SAS, mass, hydrophobicity, and polarity or by uploading the file (.txt) that contains user-self defined property. Lastly, the e-mail address can be optionally provided (C in **Figure 2**) that will be used to receive the results page link from ANCA portal. More detailed description is available at http://sysbio.suda.edu.cn/anca/

Step 3: Output Description

The output of the ANCA is composed of three parts: visualization, network information, and network topological properties. For visualization, the protein structure and corresponding network are shown in the results page of the Web server, and both of them are colored by the chain of structure. For network information, the results page provides files with adjacent matrix, edge list, and node list of the network. For network topological properties, the ANCA provides the parameters from the node level, edge level, and network level, which have been shown in the Analysis and Visualization of the Amino Acid Network section.

CASE STUDY

To evaluate the performance of ANCA, we carried out case studies for the single-structure module and multiple-structures module separately, as follows:

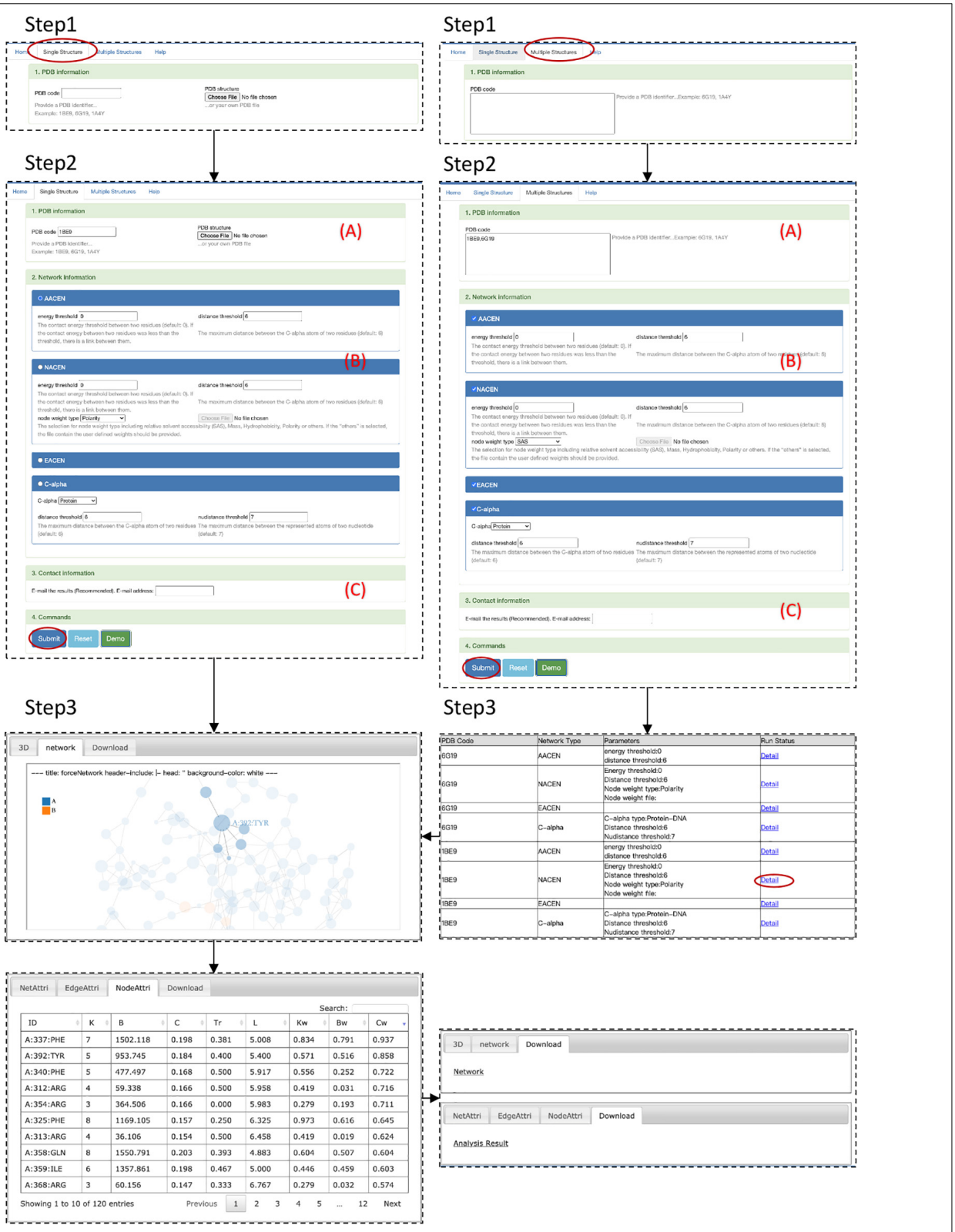
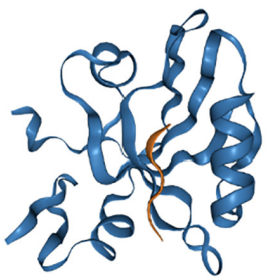


FIGURE 2 | Step-by-step workflow of ANCA.

ANCA
Amino acids Network Construction and Analysis

Parameters	
PDB Code:	1BE9
PDB Structure File:	not set
Network type:	NACEN
AACEN Energy threshold:	not set
AACEN Distance threshold:	not set
NACEN Energy threshold:	0
NACEN Distance threshold:	6
NACEN Node weight type:	Polarity
C-alpha:	not set
C-alpha distance threshold:	not set
C-alpha nudistance threshold:	not set
Email	wyyan@suda.edu.cn

3D | network | Download



NetAttri | EdgeAttri | NodeAttri | Download

Search:

ID	K	B	C	Tr	L	Kw	Bw	Cw
A:337:PHE	7	1502.118	0.198	0.381	5.008	0.834	0.791	0.937
A:392:TYR	5	953.745	0.184	0.400	5.400	0.571	0.516	0.858
A:340:PHE	5	477.497	0.168	0.500	5.917	0.556	0.252	0.722
A:312:ARG	4	59.338	0.166	0.500	5.958	0.419	0.031	0.716
A:354:ARG	3	364.506	0.166	0.000	5.983	0.279	0.193	0.711
A:325:PHE	8	1169.105	0.157	0.250	6.325	0.973	0.616	0.645
A:313:ARG	4	36.106	0.154	0.500	6.458	0.419	0.019	0.624
A:358:GLN	8	1550.791	0.203	0.393	4.883	0.604	0.507	0.604
A:359:ILE	6	1357.861	0.198	0.467	5.000	0.446	0.459	0.603
A:368:ARG	3	60.156	0.147	0.333	6.767	0.279	0.032	0.574

Showing 1 to 10 of 120 entries Previous 1 2 3 4 5 … 12 Next

FIGURE 3 | Results page for the case study.

Case 1 Single-Structure Module: Node-Weighted Amino Acid Contact Energy Network for Human PDZ Domain

Postsynaptic density-95/Discs large/Zonula occludens 1 (PDZ) protein domain family is a protein–protein interaction module, which is involved in dynamic regulation of signaling pathways and scaffolding and has emerged as a paradigmatic model system for intra-domain allostery (Reynolds et al., 2011; McLaughlinJr., Poelwijk et al., 2012). Here we tried to use our portal ANCA to investigate the allosteric residues of the third PDZ domain of PSD-95 (PDB 1BE9) by the NACEN model. As shown in **Figure 2**, we chose the single-structure module and NACEN network type with default threshold and selected polarity as node weights. Then, the network and protein structure can be visualized and the topological parameters of the network were listed. At last, the residues were ordered by the weighted closeness centrality (C_w), and the results showed that the top 3 residues were PHE337, TYR392, and PHE340 as shown in **Figure 3**. Among the top 3 residues, two of them, TYR392 and PHE340, have been validated as allosteric residues by double-mutant cycle analysis (Gerek and Ozkan, 2011).

Case 2 Multiple-Structures Module: All Network Types for Multiple Structures

At this part, we constructed and analyzed all the four types of network, including AACEN, NACEN, EACEN, and C-alpha network for two structures, PDZ3 in case 1 and protein–RNA complex MDA5 double-stranded RNA Filament (PDB 6G19) with default parameters at one time. Then, ANCA portal gave the results page with summary information for each network and the link of the network in the column "Run Status." The links point to the page containing detailed information for each network as described in case 1.

CONCLUSION

ANCA is a comprehensive portal for the construction and analysis of network representations of protein and protein–protein/DNA/RNA complexes to explore and understand the macromolecules at different levels of organization. It can help in the management of heterogeneous information sources, such as structural, sequence, physicochemical, and dynamical information of residues. Another advantage of our portal is that it also allows scientists to address diverse questions by choosing different network models. For example, NACEN is more suitable to identify the functional residues in the structures, while EACEN can capture the intramolecular information flow to help in understanding the allosteric regulation.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://sysbio.suda.edu.cn/anca/.

AUTHOR CONTRIBUTIONS

WY and BS conceived and designed the Web server. CY performed the server front-end and back-end. WY carried out the case studies. JZ and JC contributed to the Web server testing. WY, CY, and JC wrote the manuscript. JC polished the language and gave many constructive suggestions. BS critically reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The National Natural Science Foundation of China (Nos. 31670851 and 31770903). A project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions. The Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 18KJD520003). Training objects of young outstanding backbone teachers of Jiangsu Blue Project.

REFERENCES

Allaire, J. J., Gandrud, C., Russell, K., and Yetman, C. J. (2017). *"networkD3: D3 JavaScript Network Graphs From R". R Package Version 0.4.*

Chakrabarty, B., and Parekh, N. (2016). NAPS: network analysis of protein structures. *Nucleic Acids Res.* 44, W375–W382. doi: 10.1093/nar/gkw383

Contreras-Riquelme, S., Garate, J. A., Perez-Acle, T., and Martin, A. J. M. (2018). RIP-MD: a tool to study residue interaction networks in protein molecular dynamics. *PeerJ* 6:e5998. doi: 10.7717/peerj.5998

Delarue, M., and Sanejouand, Y. H. (2002). Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the Elastic Network Model. *J. Mol. Biol.* 320, 1011–1024. doi: 10.1016/S0022-2836(02)00562-4

Di Paola, L., De Ruvo, M., Paci, P., Santoni, D., and Giuliani, A. (2013). Protein contact networks: an emerging paradigm in chemistry. *Chem. Rev.* 113, 1598–1613. doi: 10.1021/cr3002356

Di Paola, L., and Giuliani, A. (2015). Protein contact network topology: a natural language for allostery. *Curr. Opin. Struct. Biol.* 31, 43–48. doi: 10.1016/j.sbi.2015.03.001

Di Paola, L., Platania, C. B., Oliva, G., Setola, R., Pascucci, F., and Giuliani, A. (2015). Characterization of protein-protein interfaces through a protein contact network approach. *Front. Bioeng. Biotechnol.* 3:170. doi: 10.3389/fbioe.2015.00170

Felline, A., Seeber, M., and Fanelli, F. (2020). webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules. *Nucleic Acids Res.* 48, W94–W103. doi: 10.1093/nar/gkaa397

Gerek, Z. N., and Ozkan, S. B. (2011). Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. *PLoS Comput. Biol.* 7:e1002154. doi: 10.1371/journal.pcbi.1002154

Hu, G., Di Paola, L., Pullara, F., Liang, Z., and Nookaew, I. (2017). Network proteomics: from protein structure to protein-protein interaction. *Biomed. Res. Int.* 2017:8929613. doi: 10.1155/2017/8929613

Kayikci, M., Venkatakrishnan, A. J., Scott-Brown, J., Ravarani, C. N. J., Flock, T., and Babu, M. M. (2018). Visualization and analysis of non-covalent contacts

using the Protein Contacts Atlas. *Nat. Struct. Mol. Biol.* 25, 185–194. doi: 10.1038/s41594-017-0019-z

Liang, Z., Verkhivker, G. M., and Hu, G. (2020). Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *Brief. Bioinform.* 21, 815–835. doi: 10.1093/bib/bbz029

Liu, Q., Xun, G., and Feng, Y. (2019). The state-of-the-art strategies of protein engineering for enzyme stabilization. *Biotechnol. Adv.* 37, 530–537. doi: 10.1016/j.biotechadv.2018.10.011

McLaughlin, R. N. Jr., Poelwijk, F. J., Raman, A., Gosal, W. S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. *Nature* 491, 138–142. doi: 10.1038/nature11500

Piovesan, D., Minervini, G., and Tosatto, S. C. E. (2016). The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* 44, W367–W374. doi: 10.1093/nar/gkw315

Reynolds, K. A., McLaughlin, R. N., and Ranganathan, R. (2011). Hot spots for allosteric regulation on protein surfaces. *Cell* 147, 1564–1575. doi: 10.1016/j.cell.2011.10.049

Ribeiro, A. A., and Ortiz, V. (2015). MDN: a web portal for network analysis of molecular dynamics simulations. *Biophys. J.* 109, 1110–1116. doi: 10.1016/j.bpj.2015.06.013

Romero-Rivera, A., Garcia-Borras, M., and Osuna, S. (2016). Computational tools for the evaluation of laboratory-engineered biocatalysts. *Chem. Commun.* 53, 284–297. doi: 10.1039/c6cc06055b

Rose, A. S., and Hildebrand, P. W. (2015). NGL viewer: a web application for molecular visualization. *Nucleic Acids Res.* 43, W576–W579. doi: 10.1093/nar/gkv402

Sequeiros-Borja, C. E., Surpeta, B., and Brezovsky, J. (2020). Recent advances in user-friendly computational tools to engineer protein function. *Brief. Bioinform.* doi: 10.1093/bib/bbaa150. [Epub ahead of print].

Shen, B., and Vihinen, M. (2003). RankViaContact: ranking and visualization of amino acid contacts. *Bioinformatics* 19, 2161–2162. doi: 10.1093/bioinformatics/btg293

Yan, W., Hu, G., Liang, Z., Zhou, J., Yang, Y., Chen, J., et al. (2018). Node-weighted amino acid network strategy for characterization and identification of protein functional residues. *J. Chem. Inf. Model.* 58, 2024–2032. doi: 10.1021/acs.jcim.8b00146

Yan, W., Hu, G., and Shen, B. (2016). Network analysis of protein structures: the comparison of three topologies. *Curr. Bioinform.* 11, 480–489. doi: 10.2174/1574893611666160602124707

Yan, W. Y., Sun, M. M., Hu, G., Zhou, J. H., Zhang, W. Y., Chen, J. J., et al. (2014). Amino acid contact energy networks impact protein structure and evolution. *J. Theor. Biol.* 355, 95–104. doi: 10.1016/j.jtbi.2014.03.032

Zhang, C., and Kim, S. H. (2000). Environment-dependent residue contact energies for proteins. *Proc. Natl. Acad. Sci. U.S.A.* 97, 2550–2555. doi: 10.1073/pnas.040573597

Zhou, J., Yan, W., Hu, G., and Shen, B. (2016). Amino acid network for prediction of catalytic residues in enzymes: a comparison survey. *Curr. Protein Pept. Sci.* 17, 41–51. doi: 10.2174/1389203716666150923105312

Zhou, J. H., Yan, W. Y., Hu, G., and Shen, B. R. (2014). SVR_CAF: an integrated score function for detecting native protein structures among decoys. *Protein Struct. Funct. Bioinform.* 82, 556–564. doi: 10.1002/prot.24421

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

REVIEW
published: 27 November 2020
doi: 10.3389/fmolb.2020.582966

Wrangling Shape-Shifting Morpheeins to Tackle Disease and Approach Drug Discovery

*Eileen K. Jaffe**

Fox Chase Cancer Center, Philadelphia, PA, United States

Homo-multimeric proteins that can come apart, change shape, and reassemble differently with functional consequences have been called morpheeins and/or transformers; these provide a largely unexplored context for understanding disease and developing allosteric therapeutics. This article describes such proteins within the context of protein structure dynamics, provides one detailed example related to an inborn error of metabolism and potential herbicide development, and describes the context for applying these ideas for understanding disease and designing bioactive molecules, such as therapeutics.

Keywords: protein structure dynamics, allostery, morpheein, fifth level of protein structure, drug discovery

OPEN ACCESS

Edited by:
Guang Hu,
Soochow University, China

Reviewed by:
Jian Zhang,
Shanghai Jiao Tong University, China
Claire Lesieur,
UMR 5005 Laboratoire Ampère (Ampère), France
Athi N. Naganathan,
Indian Institute of Technology Madras, India

***Correspondence:**
Eileen K. Jaffe
Eileen.Jaffe@fccc.edu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *13 July 2020*
Accepted: *26 October 2020*
Published: *27 November 2020*

Citation:
Jaffe EK (2020) Wrangling Shape-Shifting Morpheeins to Tackle Disease and Approach Drug Discovery. Front. Mol. Biosci. 7:582966. doi: 10.3389/fmolb.2020.582966

INTRODUCTION

A great number of medically relevant proteins are homo-multimers, some of which exist as an equilibrium of alternate assemblies that are both non-additive and functionally distinct. The phenomenon wherein protein homo-multimers can come apart, change shape while dissociated, and reassemble into an architecturally and functionally different assembly has been called the morpheein model of protein allostery (Jaffe, 2005; see Morpheein in Wikipedia[1]). A key to this protein structure dynamic is that the required conformational change is *spatially forbidden* within the context of either assembly. Proteins with this capacity can be called morpheeins and the alternate assemblies can be called morpheein forms. The dynamic process of dissociation and association makes this mode of allostery distinct from the classic Monod-Wyman-Changeux and Koshland-Nemethy-Filmer models; it provides a conceptually distinct approach to understanding normal protein function, disease-associated protein dysfunction, drug action, and approaches to drug design. This article describes the morpheein model for allosteric regulation, provides a disease relevant example in the protein porphobilinogen synthase, and considers current and future research intended to capitalize on targeting quaternary structure shape shifting in many different proteins as a way to understand disease and develop therapies. Because there are so few well characterized examples, even the most comprehensive treatments of allosteric drug discovery do not address proteins that are established to sample a dynamic equilibrium of assemblies comprised of alternate protomer conformations whose interconversion is forbidden within the assemblies (Lu et al., 2019a,b).

MORPHEEINS WITHIN THE CONTEXT OF PROTEIN STRUCTURE DYNAMICS

The existence of morpheeins is one of many protein structure dynamic phenomena that falls outside the classic one sequence – one structure – one function paradigm. Consequently, such discoveries

[1]http://en.wikipedia.org/wiki/Morpheein

have been surprising (e.g., Breinig et al., 2003; Kashlan and Cooperman, 2003; Bornholdt et al., 2013), often serendipitous, and have been accompanied by the introduction of alternate related nomenclature such as transformers and metamorphic proteins (e.g., Murzin, 2008; Lopez-Pelegrin et al., 2014; Wasserman and Saphire, 2016; Dishman and Volkman, 2018). These refer to a continuum of quaternary structure dynamics which expand our view of protein structure beyond the level of primary, secondary, tertiary and quaternary. In the study of ribonucleotide reductase as a drug target, the investigator Aye has referred to going beyond quaternary structure as "breaking the fourth wall"; in the study of Ebola virology, the investigator Ollmann-Saphire has termed it the "fifth level of protein structure" (Wasserman and Saphire, 2016; Long et al., 2019). Herein, we use the term "fifth level of protein structure" to refer to equilibria of alternate assemblies comprised of alternate protomer conformations. This builds on the established concept that protein function is a consequence of an equilibrium of protein structures (Parisi et al., 2015). Both Aye's and Ollmann-Saphire's treatments highlight that normal protein structure dynamics can include architecturally distinct assemblies with alternate functions that are comprised of different protomer conformations. These assemblies exist as equilibria in the absence of chemical modification. The populations (e.g., mole fraction) of alternate morpheein forms responds to environmental factors (e.g., ionic strength, pH) and most significantly to ligand binding. These factors may govern the predominance of alternate morpheein forms in different cellular locations. Single amino acid substitutions that alter the mole fractions of alternate morpheein forms can cause disease (e.g., Jaffe and Stith, 2007). The morpheein model of protein allostery is a dissociative allosteric model most closely related to the equilibrium models of Nussinov (e.g., Kar et al., 2010) and Hilser (e.g., Motlagh et al., 2014), with the added dimension of quaternary structure. In the prototype morpheein described below, porphobilinogen synthase, the alternate functions are high activity (on) vs. low activity (off) (Jaffe and Stith, 2007; Jaffe and Lawrence, 2014; Jaffe, 2016, 2020). In the Ebola virus VP40 protein, the alternate functions are entirely separate activities, each one of which is essential for the viral life cycle (Bornholdt et al., 2013). Proteins that can moonlight (carry out unrelated functions, like VP40) were first discovered in the 1980's (e.g., Gurney et al., 1986), and often arose from cloning the gene responsible for a biological function only to discover that the cloned protein sequence was already known to have a different function. A fascinating example is the protein originally identified as the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase, which now has many documented functions, many of which can be targeted for drug discovery (e.g., Kopeckova et al., 2020; Lazarev et al., 2020). The known moonlighting proteins have recently been assembled by the investigator Jeffery into a MoonProt® database, which currently has ∼400 listings (Mani et al., 2015; Chen et al., 2018). In most instances it remains to be determined if alternate moonlighting functions are associated with alterations at the fifth level of protein structure. A related fifth level phenomenon is the reversible filamentation of some enzymes, recently reviewed by Horton (Park and Horton, 2019, 2020). Outstanding questions for many filament-forming proteins is whether they are morpheeins (with alternate protomer conformations), moonlighting proteins (with more than one function), or both. Two related enzymes, CTP synthase and IMP dehydrogenase are exemplars of this unknown. Each, separately and together, undergo changes in multimerization or filament formation in response to the state of the cell, but functional distinctions among these assemblies are yet unknown (Simonet et al., 2020).

Figure 1 illustrates the morpheein phenomenon using differently shaped dice to represent different conformations of the protomer. Monod first used dice assemblies to illustrate quaternary structures (Monod, 1965). **Figure 1** shows equilibration between two alternate conformations of the protomer, where one is represented by a cubic die and the other is represented by a pyramidal die. Although not obvious from the representative shapes, *the interconversion of these conformations does not require any substantial changes in the protein fold at the level of secondary or tertiary structure. The interconversion is spontaneous; it does not require any external input of energy.* It may involve small regions of order – disorder transition. For example, interconversion between the protomer conformations represented by the cube and pyramid could be a hinge motion

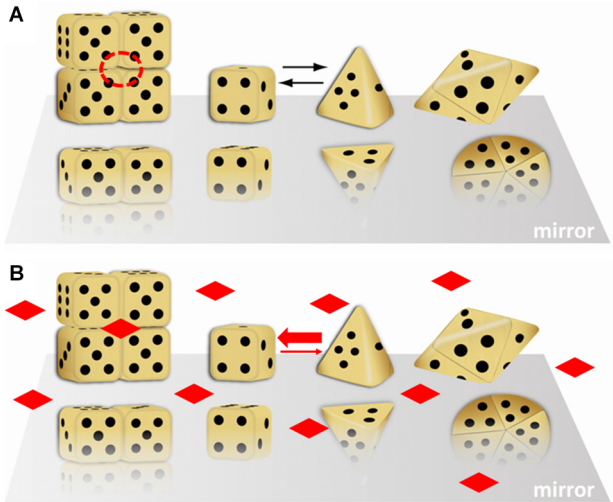


FIGURE 1 | A dice based illustration of the morpheein phenomenon. That the alternate assemblies have different functions makes this a dissociative model for protein allostery. In this illustration, the stoichiometry of the alternate assemblies is different; this is just an example, not a prerequisite. **(A)** Cubic and pyramidal dice are used as symbolic (not structural) representations of alternate conformations of a protomer that can self-assemble through association of two complementary surfaces. In this case, the two surfaces are represented by the die face with one and with four dots. The tetramer resembles a stack of boxes; the pentamer resembles a flying saucer. The red dashed circle is a multimer-specific surface cavity that can serve as a ligand binding site. Note that the orientation of the protomer is not retained in the illustration of the multimer. **(B)** The diamond shaped ligands can bind to the multimer-specific binding site and draw the equilibrium toward the assembly of cubic dice, thus dictating protein function. This figure was adapted (with permission) from an image first printed as the journal cover associated with (Lawrence et al., 2008).

between two folded domains of each protomer that allows the die face with five dots to associate with the die face with six dots, burying these surfaces. Hinge motions allow protomer shape change without requiring a change in protein fold. In **Figure 1**, the higher order multimers form by association of the die face with one dot to the die face with four dots. The cubic die forms a symmetric tetramer; the pyramidal die forms a symmetric pentamer. It is easy to imagine how these two assemblies, though made up of chemically identical components, will interact with different cellular partners and potentially have different functions. Note that the surface of the tetramer contains a multimer-specific surface cavity that can serve as a ligand binding site (dashed circle in **Figure 1A**). The ligand could be a natural allosteric effector molecule, a drug, or another cellular entity (protein, nucleic acid, lipid, membrane surface). The pentamer does not have this same surface cavity and will not interact with the same ligand. In **Figure 1B**, addition of the imagined ligand causes stabilization of the tetramer, which will draw the structural equilibrium toward the tetramer and alter the protein's function to that of the tetramer. All of this happens in the absence of post-translational modifications or any other covalent changes to the protein. We note, however, that changes in protein sequence, post-translational modification, or the presence of purification tags can shift the position of the equilibrium (mole fraction of alternate morpheein forms) and enhance or inhibit allosteric ligand binding.

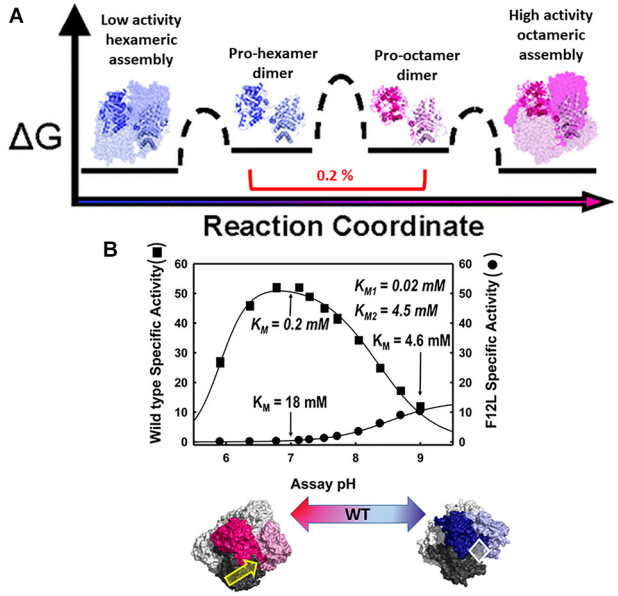


FIGURE 2 | The equilibrium of alternate morpheein forms of PBGS. In all panels, octamer (or pro-octamer dimer) components are in shades of pink; hexamer (or pro-hexamer dimer) components are in shades of blue. **(A)** The reaction coordinate diagram for the interconversion of human PBGS morpheein forms [adapted with permission from Selwood et al. (2008), copyright 2008, American Chemical Society]. In this illustration, one dimer is shown in shaded ribbons (darker/lighter) while the other dimers are in space filling with the same coloring. Note the low mole fraction of the essential dimeric intermediates (Selwood et al., 2008). **(B)** The pH rate profile (10 mM substrate, Bis-tris propane buffer) for WT human PBGS and the disease-associated F12L variant, which strongly favors the hexameric assembly [images adapted from Breinig et al. (2003)]. K_M and V_{MAX} determinations varied substrate from 10 μM to 10 mM at pH 7 and pH 9. Below illustrates that the position of the quaternary structure equilibrium is a function of pH; at neutral pH the WT protein is predominantly octamer while at pH 9, it is predominantly hexamer. The pH dependent transition from octamer to hexamer for WT human PBGS accounts for the basic arm of the pH rate profile (Selwood et al., 2008). A key component of this pH dependence is the protonation of an arginine residue, which sits at a multimer specific interface between the N-terminal arm of the dark pink subunit and the αβ-barrel of the dark gray subunit, as indicated by the yellow arrow (Tang et al., 2006). This arginine is spatially equivalent to an allosteric magnesium ion found in PBGS from species that are neither metazoan nor fungal. The octamer-specific interface stabilizes a closed active site lid. The white diamond indicates the hexamer-specific surface cavity that was successfully targeted for identification of species-specific (plant) PBGS inhibitors (Lawrence et al., 2008).

THE PROTOTYPE MORPHEEIN – PORPHOBILINOGEN SYNTHASE (PBGS)

PBGS Provides a Physiologic Relevance to the Morpheein Model of Allostery

The physiologic relevance of the morpheein model of allostery was first realized for the protein porphobilinogen synthase (PBGS), whose quaternary structure dynamic is illustrated in **Figure 2A** (Breinig et al., 2003; Selwood et al., 2008). Here I paint a broad picture of PBGS, highlighting key aspects of its fifth level of protein structure and refer the reader to recent reviews for more details (Jaffe and Lawrence, 2014; Jaffe, 2016). In the PBGS example, optimal enzyme activity requires controlled access to the enzyme active site, which is gated by the opening and closing of an active site lid. Each protomer has a complete active site, but securely closing the active site lid depends upon a network of molecular interactions that can only be achieved in the octamer (Jaffe, 2016). In the PBGS example, the alternate assemblies have different multimer-specific surface cavities that can be used for the development of bioactive molecules. The alternate assemblies also have different size, shape, and surface charge, which allows them to be separated by biochemical and biophysical methods such as native PAGE and ion exchange chromatography (Breinig et al., 2003).

The PBGS catalyzed reaction is essential for all organisms that rely on methanogenesis, photosynthesis, and/or respiration, thus covering every branch of cellular life. With all this evolutionary time to adapt to the organism's needs and to function in different cellular environments (e.g., cytoplasm, chloroplast, apicoplast), factors governing the fifth level of PBGS protein structure are not evolutionarily conserved. Additionally, the amino acid composition of the targeted surface cavities is not conserved, unlike active site residues. This makes the allosteric regulation of PBGS a potential target for the development of antimicrobials and herbicides. In some species (e.g., plants, bacteria, archaea) the equilibrium position depends upon an allosteric magnesium binding at an interface only present in the octamer (see yellow arrow in **Figure 2B**). In PBGS from metazoa and fungi, which lack the magnesium binding site, in its place is the guanidinium

group of an arginine residue. In the human PBGS variant where a leucine is substituted for Phe12 (F12L), from which we obtained the crystal structure of the hexameric assembly (Breinig et al., 2003), this conserved arginine does not contact the neighboring N-terminal arm. In an apicoplast, PBGS has evolved to contain a C-terminal extension that prevents the hinge motion necessary to convert pro-octamer dimer to pro-hexamer dimer; this hinge motion is illustrated in **Figure 2A**. Thus, in apicoplast PBGS the equilibrium components are limited to octamer and pro-octamer dimer (Jaffe et al., 2011). This phylogenetic variation demonstrates that evolution of the fifth level of protein structure provides another opportunity for the adaptation of protein functional control.

Control of PBGS Morpheein Forms by pH and Ligand Binding

The human PBGS pH rate profile (**Figure 2B**) helped reveal a pH dependence to the quaternary structure equilibrium of wild type human PBGS (see the bottom panel of **Figure 2B**; Selwood et al., 2008). The crystal structures of the PBGS octamer (e.g., Jaffe, 2004; Jaffe et al., 2011; Mills-Davies et al., 2017) show that a non-reacting moiety of the K_M-determining substrate interacts with the active site lid, essentially closing the lid and allowing deprotonation of an active site lysine required for essential Schiff base formation. With an open active site, this deprotonation requires a high solvent pH, as is seen in the pH rate profile of the constitutively hexameric F12L variant. The crystal structures of the protomers in the human PBGS octamer vs. hexamer superpose remarkably well, but differ in two key ways. First is the backbone hinge between the αβ-barrel and the N-terminal arm domains; this dictates assembly to octamer vs. hexamer. Second is the presence of an ordered active site lid, which is only present in one hemisphere of the octamer. Both structures contain active site ligands in a half-of-the-sites stoichiometry (one hemisphere) (PDB: 1E51, 1PV8). Only the ligand-containing octamer active sites contains atoms that derive from the K_M-determining substrate, which are securing a closed conformation of the active site lid through a network of bonds between the substrate's carboxyl group and basic residues on the lid.

Although the PBGS quaternary structure equilibrium is controlled by different factors in different branches of life, there are unifying characteristics in the pH rate profiles of alternate PBGS morpheein forms. Mammalian PBGS at neutral pH (see **Figure 2B**, bottom), and plant/bacterial PBGS with magnesium present are predominantly octameric and are documented to have K_M values in the range of ∼150 μM (Mitchell and Jaffe, 1992; Jaffe et al., 1995; Breinig et al., 2003), which is the range of the cellular substrate concentration. The isolated hexameric F12L variant (see **Figure 2B**), and the bacterial *Escherichia coli* PBGS without magnesium exhibit K_M values (at neutral pH) that are well above the physiological substrate concentration, in the range of 5–20 mM. Addition of magnesium to a magnesium-free *E. coli* PBGS sample has been shown to shift the quaternary structure equilibrium to octamer (**Figure 3A**) and reduce the K_M to that characteristic

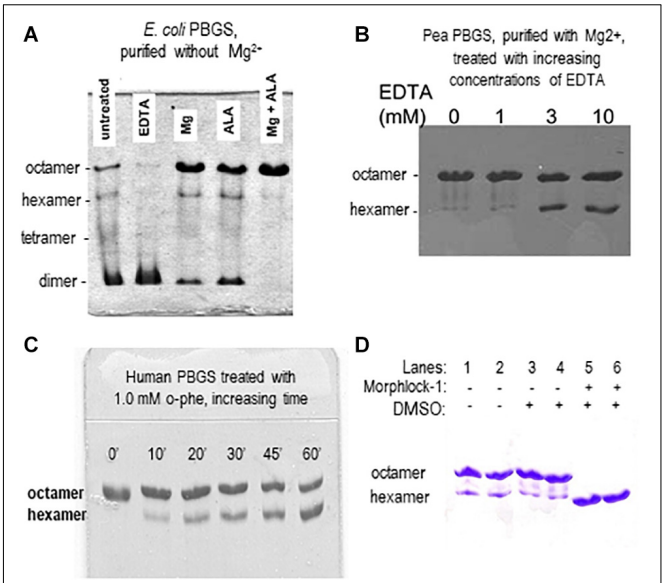


FIGURE 3 | Native PAGE illustrates how allosteric ligands control the distribution of PBGS morpheein forms. **(A)** The distribution of morpheein forms of *E. coli* PBGS responds to magnesium and substrate (ALA), both of which stabilize the octamer [image adapted with permission from Jaffe et al. (1995), copyright 1995 American Chemical Society]. In this example the protein was purified with zinc, but not magnesium (Jaffe et al., 1995). The *E. coli* PBGS octamer and the position of the allosteric magnesium are established by crystal structure (Kervinen et al., 2001). Hexamer, tetramer, and dimer positions are modeled on the assumption that mobility of all assemblies is governed by the same charge/mass ratio. **(B)** The distribution of morpheein forms of plant PBGS (*Pisum sativum*) responds to the addition of EDTA, which strips the magnesium, destabilizing the octamer and favoring accumulation of hexamer [image adapted from Breinig et al. (2003)]. In this example the protein was purified in the presence of magnesium (Kervinen et al., 2000; Breinig et al., 2003) and multimer size was determined by analytical ultracentrifugation (Kokona et al., 2008). **(C)** Binding the K_M determining substrate of human PBGS relies upon an active site zinc ion. Removing the zinc with 1,10-phenanthroline destabilizes the octamer [image adapted with permission from Jaffe (2016) copyright 2016 American Chemical Society]. **(D)** A ligand discovered through computational docking to a hexamer-specific binding site on plant PBGS stabilizes the hexamer [image adapted with permission from Lawrence et al. (2008)]. This ligand inhibits the plant enzyme, but does not affect the activity of the human enzyme [not shown, Lawrence et al. (2008)].

of octamer (Mitchell and Jaffe, 1993; Jaffe et al., 1995). Similarly, treatment of a plant PBGS with EDTA causes a shift from octamer to hexamer (**Figure 3B**; Breinig et al., 2003). For human PBGS, this same phenomenon can be demonstrated by removal of a catalytic zinc ion that is essential for binding the K_M-determining substrate (**Figure 3C**; Jaffe, 2016). The high K_M values for PBGS that are not octameric derives from an inability to secure the closed active site lid, leaving the K_M-determining substrate loosely bound. What is striking about the pH rate profile of WT human PBGS is the shift in K_M values between pH 7 and pH 9, where the kinetic parameters resemble that of hexameric F12L. In fact, WT human PBGS at pH 9 exhibits a double hyperbolic kinetic behavior. At low substrate concentration, the observed reaction rate is dominated by the low K_M (high V_{MAX}) octameric component, which is at

a relatively low mole fraction. As the substrate concentration starts to approach the K_M of the hexamer, the observed enzyme catalyzed reaction rate become dominated by the high K_M (low V_{MAX}) hexameric component (which is at a high mole fraction). This double hyperbolic kinetic phenomenon, caused by a mixture of species with different kinetic constants follows a classic treatment of isozymes. The alternate morpheein forms of PBGS are not isozymes. Rather, they are a slowly exchanging mixture octamer and hexamer.

An Inborn Error of Metabolism Is Linked to Perturbation of an Equilibrium of Morpheein Forms

ALAD porphyria is a rare inborn error of metabolism caused by dysfunctional PBGS and inherited as a recessive disease (Maruno et al., 2001). ALAD, an abbreviation for amino-levulinic acid dehydratase, is an alternate name for PBGS; it remains in clinical use. There are eight known disease-associated variants and all patients are compound heterozygotes. Only one disease-associated variant alters an amino acid at the enzyme active site. When heterologously expressed and purified from *E. coli*, the disease-associated human PBGS variants all show an increased propensity to populate the hexameric assembly at neutral pH (Jaffe and Stith, 2007). This can be illustrated by the appearance of the double hyperbolic kinetic phenomenon at neutral pH (**Figure 4**), which is indicative of a mixture of octamer and hexamer. It is also seen by native PAGE and ion exchange chromatography, both of which separate the octamer from the hexamer, as confirmed by the established crystal structures of the wild-type and the constitutively hexameric F12L variant (Breinig et al., 2003).

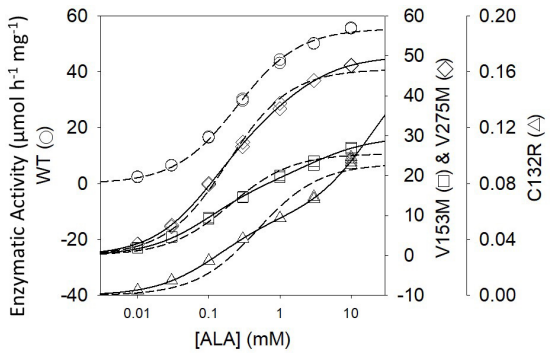


FIGURE 4 | The activity vs. substrate relationship for disease-associated human PBGS proteins indicate the presence of two morpheein forms at neutral pH. The dashed lines are fitted to the hyperbolic Michaelis–Menten equations, which is an excellent fit for the WT protein. The solid lines are double hyperbolic fits (the sum of two species with different K_M and V_{MAX} values) [image adapted with permission from Jaffe and Stith (2007)]. The kinetic behavior of each variant shows evidence for a mixture of low K_M and high K_M morpheein forms, corresponding to octamer and hexamer (Jaffe and Stith, 2007). The wild type protein shows this double hyperbolic behavior at pH 9 (see **Figure 2B**).

We have posited that the disequilibrium of alternate assemblies contributes to the complex phylogenetic patterns seen in the most common inborn error of amino acid metabolism, phenylketonuria, which is caused by dysfunctional phenylalanine hydroxylase (Jaffe et al., 2013; Jaffe, 2017). Although it is now established that phenylalanine hydroxylase can equilibrate between architecturally distinct high activity and low activity tetrameric assemblies, it is not yet established whether tetramer dissociation is a required component their interconversion (Jaffe, 2017).

WE COULD EASILY HAVE MISSED THE FIFTH LEVEL OF PROTEIN STRUCTURE WHEN STUDYING PBGS

The four levels of protein structure are introduced in every biochemistry text. Additionally, the relationship between sequence, structure, and function is the foundation of broad applications of bioinformatics, which drive much biomedical research. For the first 20 years that we studied PBGS, we were not looking for the fifth level of protein structure. All of our data prior to ∼2003 was interpreted within the context that PBGS has one fixed quaternary architecture, which is an octamer (e.g., Wu et al., 1974), in possible equilibrium with tetramers or dimers comprised of the same protomer conformation. For example, our published model of the protein concentration dependence of a plant PBGS included only dimer, tetramer, and octamer (Kervinen et al., 2000). A few publications had suggested that PBGS is a hexamer (e.g., Stolz and Dornemann, 1996). But, we believed that one of these interpretations must be incorrect.

The first PBGS crystal structure, published in 1997, showed the architecture of the octamer, establishing precedent, as crystal structures often do (Erskine et al., 1997). As described above, the third decade of our work with PBGS revealed the protein as the prototype morpheein. We would not have seen this but for a confluence factors. (1) With a focus on the evolution of metal ion usage, we had correlated sequence variations with *in vitro* behaviors looking at PBGS from mammals, bacteria, and plants (e.g., Jaffe, 2003). By the time we fully evaluated this relationship, the first crystal structure of a bacterial PBGS showed the location of the allosteric magnesium binding site (Frankenberg et al., 1999). It is present in the octamer and absent in the hexamer. (2) We had not used purification tags, which allowed ion exchange chromatography to reveal the separation of PBGS morpheein forms during purification. It was a surprise when the disease-associated PBGS variant F12L had a very different mobility on an ion exchange column relative to the charge-equivalent WT protein (Breinig et al., 2003). (3) Had we not chosen to study the F12L variant, we would not have had the stable hexamer as a reagent for comparative analysis. For decades we had been discarding the small amount of low activity hexameric mammalian PBGS that is present at neutral pH and separated during our purification of wild type proteins. When studying a human PBGS variant from which the catalytic zinc binding site had been removed, we consistently

saw two peaks on an ion exchange column, but could not discriminate these by circular dichroism or enzyme kinetics. By kinetic criteria they were the same, because addition of substrate converted the hexamer to the octamer during the long time-course assays for these very low activity variants. Substrate stabilization of the octameric assembly can be illustrated using 2-dimensional native PAGE during which the gel is incubated in assay mixture between the two dimensions (e.g., Jaffe et al., 1995; Lawrence et al., 2008; see also **Figure 3A**). (4) Had we not obtained the crystal structure of F12L, we would not have seen the structure of the hexamer. From this, and knowledge of the location of the allosteric magnesium from prior crystal structures, we realized the octamer ↔ hexamer equilibrium as the basis for allosteric regulation of PBGS. Our extensive work on *E. coli* PBGS (including the 1995 documentation of the native PAGE phenomenon shown in **Figure 3A**) could now be reinterpreted within the context of this new allosteric model. Having previously interpreted our data in terms of a classic primary, secondary, tertiary, and quaternary structure paradigm, we had overlooked hints that the PBGS octamer was part of an equilibrium of differently sized morpheein forms. For example, a ^{13}C NMR study of labeled substrate bound to *E. coli* PBGS (in the absence of magnesium) had unexpectedly yielded NMR line widths significantly narrower relative to an octameric mammalian PBGS (Jaffe and Markham, 1988; Mitchell and Jaffe, 1993), suggesting a faster rotational correlation time indicative of a smaller multimeric size.

FINDING PROTEINS THAT EXPERIENCE THE FIFTH LEVEL OF PROTEIN STRUCTURE

Since the fifth level of protein structure provides an additional way to manipulate protein function, it also provides an additional approach to allosteric drug discovery. Consequently, there is value in identifying proteins that behave as morpheeins. This goal remains difficult. The initial discovery of the morpheein character of PBGS was serendipitously based on *in vitro* protein behavior coupled with X-ray crystal structures showing architecturally different assemblies comprised of different protomer conformations, but not different folds. This discovery was not based on a bioinformatics approach. The PBGS example presented itself because there happened to be a naturally occurring variant, F12L, that sufficiently stabilized the hexameric assembly to obtain its crystal structure. Although the wild-type human PBGS was later realized to readily equilibrate between octamer and hexamer in a pH-dependent fashion (**Figure 2**), to date all conditions that have yielded diffraction quality crystals of the wild-type protein favored crystals comprised of octamer. Serendipity favored our studying F12L; we could not have predicted the effect of this mutation. Well-established computational approaches (e.g., the program FoldX; Schymkowitz et al., 2005) which analyze the overall thermodynamic effect of amino acid substitutions will fail to predict substitutions that shift the position of an equilibrium of morpheein forms. A similar story applies to the discovery of alternate morpheein forms and alternate functions of the Ebola virus VP40 protein (Bornholdt et al., 2013).

Our initial approach to identifying proteins that use the fifth level of protein structure was to manually search the literature for proteins that have one or more characteristics that we had documented for PBGSs. This approach was especially challenging as most of these characteristics were not contained in the searchable abstract or keywords and much of the older literature was not yet available as text-searchable PDF documents. Some of the characteristics we focused on are listed in **Table 1**; the characteristics included in **Table 1** are each consistent with an equilibrium of alternate assemblies, but none are strictly diagnostic. Each of these behaviors can be otherwise attributed. Trevor Selwood's herculean efforts generated boxes of reprints and a list of putative morpheeins (Selwood and Jaffe, 2012). This list, originally posted on Wikipedia (Morpheein[1]), to date has not been actively edited by the community. Failure to gain research support for using the literature to identify proteins that function

TABLE 1 | Protein behaviors that might indicate an equilibrium of morpheein forms.

Characteristic	Comments
SDS-pure protein separating into alternate bands on native PAGE can indicate multimers of alternate stoichiometry (or conformation).	Native PAGE can be routinely incorporated into the final stages of protein characterization.
SDS-pure protein that separates into alternate forms using ion exchange chromatography	This has been used to monitor the interconversion of alternate assemblies as a function of ligand.
The enzyme-kinetic phenomenon known as hysteresis.	This can be observed if the transition from a low activity form to a high activity form occurs during the assay [as we have shown for the R240A variant of PBGS (Tang et al., 2006].
Double hyperbolic kinetics (the sum of two hyperbola with different kinetic constants).	This can indicate alternate morpheein forms with different kinetic constants. Seeing this may require using a broad range of substrate concentration (see **Figure 4**).
X-ray crystal structures of multi-domain proteins that cannot be superposed without clashes. This observation is often dismissed as an artifact of crystal packing.	Many multi-domain proteins do not produce diffraction quality crystals. A common approach is to truncate one or more domains. If overlaying the common elements of such structures causes domains to clash, this could be a sign of alternate assemblies. This was observed for alternate truncated constructs of HIV integrase (Andrake and Skalka, 2015).
Protein concentration dependent specific activity can indicate alternate activities associated with different multimeric stoichiometries.	This is seen for all PBGS that use only magnesium [e.g., (Petrovich et al., 1996; Kervinen et al., 2000; Shanmugam et al., 2010)].
Evidence for soluble protein multimers that dissociate along a hydrophilic protein-protein interface.	This is a difficult characteristic to search for as many crystal structure files are at insufficient resolution to position water molecules. PBGS crystal structures contain phylogenetically conserved water molecules at the subunit-subunit interfaces that dissociate upon formation of the dimeric morpheein forms (Selwood et al., 2008).

as morpheeins turned this to a back-burner approach. Yet, the relatively small number of established morpheeins precludes designing a bioinformatics method for their identification. Jeffrey has described the same dilemma for the development of a bioinformatics approach to identifying moonlighting proteins (Mani et al., 2015).

Nevertheless, a key question remains as to how we can use what we have learned from PBGS to harness the fifth level of protein structure for drug discovery. In the PBGS example, the difference between the protomer that forms a hexamer and the protomer that forms an octamer is a hinge between two domains, without significantly altering the fold of these domains. This is also the case for alternate assemblies of the HIV integrase protein, where a hinge motion dictates formation of a core-core-dimer vs. a reaching dimer; and where it has been pointed out that small molecule stabilization of one or the other dimer, to prevent their interconversion, could yield a therapeutic (Bojja et al., 2013). For integrase, this promise has not yet materialized. In the case of VP40, alternate multimerization architectures also appear to arise from hinge motions between two domains, without significant refolding (Bornholdt et al., 2013). In the VP40 example, ligand association (RNA or membrane) dramatically stabilize one assembly relative to another allowing the protein to fulfill different essential function in the viral life cycle. Although not yet realized, drugs that prevent the interconversion of such assemblies could form the basis of a therapeutic. Research to facilitate this approach is ongoing (e.g., Buzon et al., 2020). In other cases wherein allosteric drugs could be imagined to work by stabilizing one of alternate assemblies (such as ribonucleotide reductase) the detailed molecular structures of alternate assemblies are only beginning to be revealed. Some of these are described in two recent reviews on the rich oligomeric and functional repertoire of both mammalian and bacterial ribonucleotide reductase enzymes (Thomas et al., 2019; Long et al., 2020).

IDENTIFYING ALLOSTERIC REGULATORS (E.G., THERAPEUTICS)

Using PBGS as an example, we have demonstrated how an equilibrium of morpheein forms can be manipulated in ways related to drugs. Realistically, ALAD porphyria is such a rare disorder that finding an octamer-stabilizing allosteric effector is, at best, an academic exercise. However, stabilization of a PBGS hexamer could form the basis for an antimicrobial or herbicide. To test this hypothesis, we targeted a hexamer-specific surface cavity on the model of a plant PBGS hexamer (*in silico* docking/*in vitro* testing) and found a hexamer-stabilizing inhibitor that did not affect human PBGS (Lawrence et al., 2008). **Figure 3D** illustrates the hexamer-stabilizing effect of the discovered compound (named morphlock-1). Unlike the enzyme active site, the multimer specific surface cavities is not phylogenetically conserved. We also used native PAGE to screen libraries of approved drugs as well as environmental contaminants (Lawrence et al., 2011, 2013). These studies revealed an explanation for some porphyria-promoting drug side effects as well as the potential for environmental contributions to confound genotype/phenotype correlations. These demonstrations indicate that small molecule modulation of equilibria of morpheein forms is a viable approach to drug discovery. Nevertheless, a computational approach to identifying such modulators requires molecular resolution models and/or structures of alternate assemblies; these are not yet available for most putative morpheeins.

CONCLUSION AND FUTURE OUTLOOK

Proteins that can come apart, change shape, and reassemble differently with functional consequences provide expanded opportunities for understanding disease and designing therapeutics. This fifth level of protein structure provides another example where the one sequence/one structure/one function rule fails to provide the correct framework for data interpretation. Biophysical techniques that are becoming more widely available, (e.g., SEC-MALS, SEC-SAXS, cryo-EM, light scattering, to name just a few) are revealing the shape changing behavior of many disease-associated multimeric proteins. Other forms of microscopy are revealing changes in protein locations within cells. In some instances, these are associated with protein filamentation. All of these observations suggest the potential for harnessing the fifth level of protein structure for therapeutic advantage. As described, a few notable examples include ribonucleotide reductase (Thomas et al., 2019; Long et al., 2020), HIV integrase (Andrake and Skalka, 2015; Gupta et al., 2016), and Ebola VP40 protein (Bornholdt et al., 2013; Del Vecchio et al., 2018).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work has been funded by grants from the National Institutes of Health 5R01 ES003654; 5R21 AI063324; 5R01 NS100081; and in part by the National Institutes of Health NCI Cancer Center Support Grant P30 CA006927. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health. Continuous support is acknowledged from the Fox Chase Cancer Center, now a part of the Temple University Health Systems.

ACKNOWLEDGMENTS

The author acknowledges many who have contributed to our deciphering the quaternary structure dynamics of PBGS, most notably Drs. Sabine Breinig, Sarah H. Lawrence, Trevor Selwood, and Lei Tang, and Ms. Linda Stith. Special acknowledgment is given to Emilia Arturo who has contributed to expanding our general appreciation for the fifth level of protein structure.

REFERENCES

Andrake, M. D., and Skalka, A. M. (2015). Retroviral integrase: then and now. *Annu. Rev. Virol.* 2, 241–264. doi: 10.1146/annurev-virology-100114-055043

Bojja, R. S., Andrake, M. D., Merkel, G., Weigand, S., Dunbrack, R. L. Jr., and Skalka, A. M. (2013). Architecture and assembly of HIV integrase multimers in the absence of DNA substrates. *J. Biol. Chem.* 288, 7373–7386. doi: 10.1074/jbc.m112.434431

Bornholdt, Z. A., Noda, T., Abelson, D. M., Halfmann, P., Wood, M. R., Kawaoka, Y., et al. (2013). Structural rearrangement of ebola virus VP40 begets multiple functions in the virus life cycle. *Cell* 154, 763–774. doi: 10.1016/j.cell.2013.07.015

Breinig, S., Kervinen, J., Stith, L., Wasson, A. S., Fairman, R., Wlodawer, A., et al. (2003). Control of tetrapyrrole biosynthesis by alternate quaternary forms of porphobilinogen synthase. *Nat. Struct. Biol.* 10, 757–763. doi: 10.1038/nsb963

Buzon, P., Ruiz-Sanz, J., Martinez, J. C., and Luque, I. (2020). Stability, conformational plasticity, oligomerization behaviour and equilibrium unfolding intermediates of the Ebola virus matrix protein VP40. *J. Biomol. Struct. Dyn.* 38, 4289–4303. doi: 10.1080/07391102.2019.1671226

Chen, C., Zabad, S., Liu, H., Wang, W., and Jeffery, C. (2018). MoonProt 2.0: an expansion and update of the moonlighting proteins database. *Nucleic Acids Res.* 46, D640–D644.

Del Vecchio, K., Frick, C. T., Gc, J. B., Oda, S. I., Gerstman, B. S., Saphire, E. O., et al. (2018). A cationic, C-terminal patch and structural rearrangements in Ebola virus matrix VP40 protein control its interactions with phosphatidylserine. *J. Biol. Chem.* 293, 3335–3349. doi: 10.1074/jbc.m117.816280

Dishman, A. F., and Volkman, B. F. (2018). Unfolding the mysteries of protein metamorphosis. *ACS Chem. Biol.* 13, 1438–1446. doi: 10.1021/acschembio.8b00276

Erskine, P. T., Senior, N., Awan, S., Lambert, R., Lewis, G., Tickle, L. J., et al. (1997). X-ray structure of 5-aminolaevulinate dehydratase, a hybrid aldolase. *Nat. Struct. Biol.* 4, 1025–1031. doi: 10.1038/nsb1297-1025

Frankenberg, N., Erskine, P. T., Cooper, J. B., Shoolingin-Jordan, P. M., Jahn, D., and Heinz, D. W. (1999). High resolution crystal structure of a Mg2+-dependent porphobilinogen synthase. *J. Mol. Biol.* 289, 591–602. doi: 10.1006/jmbi.1999.2808

Gupta, K., Turkki, V., Sherrill-Mix, S., Hwang, Y., Eilers, G., Taylor, L., et al. (2016). Structural basis for inhibitor-induced aggregation of HIV integrase. *PLoS Biol.* 14:e1002584. doi: 10.1371/journal.pbio.1002584

Gurney, M. E., Heinrich, S., Lee, M. R., and Yin, H. S. (1986). Molecular cloning and expression of neuroleukin, a neurotrophic factor for spinal and sensory neurons. *Science* 234, 566–574. doi: 10.1126/science.3764429

Jaffe, E. K. (2003). An unusual phylogenetic variation in the metal ion binding sites of porphobilinogen synthase. *Chem. Biol.* 10, 25–34. doi: 10.1016/s1074-5521(02)00296-x

Jaffe, E. K. (2004). The porphobilinogen synthase catalyzed reaction mechanism. *Bioorg. Chem.* 32, 316–325. doi: 10.1016/j.bioorg.2004.05.010

Jaffe, E. K. (2005). Morpheeins–a new structural paradigm for allosteric regulation. *Trends Biochem. Sci.* 30, 490–497. doi: 10.1016/j.tibs.2005.07.003

Jaffe, E. K. (2016). The remarkable character of porphobilinogen synthase. *Acc. Chem. Res.* 49, 2509–2517. doi: 10.1021/acs.accounts.6b00414

Jaffe, E. K. (2017). New protein structures provide an updated understanding of phenylketonuria. *Mol. Genet. Metab.* 121, 289–296. doi: 10.1016/j.ymgme.2017.06.005

Jaffe, E. K. (2020). "Chapter three - porphobilinogen synthase: an equilibrium of different assemblies in human health, in progress," in *Molecular Biology and Translational Science*, eds J. Giraldo, and F. Ciruela (Cambridge, MA: Academic Press), 85–104. doi: 10.1016/bs.pmbts.2019.11.003

Jaffe, E. K., Ali, S., Mitchell, L. W., Taylor, K. M., Volin, M., and Markham, G. D. (1995). Characterization of the role of the stimulatory magnesium of *Escherichia coli* porphobilinogen synthase. *Biochemistry* 34, 244–251. doi: 10.1021/bi00001a029

Jaffe, E. K., and Lawrence, S. H. (2014). "The dance of porphobilinogen synthase in the control of tetrapyrrole biosynthesis," in *Handbook of Porphyrin Science*, ed. G. C. Ferreira (Toh Tuck: World Scientific Publishing Co. Pte. Ltd), 79–128. doi: 10.1142/9789814407755_0002

Jaffe, E. K., and Markham, G. D. (1988). C-13 Nmr-Studies of methylene and methine carbons of substrate-bound to a 280000-Dalton protein, porphobilinogen synthase. *Biochemistry* 27, 4475–4481. doi: 10.1021/bi00412a039

Jaffe, E. K., Shanmugam, D., Gardberg, A., Dieterich, S., Sankaran, B., Stewart, L. J., et al. (2011). Crystal structure of *Toxoplasma gondii* porphobilinogen synthase: insights on octameric structure and porphobilinogen formation. *J. Biol. Chem.* 286, 15298–15307. doi: 10.1074/jbc.m111.226225

Jaffe, E. K., and Stith, L. (2007). ALAD porphyria is a conformational disease. *Am. J. Hum. Genet.* 80, 329–337. doi: 10.1086/511444

Jaffe, E. K., Stith, L., Lawrence, S. H., Andrake, M., and Dunbrack, R. L. Jr. (2013). A new model for allosteric regulation of phenylalanine hydroxylase: implications for disease and therapeutics. *Arch. Biochem. Biophys.* 530, 73–82. doi: 10.1016/j.abb.2012.12.017

Kar, G., Keskin, O., Gursoy, A., and Nussinov, R. (2010). Allostery and population shift in drug discovery. *Curr. Opin. Pharmacol.* 10, 715–722. doi: 10.1016/j.coph.2010.09.002

Kashlan, O. B., and Cooperman, B. S. (2003). Comprehensive model for allosteric regulation of mammalian ribonucleotide reductase: refinements and consequences. *Biochemistry* 42, 1696–1706. doi: 10.1021/bi020634d

Kervinen, J., Dunbrack, R. L. Jr., Litwin, S., Martins, J., Scarrow, R. C., Volin, M., et al. (2000). Porphobilinogen synthase from pea: expression from an artificial gene, kinetic characterization, and novel implications for subunit interactions. *Biochemistry* 39, 9018–9029. doi: 10.1021/bi000620c

Kervinen, J., Jaffe, E. K., Stauffer, F., Neier, R., Wlodawer, A., and Zdanov, A. (2001). Mechanistic basis for suicide inactivation of porphobilinogen synthase by 4,7-dioxosebacic acid, an inhibitor that shows dramatic species selectivity. *Biochemistry* 40, 8227–8236. doi: 10.1021/bi010656k

Kokona, B., Rigotti, D. J., Wasson, A. S., Lawrence, S. H., Jaffe, E. K., and Fairman, R. (2008). Probing the oligomeric assemblies of pea porphobilinogen synthase by analytical ultracentrifugation. *Biochemistry* 47, 10649–10656. doi: 10.1021/bi801128d

Kopeckova, M., Pavkova, I., and Stulik, J. (2020). Diverse localization and protein binding abilities of Glyceraldehyde-3-Phosphate dehydrogenase in pathogenic bacteria: the key to its multifunctionality? *Front. Cell. Infect. Microbiol.* 10:89.

Lawrence, S. H., Ramirez, U. D., Tang, L., Fazliyez, F., Kundrat, L., Markham, G. D., et al. (2008). Shape shifting leads to small-molecule allosteric drug discovery. *Chem. Biol.* 15, 586–596. doi: 10.1016/j.chembiol.2008.04.012

Lawrence, S. H., Selwood, T., and Jaffe, E. K. (2011). Diverse clinical compounds alter the quaternary structure and inhibit the activity of an essential enzyme. *Chem. Med. Chem.* 6, 1067–1073. doi: 10.1002/cmdc.201100009

Lawrence, S. H., Selwood, T., and Jaffe, E. K. (2013). Environmental contaminants perturb fragile protein assemblies and inhibit normal protein function. *Curr. Chem. Biol.* 7, 196–206. doi: 10.2174/2212796811307020011

Lazarev, V. F., Guzhova, I. V., and Margulis, B. A. (2020). Glyceraldehyde-3-phosphate dehydrogenase is a multifaceted therapeutic target. *Pharmaceutics* 12:416. doi: 10.3390/pharmaceutics12050416

Long, M. J. C., Hall-Beauvais, A. Van, and Aye, Y. (2020). The more the merrier: how homo-oligomerization alters the interactome and function of ribonucleotide reductase. *Curr. Opin. Chem. Biol.* 54, 10–18. doi: 10.1016/j.cbpa.2019.09.003

Long, M. J. C., Hnedzko, D., Kim, B. K., and Aye, Y. (2019). Breaking the fourth wall: modulating quaternary associations for protein regulation and drug discovery. *Chembiochem* 20, 1091–1104. doi: 10.1002/cbic.201800716

Lopez-Pelegrin, M., Cerda-Costa, N., Cintas-Pedrola, A., Herranz-Trillo, F., Peinado, J. R., et al. (2014). Multiple stable conformations account for reversible concentration-dependent oligomerization and autoinhibition of a metamorphic metallopeptidase. *Angew. Chem. Int. Ed. Engl.* 53, 10624–10630. doi: 10.1002/anie.201405727

Lu, S., He, X., Ni, D., and Zhang, J. (2019a). Allosteric modulator discovery: from serendipity to structure-based design. *J. Med. Chem.* 62, 6405–6421. doi: 10.1021/acs.jmedchem.8b01749

Lu, S., Shen, Q., and Zhang, J. (2019b). Allosteric methods and their applications: facilitating the discovery of allosteric drugs and the investigation of allosteric mechanisms. *Acc. Chem. Res.* 52, 492–500. doi: 10.1021/acs.accounts.8b00570

Mani, M., Chen, C., Amblee, V., Liu, H., Mathur, T., Zwicke, G., et al. (2015). MoonProt: a database for proteins that are known to moonlight. *Nucleic Acids Res.* 43, D277–D282.

Maruno, M., Furuyama, K., Akagi, R., Horie, Y., Meguro, K., Garbaczewski, L., et al. (2001). Highly heterogeneous nature of delta-aminolevulinate dehydratase

(ALAD) deficiencies in ALAD porphyria. *Blood* 97, 2972–2978. doi: 10.1182/blood.v97.10.2972

Mills-Davies, N., Butler, D., Norton, E., Thompson, D., Sarwar, M., Guo, J., et al. (2017). Structural studies of substrate and product complexes of 5-aminolaevulinic acid dehydratase from humans, *Escherichia coli* and the hyperthermophile Pyrobaculum calidifontis. *Acta Crystallogr. D. Struct. Biol.* 73(Pt 1), 9–21.

Mitchell, L. W., and Jaffe, E. K. (1992). *Escherichia*-Coli porphobilinogen synthase is a Zn(Ii) metalloenzyme. *FASEB J.* 6, A459–A459.

Mitchell, L. W., and Jaffe, E. K. (1993). Porphobilinogen synthase from *Escherichia*-coli is a Zn(Ii) metalloenzyme stimulated by Mg(Ii). *Arch. Biochem. Biophys.* 300, 169–177. doi: 10.1006/abbi.1993.1024

Monod, J. (1965). [Reflections on the relationship between the structure and function of globular proteins]. *Annee Biol.* 59, 231–240.

Motlagh, H. N., Wrabl, J. O., Li, J., and Hilser, V. J. (2014). The ensemble nature of allostery. *Nature* 508, 331–339. doi: 10.1038/nature13001

Murzin, A. G. (2008). Metamorphic proteins. 320, 1725–1726. doi: 10.1126/science.1158868

Parisi, G., Zea, D. J., Monzon, A. M., and Marino-Buslje, C. (2015). Conformational diversity and the emergence of sequence signatures during evolution. *Curr. Opin. Struct. Biol.* 32, 58–65. doi: 10.1016/j.sbi.2015.02.005

Park, C. K., and Horton, N. C. (2019). Structures, functions, and mechanisms of filament forming enzymes: a renaissance of enzyme filamentation. *Biophys. Rev.* 11, 927–994. doi: 10.1007/s12551-019-00602-6

Park, C. K., and Horton, N. C. (2020). Novel insights into filament-forming enzymes. *Nat. Rev. Mol. Cell Biol.* 21, 1–2. doi: 10.1038/s41580-019-0188-1

Petrovich, R. M., Litwin, S., and Jaffe, E. K. (1996). Bradyrhizobium japonicum porphobilinogen synthase uses two Mg(II) and monovalent cations. *J. Biol. Chem.* 271, 8692–8699. doi: 10.1074/jbc.271.15.8692

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: an online force field. *Nucleic Acids Res.* 33, W382–W388.

Selwood, T., and Jaffe, E. K. (2012). Dynamic dissociating homo-oligomers and the control of protein function. *Arch Biochem. Biophys.* 519, 131–143. doi: 10.1016/j.abb.2011.11.020

Selwood, T., Tang, L., Lawrence, S. H., Anokhina, Y., and Jaffe, E. K. (2008). Kinetics and thermodynamics of the interchange of the morpheein forms of human porphobilinogen synthase. *Biochemistry* 47, 3245–3257. doi: 10.1021/bi702113z

Shanmugam, D., Wu, B., Ramirez, U., Jaffe, E. K., and Roos, D. S. (2010). Plastid-associated porphobilinogen synthase from *Toxoplasma gondii*: kinetic and structural properties validate therapeutic potential. *J. Biol. Chem.* 285, 22122–22131. doi: 10.1074/jbc.m110.107243

Simonet, J. C., Burrell, A. L., Kollman, J. M., and Peterson, J. R. (2020). Freedom of assembly: metabolic enzymes come together. *Mol. Biol. Cell* 31, 1201–1205. doi: 10.1091/mbc.e18-10-0675

Stolz, M., and Dornemann, D. (1996). Purification, metal cofactor, N-terminal sequence and subunit composition of a 5-aminolevulinic acid dehydratase from the unicellular green alga Scenedesmus obliquus, mutant C-2A'. *Eur. J. Biochem.* 236, 600–608. doi: 10.1111/j.1432-1033.1996.00600.x

Tang, L., Breinig, S., Stith, L., Mischel, A., Tannir, J., Kokona, B., et al. (2006). Single amino acid mutations alter the distribution of human porphobilinogen synthase quaternary structure isoforms (morpheeins). *J. Biol. Chem.* 281, 6682–6690. doi: 10.1074/jbc.m511134200

Thomas, W. C., Brooks, F. III, Burnim, A. A., Bacik, J., Stubbe, J., Kaelber, J. T., et al. (2019). Convergent allostery in ribonucleotide reductase. *Nat. Commun.* 10:2653.

Wasserman, H., and Saphire, E. O. (2016). More than meets the eye: hidden structures in the proteome. *Annu. Rev. Virol.* 3, 373–386. doi: 10.1146/annurev-virology-100114-054923

Wu, W. H., Shemin, D., Richards, K. E., and Williams, R. C. (1974). The quaternary structure of delta-aminolevulinic acid dehydratase from bovine liver. *Proc. Natl. Acad. Sci. U.S.A.* 71, 1767–1770. doi: 10.1073/pnas.71.5.1767

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

REVIEW
published: 18 December 2020
doi: 10.3389/fmolb.2020.596945

Surveying the Side-Chain Network Approach to Protein Structure and Dynamics: The SARS-CoV-2 Spike Protein as an Illustrative Case

Anushka Halder[1], Arinnia Anto[2], Varsha Subramanyan[3], Moitrayee Bhattacharyya[1], Smitha Vishveshwara[3]* and Saraswathi Vishveshwara[2]**

[1] Department of Pharmacology, Yale University, New Haven, CT, United States, [2] Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India, [3] Department of Physics, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Network theory-based approaches provide valuable insights into the variations in global structural connectivity between different dynamical states of proteins. Our objective is to review network-based analyses to elucidate such variations, especially in the context of subtle conformational changes. We present technical details of the construction and analyses of protein structure networks, encompassing both the non-covalent connectivity and dynamics. We examine the selection of optimal criteria for connectivity based on the physical concept of percolation. We highlight the advantages of using side-chain-based network metrics in contrast to backbone measurements. As an illustrative example, we apply the described network approach to investigate the global conformational changes between the closed and partially open states of the SARS-CoV-2 spike protein. These conformational changes in the spike protein is crucial for coronavirus entry and fusion into human cells. Our analysis reveals global structural reorientations between the two states of the spike protein despite small changes between the two states at the backbone level. We also observe some differences at strategic locations in the structures, correlating with their functions, asserting the advantages of the side-chain network analysis. Finally, we present a view of allostery as a subtle synergistic-global change between the ligand and the receptor, the incorporation of which would enhance drug design strategies.

Keywords: protein side-chain network, allostery, network theory, SARS-CoV-2 spike protein, conformational dynamics, network parameters, molecular dynamics simulations

OPEN ACCESS

Edited by:
Hongchun Li,
Shenzhen Institutes of Advanced Technology (CAS), China

Reviewed by:
Alessandro Giuliani,
National Institute of Health (ISS), Italy
Ozlem Tastan Bishop,
Rhodes University, South Africa

****Correspondence:***
Moitrayee Bhattacharyya
moitrayee.bhattacharyya@yale.edu
Smitha Vishveshwara
smivish@illinois.edu
Saraswathi Vishveshwara
saraswathi@iisc.ac.in

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *20 August 2020*
Accepted: *04 November 2020*
Published: *18 December 2020*

Citation:
Halder A, Anto A, Subramanyan V, Bhattacharyya M, Vishveshwara S and Vishveshwara S (2020) Surveying the Side-Chain Network Approach to Protein Structure and Dynamics: The SARS-CoV-2 Spike Protein as an Illustrative Case. Front. Mol. Biosci. 7:596945. doi: 10.3389/fmolb.2020.596945

INTRODUCTION

The concept of allostery has evolved for more than half a century (Monod et al., 1965; Koshland et al., 1966; Cooper and Dryden, 1984; Cui and Karplus, 2008; Changeux, 2011; Motlagh et al., 2014). This word in simple terms means "action at a distance" and implies long-distance communication within and across the three-dimensional structures of proteins. Fundamental understanding of the principles guiding allostery in proteins came from two classical models, the concerted Monod–Wyman–Changeux (MWC) (Monod et al., 1965) model and the sequential Koshland–Nemethy–Filmer (KNF) model (Koshland et al., 1966), with the structural insights coming from one of the earliest crystal

structures of hemoglobin (Perutz, 1970). An exponential increase in the availability of protein structures in different functional states has improved our comprehension of the phenomenon of allostery (Liu and Nussinov, 2016; Greener and Sternberg, 2018). Studies over the past decades have associated allostery in proteins with accompanying conformational variations. Such conformational changes range from dramatic alterations at the protein backbone level to subtle re-orchestrations involving protein side chains in the absence of appreciable backbone variations (Bhattacharyya and Vishveshwara, 2011; Motlagh et al., 2014; Tsai and Nussinov, 2014; Salamanca Viloria et al., 2017). It is this latter mode of conformational fluctuations and long-range signaling that are more challenging to capture.

Current advances in both experimental and theoretical techniques have started shedding light into these subtle conformational variations. In particular, long-range molecular dynamics (MD) simulations, providing equilibrium conformational ensembles, have offered extensive computational characterization of the conformational dynamics in proteins/protein complexes (Lindorff-Larsen et al., 2010, 2016; Karandur et al., 2020; Mysore et al., 2020). The goal of these studies ranges from understanding the fundamental biophysical principles to understanding more practical applications for drug design (Borhani and Shaw, 2012; de Vivo et al., 2016). Over the past decades, this topic has been extensively discussed in many excellent articles and reviews (Bagler and Sinha, 2005; Ghosh and Vishveshwara, 2007; Cui and Karplus, 2008; de Ruvo et al., 2012; Bhattacharyya et al., 2016; Astl et al., 2019; Zhang and Nussinov, 2019; Verkhivker et al., 2020; Zhang et al., 2020).

Network theory-based analyses of protein structure (Bagler and Sinha, 2005; Atilgan et al., 2012) and dynamics have provided unprecedented insights into the global structural connectivity of proteins and its complexes in the context of allostery and other biological processes (Atilgan et al., 2012; Di Paola et al., 2013; Bhattacharyya et al., 2016; Gadiyaram et al., 2019; Krieger et al., 2020; Verkhivker et al., 2020). When combined with information on conformational variations, as obtained from molecular dynamics, network approaches have elucidated several examples of protein structure–function relationships (Doruker et al., 2000; Sethi et al., 2009; Bhattacharyya et al., 2013; Papaleo, 2015; Tse and Verkhivker, 2015; Doshi et al., 2016).

In essence, it has become possible to obtain a better perception of biological phenomena at the molecular level, such as allostery, evolutionary effects, and transport phenomena, mediated through macromolecules, by employing two major concepts: (1) viewing macromolecules, such as proteins, as one single connected entity, where perturbations can affect the conformations at the local or global level, and (2) considering the dynamically accessible conformations of proteins, and the interconversion of their populations under different conditions as a key to biological functions. Regarding the concept of viewing proteins as a single unit, the connections at the non-covalent level play an important role since these are pliable for minor perturbations that are encountered at normal physiological conditions, unlike the covalently stitched polymer chains.

A number of approaches available in the literature (some of them referenced above) differ in how we view the protein structure as a single unit, connected through non-covalent interactions. One can focus on backbone connectivity alone, or connectivity at the level of side chains (explicit atoms or a representation through centroids) (Greene, 2012; Bhattacharyya et al., 2016; Kayikci et al., 2018). There are a number of ways to define the connectivity criteria and assign strengths of interactions. Similarly, the dynamical conformational landscape can be obtained at the explicit atomic level or indirectly achieved through methods like ENM, which provides cooperative modes of motion (Krieger et al., 2020; Zhang et al., 2020). The atomic level description can be obtained experimentally through biophysical techniques such as X-ray crystallography, cryo-EM, and NMR, and computationally through molecular dynamics (MD) simulations.

The identification of specific regions responsible for the overall perturbation and the reorganization of interactions to yield a different conformation in the landscape has received great attention. This is a crucial step in the process of making the connection between molecular events and protein functions. The methods range from direct analysis of the structures to ones developed based on the physical, mathematical, and engineering principles. Many such concepts are integrated together in computational programs to obtain critical biological insights (Greener and Sternberg, 2018; Verkhivker et al., 2020; Zhang et al., 2020). Network theory is a widely used approach which provides explicit information on the role of constituent amino acids on the stability of structure networks at a global level (Atilgan et al., 2010; Brown et al., 2017; Gadiyaram et al., 2018). A vast range of experimental and computational studies have taken up the challenge of correlating biological cellular functions to the molecular level changes. Understanding protein connectivity and dynamics can provide molecular mechanistic insights into the various biological processes, like allostery (Atilgan et al., 2007; Verkhivker et al., 2020; Wang et al., 2020a), protein–protein or protein–nucleic acid interactions (Brinda and Vishveshwara, 2005; Keskin et al., 2005; Sathyapriya et al., 2008; Sethi et al., 2009), and ligand/perturbation-induced conformational variations (Csermely et al., 2013; Bandaru et al., 2017; Creixell et al., 2018). Such calculations may also aid the identification of epitopes for drug-binding and capture drug-induced conformational changes in proteins and protein complexes (Csermely et al., 2013; Krieger et al., 2020).

The focus of this review is to provide a brief account of the different network theory-based techniques targeted at (i) characterizing protein structures as a single entity connected by non-covalent interactions and (ii) integrating with conformational dynamics, for which several comprehensive reviews are available (Atilgan et al., 2012; Hu et al., 2017; Verkhivker et al., 2020). The main emphasis here is on the development and application of protein side-chain network approaches (Bhattacharyya et al., 2016; Salamanca Viloria et al., 2017; Kayikci et al., 2018), which have been shown to capture subtle conformational differences that are sometimes elusive to conventional analyses, such as the root mean square deviation (RMSD) at the backbone level. Here, we have considered

the SARS-CoV-2 Spike glycoprotein (Zhu et al., 2020) as an illustrative example to demonstrate the capabilities of side-chain network studies. Our focus on analyzing SARS-CoV-2 spike protein, in particular, stems from its critical role in COVID-19 and the immediacy posed by the global pandemic caused by this highly infectious coronavirus.

In order to appreciate the relevance of side-chain network studies on the SARS-CoV-2 spike protein, here we provide an introduction to this protein in the context of its structure dynamics and function. SARS-CoV-2 belongs to the family of β-coronaviruses, and is closely related to the earlier pathogens, such as SARS-CoV and MERS-CoV, which caused severe respiratory diseases in 2004 and 2013, respectively. To develop promising therapeutic strategies, we need a clear understanding of the mechanism of action of the SARS-CoV-2 virus. A succinct summary of the structures of the SARS-CoV-2 spike protein and its interactions with the host cell membrane has been recently provided (Wang et al., 2020b; Xia et al., 2020a; Zhu et al., 2020). These studies highlight how the recognition of the human ACE2 receptor by the spike protein mediates viral entry into the host cell. A simplified version of the interaction between the human ACE2 receptor and the SARS-CoV-2 spike protein, with an emphasis on the structure of the spike protein, shows the steps that lead to viral fusion to the host cell membrane (**Figure 1**). Long-timescale MD simulations of the viral spike protein in different conformations have been recently made available under the Creative Commons Attribution 4.0 International Public License (D. E. Shaw Research., 2020). Further, a few computational studies on the SARS-CoV-2 spike protein, to explore putative allosteric binding sites (Di Paola et al., 2020) and the role of glycans (Casalino et al., 2020) have also been recently published.

The SARS-CoV-2 spike protein is a homotrimeric complex that is pivotal to the coronavirus entry into host cells and one of the key drug targets for COVID-19 (Hoffmann et al., 2020; Huang et al., 2020). In this article, we have selected the closed (PDB_ID: 6VXX) and partially open (PDB_ID: 6VYB) structures of the spike protein (Walls et al., 2020) as examples to explicitly elucidate the protein side-chain network concepts. Each subunit in the spike protein is organized into an S1 and S2 domain (**Figures 2A,B**) (Xia et al., 2020a; Zhu et al., 2020). The S1 domain hosts the receptor-binding domain (RBD) that recognizes the human ACE2 receptor (**Figures 2A,B**) and the N-terminal domain (NTD). In order to engage the ACE2 receptor, the RBD undergoes a conformational change much like the opening and closing of a hinge (**Figure 2C**). It is either in the receptor-inaccessible state (closed state) or receptor-accessible state (open state), governing access to the factors that control ACE2 binding (**Figure 2C**). The S2 domain hosts the TMPRSS2 cleavage site and the heptad repeat 1 and heptad repeat 2 (HR1/HR2) domains, which are the targets for fusion inhibitors (Xia et al., 2020a,b).

Backbone alignment of a closed structure (PDB_ID: 6VXX) and a partially open structure (PDB_ID: 6VYB) reveals small structural differences except that in the partially opened state the receptor-binding domain of one subunit swings outward as compared to the closed state (**Figure 2C**). This is an ideal model system to apply protein side-chain-based network calculations, as the observed backbone changes are small, but carries important functional information. The availability of long-scale MD simulation trajectories of the closed and partially open states of the spike protein (D. E. Shaw Research., 2020) further emboldened our choice of using the spike protein as our model system. This data allows us to demonstrate the capabilities of the dynamically stable protein side-chain network, correlating structural connectivity with conformational dynamics.

In the section *Protein Structure Network (PSN) Perspective Into Structural Organization*, an overview of network theory in the context of protein side-chain networks, connectivity criteria for the protein backbone (PBN) and the side-chain (PScN) networks, the selection of optimal strength of interaction from percolation theory perspective, and the cluster identification from graph spectral analysis are presented. In the section *Protein Structure Network (PScN) for Dynamically Accessible Conformational Ensembles*, the method for integrating network analysis with dynamically stable conformational landscapes is introduced. Additionally, the differences between the closed and open trimeric states of the SARS-CoV-2 spike protein are elucidated through chosen network metrics. This is followed by a Summary and Outlook section.

PROTEIN STRUCTURE NETWORK (PSN) PERSPECTIVE INTO STRUCTURAL ORGANIZATION

Network Theory-Based Representation of Protein Structures

The overall shape of protein structures at a molecular level is captured elegantly through secondary structures, such as helices, beta strands and sheets, and loops, formed by the backbone atoms of the polypeptide chain. Based on non-covalent interactions, the Ramachandran map characterized the allowed regions of the backbone torsion angles (φ,ψ) and demonstrated that the allowed conformational space of the polypeptide chain mainly consists of compact secondary and super-secondary structures, stabilized by backbone hydrogen bonds (Ramachandran et al., 1963). However, there are also numerous examples where despite insignificant differences at the protein backbone level, subtle conformational changes at the protein side-chain level guide a plethora of biological functions (Ghosh and Vishveshwara, 2007; Sethi et al., 2009). Such examples have motivated the development of techniques to completely map structural connectivity of proteins at both the backbone and side-chain interaction levels, enabling us to correlate even subtle structural variations that elude backbone-based alignments with biological functions.

The study based on the mathematical principles of network theory enables us to view the protein structures with non-covalent interactions as a single, global network. Numerous studies have formulated the backbone and the side-chain structural connectivity in proteins using adjacency matrices and analyzed them using various network metrics (del Sol et al., 2006; de Ruvo et al., 2012; Bhattacharyya et al., 2016; Kayikci

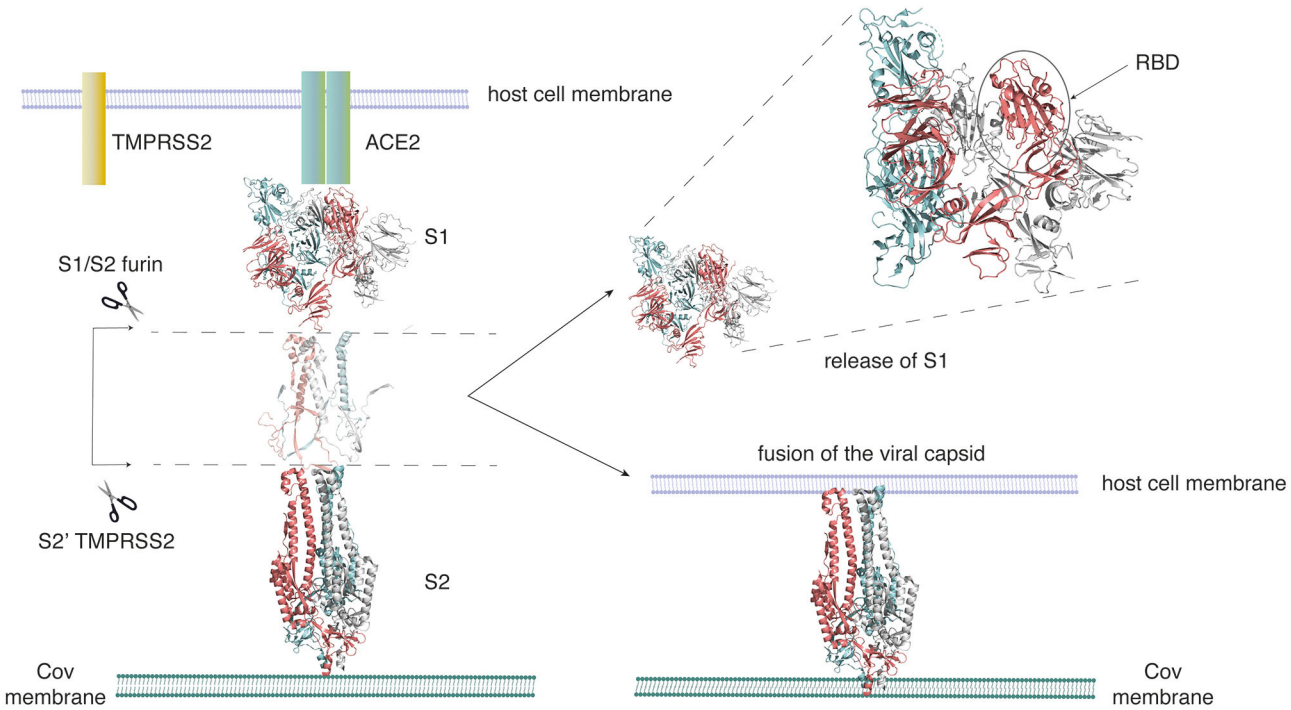


FIGURE 1 | SARS-CoV-2 employs the spike glycoprotein to enter its host cell. The spike protein is composed of two domains, the S1 domain that hosts the receptor-binding domain (RBD), and the S2 domain. The S2 domain arbitrates the fusion of the viral and host cell membranes. Activation of the spike protein happens by cleavage at two sites (S1/S2 and S2′) by the Furin TMPRSS2 protease. The spike protein initially binds to the ACE2 receptor on the host cell through its RBD. On activation, it sheds the S1 domain, enabling S2 to fuse to the host cell membrane. This figure was adapted from Structural and functional mechanism of SARS-CoV-2 | Abcam (https://www.abcam.com/content/structural-and-functional-mechanism-of-sars-cov-2-cell-entry).

et al., 2018; Astl et al., 2019; Krieger et al., 2020; Verkhivker et al., 2020). While protein backbone networks (PBN) capture the non-covalent connectivity at the level of the backbone atoms, protein side-chain networks (PScN/PSN) capture the structural connectivity at the level of non-covalent interactions between side-chain atoms. A representation of the global connectivity across the protein structure in terms of networks can capture the effect of perturbations at the local level and also across the entire protein structure network. This property is key to the understanding of how ligand or mutation-induced local conformational changes affect the global structure of a protein, and therefore its function.

An analysis of the network metrics from such a connectivity matrix allows the identification of allosteric communication pathways within a protein structure network by affording insights into the interconnected global architecture of proteins. A variety of network metrics can be used to analyze these protein structure networks (both PBN and PScN). The choice of the network metric being used depends on the question of interest. Briefly, metrics such as hubs and clustering coefficient indicate the degree of a residue and its connectivity to neighboring residues. The percolation behavior of a network can be captured in terms of clusters and cliques/communities (Palla et al., 2005; Deb et al., 2009). The molecular details of pathways responsible for allosteric signaling can be examined using the algorithms to identify shortest paths of communication (Ghosh et al., 2011; Tse and Verkhivker, 2015).

Protein Structure Network Based on Backbone (PBN) and the Side-Chain (PScN) Connectivity

Protein Structure Network refers to the representation of the three-dimensional structural connectivity in a protein in terms of a connectivity or adjacency matrix. In network theory language, individual amino acid residues are termed as nodes and the connections between them are defined as edges. In case of the protein backbone network (PBN), Cα atoms are generally considered as the representative of nodes and a distance of about 6.5 Å or less (based on the radial distribution of Cα atoms in protein structures) between any two sequentially non-adjacent residues are considered as an edge (Miyazawa and Jernigan, 1985; Patra and Vishveshwara, 2000). The construction and application of PBN have been extensively discussed in earlier reviews (Greene, 2012; Di Paola et al., 2015). Here, our focus is on the technical details of construction and the subsequent application of amino acid side-chain-based protein structure networks denoted as PScN (or PSN). There are different ways of measuring side-chain connectivity, such as the distance between the centroids of the side-chains, or all atom–atom pairwise

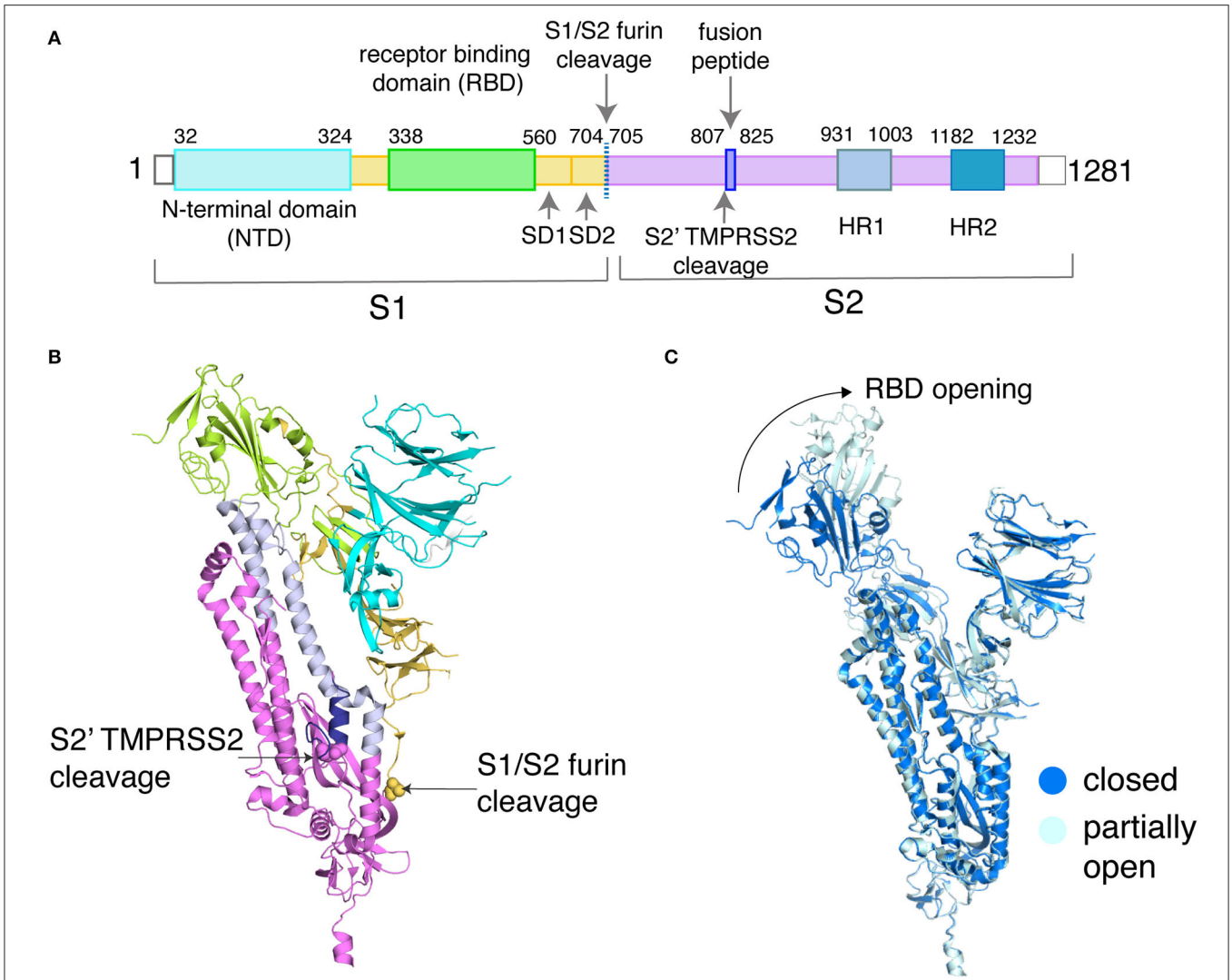


FIGURE 2 | Sequence and structural organization of the SARS-CoV-2 spike protein. **(A)** Domains of SARS-CoV-2 spike protein are depicted along with the two critical sites of cleavage. **(B)** These domains are shown on the structure of the spike protein (PDB_ID: 6VXX, only chain A is shown for clarity). The backbone is represented as a cartoon, and the domains are color-coded based on **(A)**. The first residue at the two cleavage sites is highlighted as spheres and labeled. **(C)** Backbone alignment (chain B only) of the closed and partially open states of the spike protein reveals conformational changes at the RBD (shown by the arrow), with the rest of the domains showing RMSD <0.5 Å (a low backbone RMSD of <0.5 Å is also observed for chain A/C between the closed and partially open states).

distances. Pairwise distances between all atoms of the side-chain of residues i and j (n_{ij}) and values within a distance of 4.5 Å (related to the sum of atomic radii; Heringa and Argos, 1991) capture explicit atomic-level connectivity. A normalization value of the total number of interactions (n_{ij}) with respect to the maximum values (N_i and N_j) observed from a large dataset of high-resolution crystal structures of proteins provides a uniform basis of evaluation as shown in Equation 1 (Kannan and Vishveshwara, 1999; Sathyapriya et al., 2008).

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} \times 100 \quad (1)$$

This expression allows us to weigh the strength of the interactions (edge weights) in a systematic manner, which can be uniformly applied to all protein structures. Edge weight (I_{ij}) can range from a value of zero to one. Values close to zero and one represent weak and strong side-chain connections, respectively. In general, strong connections can be related to nucleation centers formed by the interactions between the residue pairs such as oppositely charged, stacked aromatic residues, or polar residues involved in hydrogen bonds. The weak interactions, on the other hand, usually emerge from a smaller number of non-covalent interactions (n_{ij}) between pairs of hydrophobic amino acid residues. Generally, these interactions aid in bridging the strongly connected nucleation centers and in organizing the overall tertiary structure of the proteins. A PScN is constructed

based on a user-defined value of I_{ij} (termed I_{min}), and an edge is drawn when the calculated I_{ij} between a pair of residues exceeds I_{min}.

Percolation Profile for the Largest Connected Subnetwork as a Function of Edge Weight

To formalize the effect of the edge weight cutoff (I_{min}) on the properties of PScN for protein structures, the concept of the largest connected subnetwork (cluster or cliques/communities) transition profile was established and has been applied to a wide range of biological problems (Deb et al., 2009; Brinda et al., 2010). The PScN constructed from low values of I_{min} results in a dense matrix with a large fraction of the residues in the protein getting connected, yielding the largest cluster of the size ∼80–90% of the amino acid residues in the protein. The PScN at higher I_{min} values consists only of strongly connected edges, leading to a sparse matrix. The largest cluster in such a case does not cover a major fraction of the residues in the protein structure. On the other hand, in the largest cluster from a PScN of low I_{min}, although it encompasses a large fraction of residues, the ratio of signal/noise is low in this network. It is therefore important to identify an optimal I_{min} to construct the PScN, without losing information from a sparse graph or encountering low signal/noise from a dense graph.

The identification of the optimal I_{min} to construct the PScN has been addressed from the concepts of percolation within a system, as defined by percolation transition point. In these studies, the PScN is characterized by a macroscopically connected subnetwork obtained from I_{min}, around the transition point. The sizes of the largest cluster (L_{Clu}) or the largest clique/community (L_{Cli}) in the protein structure network are measured as a function of network connectivity at various I_{min} values. Plotting the values of L_{Clu} or L_{Cli} as a function of I_{min} leads to a sigmoidal curve. The transition point of this sigmoidal curve is identified as the percolation transition point at which a giant connected cluster still permeates the protein structure network. Interaction strength (I_{min}) around this transition point balances the problems of identifying non-specific, weak interactions at smaller I_{min} values, and discontinuous, sparse network connectivity across the structure at high I_{min} values. From earlier studies, it is shown that generally the transition point occurs in the range of I_{min} values 0.2 to 0.4 (Brinda et al., 2010). This transition point is noted to be a common feature in most protein structures.

In this study, we have analyzed the largest cluster percolation profile for the partially open and closed states of the SARS-CoV-2 spike protein. The plots of I_{min} vs. L_{Clu} are generated from the dynamically stable adjacency matrices (the generation of which is described in the section *Protein Structure Network (PScN) for Dynamically Accessible Conformational Ensembles*) corresponding to the closed (PDB_ID: 6VXX) and partially open structure (PDB_ID: 6VYB) of the spike protein (**Figure 3**). A noteworthy feature is that the profiles of the closed and partially open states show some differences in the percolation transition point. These differences are specifically located in the transition regions of I_{min} (between 0.2 and 0.3), with the closed state exhibiting a plateau of the LClu consisting of about 2000 residues, whereas, in the partially open state, the plateau is around a 1,000-residue cluster. Structurally, this is reminiscent of the conformational variations between the two states, with more residues held together as the largest cluster in the closed state, in comparison with the partially open state in the transition region. Based on earlier studies, as well as this one, we infer that the results obtained from the interaction strengths around the percolation transition point, I_{min} value of 0.3, are more sensitive and also provide a global view of the structural connectivity in proteins. In the analysis described in subsequent sections, we have used an I_{min} value of 0.3 to generate the networks.

Network Parameters of the PScN

With an increase in availability of data in various fields, the advancement in the research area of large-complex network studies has moved in different directions, such as problem and data-driven approaches, development of efficient algorithms, and availability of computational packages (Newman, 2001; Newman and Girvan, 2004; Palla et al., 2005). As we have seen above, the crucial input to obtain a solution to the graph is the connectivity or adjacency matrix in which the nodes and edges are defined based on the chosen application. The connectivity matrices (PBN/PScN) can be analyzed using well-established algorithms and network metrics, which can be used to describe various structural and functional properties of the protein. Some of the frequently used network metrics for analyzing protein/macromolecular structures are hub, clustering coefficient, cluster, cliques, and communities, and the shortest paths of communication rely on well-established mathematical algorithms (Dijkstra, 1959; Newman, 2001, 2004; Palla et al., 2005). Details of these individual parameters and their physical significance have been extensively discussed in past reviews. A brief description of the parameters and their physical significance follows.

Hubs represent highly connected residues in the protein structure network, which essentially refer to the degree of a node. In some general networks, the degree of certain nodes can be very large (Newman, 2001; Tsai et al., 2009). However, the degree of any residue in PScN generally does not exceed 10 due to the steric constraints. The hubs are key in maintaining structural stability and information flow in the protein structure network and are often termed as "hot spots" in the structure (Amitai et al., 2004; del Sol et al., 2006). ***Clusters***, commonly identified using the depth first search (DFS) algorithm (Cormen et al., 2001), are a set of residues that are connected in a way that the number of intra-cluster connections is higher than the number of intercluster connections, involving these residues. Clusters evaluated at different I_{min} values are used to predict the strength of connectivity within a network as well as to study interfacial interactions in protein complexes (Brinda and Vishveshwara, 2005).

Cliques are defined as completely connected subgraphs in a network such that every residue is connected to every other residue in this subgraph. An assemblage of cliques that share common edges are termed ***communities*** (Palla et al., 2005).

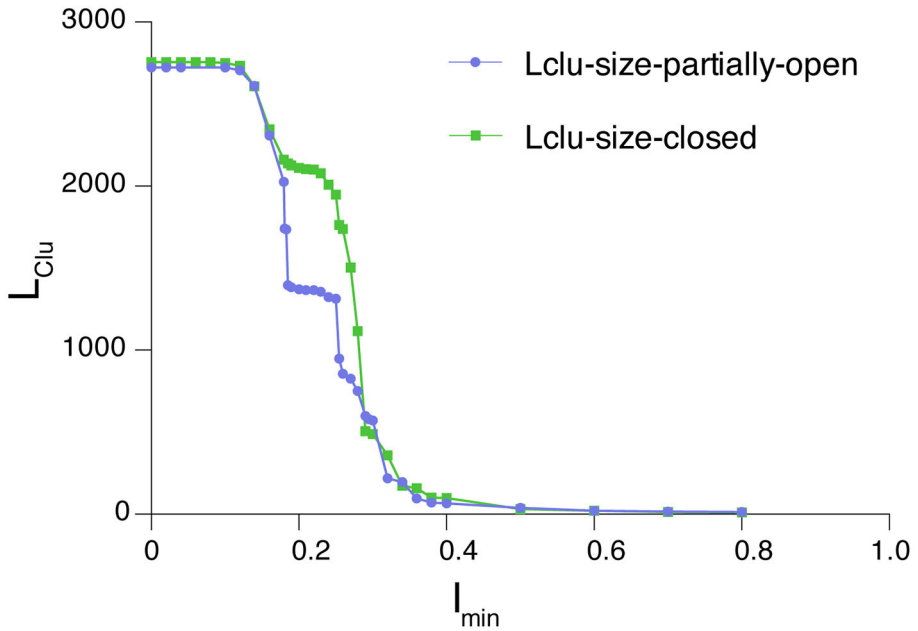


FIGURE 3 | Percolation profile for the largest connected cluster as a function of edge weight for the closed and partially open structures of the SARS-CoV-2 spike protein. The differences between these conformational states manifest as deviations in the connectivity profile in the region around the percolation point (0.2–0.3).

Cliques and communities are used to identify regions of rigidity and higher-order connectivity in protein structures (Ghosh and Vishveshwara, 2008). Like L_{Clu}, described in the section *Percolation Profile for the Largest Connected Subnetwork as a Function of Edge Weight*, the largest identified communities (L_{Cli}) can also provide insights into the percolation behavior of strongly connected components within a protein as a function of I_{min} (Deb et al., 2009). Floyd-Warshall and Dijkstra algorithms for computing the shortest paths of communication have been widely used to determine the critical residues involved in allosteric communication in proteins, and for mapping ligand-induced conformational changes (Atilgan et al., 2007; Bhattacharyya and Vishveshwara, 2011; Pandini et al., 2012). The specific choice of a network metric used for analyzing a protein structure network is therefore determined by the structural–biological insight we plan to seek. In the section *Protein Structure Network (PScN) for Dynamically Accessible Conformational Ensembles*, we will demonstrate the application of some of these parameters, by comparing the hubs and cliques/communities between the closed and partially open states of the SARS-CoV-2 spike protein, in their dynamical equilibrium states.

Graph Spectra of PScN

The graph spectral methods based on analyzing eigenvalues and eigenvector components of the connectivity matrices are another approach that has been extensively used to analyze protein structure networks (Hall, 1970). Graph spectral analysis on such a network is performed by studying the eigenspace of the Laplacian matrix associated with it. For a network with n nodes, the Laplacian L is an $n \times n$ matrix that satisfies equation 2.

$$X^T L X = \sum_{u \sim v} w_{uv} \left(x(u) - x(v) \right)^2 \tag{2}$$

where the summation is over every pair of nodes (u, v) connected by an edge with weight w_{uv} for some vector X in the space of nodes. It is shown that the Laplacian may be expressed in terms of the degree matrix D and the adjacency matrix A as equation 3 (Hall, 1970; Chung, 1997).

$$L_{uv} = D_{uv} - A_{uv} \tag{3}$$

The eigenvalues and eigenvectors of the Laplacian matrix contain information about the connected components or clusters of the network (Gadiyaram et al., 2016). The eigenvector corresponding to the lowest non-zero eigenvalue of the Laplacian, called the Fiedler vector, contains the clustering information. Sorting the Fiedler vector by value (FVC) (the components range from values -1 to $+1$) identifies nodes that are part of the same cluster. In this manner, all the clusters in the graph, ranging from the largest cluster with maximum number of residues to isolated edges, can be obtained as an analytical solution to the Laplacian matrix of the graph.

We have considered the example of the SARS-CoV-2 spike protein receptor-binding domain (RBD) (**Figures 2A,B**) to demonstrate the capability of the graph spectral analysis from the Laplacian matrix. Specifically, we have extracted the clusters from the sorted Fiedler vector of the receptor-binding domain (RBD) of the spike protein (PDB_ID: 6LZG), which is complexed with the target ACE2 receptor (Wang et al., 2020b). The adjacency

matrix is constructed as a binary matrix with $I_{ij} \geq 0.3$, with the elements taking values one and zero, respectively. A plot of sorted Fiedler components (FVC) as a function of the nodes (residue details given in **Supplementary Table S1**) is shown in **Figure 4A**. The slope of the FVC plot is also shown in this figure, which provides a clearer indication of cluster separation (Sistla et al., 2005). The clusters with the number of residues four and above are plotted on the structure of the RBD (**Figure 4B**). Thus, the graph spectral analysis is a powerful analytical method to extract the clustering information in protein structure networks.

It should be noted that there are limitations of performing graph spectral calculations, such as on large datasets (e.g., long MD simulation trajectories), as they are computationally expensive. However, the method provides unique information which is difficult to obtain directly from other methods. For instance, information can be extracted not only for clusters within a protein but also on the interfaces between domains in a single protein or across subunits in multimeric proteins (Brinda et al., 2005; Sistla et al., 2005). Graph spectral studies can also be performed on weighted networks, improving the accuracy. Further, it lends itself for quantitative comparison of networks by providing a score and allows us to identify the regions of the network which are dissimilar. A brief review of these aspects has been presented earlier (Gadiyaram et al., 2019).

PROTEIN STRUCTURE NETWORK (PScN) FOR DYNAMICALLY ACCESSIBLE CONFORMATIONAL ENSEMBLES

Biological systems exist in a dynamic equilibrium which alters under different conditions of temperature and ionic strengths, in complex with endogenous ligands/small molecules/drugs/interacting proteins, and so on. A glimpse of the accessible conformational landscape can be obtained by studying a large number of experimentally solved structures in different conditions or through long-timescale molecular dynamics (MD) simulations. The network properties that we described above for a single structure of proteins can also be extended to study the dynamically average properties of conformational ensembles. Depending on the objective, a judicious choice has to be made as to whether to get the averages from all the structural snapshots along the MD trajectory or from selected structures representing various local minima along the trajectory.

Dynamically Stable Protein Structure Network (PScN) for Conformational Ensembles

Analyses of protein conformational ensembles have been facilitated by the development of multiple open-source program packages (Eargle and Luthey-Schulten, 2012; Bhattacharyya et al., 2013; Chakrabarty and Parekh, 2016; Brown et al., 2017; Felline et al., 2020) that analyze multiple structural snapshots in dynamically accessed conformational states. The critical element in many of these open-source software is the ability to implement network theory-based calculations to analyze MD simulation trajectories. Some of these packages (PSN-Ensemble, webPSN v2.0, and NetworkView) also enable the use of residue pairwise interaction energies to weigh the connectivity matrix.

In this review, we discuss the general concepts of network theory-based analysis of protein conformational ensembles, specifically using PSN-Ensemble as an example software package. The basic principles and capabilities of PSN-Ensemble have been described before (Bhattacharyya et al., 2013). Briefly, PSN-Ensemble provides a consolidated and automated analysis platform, bridging network studies with protein conformational dynamics. Taking the coordinates of structural snapshots from conformational ensembles (MD simulations, NMR studies, or multiple crystal structures) as input, the program computes protein side-chain connectivity matrices (PScN). The individual matrices can be averaged by imposing a user-defined cutoff for dynamic stability (say X%). This "dynamically stable" PScN retains any interaction that appears in greater than X% of the structural ensemble, highlighting interactions that persist in a user-defined fraction of the ensemble.

Network parameters and metrics, as described in the section *Protein Structure Network (PSN) Perspective Into Structural Organization*, can be used to analyze the dynamically stable PScN. Using the dynamically stable matrix, PSN-Ensemble can compute structural hotspots (e.g., hubs/cliques) and analyze structural rigidity or flexibility (e.g., cliques/communities) (Ghosh and Vishveshwara, 2008; Bhattacharyya and Vishveshwara, 2011), percolation properties of the network (e.g., clusters and largest cluster transition profile) (Deb et al., 2009; Brinda et al., 2010), molecular determinants of allosteric signaling (e.g., shortest paths of communication) (Ghosh and Vishveshwara, 2007), and ligand/perturbation-induced conformational variations (e.g., hubs/cliques/communities) (Sukhwal et al., 2011; Creixell et al., 2018).

Here we provide an example of the application of network theory to analyze MD simulation trajectories. Using PSN-Ensemble, we analyze the long-timescale MD simulation trajectories (10 μs each) of SARS-CoV-2 spike protein in the closed and partially open states (D. E. Shaw Research., 2020). Based on the fact that the interaction strength around the percolation transition point is most sensitive while providing a global view of the structural connectivity (**Figure 3**), an I_{min} value of 0.3 is chosen to construct the PScN for the two states of the spike protein. Further, the dynamically stable PScN for the MD conformational ensemble is computed at a cutoff of 50%. We compared the hubs and cliques/communities from the dynamically stable PScN for the closed and partially open states of the spike protein. The results of these analyses on the spike protein are summarized in the subsequent sections as an example of network theory-based comparison of different conformational states.

Analysis of Dynamically Stable Metrics of PScN for the Closed and Partially Open States of the SARS-CoV-2 Spike Protein

In this section, we compare the side chain network properties related to rigidity/flexibility (hubs, cliques/communities) from

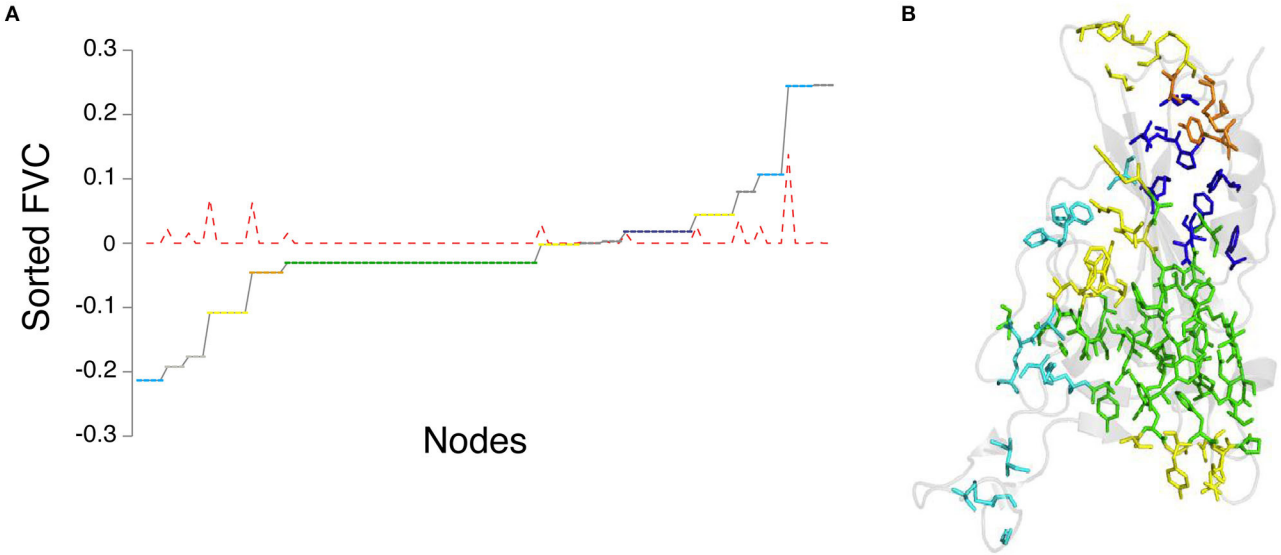


FIGURE 4 | Graph spectral analysis of RBD from SARS-CoV-2 spike protein. **(A)** Side-chain cluster plot of the RBD obtained from the sorted Feidler vector components (FVC-at lij = 0.3). Y-axis: sorted FVC; x-axis: nodes corresponding to FVC (node details in **Supplementary Table S1**). Nodes with identical FVC belong to a cluster. The red dotted line represents the slope of the FVC plot. The constant zero values of slope correspond to clusters, and spikes in the slope indicate the separation between clusters. **(B)** Representation of the side-chain clusters identified through FVC on the structure of RBD. The clusters are color-coded green (largest), blue (2nd largest), yellow (6 residue cluster), orange (5 residue cluster), and cyan (4 residue cluster).

the long-timescale MD trajectories on the closed and partially open states of the SARS-CoV-2 spike protein (10 μs each) (D. E. Shaw Research., 2020). In order to engage the host cell receptor (ACE2), the receptor-binding domain (RBD) of the spike protein undergoes conformational changes, much like the opening of a hinge (Walls et al., 2020). The closed state of the spike protein (PDB_ID: 6VXX) is receptor inaccessible. A partially open structure, with one of the subunits exhibiting RBD opening (PDB_ID: 6VYB), is representative of the receptor accessible states of the protein. Using PSN-Ensemble on the MD simulation trajectories, we analyzed the global conformational changes between these closed and partially open states.

The root mean square deviation resulting from a backbone alignment of the closed (PDB_ID: 6VXX) and partially open (PDB_ID:6VYB) structures of the spike protein is ∼0.5 Å. In addition, each subunit in the trimeric spike protein shows < ∼0.5 Å root mean square deviation when compared to each other, either within or between the two conformational states (closed and partially open). This indicates highly symmetric trimeric organization in the starting structures used for the long MD simulations, in terms of backbone superposition. The two states mainly differ in the conformations of the RBD in one subunit, with the rest of the domains relatively unchanged at the backbone level (**Figure 2C**) (Walls et al., 2020). Subtle conformational changes that differentiate these two states elude backbone-based structural comparisons, which cannot efficiently capture re-orientations at the protein side-chain level. These factors make the SARS-CoV-2 spike protein conformational states an ideal model system to highlight the benefits of side-chain-based network (PScN) analysis.

To compute the dynamically stable PScN, we extracted conformational snapshots every 100 ns from the two MD simulation trajectories of the spike protein in its closed and partially open states (D. E. Shaw Research., 2020). About 100 snapshots from each trajectory are used as an input to the software package PSN-Ensemble to calculate network metrics that persist in at least 50% of the structural ensemble. A comparison of the dynamically stable hub and cliques/communities between the closed and partially open states of the spike protein shows how the conformational change in the RBD leads to global structural rearrangements, which percolates into the membrane-binding domains of the spike protein. Through this highly relevant example, we reaffirm the advantages of using protein side-chain network-based calculations in capturing the changes in structural connectivity and conformational dynamics in proteins under different conditions of activity, ligand binding, environmental stimulus, and allosteric communication.

Comparison of Dynamically Stable Hubs Between the Closed and Partially Open States of Spike Protein

As mentioned in the section *Network Parameters of the PScN*, the residues that form four or more connections with other residues are defined as hubs and these are considered as dynamically stable, if they appear as hubs in at least 50% of the MD simulation snapshots. These hubs are considered structural hotspots. A change in the number and location of the dynamically stable hub residues in the closed and partially open states of spike protein captures the differences in structural connectivity between these two states. The three subunits in the

closed and partially open states of the trimeric SARS-CoV-2 spike protein show a total of 187 common dynamically stable hubs (**Supplementary Table S2**). The common hubs between the two states represent the structural connectivity in the PScN which remains unchanged between the two conformational states. The three subunits exhibit mostly symmetrical distribution of these common hubs, in terms of both number of hubs and the participating residues (**Supplementary Table S2**).

The distinctive structural features of the closed and partially open states are shown by the hubs that are unique to each conformational state. A comparison of these unique hubs reveals striking differences between the closed (64 unique hubs) and partially open states (74 unique hubs). The number and distribution of these dynamically stable unique hubs show large variations between corresponding subunits of the trimeric spike protein, in going from the closed to the partially open state (**Figure 5**, **Supplementary Table S2**). Strikingly, the number and distribution of the unique hubs among the three subunits within a particular conformational state also show significant differences. This indicates asymmetry between the three subunits of the spike protein, in both closed and partially open states. This asymmetry is exhibited at the side-chain network level, despite the highly symmetric nature of the spike protein structure in terms of backbone alignment of every subunit with every other subunit (RMSD < 0.5 Å).

Depiction of these unique hubs on each subunit of the respective conformational states reveals structural rewiring in the entire SARS-CoV-2 spike protein as the RBD goes from the closed to the open conformation. Here, we discuss the differences observed in chain A (detailed differences for all the three subunits are summarized in **Figure 5**). In chain A, the partially open state of the spike protein shows an increased number of hubs in the NTD as well as in the region connecting the HR1 and HR2 domains (**Figure 5**). The increased number of dynamically stable hubs suggests enhanced connectivity in these regions as the spike protein transitions into a partially open state. Our results also suggest that the conformational changes in RBD between the two states induce significant reorganization in the dynamically stable PScN. These global side-chain conformational changes are reflected as differences in the distribution of hubs (**Figure 5**), especially at sites distant from the RBD, despite minimal backbone reorganization between the two states.

Comparison of Dynamically Stable Cliques and Communities Between the Closed and Partially Open States of Spike Protein

Cliques represent a subset of residues within a protein structure network where each residue is connected to every other residue (Palla et al., 2005). Cliques represent higher-order connectivity in a network, highlighting regions of structural rigidity in the context of protein structures. An assemblage of cliques through shared edges/interactions is defined as communities (details in the section *Network Parameters of the PScN*). Communities capture the percolation of structural rigidity through the protein structure network. Together, comparison of cliques/communities reflects subtle conformational changes that alter regions of rigidity/flexibility in protein structural organization.

We compared the dynamically stable cliques and communities obtained from the SARS-CoV-2 trimeric spike protein in the closed and partially open states (**Figure 6**, **Supplementary Tables S3A–C**). The conformational changes that accompany the transition between the two states of the spike protein are reflected by the cliques/communities that are unique to each state. For clarity, we will only focus on the dynamically stable cliques/communities formed at the trimeric interface between the three subunits (also see **Supplementary Tables S3A–C** for a comparison of all cliques/communities between partially open and closed states of the spike protein).

Interfacial cliques/communities are an excellent metric to measure changes in connectivity or interaction between subunits for multimeric proteins. In the closed state of the spike protein, a large number of unique interfacial cliques are seen within the RBD of the three subunits (**Figure 6**, **Supplementary Table S4**). A total of 32 unique interfacial cliques are identified in the closed state, involving residues in the RBD, NTD, SD1, S2, and HR1 domains in the spike protein. This suggests a tightly packed trimeric interface in the closed state with rigid connections between the residues across the three subunits. In contrast, only 21 unique interfacial cliques are observed in the partially open state, with marked alterations in the domains participating in the cliques. Only eight common interfacial cliques are shared between the two conformational states, indicating significant variations in the trimeric interface packing.

Interestingly, most interfacial cliques formed by the RBD residues are lost in the partially open state. This, as expected, may be due to opening of the RBD in one of the subunits in the spike protein, which leads to a weakening of the interfacial connections involving the RBD residues across the trimer. Interestingly, this conformational change percolates to domains that are distant from the RBD, with cliques altering across the entire spike protein (**Figure 6**, **Supplementary Table S4**). A slight increase in the number of interfacial cliques is noted near the HR1/HR2 domain, especially in the region connection the cores of HR1 and HR2.

In this section, we have demonstrated the utility of side-chain network metrics like hubs, clusters, and cliques/communities by correlating the function of partial RBD opening to global conformational changes at the side-chain interaction level. We have shown that the local conformational changes at the RBD lead to extensive re-orchestration of the entire spike protein side-chain network.

SUMMARY AND OUTLOOK

The term "allostery" was coined more than half a century ago, to characterize the action of proteins away from the classically identified binding site (Monod et al., 1965; Koshland et al., 1966; Changeux, 2011). The mechanism of action was described through lock-and-key or induced fit models. Our understanding of the protein structure–function relationship has increased with advancement in structural biology. Today, there is an exponential increase of structural data from experiments such as X-ray crystallography, NMR, and Cryo Electron Microscopy

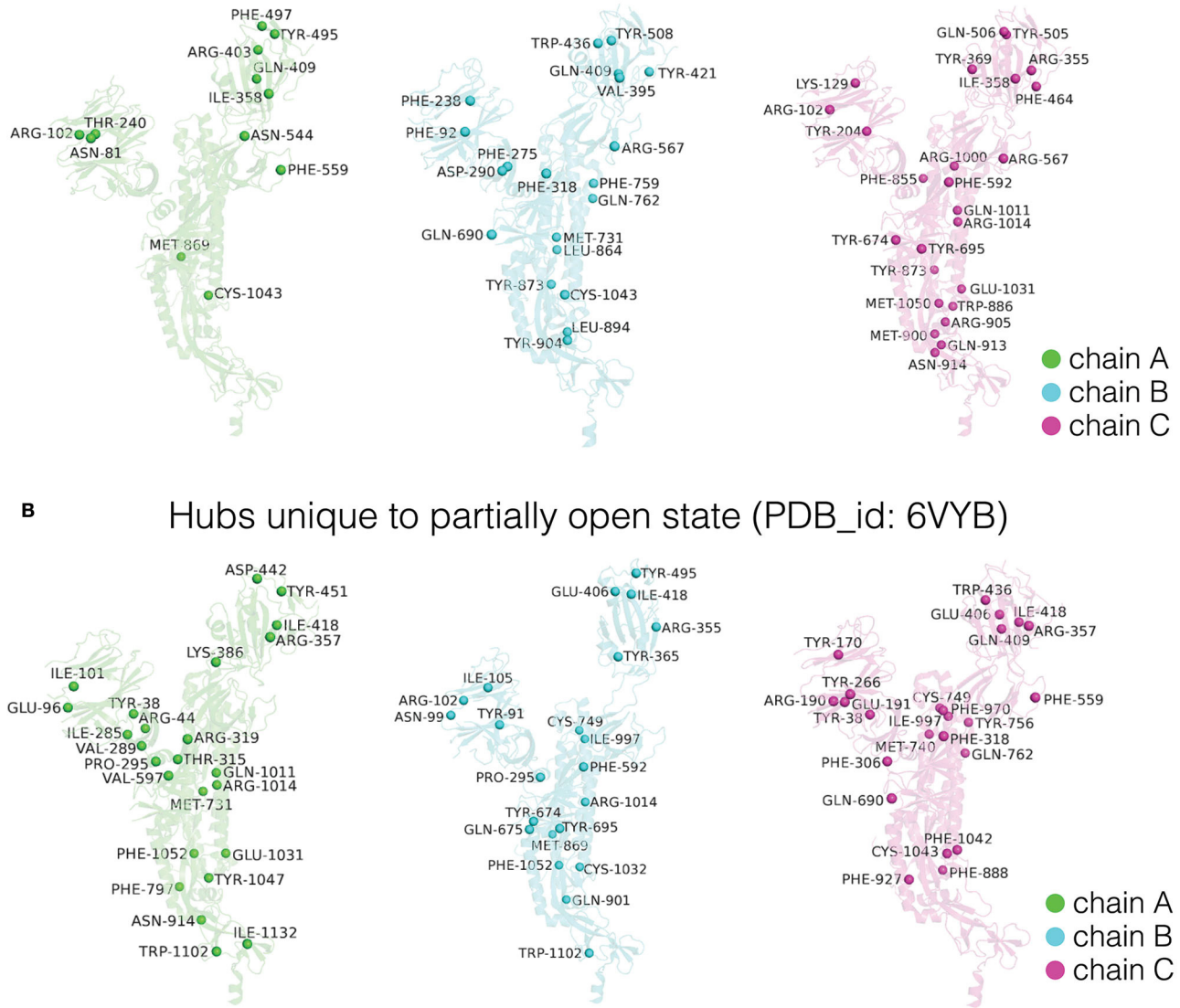


FIGURE 5 | Depiction of dynamically stable unique hub residues in the three subunits (Chain A/B/C) for the **(A)** closed and **(B)** partially open conformational states of the trimeric SARS-CoV-2 spike protein. The protein backbone is shown in cartoon representation, and each subunit is color coded. The unique hub residues in each subunit for the two conformational states are represented as spheres and these residues are labeled.

(Structural Biology Shapes Up | Science | AAAS., 2016; Nitta et al., 2018). In parallel, computational biology has reached the maturity to explore the conformational space of large protein assemblies through long timescale MD simulations (Lindorff-Larsen et al., 2016; Wang et al., 2019; Mysore et al., 2020).

The data from experimental structural biology and the MD simulations have become a rich source of information to investigate macromolecular systems in atomic details. Mining such valuable data for protein conformation and dynamics, in order to unravel biological function at a molecular level and provide meaningful and reliable predictions for experimental biologists, has been a challenge. This has led to multidisciplinary approaches and adaptation of different domain expertise to investigate the importance of specific amino acids toward the stability and functions of proteins from various perspectives. Some of the computational concepts and methods that have made their way to address biological systems are network theory, accessible modes, machine learning, and percolation phenomenon, in combination with highly valuable chemical and biological inputs.

Here we have presented a focused review of the protein structures from a network perspective. Specifically, we have focused on the networks of side-chain connectivity to highlight

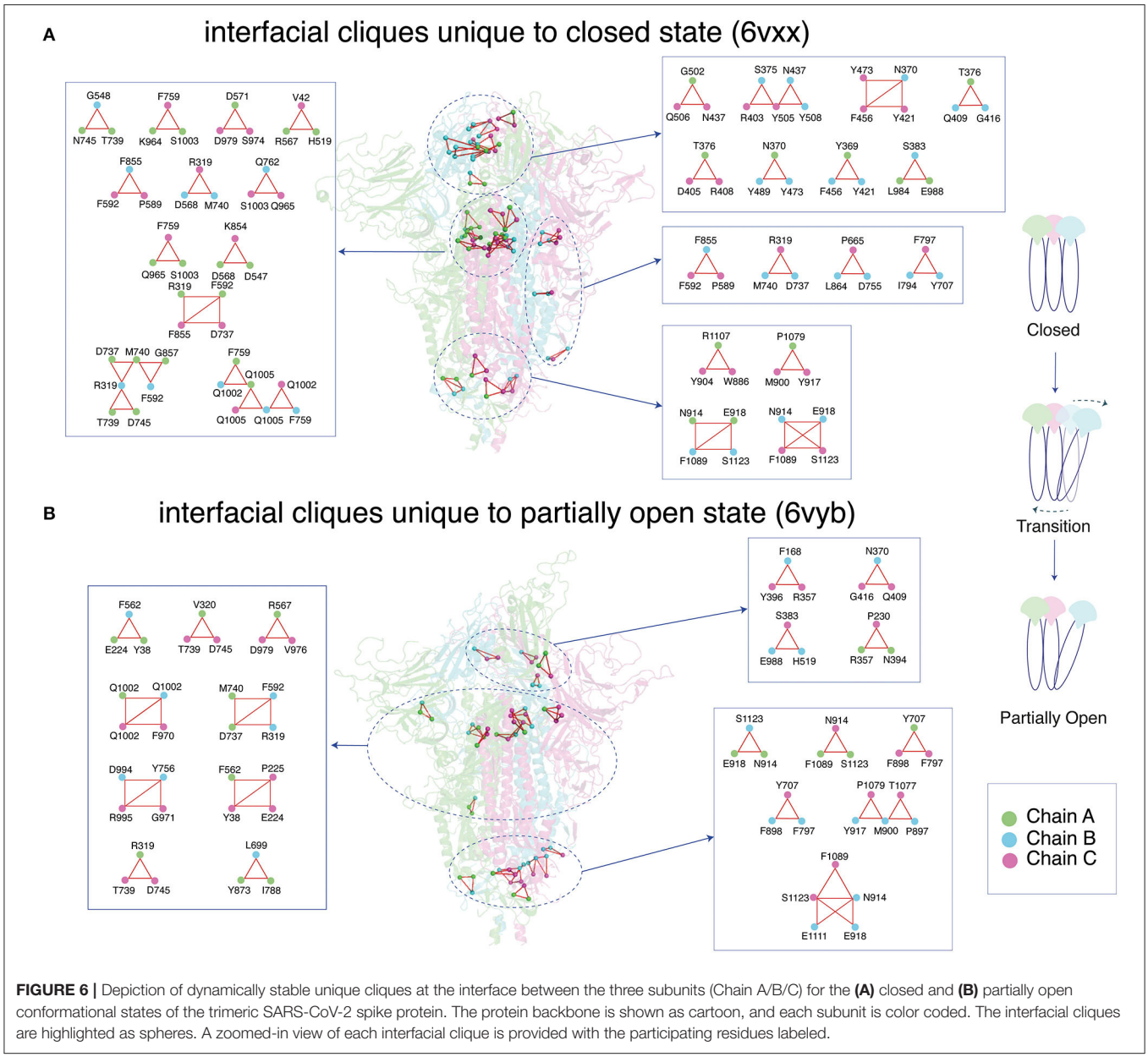


FIGURE 6 | Depiction of dynamically stable unique cliques at the interface between the three subunits (Chain A/B/C) for the **(A)** closed and **(B)** partially open conformational states of the trimeric SARS-CoV-2 spike protein. The protein backbone is shown as cartoon, and each subunit is color coded. The interfacial cliques are highlighted as spheres. A zoomed-in view of each interfacial clique is provided with the participating residues labeled.

the unique benefits of this approach. We have described the method of quantifying connectivity and identifying the optimal connectivity criteria by employing concepts from percolation theory. We have also discussed the global connectivity of the protein side-chains and clustering of interacting residues from the graph spectral perspective. We have pointed out that the highly similar backbone conformations of proteins can host a repertoire of conformational landscapes, which subtly differ in their side-chain interactions. Thus, mild perturbations to proteins can lead to side-chain reorganizations that elude backbone-based structural studies and drive allosteric communication. We have briefly touched upon a variety of approaches to investigate allostery, on which excellent recent reviews are available.

Molecular dynamics simulations can yield an ensemble of protein conformations, which can capture both the backbone and the side-chain level differences in interactions. Analysis of MD simulation trajectories using side-chain network formalism provides a global view of protein structural connectivity from a dynamic perspective. We have reviewed the methodology for such integration of MD simulation with network theory-based analyses.

Due to the global pandemic caused by the highly infectious COVID-19, we have chosen the SARS-CoV-2 spike protein as an example to illustrate the dynamic PScN perspective. We have investigated the molecular dynamics trajectories of the closed and partially open states of the trimeric spike protein that have been made available by D.E. Shaw Research (D. E.

Shaw Research., 2020). Backbone-based structural comparison between the closed and partially open states reveals minimal structural changes. Highlighting the importance of side-chain network analyses, a dynamic PScN-based comparison reveals key differences between the two conformational states. The present investigation highlights the differences at the side-chain interaction level, between the two states such as (1) the differences in the size of the largest connected clusters (L_{Clu}) in the percolation transition region, with the closed state being more stable than the open state, and (2) the differences in the network parameters such as hubs, cliques, and communities.

A comparison of the network properties of the partially open and closed forms of the SARS-CoV-2 spike protein reaffirms that different functional states of proteins can adopt very close backbone topology. While substantial side-chain network parameters like hubs, cliques, and communities are also common to both the forms, the unique ones are strategically located in various parts of this multimeric protein. For example, the local conformational changes during the RBD opening lead to extensive re-orchestration of the entire spike protein network, more pronounced in the interfacial region of the trimeric contacts. The different functional states are carefully balanced through the re-organization of side-chain connectivity to mediate interactions with the ACE2 receptor, and ultimately viral fusion to host cell membrane. A detailed study of these interactions between the SARS-CoV-2 spike protein and ACE2 receptor or relevant antibodies/drugs from a side-chain network perspective will be the subject of future investigations.

In addition to offering insights into the structure–function correlation in proteins at the side-chain connectivity level, the dynamic network-based studies also provide a new perspective of allostery. The flexibility of the protein involved in interactions with ligand/drug or other proteins is as important as their interacting partners, to have a productive signaling output. Allostery should be viewed as a synergistic–global interaction between the ligand (or the environment) and the receptor. The mechanism of long-distance communication involves specific routes and subtle changes in the communication paths, in order to signal at a distance. Analysis of PScN reveals allosteric communication paths via side-chain interactions even without substantial backbone reorganization. A stimulus at the ligand-binding pocket may be transmitted to the desired destination through subtle reorganization of the side-chain interactions that are allowed in the equilibrium dynamical state. Comparison between the two states of the SARS-CoV-2 spike protein reveals significant changes in the hubs and cliques/communities in regions distant from the RBD. The global reorganization of the side-chain connectivity between the two states of the spike protein could also influence the communication paths within and across proteins. Thus, one can consider the conformational landscape as being made up of various side-chain network paths.

Finally, in the context of treatment of infections, the antibodies and vaccines are produced in response to the global topology of the host protein or receptor. They complement the naturally evolved receptor more globally around the binding sites. The drugs, on the other hand, which are designed based mainly on the binding site information, may not be highly effective. As we have seen here, the binding site residues are held loosely or tightly by the residue clusters, firmly anchoring some of the interacting residues deep within the pocket. The drug-development strategies would benefit by incorporating the side-chain network connectivity information into their design, thus providing a rationale for incorporating the effects of variations in global structural connectivity in proteins.

AUTHOR CONTRIBUTIONS

AA, AH, and VS worked on data collection, running the programs, and writing codes, participated in data analysis, and assisted in manuscript writing. MB, SmV, and SaV designed the article concept, analyzed the data, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

SaV and AA thank NASI for the fellowships. SmV and VS acknowledge the Institute for Condensed Matter Theory at the University of Illinois at Urbana-Champaign.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.596945/full#supplementary-material

REFERENCES

Amitai, G., Shemesh, A., Sitbon, E., Shklar, M., Netanely, D., Venger, I., et al. (2004). Network analysis of protein structures identifies functional residues. *J. Mol. Biol.* 344, 1135–1146. doi: 10.1016/j.jmb.2004.10.055

Astl, L., Tse, A., and Verkhivker, G. M. (2019). Interrogating regulatory mechanisms in signaling proteins by allosteric inhibitors and activators: a dynamic view through the lens of residue interaction networks. *Adv. Exp. Med. Biol.* 1163, 187–223. doi: 10.1007/978-981-13-8719-7_9

Atilgan, A. R., Turgut, D., and Atilgan, C. (2007). Screened nonbonded interactions in native proteins manipulate optimal paths for robust residue communication. *Biophys. J.* 92, 3052–3062. doi: 10.1529/biophysj.106.099440

Atilgan, C., Gerek, Z. N., Ozkan, S. B., and Atilgan, A. R. (2010). Manipulation of conformational change in proteins by single-residue perturbations. *Biophys. J.* 99, 933–943. doi: 10.1016/j.bpj.2010.05.020

Atilgan, C., Okan, O. B., and Atilgan, A. R. (2012). Network-based models as tools hinting at nonevident protein functionality. *Annu. Rev. Biophys.* 41, 205–225. doi: 10.1146/annurev-biophys-050511-102305

Bagler, G., and Sinha, S. (2005). Network properties of protein structures. *Phys. A Stat. Mech. Appl.* 346, 27–33. doi: 10.1016/j.physa.2004.08.046

Bandaru, P., Shah, N. H., Bhattacharyya, M., Barton, J. P., Kondo, Y., Cofsky, J. C., et al. (2017). Deconstruction of the Ras switching cycle through saturation mutagenesis. *Elife* 6:e27810. doi: 10.7554/eLife.27810.040

Bhattacharyya, M., Bhat, C. R., and Vishveshwara, S. (2013). An automated approach to network features of protein structure ensembles. *Protein Sci.* 22, 1399–1416. doi: 10.1002/pro.2333

Bhattacharyya, M., Ghosh, S., and Vishveshwara, S. (2016). Protein structure and function: looking through the network of side-chain interactions. *Curr. Protein Pept. Sci.* 17, 4–25. doi: 10.2174/1389203716666150923105727

Bhattacharyya, M., and Vishveshwara, S. (2011). Probing the allosteric mechanism in pyrrolysyl-tRNA synthetase using energy-weighted network formalism. *Biochemistry* 50, 6225–6236. doi: 10.1021/bi200306u

Borhani, D. W., and Shaw, D. E. (2012). The future of molecular dynamics simulations in drug discovery. *J. Comput. Aided Mol. Des.* 26, 15–26. doi: 10.1007/s10822-011-9517-y

Brinda, K. V., Surolia, A., and Vishveshwara, S. (2005). Insights into the quaternary association of proteins through structure graphs: a case study of lectins. *Biochem. J.* 391, 1–15. doi: 10.1042/BJ20050434

Brinda, K. V., and Vishveshwara, S. (2005). Oligomeric protein structure networks: insights into protein-protein interactions. *BMC Bioinformatics* 6:296. doi: 10.1186/1471-2105-6-296

Brinda, K. V., Vishveshwara, S., and Vishveshwara, S. (2010). Random network behaviour of protein structures. *Mol. Biosyst.* 6, 391–398. doi: 10.1039/B903019K

Brown, D. K., Penkler, D. L., Sheik Amamuddy, O., Ross, C., Atilgan, A. R., Atilgan, C., et al. (2017). MD-TASK: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics* 33, 2768–2771. doi: 10.1093/bioinformatics/btx349

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent. Sci.* 6, 1722–1734. doi: 10.1021/acscentsci.0c01056

Chakrabarty, B., and Parekh, N. (2016). NAPS: network analysis of protein structures. *Nucleic Acids Res.* 44, W375–W382. doi: 10.1093/nar/gkw383

Changeux, J.-P. (2011). 50th anniversary of the word "allosteric". *Protein Sci.* 20, 1119–1124. doi: 10.1002/pro.658

Chung, F. R. K. (1997). *Spectral Graph Theory, 2nd Edn*. Providence, RI: American Mathematical Society.

Cooper, A., and Dryden, D. T. (1984). Allostery without conformational change. A plausible model. *Eur. Biophys. J.* 11, 103–109. doi: 10.1007/BF00276625

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2001). "Depth first search," in *Introduction to Algorithms, 2nd edn*, (MIT Press, McGraw-Hill), 540–549.

Creixell, P., Pandey, J. P., Palmeri, A., Bhattacharyya, M., Creixell, M., Ranganathan, R., et al. (2018). Hierarchical organization endows the kinase domain with regulatory plasticity. *Cell Syst.* 7, 371–383.e4. doi: 10.1016/j.cels.2018.08.008

Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G., and Nussinov, R. (2013). Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* 138, 333–408. doi: 10.1016/j.pharmthera.2013.01.016

Cui, Q., and Karplus, M. (2008). Allostery and cooperativity revisited. *Protein Sci.* 17, 1295–1307. doi: 10.1110/ps.03259908

de Ruvo, M., Giuliani, A., Paci, P., Santoni, D., and Di Paola, L. (2012). Shedding light on protein-ligand binding by graph theory: the topological nature of allostery. *Biophys. Chem. 165–166*, 21–29. doi: 10.1016/j.bpc.2012.03.001

de Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.* 59, 4035–4061. doi: 10.1021/acs.jmedchem.5b01684

Deb, D., Vishveshwara, S., and Vishveshwara, S. (2009). Understanding protein structure from a percolation perspective. *Biophys. J.* 97, 1787–1794. doi: 10.1016/j.bpj.2009.07.016

del Sol, A., Fujihashi, H., Amoros, D., and Nussinov, R. (2006). Residue centrality, functionally important residues, and active site shape: analysis of enzyme and non-enzyme families. *Protein Sci.* 15, 2120–2128. doi: 10.1110/ps.062249106

D. E. Shaw Research. (2020). *Molecular Dynamics Simulations Related to SARS-CoV-2*. D. E. Shaw Research Technical Data. Available online at: http://www.deshawresearch.com/resources_sarscov2.html (accessed April 8, 2020).

Di Paola, L., de Ruvo, M., Paci, P., Santoni, D., and Giuliani, A. (2013). Protein contact networks: an emerging paradigm in chemistry. *Chem. Rev.* 113, 1598–1613. doi: 10.1021/cr3002356

Di Paola, L., Platania, C. B. M., Oliva, G., Setola, R., Pascucci, F., and Giuliani, A. (2015). Characterization of protein-protein interfaces through a protein contact network approach. *Front. Bioeng. Biotechnol.* 3:170. doi: 10.3389/fbioe.2015.00170

Di Paola, L., Hadi-Alijanvand, H., Song, X., Hu, G., and Giuliani, A. (2020). The discovery of a putative allosteric site in the SARS-CoV-2 spike protein using an integrated structural/dynamic approach. *J. Proteome Res.* 19, 4576–4586. doi: 10.1021/acs.jproteome.0c00273

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Num. Math.* 1, 269–271. doi: 10.1007/BF01386390

Doruker, P., Atilgan, A. R., and Bahar, I. (2000). Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins* 40, 512–524. doi: 10.1002/1097-0134(20000815)40:3<512::AID-PROT180>3.0.CO;2-M

Doshi, U., Holliday, M. J., Eisenmesser, E. Z., and Hamelberg, D. (2016). Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4735–4740. doi: 10.1073/pnas.1523573113

Eargle, J., and Luthey-Schulten, Z. (2012). NetworkView: 3D display and analysis of protein·RNA interaction networks. *Bioinformatics* 28, 3000–3001. doi: 10.1093/bioinformatics/bts546

Felline, A., Seeber, M., and Fanelli, F. (2020). webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules. *Nucleic Acids Res.* 48, W94–W103. doi: 10.1093/nar/gkaa397

Gadiyaram, V., Ghosh, S., and Vishveshwara, S. (2016). A graph spectral-based scoring scheme for network comparison. *J. Complex Netw.* 5, 219–244. doi: 10.1093/comnet/cnw016

Gadiyaram, V., Dighe, A., and Vishveshwara, S. (2018). Identification of crucial elements for network integrity: a perturbation approach through graph spectral method. *Int. J. Adv. Eng. Sci. Appl. Math.* 11, 91–104. doi: 10.1007/s12572-018-0236-7

Gadiyaram, V., Vishveshwara, S., and Vishveshwara, S. (2019). From quantum chemistry to networks in biology: a graph spectral approach to protein structure analyses. *J. Chem. Inf. Model.* 59, 1715–1727. doi: 10.1021/acs.jcim.9b00002

Ghosh, A., Sakaguchi, R., Liu, C., Vishveshwara, S., and Hou, Y.-M. (2011). Allosteric communication in cysteinyl tRNA synthetase: a network of direct and indirect readout. *J. Biol. Chem.* 286, 37721–37731. doi: 10.1074/jbc.M111.246702

Ghosh, A., and Vishveshwara, S. (2007). A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15711–15716. doi: 10.1073/pnas.0704459104

Ghosh, A., and Vishveshwara, S. (2008). Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes. *Biochemistry* 47, 11398–11407. doi: 10.1021/bi8007559

Greene, L. H. (2012). Protein structure networks. *Brief. Funct. Genomics* 11, 469–478. doi: 10.1093/bfgp/els039

Greener, J. G., and Sternberg, M. J. (2018). Structure-based prediction of protein allostery. *Curr. Opin. Struct. Biol.* 50, 1–8. doi: 10.1016/j.sbi.2017.10.002

Hall, K. M. (1970). An *r* -dimensional quadratic placement algorithm. *Manage. Sci.* 17, 219–229. doi: 10.1287/mnsc.17.3.219

Heringa, J., and Argos, P. (1991). Side-chain clusters in protein structures and their role in protein folding. *J. Mol. Biol.* 220, 151–171. doi: 10.1016/0022-2836(91)90388-M

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052

Hu, G., Di Paola, L., Pullara, F., Liang, Z., and Nookaew, I. (2017). Network proteomics: from protein structure to protein-protein interaction. *Biomed Res. Int.* 2017:8929613. doi: 10.1155/2017/8929613

Huang, Y., Yang, C., Xu, X.-F., Xu, W., and Liu, S.-W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* 41, 1141–1149. doi: 10.1038/s41401-020-0485-4

Kannan, N., and Vishveshwara, S. (1999). Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* 292, 441–464. doi: 10.1006/jmbi.1999.3058

Karandur, D., Bhattacharyya, M., Xia, Z., Lee, Y. K., Muratcioglu, S., McAffee, D., et al. (2020). Breakage of the oligomeric CaMKII hub by the regulatory segment of the kinase. *Elife* 9:e57784. doi: 10.7554/eLife.57784.sa2

Kayikci, M., Venkatakrishnan, A. J., Scott-Brown, J., Ravarani, C. N. J., Flock, T., and Babu, M. M. (2018). Visualization and analysis of non-covalent contacts using the protein contacts Atlas. *Nat. Struct. Mol. Biol.* 25, 185–194. doi: 10.1038/s41594-017-0019-z

Keskin, O., Ma, B., and Nussinov, R. (2005). Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* 345, 1281–1294. doi: 10.1016/j.jmb.2004.10.077

Koshland, D. E., Némethy, G., and Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits*. *Biochemistry* 5, 365–385. doi: 10.1021/bi00865a047

Krieger, J. M., Doruker, P., Scott, A. L., Perahia, D., and Bahar, I. (2020). Towards gaining sight of multiscale events: utilizing network models and normal modes in hybrid methods. *Curr. Opin. Struct. Biol.* 64, 34–41. doi: 10.1016/j.sbi.2020.05.013

Lindorff-Larsen, K., Maragakis, P., Piana, S., and Shaw, D. E. (2016). Picosecond to millisecond structural dynamics in human ubiquitin. *J. Phys. Chem. B* 120, 8313–8320. doi: 10.1021/acs.jpcb.6b02024

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–1958. doi: 10.1002/prot.22711

Liu, J., and Nussinov, R. (2016). Allostery: an overview of its history, concepts, methods, and applications. *PLoS Comput. Biol.* 12:e1004966. doi: 10.1371/journal.pcbi.1004966

Miyazawa, S., and Jernigan, R. L. (1985). Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 18, 534–552. doi: 10.1021/ma00145a039

Monod, J., Wyman, J., and Changeux, J. P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12, 88–118. doi: 10.1016/S0022-2836(65)80285-6

Motlagh, H. N., Wrabl, J. O., Li, J., and Hilser, V. J. (2014). The ensemble nature of allostery. *Nature* 508, 331–339. doi: 10.1038/nature13001

Mysore, V. P., Zhou, Z.-W., Ambrogio, C., Li, L., Kapp, J. N., Lu, C., et al. (2020). A structural model of a Ras-Raf signalosome. *bioRxiv*. doi: 10.1101/2020.07.15.165266

Newman, M. E. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* 64:016132. doi: 10.1103/PhysRevE.64.016132

Newman, M. E. J. (2004). Detecting community structure in networks. *Eur. Phys. J. B Condens. Matter* 38, 321–330. doi: 10.1140/epjb/e2004-00124-y

Newman, M. E. J., and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E, Stat. Nonlin. Soft. Matter. Phys.* 69:026113. doi: 10.1103/PhysRevE.69.026113

Nitta, R., Imasaki, T., and Nitta, E. (2018). Recent progress in structural biology: lessons from our research history. *Syst. Sex. Disord. Microscopy* 67, 187–195. doi: 10.1093/jmicro/dfy022

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 814–818. doi: 10.1038/nature03607

Pandini, A., Fornili, A., Fraternali, F., and Kleinjung, J. (2012). Detection of allosteric signal transmission by information-theoretic analysis of protein dynamics. *FASEB J.* 26, 868–881. doi: 10.1096/fj.11-190868

Papaleo, E. (2015). Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics: strength in unity. *Front. Mol. Biosci.* 2:28. doi: 10.3389/fmolb.2015.00028

Patra, S. M., and Vishveshwara, S. (2000). Backbone cluster identification in proteins by a graph theoretical method. *Biophys. Chem.* 84, 13–25. doi: 10.1016/S0301-4622(99)00134-9

Perutz, M. F. (1970). Stereochemistry of cooperative effects in haemoglobin. *Nature* 228, 726–739. doi: 10.1038/228726a0

Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95–99. doi: 10.1016/S0022-2836(63)80023-6

Salamanca Viloria, J., Allega, M. F., Lambrughi, M., and Papaleo, E. (2017). An optimal distance cutoff for contact-based protein structure networks using side-chain centers of mass. *Sci. Rep.* 7:2838. doi: 10.1038/s41598-017-01498-6

Sathyapriya, R., Vijayabaskar, M. S., and Vishveshwara, S. (2008). Insights into protein-DNA interactions through structure network analysis. *PLoS Comput. Biol.* 4:e1000170. doi: 10.1371/journal.pcbi.1000170

Sethi, A., Eargle, J., Black, A. A., and Luthey-Schulten, Z. (2009). Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6620–6625. doi: 10.1073/pnas.0810961106

Sistla, R. K., Brinda, K. V., and Vishveshwara, S. (2005). Identification of domains and domain interface residues in multidomain proteins from graph spectral method. *Proteins* 59, 616–626. doi: 10.1002/prot.20444

Structural Biology Shapes Up | Science | AAAS. (2016). Available online at: https://www.sciencemag.org/features/2016/07/structural-biology-shapes (accessed October 13, 2020).

Sukhwal, A., Bhattacharyya, M., and Vishveshwara, S. (2011). Network approach for capturing ligand-induced subtle global changes in protein structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 429–439. doi: 10.1107/S0907444911007062

Tsai, C.-J., Ma, B., and Nussinov, R. (2009). Protein-protein interaction networks: how can a hub protein bind so many different partners? *Trends Biochem. Sci.* 34, 594–600. doi: 10.1016/j.tibs.2009.07.007

Tsai, C.-J., and Nussinov, R. (2014). A unified view of "how allostery works". *PLoS Comput. Biol.* 10:e1003394. doi: 10.1371/journal.pcbi.1003394

Tse, A., and Verkhivker, G. M. (2015). Molecular dynamics simulations and structural network analysis of c-Abl and c-Src kinase core proteins: capturing allosteric mechanisms and communication pathways from residue centrality. *J. Chem. Inf. Model.* 55, 1645–1662. doi: 10.1021/acs.jcim.5b00240

Verkhivker, G. M., Agajanian, S., Hu, G., and Tao, P. (2020). Allosteric regulation at the crossroads of new technologies: multiscale modeling, networks, and machine learning. *Front. Mol. Biosci.* 7:136. doi: 10.3389/fmolb.2020.00136

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 281–292.e6. doi: 10.1016/j.cell.2020.02.058

Wang, J., Jain, A., McDonald, L. R., Gambogi, C., Lee, A. L., and Dokholyan, N. V. (2020a). Mapping allosteric communications within individual proteins. *Nat. Commun.* 11:3862. doi: 10.1038/s41467-020-17618-2

Wang, Q., Pechersky, Y., Sagawa, S., Pan, A. C., and Shaw, D. E. (2019). Structural mechanism for Bruton's tyrosine kinase activation at the cell membrane. *Proc. Natl. Acad. Sci. U.S.A.* 116, 9390–9399. doi: 10.1073/pnas.1819301116

Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., et al. (2020b). Structural and functional basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* 181, 894–904.e9. doi: 10.1016/j.cell.2020.03.045

Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., et al. (2020a). Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* 30, 343–355. doi: 10.1038/s41422-020-0305-x

Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., et al. (2020b). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol. Immunol.* 17, 765–767. doi: 10.1038/s41423-020-0374-2

Zhang, J., and Nussinov, R. (2019). *Protein Allostery in Drug Discovery*. Singapore: Springer Singapore.

Zhang, Y., Doruker, P., Kaynak, B., Zhang, S., Krieger, J., Li, H., et al. (2020). Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Curr. Opin. Struct. Biol.* 62, 14–21. doi: 10.1016/j.sbi.2019.11.002

Zhu, G., Zhu, C., Zhu, Y., and Sun, F. (2020). Minireview of progress in the structural study of SARS-CoV-2 proteins. *Curr. Res. Microb. Sci.* 1, 53–61. doi: 10.1016/j.crmicr.2020.06.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 03 February 2021
doi: 10.3389/fmolb.2020.607323

Computational Ways to Enhance Protein Inhibitor Design

Robert L. Jernigan[1]*, *Kannan Sankar*[1], *Kejue Jia*[1], *Eshel Faraggi*[2,3] and *Andrzej Kloczkowski*[4,5]

[1] Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA, United States, [2] Research and Information Systems, LLC, Indianapolis, IN, United States, [3] Department of Physics, Indiana University Purdue University Indianapolis, Indianapolis, IN, United States, [4] Battelle Center for Mathematical Medicine, Nationwide Children's Hospital, Columbus, OH, United States, [5] Department of Pediatrics, The Ohio State University, Columbus, OH, United States

Two new computational approaches are described to aid in the design of new peptide-based drugs by evaluating ensembles of protein structures from their dynamics and through the assessing of structures using empirical contact potential. These approaches build on the concept that conformational variability can aid in the binding process and, for disordered proteins, can even facilitate the binding of more diverse ligands. This latter consideration indicates that such a design process should be less restrictive so that multiple inhibitors might be effective. The example chosen here focuses on proteins/peptides that bind to hemagglutinin (HA) to block the large-scale conformational change for activation. Variability in the conformations is considered from sets of experimental structures, or as an alternative, from their simple computed dynamics; the set of designe peptides/small proteins from the David Baker lab designed to bind to hemagglutinin, is the large set considered and is assessed with the new empirical contact potentials.

Keywords: protein design, peptide design, computational design, protein ensemble, protein potentials

OPEN ACCESS

Edited by:
Hongchun Li,
Shenzhen Institutes of Advanced Technology (CAS), China

Reviewed by:
Canan Atilgan,
Sabanci University, Turkey
Igor N. Berezovsky,
*Bioinformatics Institute (A*STAR), Singapore*

***Correspondence:**
Robert L. Jernigan
jernigan@iastate.edu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *16 September 2020*
Accepted: *08 December 2020*
Published: *03 February 2021*

Citation:
Jernigan RL, Sankar K, Jia K, Faraggi E and Kloczkowski A (2021) Computational Ways to Enhance Protein Inhibitor Design. Front. Mol. Biosci. 7:607323. doi: 10.3389/fmolb.2020.607323

INTRODUCTION

Influenza infection is a widespread cause of major medical concern because of rapid viral evolution, which causes both occasional pandemics and, more frequently, health problems almost every year. It has been estimated that the annual outbreaks by influenza A and B viruses over the past 100 years have had an even greater impact than all other past pandemics combined (Wilson et al., 1981; Bullough et al., 1994; Bizebard et al., 1995). The extremely high mutation rate of the virus means that any given vaccine soon becomes outdated. Thus, vaccination offers limited protection, especially when facing the highly virulent nature and rapid evolution of influenza (Chen et al., 1999). Although some effective anti-influenza drugs have been developed, drug resistance usually appears rapidly.

Hemagglutinin (HA) is a major surface glycoprotein of this virus that is involved in four of the most important aspects of influenza infection: (a) it is the target of antibodies that neutralize infectivity, (b) it undergoes antigenic drift to escape neutralization, (c) it binds to cell-surface receptors to initiate infection, and (d) it mediates the fusion of viral and host membranes essential for viral entry. The large-scale conformational changes in HA are critical for the steps in which the virus inserts itself into the host cells by fusing to the host membrane, and the residues involved in this process are highly conserved across different types and subtypes during antigenic drift. These

residues can serve as important targets for developing broad-reacting antiviral inhibitors (Jiang et al., 1993; Wild et al., 1994; Chan et al., 1998; Skehel and Wiley, 1998). Based on a set of crystal structures of the HA-antibody complex showing the conformational changes to HA during the essential activation steps, David Baker and his colleagues designed a novel HA inhibitor for Group 1 of type A virus (Fleishman et al., 2011).

Influenza HA is a homo-trimeric protein where each monomer contains two disulfide-bonded polypeptides, HA1 and HA2. HA1 is responsible for attaching to host cell-surface receptors, and HA2 mediates the fusion of the influenza envelope with the endosomal membrane thus allowing the entry of influenza RNA into the host cell. The pre- (Wilson et al., 1981) and post-fusion structures (Bizebard et al., 1995) of HA1 are essentially the same, while those of HA2 (Wilson et al., 1981; Bullough et al., 1994; Chen et al., 1999) are drastically different (see **Figure 1**).

The structural change in HA2 includes a partial unfolding of the long α-helix into a loop (dark blue) and the folding of an inter-helix loop (in red) into a part of the long α-helix, thus delivering both N- (blue) and C-terminal (pink) fragments to the same end of the molecule upon the fusion of viral and endosomal membranes.

The protein gp41 of HIV-1 is the membrane fusion protein, similar to HA2 of HA (Skehel and Wiley, 1998). In that case, peptides derived from the C-terminal region of gp41 corresponding to the outer-layer helices, referred to as C-peptides, were found to inhibit HIV-1 infection with IC_{50} in the nanomolar range (Jiang et al., 1993; Wild et al., 1994; Chan et al., 1998). C-peptides are believed to act by binding to the exposed surface of the N-terminal central three-helical bundle in a transient pre-fusion gp41 intermediate, thereby blocking membrane fusion. One such L-peptide, T-20/Enfuvirtide with 36 residues, was approved previously as a drug by the Food and Drug Administration (FDA) (FDA Notifications, 2003); it showed high efficacy in suppressing resistant HIV-1 strains. Moreover, efforts to target a prominent pocket on the surface of the central three-helical bundle have led to the discovery of small, cyclic D-peptides that inhibit HIV-1 infection, thereby validating the pocket as a potential target for small-molecule HIV-1 fusion inhibitors (Eckert et al., 1999).

To evade host antibody recognition, the HA protein on the surface of influenza virus, primarily on the globular domain, must constantly mutate. This interferes in important ways with any vaccine and reduces the vaccine's efficiency and useful lifetime. However, no matter how much the influenza virus mutates, it must maintain the ability to induce membrane fusion to ensure its propagation. Thus, the stem domain that is primarily responsible for inducing membrane fusion is the most conserved. Ian Wilson's group identified antibodies that broadly neutralize influenza A virus Group 1 (Ekiert et al., 2009) (**Figure 2A**), Group 2 (Murphy and Webster, 2001), Group 1 and 2 (Ekiert et al., 2012), and influenza type A and B viruses (Dreyfus et al., 2012) (see **Figure 2B**). All these antibodies recognize epitopes located in the stem domain. David Baker's group designed small proteins against influenza A virus Group 1 (Fleishman et al., 2011) (**Figure 2A**). In addition, they identified a conserved patch on the surface of the central helical bundle in the low-pH post-fusion state (**Figure 2C**). These three interfaces may all serve as useful targets for developing inhibitors against influenza virus.

Molecular recognition in general and protein-protein interactions in particular are essential in almost every aspect of biological function. Moreover, proteins that bind other proteins with high affinity and high specificity have numerous applications for diagnostics and therapeutics. Currently, antibodies are by far the most commonly used proteins for both

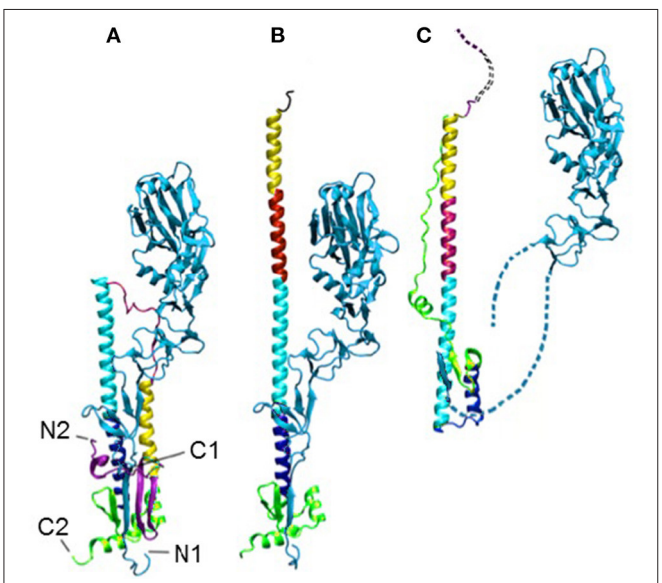


FIGURE 1 | The HA in the pre- **(A)**, intermediate **(B)**, and post-fusion **(C)** states. The termini of HA1 (light blue) and HA2 are labeled as N1, C1, N2, and C2, respectively, in **(A)**. The dotted lines in **(C)** indicate unresolved regions. The structures have all been aligned on the cyan helix, which is the only region in that domain that does not change during the transition.

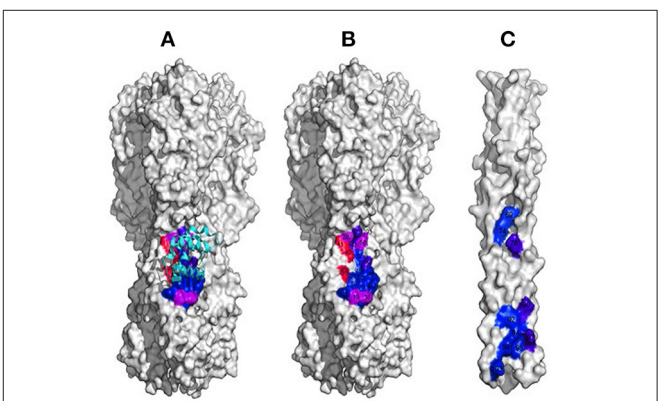


FIGURE 2 | Interfaces on HA that could likely be targeted in inhibitor design. **(A)** The interface conserved among Group 1 influenza A virus (in colors). **(B)** The interface conserved among all influenza A and B viruses. **(C)** The conserved patch on the surface of the central helical bundle at low-pH on HA.

TABLE 1 | The PDB identifiers of the 43 structures of hemagglutinin used here for extracting dynamics.

1HGD	2HMG	3FKU	4BGZ	4KPQ
1MQL	2IBX	3HMG	4BH1	4KPS
1MQM	2WR7	3LZG	4DJ6	5HMG
1MQN	2WRB	3M5G	4EDB	
1RD8	2WRD	3M6S	4F23	
1RUY	2WRE	3S11	4F3Z	
1RUZ	2WRF	3SM5	4FIU	
1RV0	2WRG	3UBE	4GXX	
1RVX	2WRH	3VUN	4JTX	
2FK0	3EYM	3ZTJ	4KDM	

detection and therapeutic intervention. However, antibodies are large proteins that are expensive to produce and difficult to deliver. Thus, it would be important progress for biomedicine to be able to design novel protein-binding modules at will.

The set of 88 proteins that were designed and tested by Baker et al. provides an excellent test set for use in the present study. Below we consider the dynamics of the structure in two different ways, from a set of experimental structures and from computed dynamics. Then we apply new knowledge-based free energies to rank the different designs, specifically predicting which designs are likely to bind. Baker and colleagues were not able to do this without experimental testing. These are empirical-free energy contact potentials developed by Jernigan, Kloczkowski, and Faraggi that have proven to be highly successfully in blind-tests at past CASP experiments. In the present paper, we aim to make some suggestions for new ways to sample conformations of a target protein and how to assess the designed structures.

MULTIPLE EXPERIMENTAL STRUCTURES CAPTURE THE IMPORTANT FUNCTIONAL MOTIONS WITHIN A HEMAGGLUTININ STRUCTURE SET

The 43 structures of hemagglutinin listed in **Table 1** were collected from the PDB with a BLAST search, retaining only those structures present as trimeric complexes of the HA1 and HA2 subunits. The individual subunits were extracted separately and aligned. This yields a total of 129 structures of the HA1 + HA2 monomers that were superimposed onto the central structure (PDB: 1 mqm) using the Combinatorial Extension (CE) algorithm, and these have a continuous distribution of RMSDs from 0 to 3.3 Å.

After these structures have been superimposed, the covariances for all pairs of positions were computed. Then Principal Component Analysis is performed on this dataset. The input is the set of all of the structures in the set (Teodoro et al., 2002, 2003). From these data, the average position of each point in the reference structure is computed as $<x_i>$, and the covariances for each pair of points, i and j, was computed according to $c_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$, where brackets $< >$ indicate averages over the set of structures. The covariance matrix

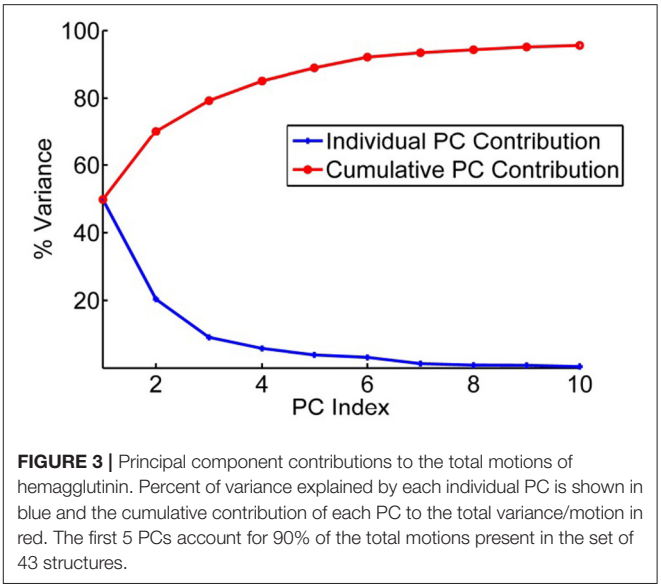


FIGURE 3 | Principal component contributions to the total motions of hemagglutinin. Percent of variance explained by each individual PC is shown in blue and the cumulative contribution of each PC to the total variance/motion in red. The first 5 PCs account for 90% of the total motions present in the set of 43 structures.

C can be decomposed as $C = P\Delta P^T$, where the eigenvectors P represent the principal components (PCs) and the eigenvalues are the elements of the diagonal matrix Δ. The eigenvalues are sorted in order. Each eigenvalue is directly proportional to the amount of the total variance it captures. The results of this analysis are shown in **Figure 3** for the set of coarse-grained hemagglutinins, which shows how truly limited the characteristic motions are within the structure set. Clearly, it does not require many of these characteristic motions to capture nearly all of the overall motions.

CHARACTERIZATION OF THE GLOBAL MOTIONS IN HEMAGGLUTININS

Based on their sequences, HAs have been subdivided into two main groups: Group 1 (H1, H2, H5, H6, H8, H9, H11, H12, H13, and H16) and Group 2 (H3, H4, H7, H10, H14, and H15) (Air, 1981). Interestingly, the first three PCs separately cluster into these two major groups, with minor exceptions. The distribution of the experimental structures over the PCs are shown in **Figure 4** for pairs of PCs. This distinctive clustering can be seen clearly.

Different conformations can bind to different partners and thus include dynamics in the process that will improve the probability of success in computational protein design. When the PCs are visualized on the structures, it can be seen that the first three PCs primarily represent motions in the B-loop (blue) that are involved in the large-scale transition. PC1, PC2, and PC3 can be interpreted as primarily involving conformations changes in the C-terminus, the central, and N-terminus parts of the B-loop (see **Figure 5**). Interestingly, the B-loop is a region with a strong tendency to form a coiled-coil and is implicated in the formation of the pre-hairpin intermediate in the "spring loaded mechanism" of HA action (Carr and Kim, 1993; Xu and Wilson, 2011). The PC3 motion also clearly demonstrates the shift in the loop necessary for it to position itself at the top of helix C. In

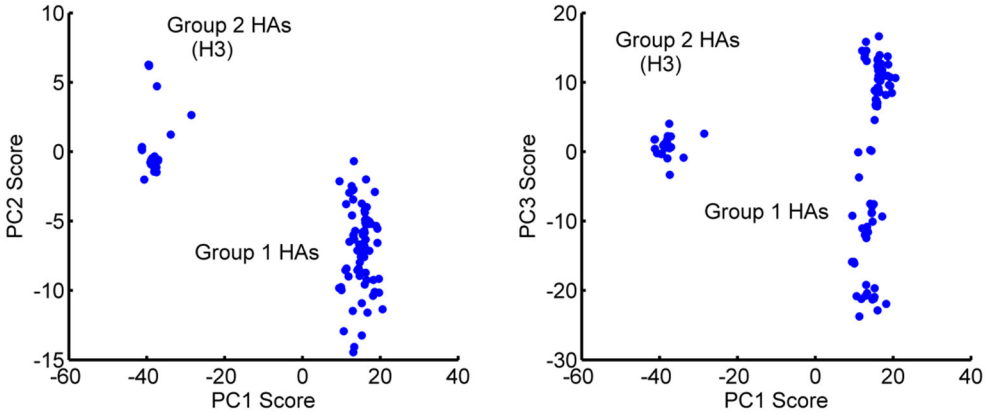


FIGURE 4 | Distribution of the 129 HA monomeric structures projected onto the first 3 PCs. **(Left)** PC2-PC1 space and **(Right)** PC3-PC1 space. PC1 separates the Group 1 and Group 2 hemagglutinins into different clusters. Outliers H13 and H16 have been eliminated from the figure for the sake of clarity. PC1 has a major gap between the two groups of clustered structures. PC3 appears to be populated in two somewhat similar clusters, with Group 1 showing a particularly wide range of PC3 values.

addition, PC2 captures a hinge motion in the head of HA with respect to the stem as well as well as a motion at the N-terminus of HA2 (fusion peptide) that is subsequently exposed for insertion into the membrane during fusion. These computed structures show a high level of variability of conformations particularly for the B-loop, which relate well to the known conformational transition, even though the full extent of motions is not shown in **Figure 5**. As shown in **Table 2**, these PCs provide a useful representation of changes present in the ensemble of structures.

ANISOTROPIC NETWORK MODELS (ANM) CAN SUBSTITUTE, OF INSUFFICIENT NUMBERS OF EXPERIMENTAL STRUCTURES ARE AVAILABLE

Elastic Network Models of proteins, such as the Gaussian Network Model (GNM) and Anisotropic Network Models (ANM) of proteins as developed by Tirion (1996), Bahar, Erman, and Jernigan (Bahar and Jernigan, 1994, 1998; Bahar et al., 1997a; Demirel et al., 1998; Atilgan et al., 2001; Doruker et al., 2002a,b; Doruker and Jernigan, 2003; Sen et al., 2006), computationally yield information about protein fluctuation dynamics, the directions of motions of the residues and atoms around their equilibrium positions. This information has already been used by Bahar, Jernigan, Kloczkowski, and many others with significant success (Bahar and Jernigan, 1994; Keskin et al., 2002a,b; Isin et al., 2012) to explain functional motions and mechanisms in proteins, nucleic acids, and large biological assemblies, such as the ribosome. ANM could be used as an alternative to calculate the normal modes from a single structure when insufficient numbers of experimental structures or structures having sufficient variability are not available to perform PC analysis, then normal modes from the elastic network models could also be used to compute entropies (Zimmermann et al., 2012) (But, as we show below, contact entropies are simpler and provide significant gains). In ANM, the potential energy V is a function of the displacement vector D of each point in the structure $V = \frac{\gamma}{2} DHD^T$, where γ is the spring constant for all closely interacting points in a structure (here we used a cutoff distance of 13 Å between alpha-carbons for coarse-grained models retaining only C^{α} atoms) to establish the spring connections between residues), and H is the Hessian matrix containing the second derivatives of the energy, with respect to each of the coordinates x, y, z. For a structure with n residues, the Hessian matrix H contains $n \times n$ super-elements each of size 3×3. The Hessian matrix H can be decomposed (Atilgan et al., 2001) as $H = M\Lambda M^T$, where Λ is a diagonal matrix comprising the eigenvalues with the eigenvectors forming the columns of the matrix M. This decomposition generates $3n$-6 normal modes (the first six modes account for the rigid body translations and rotations of the system) reflecting the vibrational fluctuations, so singular value decomposition is utilized.

COMPARING DIRECTIONS OF MOTIONS USING OVERLAPS

The alignment between the directions of a given experimental PC and a given computed normal mode can be measured by comparing the directions of motion in their overlap, as defined by Tama and Sanejouand (2001): $O_{ij} = \frac{|P_i \cdot M_j|}{\|P_i\| \|M_j\|}$, where P_i is the ith PC for and M_j is the jth normal mode. A perfect match yields an overlap value of 1, meaning these motions are in the same direction. We also define the cumulative overlap (CO) between the first k vectors M_j and P_i as $CO\,(k) = \left(\sum_{j=1}^{k} O_{ij}^2 \right)^{\frac{1}{2}}$.

The high overlaps between the two methods ensures the reliability of the computed dynamics. The 1st, 2nd, and 3rd PCs have good overlaps of 0.57, 0.43, and 0.34 with the 3rd, 2nd, and 1st individual modes, respectively. We compare the first three

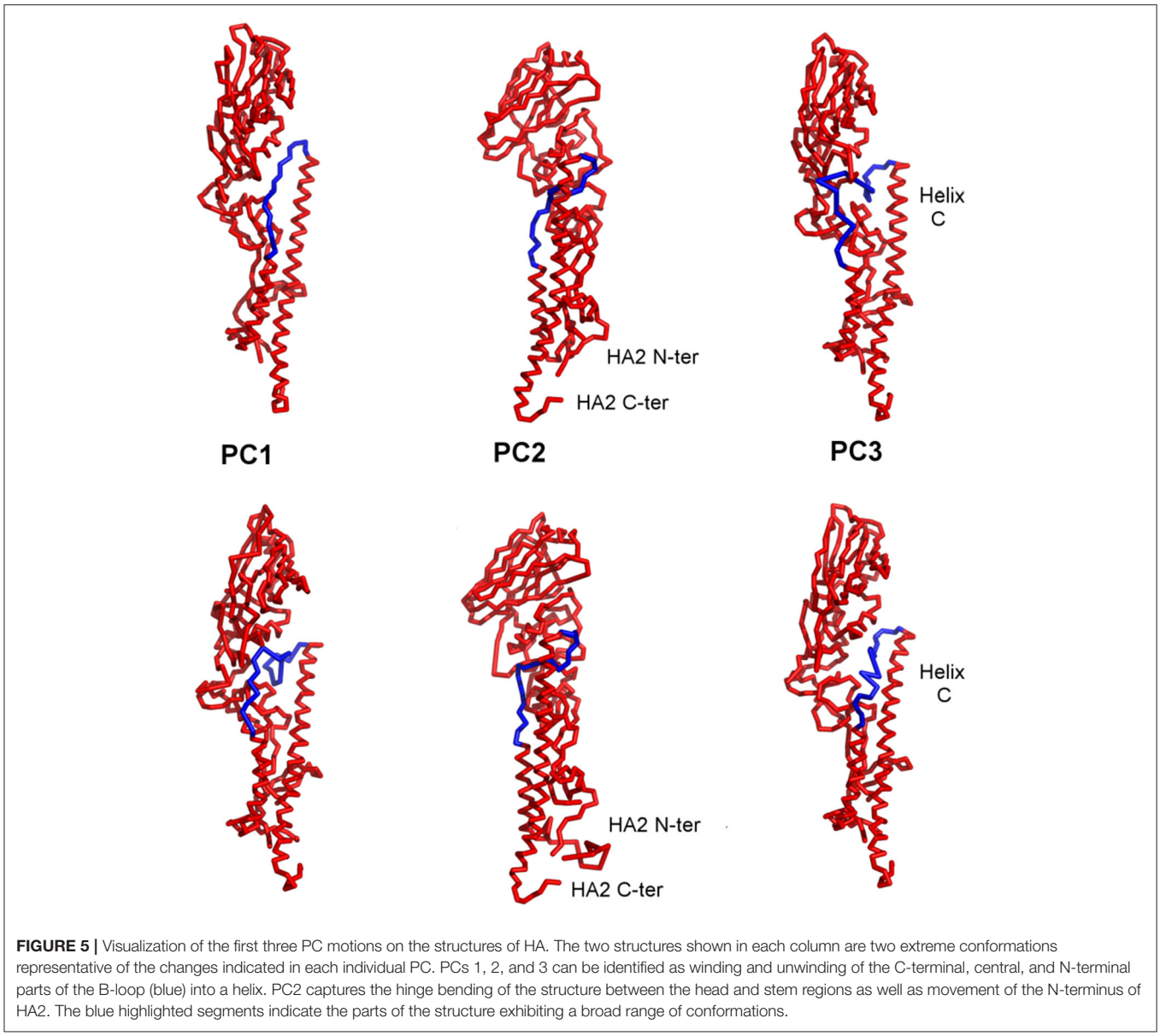


FIGURE 5 | Visualization of the first three PC motions on the structures of HA. The two structures shown in each column are two extreme conformations representative of the changes indicated in each individual PC. PCs 1, 2, and 3 can be identified as winding and unwinding of the C-terminal, central, and N-terminal parts of the B-loop (blue) into a helix. PC2 captures the hinge bending of the structure between the head and stem regions as well as movement of the N-terminus of HA2. The blue highlighted segments indicate the parts of the structure exhibiting a broad range of conformations.

PC's from the X-ray set with the first 20 normal modes from the elastic network models, and these are relatively high between all three PCs of the X-ray hemagglutinin and the set of normal modes for the computed normal modes (see **Table 2**).

TABLE 2 | Cumulative overlaps between computed ANM modes and PCs from the set of experimental hemagglutinin structures.

	CO		
	3 Modes	**6 Modes**	**20 Modes**
PC1	0.60	**0.66**	**0.71**
PC2	**0.50**	**0.57**	**0.65**
PC3	0.40	0.44	**0.60**

Values above 0.5 are in bold.

STRATEGIES FOR GENERATING AND RANKING AN ENSEMBLE OF STRUCTURES AND IDENTIFYING A STRUCTURE MODULE TARGETED FOR INHIBITOR DESIGN

Identifying the most conformationally variable part of the structure is the aim here. These are the parts of a structure that should be the most useful to use for inhibitor design. These parts can be identified simply by computing the changes in all internal distances over the ensemble. Examples of such potential binding parts to target have been extracted from the ensemble of sampled conformations for HA generated by utilizing combinations of the first several PCs (**Figure 6**). This highly variable segment should be susceptible to binding by a broader range of ligands.

FIGURE 6 | Examples of the diversity of conformations from the first 3 PCs for the B loop (blue) of hemagglutinin. The PCs can be used to generate an ensemble of conformations. Each of three shows a conformation generated from one PC. The motions showed that this loop is the most flexible part of the structure and possesses an extremely diverse set of conformations.

ASSESSING PEPTIDE/PROTEIN DESIGNS WITH NEW EMPIRICAL CONTACT POTENTIALS

Here we present new strategies for the assessment of bound ligand structures by taking as our target the designed small proteins from David Baker and his colleagues that were targeted to bind hemagglutinin (Fleishman et al., 2011; Fleishman and Baker, 2012). This provides an interesting relatively large dataset, which we can use to test our assessment method. The Baker designs, originating from small, monomeric proteins in the PDB having between 80 and 250 residues, were targeted against a hydrophobic region on the "stem" of hemagglutinin. Of the 88 designs that they tested, only two were reported to have detectable binding affinity for hemagglutinin (this affinity was subsequently improved in rounds of randomization and selection).

Four-Body Coarse-Grained Contact Potentials (Feng et al., 2007, 2010)

Four-body potentials were developed by Kloczkowski and Jernigan to account for the cooperative interactions in proteins; they take into account the coarse-grained contact interactions together with the extent of solvent exposure, and thus provide a more detailed and more cooperative representation of protein interaction energies than do pairwise potentials. Capturing this cooperativity is considered to be critical for evaluating densely packed protein structures. These potentials are highly empirical and are based simply on the observed frequency of occurrences of different types of amino acids in closely interacting quartets of amino acid types within a large set of protein structures. We have found that these four-body contact potentials can discriminate well between native structures and partially unfolded or deliberately misfolded structures. These have also included short-range backbone energies (Bahar et al., 1997b). We tested these optimized potentials at CASP9 as the prediction group 4_BODY_POTENTIALS from Iowa State University. There were 110 other human prediction groups participating in CASP9 competition, and 140 prediction servers. According to Nick Grishin, the assessor of free modeling techniques at CASP9, 4_BODY_POTENTIALS was one of most successful groups in free modeling at that time, ranking third according to the averaged z_{score} both for best models and top models. Free modeling is the most difficult and most challenge in protein structure prediction, when the sequence of the protein has only a low sequence similarity in comparison to any known protein structures. This success at CASP9 demonstrates clearly that the cooperative multibody interactions are an appropriate tool for assessing predicted structures, and we apply them here to Baker's hemagglutinin inhibitor structures. Later we added in electrostatic interactions, and these were tested at the subsequent CASP10.

Including Entropies in the Inhibitor Assessments

The Elastic Network Models (ENM) have proven themselves to be highly useful in representing the global motions for a wide variety of diverse protein structures (Bahar and Jernigan, 1997, 1998, 1999; Bahar et al., 1997a,b,c; Bahar et al., 1998, 1999; Demirel et al., 1998; Keskin et al., 1998, 2000, 2002a,b; Jernigan et al., 1999, 2000, 2008; Atilgan et al., 2001; Bahar and Rader, 2005; Sen et al., 2006; Jernigan and Kloczkowski, 2007; Yang et al., 2007, 2008, 2009; Zhu and Hummer, 2010; Bakan et al., 2011; Karaca and Bonvin, 2011; May and Brooks, 2011; Peng and Head-Gordon, 2011; Uyar et al., 2011; Wieninger et al., 2011; Zheng, 2011; Zheng and Auerbach, 2011; Zimmermann et al., 2011a,b; Duttmann et al., 2012; Gniewek et al., 2012; Isin et al., 2012; Martin et al., 2012; Ruvinsky et al., 2012; Globisch et al., 2013; Kim et al., 2013; Sanejouand, 2013; Dasgupta et al., 2014). Since they have proven to be so successful in capturing the most important motions of protein structures, it is reasonable to expect that they should also be able to estimate the conformational entropies of structures. We employ the Elastic Network Model to compute the motions of protein structures, and then these motions are then used directly to approximate the entropy of a conformation (Zimmermann et al., 2011c, 2012). We previously (Zimmermann et al., 2011c) used vibrational entropies based on the frequencies of the normal modes, but more recently have found significant gains by utilizing the mean square fluctuations computed from the ENM as a direct measure of entropy: $\Delta S \propto$

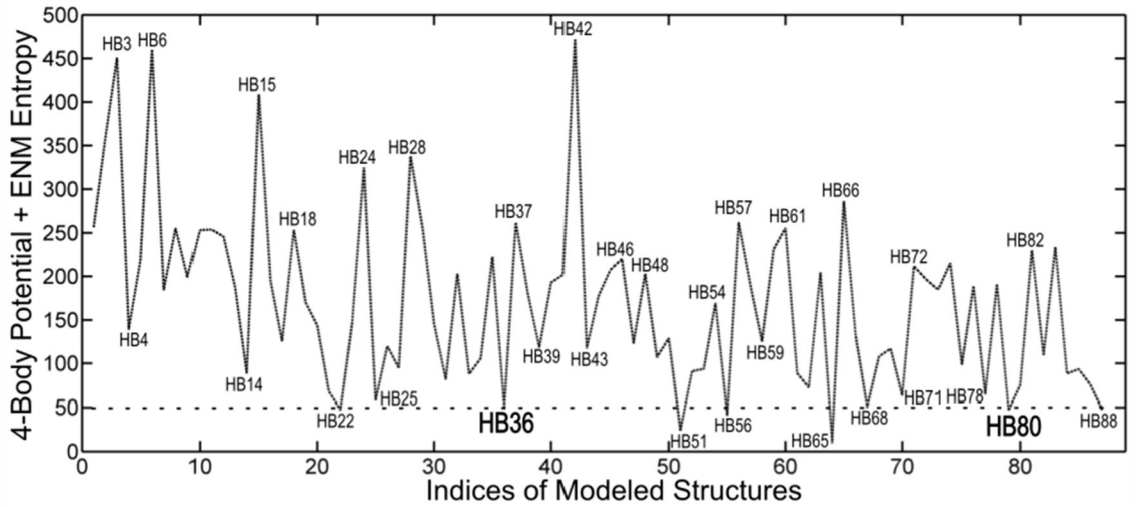


FIGURE 7 | Ranking by coarse-grained free energies of inhibitor proteins designed against hemagglutinin. Free energies are given on the ordinate axis (arbitrary scale), and the different structures (from the pdb) are indicated along the abscissa. The eight top-ranked structures with favorable free energies can be seen to be HB22, HB36, HB51, HB56, HB65, HB68, HB80, and HB88. This demonstrates the utility of the coarse-grained free energies to computationally screen for favorable structures. The two structures HB36 and HB80 were experimentally shown to be functional.

$\Gamma^{-1} = \sum_{i=2}^{N} \frac{1}{\lambda_i} \left(Q_i Q_i^T\right)$, where Q is a normal mode vector, λ the corresponding square frequency, Γ the system's Hessian, and Γ^{-1} its pseudo-inverse. We obtain the Free Energy changes from $\Delta G = \Delta E - T\Delta S$ by simply combining the four-body potential with the ENM-based entropy (Zimmermann et al., 2012). The excellent blind-tested performance of our method in CASP experiments shows that our methodology is an outstanding tool for assessing protein designs, such as the ones from Baker's hemagglutinin inhibitor designs.

THESE NEW FREE ENERGIES SUCCESSFULLY SELECT NATIVE-LIKE POSES IN PROTEIN-PROTEIN DOCKING

We have applied this method to the set of 89 inhibitor proteins designed against hemagglutinin by David Baker's group, and we find that it provides a useful screen for that set of structures. Structures having the lowest energies indicate stable favorable conformations. However, stable structures are not always functional. In this case, we tested a set of eight structures at local minimums of the energy landscape ranked by their energies. From these, it was reported that two of them were found to be functional (see **Figure 7**).

DISCUSSION

We have outlined a simple new way to use protein dynamics for peptide/protein design studies. This approach serves to identify those specific regions in the structure having particularly wide-ranging conformational variability, which could be of particular importance for targeting computational design efforts. Specifically, the highly variable segments should be able to bind to a particularly wide range of diverse ligands. Such variable conformations are well-known to be important for the promiscuous binding exhibited by disordered proteins and using this approach should have some advantage. Using such more localized protein targets might be an important new approach for targeted computational design. Another advantage of this is that more exhaustive computations can be carried out for smaller targets.

Application of the potentials described above to assess structural designs would allow ranking of sets of designed inhibitor proteins. The differences in rankings should allow to conclude the extent to which the large-scale backbone fluctuations identified in the dynamics could be utilized in the design process. This would require a significantly larger effort than has been presented here. Of course, the potentials themselves are empirical and could be modified to reflect the data from the experimental studies on the designed molecules for the specific class of targeted protein, which is one of the major advantages of the adaptability of the empirical potentials in any particular application.

Our approach can be extended by detailed analysis of allosteric sites that are important for drug design. Most drugs are designed to bind directly to the primary active sites, called orthosteric sites, to inhibit or modify the function of the protein. Binding of a drug to the active site prevents binding to a virus or other disease-related agent and most drugs are designed to fit into the primary active sites. However, adverse side effects of a drug may occur because many enzymes or receptors with related functions may have similarities in their active sites.

A new approach to drug design is based on secondary binding site effects. In this approach, small molecule drugs are designed to bind at secondary binding sites called allosteric sites (Tsai and Nussinov, 2014; Dokholyan, 2016; Guarnera and Berezovsky, 2016, 2020; Schueler-Furman and Wodak, 2016; Wodak et al.,

2019; Zhang et al., 2020). A potential drug—an allosteric modulator binds to an allosteric site and remotely modifies the conformation of the primary binding site of the protein. Allosteric sites are controlled by intrinsic protein dynamics, and the approach proposed here could also be applied to these allosteric sites.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: protein data bank.

AUTHOR CONTRIBUTIONS

RJ, KS, KJ, and AK conceptualized, carried out the research, and wrote the paper. EF engaged in discussions about this research. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by NSF grant DBI-1661391 and NIH grant R01GM127701.

REFERENCES

Air, G. M. (1981). Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A virus. *Proc. Natl. Acad. Sci. U.S.A.* 78, 7639–7643. doi: 10.1073/pnas.78.12.7639

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80, 505–515. doi: 10.1016/S0006-3495(01)76033-X

Bahar, I., Atilgan, A. R., and Erman, B. (1997a). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* 2, 173–181. doi: 10.1016/S1359-0278(97)00024-2

Bahar, I., Erman, B., Haliloglu, T., and Jernigan, R. L. (1997c). Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. *Biochemistry* 36, 13512–13523. doi: 10.1021/bi971611f

Bahar, I., Erman, B., Jernigan, R. L., Atilgan, A. R., and Covell, D. G. (1999). Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. *J. Mol. Biol.* 285, 1023–1037. doi: 10.1006/jmbi.1998.2371

Bahar, I., and Jernigan, R. L. (1994). Cooperative structural transitions induced by non-homogeneous intramolecular interactions in compact globular proteins. *Biophys. J.* 66, 467–481. doi: 10.1016/S0006-3495(94)80798-2

Bahar, I., and Jernigan, R. L. (1997). Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* 26, 195–214. doi: 10.1006/jmbi.1996.0758

Bahar, I., and Jernigan, R. L. (1998). Vibrational dynamics of transfer RNAs: comparison of the free and synthetase-bound forms. *J. Mol. Biol.* 281, 2522–2532. doi: 10.1006/jmbi.1998.1978

Bahar, I., and Jernigan, R. L. (1999). Cooperative fluctuations and subunit communication in tryptophan synthase. *Biochemistry* 38, 3478–3490. doi: 10.1021/bi982697v

Bahar, I., Kaplan, M., and Jernigan, R. L. (1997b). Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches. *Proteins* 29, 292–308. doi: 10.1002/(SICI)1097-0134(199711)29:3<292::AID-PROT4>3.0.CO;2-D

Bahar, I., and Rader, A. J. (2005). Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* 15, 586–592. doi: 10.1016/j.sbi.2005.08.007

Bahar, I., Wallqvist, A., Covell, D. G., and Jernigan, R. L. (1998). Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. *Biochemistry* 37, 1067–1075. doi: 10.1021/bi9720641

Bakan, A., Meireles, L. M., and Bahar, I. (2011). ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* 27, 1575–1577. doi: 10.1093/bioinformatics/btr168

Bizebard, T., Gigant, B., Rigolet, P., Rasmussen, B., Diat, O., Bosecke, P., et al. (1995). Structure of influenza virus haemagglutinin complexed with a neutralizing antibody. *Nature* 376, 92–94. doi: 10.1038/376092a0

Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* 371, 37–43. doi: 10.1038/371037a0

Carr, C. M., and Kim, P. S. (1993). A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* 73, 823–832. doi: 10.1016/0092-8674(93)90260-W

Chan, D. C., Chutkowski, C. T., and Kim, P. S. (1998). Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc. Natl. Acad. Sci. U.S.A.* 95, 15613–15617. doi: 10.1073/pnas.95.26.15613

Chen, J., Skehel, J. J., and Wiley, D. C. (1999). N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8967–8972. doi: 10.1073/pnas.96.16.8967

Dasgupta, B., Nakamura, H., and Kinjo, A. R. (2014). Rigid-body motions of interacting proteins dominate multispecific binding of ubiquitin in a shape-dependent manner. *Proteins* 82, 77–89. doi: 10.1002/prot.24371

Demirel, M. C., Atilgan, A. R., Jernigan, R. L., Erman, B., and Bahar, I. (1998). Identificatioin of kinetically hot residues in proteins. *Protein Sci.* 7, 871–884. doi: 10.1002/pro.5560071205

Dokholyan, N. V. (2016). Controlling allosteric networks in proteins. *Chem. Rev.* 116, 6463–6487. doi: 10.1021/acs.chemrev.5b00544

Doruker, P., and Jernigan, R. L. (2003). Functional motions can be extracted from on-lattice construction of protein structures. *Proteins* 53, 174–181. doi: 10.1002/prot.10486

Doruker, P., Jernigan, R. L., and Bahar, I. (2002a). Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.* 23, 119–127. doi: 10.1002/jcc.1160

Doruker, P., Jernigan, R. L., Navizet, I., and Hernandez, R. (2002b). Important fluctuation dynamics of large protein structures are preserved upon renormalization. *Int. J. Quantum Chem.* 90, 822–837. doi: 10.1002/qua.955

Dreyfus, C., Laursen, N. S., Kwaks, T., Zuijdgeest, D., Khayat, R., Ekiert, D. C., et al. (2012). Highly conserved protective epitopes on influenza B viruses. *Science* 337, 1343–1348. doi: 10.1126/science.1222908

Duttmann, M., Mittnenzweig, M., Togashi, Y., Yanagida, T., and Mikhailov, A. S. (2012). Complex intramolecular mechanics of G-actin–an elastic network study. *PLoS ONE* 7:e45859. doi: 10.1371/journal.pone.0045859

Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. A., and Kim, P. S. (1999). Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* 99, 103–115. doi: 10.1016/S0092-8674(00)80066-5

Ekiert, D. C., Bhabha, G., Elsliger, M. A., Friesen, R. H., Jongeneelen, M., Throsby, M., et al. (2009). Antibody recognition of a highly conserved influenza virus epitope. *Science* 324, 246–251. doi: 10.1126/science.1171491

Ekiert, D. C., Kashyap, A. K., Steel, J., Rubrum, A., Bhabha, G., Khayat, R., et al. (2012). Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature* 489, 526–532. doi: 10.1038/nature11414

FDA Notifications (2003). *FDA Approves Fuzeon, The First Fusion Inhibitor*. Rockville, MD: AIDS Alert 18.

Feng, Y., Kloczkowski, A., and Jernigan, R. L. (2007). Four-body contact potentials derived from two protein datasets to discriminate native structures from decoys. *Proteins* 68, 57–66. doi: 10.1002/prot.21362

Feng, Y., Kloczkowski, A., and Jernigan, R. L. (2010). Potentials 'R' Us web-server for protein energy estimations with coarse-grained knowledge-based potentials. *BMC Bioinformatics* 11:92. doi: 10.1186/1471-2105-11-92

Fleishman, S. J., and Baker, D. (2012). Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* 149, 262–273. doi: 10.1016/j.cell.2012.03.016

Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E. M., et al. (2011). Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332, 816–821. doi: 10.1126/science.1202617

Globisch, C., Krishnamani, V., Deserno, M., and Peter, C. (2013). Optimization of an elastic network augmented coarse grained model to study CCMV capsid deformation. *PLoS ONE* 8:e60582. doi: 10.1371/journal.pone.0060582

Gniewek, P., Kolinski, A., Jernigan, R. L., and Kloczkowski, A. (2012). Elastic network normal modes provide a basis for protein structure refinement. *J. Chem. Phys.* 136:195101. doi: 10.1063/1.4710986

Guarnera, E., and Berezovsky, I. N. (2016). Allosteric sites: remote control in regulation of protein activity. *Curr. Opin. Struct. Biol.* 37, 1–8. doi: 10.1016/j.sbi.2015.10.004

Guarnera, E., and Berezovsky, I. N. (2020). Allosteric drugs and mutations: chances, challenges, and necessity. *Curr. Opin. Struct. Biol.* 62, 149–157. doi: 10.1016/j.sbi.2020.01.010

Isin, B., Tirupula, K. C., Oltvai, Z. N., Klein-Seetharaman, J., and Bahar, I. (2012). Identification of motions in membrane proteins by elastic network models and their experimental validation. *Methods Mol. Biol.* 914, 285–317. doi: 10.1007/978-1-62703-023-6_17

Jernigan, R. L., Bahar, I., Covell, D. G., Atilgan, A. R., Erman, B., and Flatow, D. T. (2000). Relating the structure of HIV-1 reverse transcriptase to its processing step. *J. Biomol. Struct. Dyn.* 17, 49–55. doi: 10.1080/07391102.2000.10506603

Jernigan, R. L., Demirel, M. C., and Bahar, I. (1999). Relating structure to function through the dominant modes of motion of DNA topoisomerase II. *Int. J. Quant. Chem.* 175, 301–312. doi: 10.1002/(SICI)1097-461X(1999)75:3<301::AID-QUA19>3.0.CO;2-0

Jernigan, R. L., and Kloczkowski, A. (2007). Packing regularities in biological structures relate to their dynamics. *Methods Mol. Biol.* 350, 251–76. doi: 10.1385/1-59745-189-4:251

Jernigan, R. L., Yang, L., Song, G., and Doruker, P. (2008). "Elastic network models of coarse-grained proteins are effective for studying the structural control exerted over their dynamics," in *Coarse-Graining of Condensed Phase and Biomolecular Systems*, ed G. Voth (Taylor and Francis Group LLC; CRC Press), 237–254. doi: 10.1201/9781420059564.ch16

Jiang, S., Lin, K., Strick, N., and Neurath, A. R. (1993). HIV-1 inhibition by a peptide. *Nature* 365:113. doi: 10.1038/365113a0

Karaca, E., and Bonvin, A. M. (2011). A multidomain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes. *Structure* 19, 555–565. doi: 10.1016/j.str.2011.01.014

Keskin, O., Bahar, I., Badretdinov, A. Y., Ptitsyn, O. B., and Jernigan, R. L. (1998). Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. *Protein Sci.* 7, 2578–2586. doi: 10.1002/pro.5560071211

Keskin, O., Bahar, I., Flatow, D., Covell, D. G., and Jernigan, R. L. (2002a). Molecular mechanisms of chaperonin GroEL-GroES function. *Biochemistry* 41, 491–501. doi: 10.1021/bi011393x

Keskin, O., Durell, S. R., Bahar, I., Jernigan, R. L., and Covell, D. G. (2002b). Relating molecular flexibility to function: a case study of tubulin. *Biophys. J.* 83, 663–680. doi: 10.1016/S0006-3495(02)75199-0

Keskin, O., Jernigan, R. L., and Bahar, I. (2000). Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys. J.* 78, 2093–2106. doi: 10.1016/S0006-3495(00)76756-7

Kim, M. H., Seo, S., Jeong, J. I., Kim, B. J., Liu, W. K., Lim, B. S., et al. (2013). A mass weighted chemical elastic network model elucidates closed form domain motions in proteins. *Protein Sci.* 22, 605–613. doi: 10.1002/pro.2244

Martin, D. R., Ozkan, S. B., and Matyushov, D. V. (2012). Dissipative electro-elastic network model of protein electrostatics. *Phys. Biol.* 9:036004. doi: 10.1088/1478-3975/9/3/036004

May, E. R., and Brooks, C. L. III. (2011). Determination of viral capsid elastic properties from equilibrium thermal fluctuations. *Phys. Rev. Lett.* 106:188101. doi: 10.1103/PhysRevLett.106.188101

Murphy, B. R., and Webster, R. G. (2001). "Orthomyxoviruses," in *Fields Virology*, eds B. N. Fields, D. M. Knipe, P. M. Howley, and D. E. Griffini (Philadelphia, PA: Lippincott Williams & Wilkins).

Peng, C., and Head-Gordon, T. (2011). The dynamical mechanism of auto-inhibition of AMP-activated protein kinase. *PLoS Comput. Biol.* 7:e1002082. doi: 10.1371/journal.pcbi.1002082

Ruvinsky, A. M., Kirys, T., Tuzikov, A. V., and Vakser, I. A. (2012). Structure fluctuations and conformational changes in protein binding. *J. Bioinform. Comput. Biol.* 10:1241002. doi: 10.1142/S0219720012410028

Sanejouand, Y. H. (2013). Elastic network models: theoretical and empirical foundations. *Methods Mol. Biol.* 924, 601–616. doi: 10.1007/978-1-62703-017-5_23

Schueler-Furman, O., and Wodak, S. J. (2016). Computational approaches to investigating allostery. *Curr. Opin. Struct. Biol.* 41, 159–171. doi: 10.1016/j.sbi.2016.06.017

Sen, T. Z., Feng, Y., Garcia, J. V., Kloczkowski, A., and Jernigan, R. L. (2006). The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. *J. Chem. TheoryComput.* 2, 696–704. doi: 10.1021/ct600060d

Skehel, J. J., and Wiley, D. C. (1998). Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell* 95, 871–874. doi: 10.1016/S0092-8674(00)81710-9

Tama, F., and Sanejouand, Y. H. (2001). Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14, 1–6. doi: 10.1093/protein/14.1.1

Teodoro, M. L., Philips, G. N. Jr., and Kavraki, L. E. (2002). "A dimensionality reduction approach to modeling protein flexibility," in *International Conference on Computational Molecular Biology (RECOMB)*, 299–308. doi: 10.1145/565196.565235

Teodoro, M. L., Philips, G. N. Jr., and Kavraki, L. E. (2003). Understanding protein flexibility through dimensionality reduction. *J. Comput. Biol.* 10, 617–634. doi: 10.1089/10665270360688228

Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77, 1905–1908. doi: 10.1103/PhysRevLett.77.1905

Tsai, C. J., and Nussinov, R. (2014). A unified view of "how allostery works." *PLoS Comput. Biol.* 10:e1003394. doi: 10.1371/journal.pcbi.1003394

Uyar, A., Kurkcuoglu, O., Nilsson, L., and Doruker, P. (2011). The elastic network model reveals a consistent picture on intrinsic functional dynamics of type II restriction endonucleases. *Phys. Biol.* 8:056001. doi: 10.1088/1478-3975/8/5/056001

Wieninger, S. A., Serpersu, E. H., and Ullmann, G. M. (2011). ATP binding enables broad antibiotic selectivity of aminoglycoside phosphotransferase(3′)-IIIa: an elastic network analysis. *J. Mol. Biol.* 409, 450–465. doi: 10.1016/j.jmb.2011.03.061

Wild, C. T., Shugars, D. C., Greenwell, T. K., McDanal, C. B., and Matthews, T. J. (1994). Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc. Natl. Acad. Sci. U.S.A.* 91, 9770–9774. doi: 10.1073/pnas.91.21.9770

Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. *Nature* 289, 366–373. doi: 10.1038/289366a0

Wodak, S. J., Paci, E., Dokholyan, N. V., Berezovsky, I. N., Horovitz, A., Li, J., et al. (2019). Allostery in its many disguises: from theory to applications. *Structure* 27, 566–578. doi: 10.1016/j.str.2019.01.003

Xu, R., and Wilson, I. A. (2011). Structural characterization of an early fusion intermediate of influenza virus hemagglutinin. *J. Virol.* 85, 5172–5182. doi: 10.1128/JVI.02430-10

Yang, L., Song, G., Carriquiry, A., and Jernigan, R. L. (2008). Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure* 16, 321–330. doi: 10.1016/j.str.2007.12.011

Yang, L., Song, G., and Jernigan, R. L. (2007). How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* 93, 920–929. doi: 10.1529/biophysj.106.095927

Yang, L., Song, G., and Jernigan, R. L. (2009). Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12347–12352. doi: 10.1073/pnas.0902159106

Zhang, Y., Doruker, P., Kaynak, B., Zhang, S., Krieger, J., Li, H., et al. (2020). Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Curr. Opin. Struct. Biol.* 62, 14–21. doi: 10.1016/j.sbi.2019.11.002

Zheng, W. (2011). Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization. *Biophys. J.* 100, 478–488. doi: 10.1016/j.bpj.2010.12.3680

Zheng, W., and Auerbach, A. (2011). Decrypting the sequence of structural events during the gating transition of pentameric ligand-gated ion channels based on an interpolated elastic network model. *PLoS Comput. Biol.* 7:e1001046. doi: 10.1371/journal.pcbi.1001046

Zhu, F., and Hummer, G. (2010). Pore opening and closing of a pentameric ligand-gated ion channel. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19814–19819. doi: 10.1073/pnas.1009313107

Zimmermann, M. T., Kloczkowski, A., and Jernigan, R. L. (2011b). MAVENs: motion analysis and visualization of elastic networks and structural ensembles. *BMC Bioinformatics* 12:264. doi: 10.1186/1471-2105-12-264

Zimmermann, M. T., Leelananda, S. P., Gniewek, P., Feng, Y., Jernigan, R. L., and Kloczkowski, A. (2011c). Free energies for coarse-grained proteins by integrating multibody statistical contact potentials with entropies from elastic network models. *J. Struct. Funct. Genomics* 12, 137–147. doi: 10.1007/s10969-011-9113-3

Zimmermann, M. T., Leelananda, S. P., Kloczkowski, A., and Jernigan, R. L. (2012). Combining statistical potentials with dynamics-based entropies improves selection from protein decoys and docking poses. *J. Phys. Chem. B* 116, 6725–6731. doi: 10.1021/jp2120143

Zimmermann, M. T., Skliros, A., Kloczkowski, A., and Jernigan, R. L. (2011a). Immunoglobulin structure exhibits control over CDR motion. *Immunome Res.* 7:5. doi: 10.4172/1745-7580.1000047

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 04 May 2021
doi: 10.3389/fmolb.2021.587151

DESP: Deep Enhanced Sampling of Proteins' Conformation Spaces Using AI-Inspired Biasing Forces

Emmanuel Oluwatobi Salawu *

Machine Learning Solutions Lab, Amazon Web Services (AWS), Herndon, VA, United States

The molecular structures (i.e., conformation spaces, CS) of bio-macromolecules and the dynamics that molecules exhibit are crucial to the understanding of the basis of many diseases and in the continuous attempts to retarget known drugs/medications, improve the efficacy of existing drugs, or develop novel drugs. These make a better understanding and the exploration of the CS of molecules a research hotspot. While it is generally easy to computationally explore the CS of small molecules (such as peptides and ligands), the exploration of the CS of a larger biomolecule beyond the local energy well and beyond the initial equilibrium structure of the molecule is generally nontrivial and can often be computationally prohibitive for molecules of considerable size. Therefore, research efforts in this area focus on the development of ways that systematically favor the sampling of new conformations while penalizing the resampling of previously sampled conformations. In this work, we present *Deep Enhanced Sampling of Proteins' Conformation Spaces Using AI-Inspired Biasing Forces* (DESP), a technique for enhanced sampling that combines molecular dynamics (MD) simulations and deep neural networks (DNNs), in which biasing potentials for guiding the MD simulations are derived from the KL divergence between the DNN-learned latent space vectors of [a] the most recently sampled conformation and those of [b] the previously sampled conformations. Overall, DESP efficiently samples wide CS and outperforms conventional MD simulations as well as accelerated MD simulations. We acknowledge that this is an actively evolving research area, and we continue to further develop the techniques presented here and their derivatives tailored at achieving DNN-enhanced steered MD simulations and DNN-enhanced targeted MD simulations.

Keywords: conformation space, deep neural network, protein, molecular dynamics simulation, variational autoencoder

OPEN ACCESS

Edited by:
Guang Hu,
Soochow University, China

Reviewed by:
Min Zeng,
Central South University, China
Jinwei Duan,
Chang'an University, China

***Correspondence:**
Emmanuel Oluwatobi Salawu
esalawu@amazon.com

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *25 July 2020*
Accepted: *15 February 2021*
Published: *04 May 2021*

Citation:
Salawu EO (2021) DESP: Deep Enhanced Sampling of Proteins' Conformation Spaces Using AI-Inspired Biasing Forces. Front. Mol. Biosci. 8:587151. doi: 10.3389/fmolb.2021.587151

INTRODUCTION

The functions of biomolecules are encoded in their structures and dynamics (Council and others 1989; Karplus and Kuriyan, 2005; Yang et al., 2014). And there are innumerable pieces of evidence linking the basis of many diseases to anomalies in the structures and the dynamics of the molecules that are involved in the biological systems that the diseases affect (McCafferty and Sergeev, 2016; Chiti and Dobson, 2017; Guo et al., 2017; Hartl, 2017; Tramutola et al., 2017; Salawu, 2018a; Ittisoponpisan et al., 2019; Laskowski et al., 2020) because the normal functioning of the biological systems depends on the molecules' proper structures and dynamics. Furthermore, the various structures that a molecule can take (i.e., the molecule's conformation space, CS) and their associated MD are not only of vital

importance in deciphering of many diseases (Salawu, 2018a; Salawu, 2018b) but are also crucial in the drug development efforts targeted at curing or managing many diseases (Carlson and McCammon, 2000; Lee et al., 2018; Paweland and Caflisch, 2018; Wang et al., 2018; Lin et al., 2020). These recognitions have motivated extensive efforts in the field of structural biochemistry and form the rationale for many structural biology studies (such as through X-ray crystallography, NMR, and Cryo-EM) and the creation of the Protein Data Bank (Berman et al., 2000) as well as other databases for molecular structures. Nonetheless, considerable challenges exist because the solely static molecular structures obtained through the wet laboratory approaches alone (such as the ones listed above) often fall short of providing enough insights into the dynamics of the molecules of interest. These challenges have led to the growing roles and the increasing importance of computational approaches, such as molecular dynamics (MD) simulations, that are often used for studying the dynamic behaviors of molecules and their interactions with other molecules as well as for exploring much wider CS of the molecules of interest.

While it is generally easy to computationally explore the CS of small molecules (such as peptides and ligands), the exploration of the CS of larger a biomolecule beyond the local energy well and beyond the initial equilibrium structure of the molecule is generally nontrivial (Shaw et al., 2008; Shaw et al., 2009) and can often be computationally prohibitive for a molecule of considerable size. These difficulties arise from the existence of energy barriers between different states that the molecule could assume, thereby hindering the movement of the molecule from one structural state to another (Hamelberg et al., 2004; Hénin and Chipot, 2004; Salawu, 2020). At this point, it is important to acknowledge existing efforts targeted at removing, avoiding/sidestepping, lowering, or surmounting these energy barriers, thereby achieving enhanced sampling of the CS of molecules. Therefore, we recognize some of the previous publications in this domain and highlight them in the next paragraphs.

Most of the existing popular approaches for achieving enhanced sampling may be broadly viewed in two categories, namely: those that require the user to specify well-defined collective variables (CVs)/reaction coordinates (RCs) (Babin et al., 2008; Laio and Gervasio, 2008; Bussi and Laio, 2020) and those that do not require the user to explicitly specify the CV/RC (Sugita and Okamoto, 1999; Hamelberg et al., 2004; Moritsugu et al., 2012; Harada and Kitao, 2013; Miao et al., 2015; Chen and Ferguson, 2018; Salawu, 2020). Reconnaissance meta-dynamics uses a self-learning algorithm for accelerated dynamics and is capable of handling a large number of collective variables by making use of bias potentials created as a function of individual locally valid CVs that are then patched together to obtain the sampling across a large number of collective variables (Tribello et al., 2010). Some of the challenges of reconnaissance meta-dynamics such as those associated with the creation of bias potentials as a function of individual locally valid CVs could be addressed by any technique that could potentially learn a compressed representation of those CVs and efficiently explore the combined CVs together in the compressed space. This is the subject of an actively growing research area that leverages the powers of deep neural networks (DNNs)/machine learning (ML). Bridging the fields of enhanced sampling and ML, Bonati et al. (2019) developed DNN-based variationally enhanced sampling that uses neural networks to represent the bias potential in a variational learning scheme that makes it possible for the efficient exploration of even high-dimensional free energy surfaces. In a similar way, reweighted autoencoded variational Bayes (RAVE) models MD simulation trajectories using the VAE whereby the learned distribution of the latent space variable is used to add biasing potentials, thereby penalizing the repeated sampling of the most favorable frequently visited states (Ribeiro et al., 2018). Although other enhanced sampling methods implement the biasing protocol in two steps, RAVE's identification of the RC and its derivation of unbiased probability distribution occur simultaneously. And through the systematic use of the Kullback–Leibler (KL) divergence metric, RAVE can identify physically meaningful RCs from among a group of RCs explored.

In addition to the efforts mentioned above, the combination of well-tempered meta-dynamics and time-lagged independent component analysis to study rare events and explore complex free energy landscapes have also been looked into (McCarty and Parrinello, 2017). Since the initial choice of CVs for meta-dynamics is often suboptimal, the work shows the finding of new and optimal CVs with better convergence properties by the analysis of the initial trajectory using time-lagged independent component analysis (McCarty and Parrinello, 2017). However, a more recent study has shown that rather than using linear dimension reduction methods (such as independent component analysis) a modified autoencoder could more accurately encode the low dynamics of the underlying stochastic processes of MD simulations better than linear dimension reduction methods (Wehmeyer and Noé, 2018). Indeed, there are continuous and growing efforts in the combinations of DNN models and MD simulations in the enhancement of the sampling of molecules' CS and other various aspects of molecular sciences (Allison, 2020; Salawu, 2020; Sidky et al., 2020).

In this work, we present *Deep Enhanced Sampling of Proteins' Conformation Spaces Using AI-Inspired Biasing Forces* (DESP), which also combines DNNs and MD simulations to create a robust technique for enhanced sampling of CS of molecules. Here, a DNN model is trained alongside MD simulations of the molecule of interest such that the models learn a compressed representation of the sampled structures of the molecule. The latent space vectors of the DNN model are then used in ways that provide useful information for inferring appropriate biasing potentials that are then used for guiding the MD simulations, thereby allowing efficient sampling of the molecule's CS. More specifically, the use of the KL divergence between the VAE's latent vectors of the current conformation (obtained from the MD simulations) and the VAE's latent vectors of the known, previously sampled, conformations makes it possible to bias the MD simulation away from visiting previously sampled conformations and rather toward visiting previously unsampled conformations.

The AI-based enhanced sampling approach presented in this work is not dependent on having prior knowledge of the molecule's CS distribution and does not require any careful

selection of collective variables. Therefore, this approach is very promising, given that the selection of appropriate collective variables is often very challenging (Tribello et al., 2010), and there is no well-defined solution that can fit all situations/all molecular systems. Rather than requiring manual specification of the collective variables to use, DESP, by itself, learns the compressed representation of the molecular system of interest and derives biasing potentials based on the distribution of the molecule's conformations in that compressed representation space. The results obtained show that DESP outperforms both conventional and accelerated MD simulations, and efficiently samples wider CS than conventional and accelerated MD simulations. Furthermore, the ideas in DESP are generalizable and may be used for implementing other forms of biased MD simulations including targeted and steered MD simulations. In the next section, we present the methods that make DESP possible and thereafter the overall DESP algorithm.

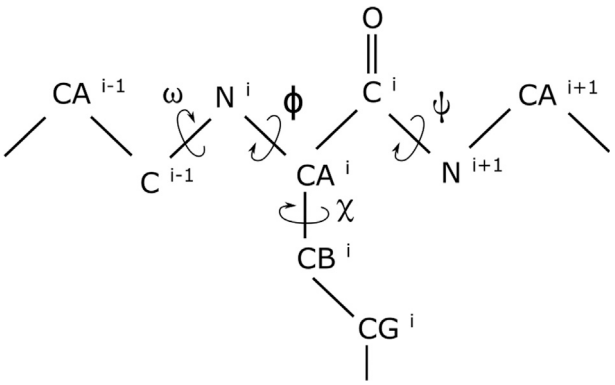


FIGURE 1 | Dihedral angles in a short segment of a protein. While the omega (ω), phi (ϕ), and psi (ψ) dihedral angles are in the proteins backbone, the chi_1 (χ1) dihedral angle is at the beginning of an amino acid's side chain.

MATERIALS AND METHODS

Protein Molecules Used

We began with a smaller protein/peptide (alanine dodecapeptide with 12 alanine residues, A_{12}) and modeled its 3D structure using RPBS (Alland et al., 2005). The small size of alanine dodecapeptide helped in the initial testing and fine-tuning of DESP. In addition to A_{12}, we obtained a solution nuclear magnetic resonance (NMR) structure of GB98 that was expressed in *Escherichia coli* BL21 (DE3) from the Protein Data Bank (Berman et al., 2000), PDB ID: 2lhd (He et al., 2012). GB98 was selected because of its relatively small/medium size and because of the presence of the various secondary structure types (namely, alpha-helix, beta-sheet, and coils) in it. On the other hand, any protein could be used for the demonstration of the functionality of DESP, and the ones used here are just examples.

Creation of the Initial Molecular Systems

Assignment of appropriate residues' charges and protonation states were handled using PDB2PQR (Dolinsky et al., 2007; Jurrus et al., 2018). Using AmberTools18's tLeap (Pearlman et al., 1995; Case et al., 2005; Salomon-Ferrer et al., 2013a; Salomon-Ferrer et al., 2013b), ff14SB (Maier et al., 2015) force-fields for the proteins, and ions234lm_126_tip3*p* for the ions and the water molecules (Li and Merz, 2014), we created explicitly solvated molecular systems for A_{12}'s and GB98's molecular dynamics (MD) simulations with OpenMM (Eastman et al., 2017) containing 2068 TIP3P water molecules (42.38Å × 48.80Å × 47.15Å box size, for A_{12}) or 9,981 (101.10Å × 94.35Å × 98.34Å box size, for GB98) TIP3P water molecules.

Energy Minimization and Heating

Each of the molecular systems was energy-minimized using OpenMM (Eastman et al., 2017). The energy minimizations were done in two stages—weakly (2.5 kcal/mol/Å^2) restraining all the alpha carbon atoms in the first stage, and without any restraints in the second stage. With the weak restraints (2.5 kcal/mol/Å^2) reapplied on the alpha carbon atoms, the molecular systems were steadily heated to a temperature of 310 K in a canonical ensemble using the Langevin thermostat (Pastor et al., 1988).

Conventional Molecular Dynamics Simulations

During both the equilibration and production runs, we controlled the systems' temperatures and pressures using the Langevin thermostat (Pastor et al., 1988) with a collision frequency of 2ps^{-1} and the Monte Carlo barostat (Chow and Ferguson, 1995; Åqvist et al., 2004), respectively. Full electrostatic interaction energies were calculated using the particle mesh Ewald method (Darden et al., 1993). A cutoff distance of 10Å and a cubic spline switch function were used when calculating nonbonded interactions. All bonds in which at least one atom is hydrogen are constrained using the SHAKE algorithm (Ryckaert et al., 1977). All production run MD simulations were performed at 2 femtoseconds time step. Overall, the results from 800 ns of conventional MD simulations, 800 ns of accelerated MD simulations, and 280 ns of DESP MD simulations are presented in this work for each of the A_{12} and the GB98 molecular systems.

Representations of the Molecules for Deep Learning Modeling

Since considerable changes in the conformation of biomolecules can be captured by variations in the dihedral angles of the molecules (Salvador, 2014; Cukier, 2015; Ostermeir and Zacharias, 2014; Lemke and Peter, 2019), we represent a molecule's conformation by the cosine and the sine of the dihedral angles (Mu et al., 2005) of that conformation. For these, we make use of the omega (ω), phi (ϕ), psi (ψ), and chi_1 (χ1) dihedral angles (with examples illustrated in **Figure 1**). Although using both the cosine and the sine of each of the dihedral angles doubles the dimensionality, it helps

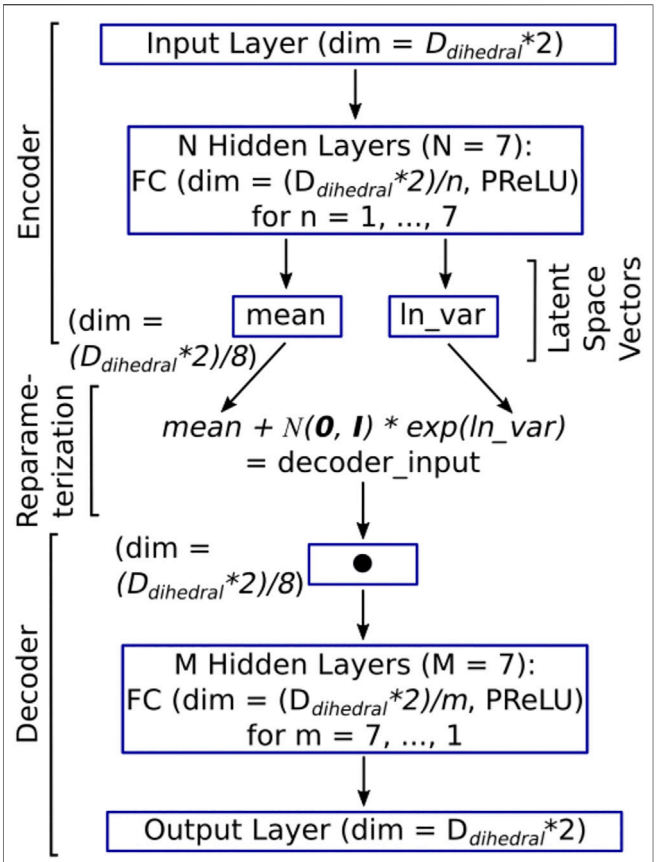


FIGURE 2 | The architecture of the DESP's VAE model. The dimension of the input layer (as well as the output layer) is two times the number of dihedral angles because both the cosine and the sine of each of the dihedral angles are used to deal with periodicity issues (Mu et al., 2005). Each of the hidden layers is a fully connected (FC) layer, followed by parameterized rectified linear units (PReLUs). The latent space between the encoder and the decoder has a dimension that is one-eighth of the input dimension to learn a compressed representation of the molecules in a reduced dimension. The encoder and the decoder are connected through the re-parametrization trick wherein samples are selected from the standard normal distribution, $\mathcal{N}(0, I)$, and then scaled by the variance and shifted by the mean.

in removing the adverse effect that the periodicity of the dihedral angles would have had on the modeling. Extensive details of the benefits of using the dihedral angles (Lemke and Peter, 2019) and of simultaneously using both their cosines and sines have been documented elsewhere (Mu et al., 2005) and, for brevity, are not repeated here.

DNN Architecture: Variational Autoencoder

Our DNN of the type variational autoencoder (VAE) has a simple architecture, as shown in **Figure 2**. The input layer takes both the cosine and the sine of the dihedral angles' representation of the molecular conformation (giving rise to a vector of dimension $D_{Dihedrals}$*2) as input. The input layer is followed by N hidden layers (where $N = 7$ in the current case). Each of the hidden layers, numbered $n = [1, 2 \ldots N]$, has $D_{Dihedrals}$/n nodes. The next layer is made up of two latent space vectors, each of size ($D_{Dihedrals}$*2)/(N + 1), which is ($D_{Dihedrals}$*2)/8 in the current case. The first latent space vector represents the mean for the Gaussian distribution that the latent space encodes (i.e., mean in **Figure 2**), while the second vector represents the natural logarithm of the variance for the Gaussian distribution that the latent space encodes (i.e., ln_var in **Figure 2**). The DNN architecture up to this point is the encoder (**Figure 2**).

The decoder, which is like a mirror image of the encoder, begins with an input layer with ($D_{Dihedrals}$*2)/(N + 1) nodes and is followed by M hidden layers (where M = 7 in the current case). Each of the hidden layers, numbered m = [M, M-1 ... 1], has ($D_{Dihedrals}$*2)/m nodes. The output (which is the final) layer emits the reconstructed cosine and sine of the dihedral angles of the molecular conformation that was passed in as input. To allow the passage of backpropagation signals through the entire VAE (i.e., from the decoder to the encoder), we connect the encoder and the decoder by a re-parameterization trick that is made up of an equation that takes the output of the encoder (namely, the vector of mean, and the vector of the logarithm of variance) as an input and uses it to sample from the corresponding normal distributions. This is done indirectly by initially drawing samples from the standard normal distribution. The samples drawn are then scaled and shifted accordingly using the variance vector and the mean vector, thereby obtaining the intended distribution (see the re-parameterization expression in **Figure 2**). The output of the re-parameterization is then fed into the decoder's input layer (**Figure 2**). We used PyTorch (Paszke et al., 2019) with CUDA support for building all deep neural network models in this study. Given the architecture of the DNN and its inputs and outputs, we can now examine how the DNN is trained.

DNN Training

We defined the model's loss function as a weighted combination (**Eq. 1**) of reconstruction loss captured by mean square error (MSE) loss (**Eq. 2**) and the Kullback–Leibler (KL) divergence loss (**Eq. 3**). We set the weighting parameter, w, to 0.1 so that the MSE loss has a higher weight (1–0.1 = 0.9) than the KL divergence loss (0.1). We arrived at this weighting scheme from our preliminary experiments through grid search, wherein we observed that setting the KL divergence's weight to 0.1 helped in the faster convergence of the model loss and in achieving a much better reconstruction accuracy for the trained model, on both the training dataset and validation dataset.

$$Loss_{model} = (1 - w) * Loss_{MSE} + (w) * Loss_{KL} \tag{1}$$

$$Loss_{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{2}$$

$$\begin{aligned} Loss_{KL} &= D_{KL} \left\{ N \left[(\mu_1, \ldots, \mu_n)^T, diag(\sigma_1^2, \sigma_n^2) \right] \middle\| \mathcal{N}(0, \boldsymbol{I}) \right\} \\ &= \frac{1}{2} \sum_{i=1}^{n} \left[\sigma_i^2 + \mu_i^2 - \ln(\sigma_i^2) - 1 \right] \end{aligned} \tag{3}$$

The KL Divergence upon which **Eq. 3** is based represents a special case involving the KL-divergence between a multivariate normal distribution, $N[(\mu_1, \ldots, \mu_n)^T, diag(\sigma_1^2, \ldots, \sigma_n^2)]$ with means $\mu_1, \ldots, \mu_n$ and variances $\sigma_1^2, \ldots, \sigma_n^2$, and a standard normal distribution, $\mathcal{N}(0, \boldsymbol{I})$.

Initialize $N_{Completed} \leftarrow 0;\ N_{Needed} \leftarrow 1e9;\ N_{Short} \leftarrow 1e7;\ N_{Saving} \leftarrow 1e4;$
$N_{Biasing} \leftarrow 50;\quad N_{BeforeRetraining} \leftarrow 100;$
$S_{Frames} \leftarrow \{\} = Empty\ Set;\ N_{Strata} \leftarrow 10$
For i *in range* $\left(1\ to\ \frac{N_{Short}}{N_{Saving}}\right)$ **do**
  Run *conventional MD simulations for* N_{Saving} *steps*
  Save *the conventional MD simulations frame to* S_{Frames}
  Update $N_{Completed} \leftarrow N_{Completed} + N_{Saving}$
Train *the VAE using the frames in* S_{Frames} *(or a subset of it)*
While $N_{Completed} < N_{Needed}$ **do**
  Set $N_{AvailableFrames} \leftarrow \frac{N_{Completed}}{N_{Saving}};\ N_{EachStartum} \leftarrow \frac{N_{AvailableFrames}}{N_{Strata}}$
  For i *in range* $\left(1\ to\ \frac{(N_{Needed} - N_{Short})}{(N_{Saving} * N_{BeforeRetraining})}\right)$ **do**
    For j *in range* $(1\ to\ N_{BeforeRetraining})$ **do**
      For k *in range* $\left(1\ to\ \frac{N_{Saving}}{N_{Biasing}}\right)$ **do**
        Set $V_{KL} \leftarrow$
        $$\frac{\sum_{l=1}^{N_{Strata}} KL_{Div}\left(Encoder(Frame_{Last}),\ Encoder\left(Frame_{(l-1)*N_{EachStartum}}\right)\right)}{N_{Strata}}$$
        Set $V_{Biasing} \leftarrow \left(\frac{V_{KL\_Upper}}{V_{KL}}\right)^2$
        Run *biased MD simulations for* $N_{Biasing}$ *steps while adding* $V_{Biasing}$ *to the protein atoms'potential energy*
        Keep *track of the* $V_{Biasing}$ *used*
        Update $N_{Completed} \leftarrow N_{Completed} + N_{Biasing}$
      Save *the biased MD simulations frame to* S_{Frames}
    Train *the VAE further using the frames in* S_{Frames} *(or a subset of it)*
Reweigh *the biased MD simulations trajectory based on* $V_{Biasing}$

FIGURE 3 | DESP algorithm. The DESP combines DNNs with MD simulations to achieve enhanced sampling of molecules' conformation spaces.

For the minimization of the loss and, thus, the training of the model, we used the Adam optimizer proposed by (Kingma and Ba, 2014) and with the modifications proposed by (Reddi et al., 2019). We initialized the learning rate to 1e-4, the betas [which are used for computing the running averages of gradient and its square (Paszke et al., 2019)] to 0.9 and 0.999, and the weight decay (which is a form of L2 regularization penalty) to 0.01. We used a multistep learning rate scheduler to gradually reduce the learning rate as the training proceeds through 50 equally distributed epoch milestones. At each of the milestones, the new learning rate is obtained by multiplying the current learning rate by 0.99. We used a batch size of 512 and set out to run 5,000 epochs in the initial training of the model. We adopt early stopping if the model does not improve over 250 consecutive epochs, in which case we would retain the last known best model and stop further training of the model.

DESP: Deep Enhanced Sampling of Proteins' Conformation Space Algorithm

Having described the individual components of the DESP above, we now present the overall DESP algorithm (**Figure 3**) that combines DNN with MD simulations to achieve enhanced sampling of the conformation space of macromolecules. It begins with the initialization of the total number of MD simulation steps needed (e.g., N_{Needed} = 1e9), the number of MD simulation steps for the initial short MD run (N_{Short} = 1e7) that will be used for the initial DNN model training, the total number of steps completed ($N_{Completed}$ = 0), the number of steps to run before saving a frame (N_{Saving} = 1e4), and the total number of steps completed before updating the biasing potentials ($N_{Biasing}$ = 50). While $N_{Completed}$ essentially ranges from 0 to N_{Needed} over time, the other variables are relatively as follows:

$$N_{Biasing} \ll N_{Saving} \ll N_{Short} \ll N_{Needed} \tag{4}$$

We use "<<" to signify a difference of one order of magnitude or more. $N_{Biasing}$, N_{Saving}, N_{Short}, N_{Needed}, and $N_{Completed}$ are natural numbers.

We run N_{Short} steps of unbiased MD simulation, saving frames for every N_{Saving} steps. The saved frames are added to a pool of frames (i.e., set S_{Frames}): increase $N_{Completed}$ by N_{short} (i.e., $N_{Completed} \leftarrow N_{Completed} + N_{Short}$). We use the MD simulation's frames in S_{Frames} (or its subset, selected randomly) to train the VAE and save the trained VAE ($VAE_{Trained}$). While $N_{Completed}$ is less than N_{Needed}, we continue the biased MD simulations coupled with the usage of the $VAE_{Trained}$ and its further training as follows. 1) Calculate the KL divergence (using latent vectors of the $VAE_{Trained}$'s means and variance, based on **Eq. 5**) between the last frame of the MD simulations and the representative/sampled structures from pool S_{Frames}. 2) Run the ongoing MD simulation for $N_{Biasing}$ steps, but now by adding a biasing potential ($V_{Biasing}$, as defined in **Eq. 6**) that is based on the KL divergence. Keep track of the $V_{Biasing}$. Increase $N_{Completed}$ by $N_{Biasing}$ (i.e., $N_{Completed} \leftarrow N_{Completed} + N_{Biasing}$). 3) For every N_{Saving} steps of the MD simulations, add the new frame to pool S_{Frames}. And for every N_{Saving} * 100 steps of the MD simulations (which means that additional 100 new frames would have been added to S_{Frames}), we use the frames in S_{Frames} (or its subset, selected randomly) to further train the $VAE_{Trained}$. When the $N_{Completed}$ is equal to N_{Needed}, we stop the MD simulations and use the trajectory of $V_{Biasing}$ to reweigh the MD simulation trajectory.

The KL Divergence upon which the biasing potential is based involves pairs of multivariate normal distributions of the same dimension and can be represented by **Eq. 5**, which denotes the KL divergence of $\mathcal{N}_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ from $\mathcal{N}_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$.

$$V_{KL} = D_{KL}(\mathcal{N}_0 \,||\mathcal{N}_1) = \frac{1}{2}\left(\mathrm{tr}\left(\Sigma_1^{-1}\Sigma_0\right) + \left(\mu_1 - \mu_0\right)^T \sum_1^{-1}\left(\mu_1 - \mu_0\right) - \mathrm{k} + \ln\left(\frac{det\Sigma_1}{det\Sigma_0}\right)\right) \quad (5)$$

$$V_{biasing} = \left(V_{KL_{upper}} \,/\, V_{KL}\right)^2 \quad (6)$$

where $V_{KL_{upper}}$, which is set to 1e-5 in this work, is a weak upper bound of V_{KL}. $V_{KL_{upper}}$ is a settable parameter but can be left at this default value obtained from our preliminary experiments where this provided optimal enhanced sampling without making the system unstable. This value can be tuned up or down to modulate how aggressive (high $V_{KL_{upper}}$) or conservative (low $V_{KL_{upper}}$) the enhancement of the sampling should be. The obtained $V_{Biasing}$ is added to the potential energy term involving the protein atoms.

At this point, we find it important to further clarify that the use of dihedral angles as input to the VAE in DESP does not mean that dihedral angles are being used directly as the reaction coordinates for biasing the MD simulations. Using all the dihedral angles by themselves would be overwhelming (especially for medium-sized to large-sized molecules) and, more importantly, will not work if used directly even with existing enhanced sampling methods. On the other hand, the VAE learns the compressed representation of the molecular system, and it is the compressed representation (obtainable from the latent space vectors of the VAE, see **Figure 2**) that is used for achieving the biasing, as presented in the algorithm (see **Figure 3**). In other words, generally, a bias potential V(R) used in the MD simulation would depend on R, the atomistic coordinate, usually through some collective variables. The same is, in principle, true in the current work, except that the bias potential V(R) used in DESP depends on R', where R' is a compressed representation of R that is obtained from the DNN.

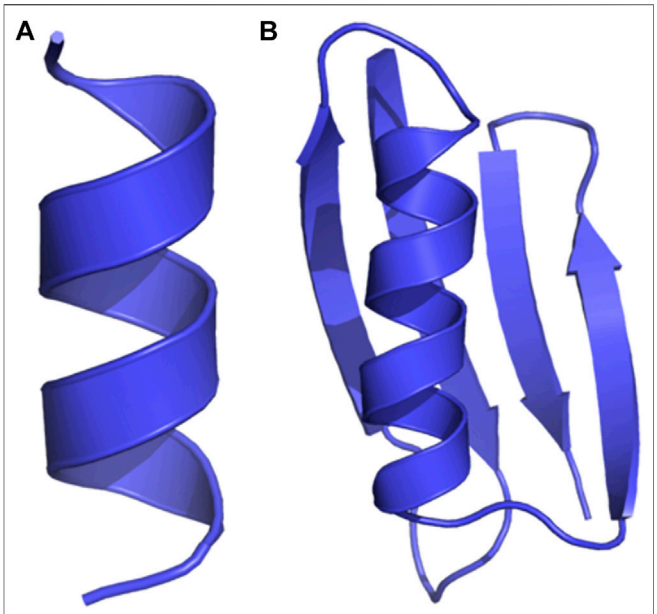


FIGURE 4 | Initial structures of the proteins studied. **(A)** A_{12} is a dodecapeptide with 12 alanine residues, and **(B)** GB98 is a small protein with four beta-sheets and one alpha helix.

Reweighing of the Probability Distribution

The probability, p' (RC), along a reaction coordinate of interest, RC (r), where r represents the atomic coordinates $r_1^3, \ldots, r_n^3$, based on the biased MD simulations can be reweighed using $V_{Biasing}$ to obtain the un-normalized probability distribution, p (RC), of the canonical ensemble (Sinko et al., 2013; Miao et al., 2015; Salawu, 2018a) as shown in **Eq. 7**. And the reweighed free energy change can be obtained from **Eq. 8**.

$$p(RC_a) = p'(RC_a) * \frac{e_a^{\beta V_{biasing}}}{\sum_{a=1}^{M} e_a^{\beta V_{biasing}}} \; for \; a = 1, \ldots, M \quad (7)$$

where $\beta = -\frac{1}{k_B T}$.

$$F(RC_a) = \beta \ln p(RC_a) \quad (8)$$

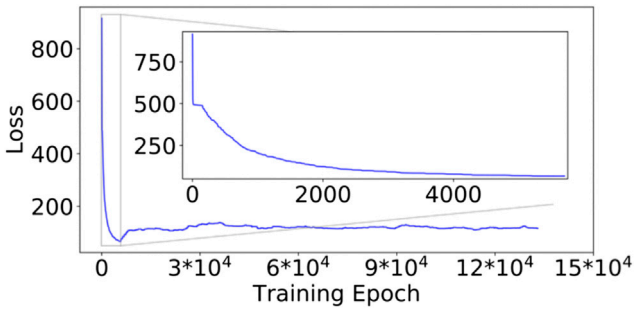


FIGURE 5 | Model loss values during the DESP for the GB98 molecular system. The loss decreased steadily in the first segment of the DESP (see the inset). The model's loss is slightly higher in the subsequent training of the model because the model was exposed to a more diverse molecular structure. The trajectory of the loss for the A_{12} molecular system is similar in the overall structure/trend to that of the GB98 molecular system and is not shown here for brevity. DESP systematically modifies the molecular system's energy surface.

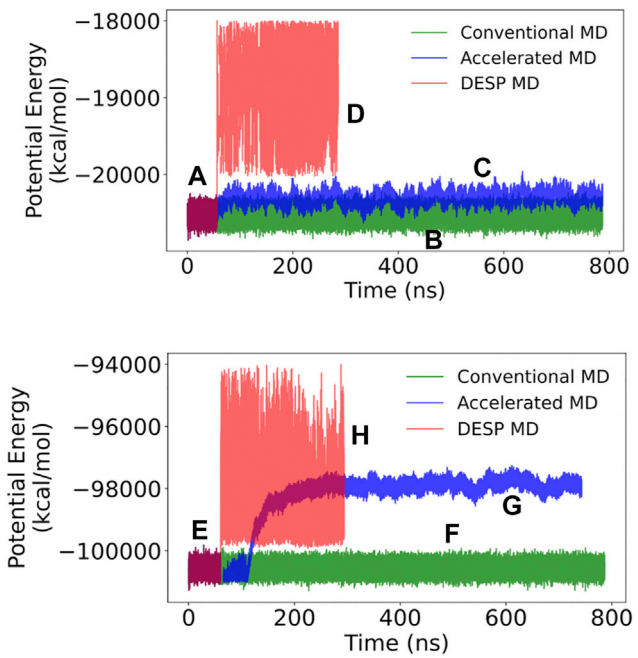


FIGURE 6 | The potential energy for A_{12} (top) and GB98 (bottom) molecular systems. **(A and E)** The initial stages of DESP (as well as the initial stages of accelerated) MD simulations are identical to conventional MD simulations and have identical systems' potential energies. **(B and F)** The trajectories of the potential energy for the conventional MD simulations are shown in green; **(C and G)** those for the accelerated MD simulations are shown in blue; while **(D and H)** those for the DESP MD simulations are shown in red.

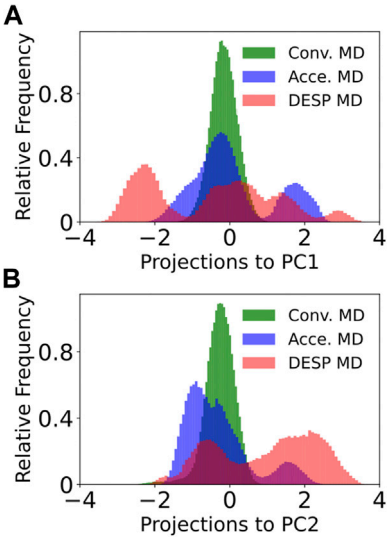


FIGURE 7 | Projections of the trajectory to the first principal component for GB98 (bottom). The projection of each of the frames in the DESP trajectory into the PC1's space **(A)** and PC2's space **(B)** are shown for GB98. Similar projections for the A_{12} molecular system do not offer additional information and are now shown here for brevity.

RESULTS AND DISCUSSION

GB98 is a Small Protein With One α- and Four β-Folds, While A_{12} is a Dodecalanine

We show the initial 3D structure of the studied molecules, A_{12} and GB98, in **Figures 4A,B**, respectively. A_{12} is a peptide with 12 alanine amino acids, while GB98 is a small protein with four beta-sheets, and one alpha-helix (He et al., 2012; Salawu, 2016). These small-sized and medium-sized molecules helped in illustrating the capabilities of DESP.

DNN Model Loss During the DESP

The initial training of the DESP's VAE started with a high model total loss ($Loss_{Model}$, **Eq. 1**) of approximately 915.71 (**Figure 5**, for the GB98 molecular system), which decreased steadily as the model continued to learn the compressed representation of the molecule under study (inset of **Figure 5**). The initial model training was stopped when the $Loss_{Model}$ reached 66.65 after 5,664 epochs and would not further decrease for the next 250 epochs. The $Loss_{Model}$ during the subsequent training of the DNN alongside the DNN-biased MD simulations (using the MD simulation's newly generated molecular structures) is shown in the rest of **Figure 5** from epoch 5,664 to the end.

The reader would notice that the $Loss_{Model}$ obtained during the subsequent training of the model alongside the DNN-biased MD simulations is slightly higher than the smallest $Loss_{Model}$ obtained in the initial model training. This is interesting and understandable because the initial training of the DESP's DNNs was done using only the structures/conformations of the molecule obtained from conventional MD simulations in the first segment of the DESP (**Figure 6A**), while the subsequent training of the DNN was done using the more structurally diverse conformations of the molecule obtained during the biasing segment of the DESP (**Figure 6B**).

The initial stages of DESP (as well as the initial stage of the accelerated MD) simulations are identical to those of

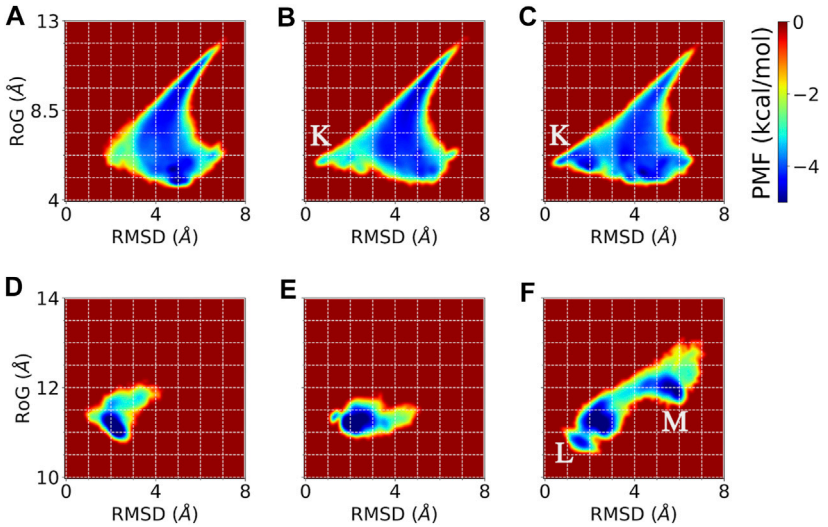


FIGURE 8 | Potentials of mean force (PMF) showing the distribution of the sampled conformations by conventional MD, accelerated MD, and DESP MD simulations of A_{12} (top) and GB98 (bottom) molecular systems. Reweighting has been done wherever necessary. Comparison using PMF based on physically meaningful collective variables, namely, the root mean square deviation from a known experimental/initial structure (RMSD) and the radius of gyration (RoG) are shown for conventional **(A and D)**, accelerated **(B and E)**, and DESP **(C and F)** MD simulations for the A_{12} (top/A, B, C) and the GB98 (bottom/D, E, F) molecular systems. Overall, one would notice that the rightmost panels **(C and D)** show wider and more diverse regions visited by the molecular system, which means that the DESP can explore more conformation spaces than either the conventional **(A and B)** or the accelerated **(B and E)** MD simulations for these collective variables. The regions with stable conformations sampled by both DESP and accelerated MD simulations but not sampled by the conventional MD simulations are marked with "K," while the regions sampled by DESP alone but not sampled by either the conventional or the accelerated MD simulations are marked with "L" and "M."

conventional MD simulations, and the molecular systems are observed to have identical systems' energy surfaces/distributions as conventional MD simulations. Indeed, in the current work, for a given molecular system, after equilibration, the ongoing conventional MD simulation is forked/copied into three: one for continuation as a conventional MD simulation, one for continuation as an accelerated MD simulation, and one for continuation as a DESP MD simulation. At the start of the biasing phase of DESP, we observed that the molecular systems' energies are modified and the potential energies increase and change based on the conformation being sampled (**Figures 6A–D** for the A_{12} molecular system, and **Figures 6E–H** for the GB98 molecular system; not drawn to the same scale for all the MD simulations). The modification of the systems' potential energies makes it possible for the system to escape possible energy barriers, thereby encouraging the sampling of wider conformation spaces (**Figure 7**; **Figure 8**).

DESP Efficiently Samples a Wider Range of a Molecule's Conformation Space Than Both Conventional and Accelerated MD Simulations

To compare the conformation spaces sampled by DESP to that sampled by conventional and accelerated MD simulations, we carried out dihedral principal components analysis (dPCA) on the molecule's dihedral angles (namely, phi, psi, omega, and chi_1) by making use of both the cosine and the sine of each of the dihedral angles (Mu et al., 2005) and projected each of the sampled structures from the DESP and from both the conventional and accelerated MD simulations into the principal components' (PC) space. A visualization of the trajectory in the PC space (see **Figure 7** for the first two PCs, PC1 and PC2) shows that DESP samples a wider range of the molecule's conformation spaces than the conventional and the accelerated MD simulations (**Figure 7**) despite that the DESP is just about one-third as long (i.e., ~280 ns) as the conventional and the accelerated MD simulations (i.e., ~800 ns, **Figure 6**; **Figure 7**).

It is worthy of note that the distributions shown in **Figure 7** are unweighed and cannot be strictly interpreted in the probability sense most especially for the DESP and for the accelerated MD simulations that involve the use of biasing potentials. It is, therefore, important to reweigh any DESP-obtained (or accelerated-MD-obtained) distribution while considering the biasing potentials (Salawu, 2018a; Sinko et al., 2013; Miao et al., 2015). Such reweighting can be achieved through **Eqs 7**, **8** or as described in previous publications (Sinko et al., 2013; Miao et al., 2015; Salawu, 2018a).

For the potentials of mean force (PMF) obtainable through the reweighting of the trajectories, we use two physically interpretable/physically meaningful reaction coordinates, namely, the molecule's radius of gyration (RoG) and the molecule's root mean square deviation (RMSD) from the experimentally solved structure (i.e., the NMR structure in the case of GB98) or the initial structure (i.e., the energy minimized modeled structure in the case of A_{12}). The PMF obtained from the reweighed trajectory (**Figure 8**) further establishes that DESP

samples have much wider conformation spaces than the conventional and the accelerated MD simulations. We show the PMF obtained from collective variables (CVs) defined by the combination of the RMSD and the RoG in **Figure 8**.

From the energy landscape one sees regions with stable conformations that are sampled by DESP but are not sampled by the conventional MD simulations (see **Figures 8A–C** for the A_{12} molecular system, and **Figures 8D–F** for the GB98 molecular system). For the sake of illustration, we mark the regions sampled by both DESP and accelerated MD simulations but not sampled by the conventional MD simulations with "K" (**Figures 8B,C**), and we mark the regions sampled by DESP alone but not sampled by either the conventional MD simulations or the accelerated MD simulations with "L" and "M" (**Figure 8F**). Overall, one would notice that the rightmost panels (C, D) show wider and more diverse regions visited by the molecular system, which means that the DESP can explore more conformation spaces than either the conventional (A, B) or the accelerated (B, E) MD simulations, to the extent of capturing a few global but moderate unfolding and refolding events. The comparison of the energy landscapes shows that while DESP shows a moderately better sampling of a wider range of conformation spaces than both the conventional and the accelerated MD simulations for a small molecular system (namely, A_{12}, **Figures 8A–C**), the superiority of the sampling efficiency of DESP is more remarkably evident for larger molecules as shown by the medium-sized GB98 molecular system wherein DESP samples much wider regions/conformations spaces than both the conventional and the accelerated MD simulations (**Figures 8D–F**). This is desirable because it is with the larger molecules that highly efficient conformation space sampling methods, such as DESP, are most needed.

CONCLUSION

In this work, 1) It has been shown, with computational experiments and pieces of evidence obtained therefrom, that it is possible to enhance the MD simulation sampling of molecules' conformation spaces using deep learning techniques (VAE in the current case). 2) It has been shown one of the possible ways with which it could be achieved, namely, by biasing the MD simulations based on the VAE's latent space vectors. 3) The use of the KL divergence of the DNN-learned latent space vectors of the most recently sampled conformation from the previously sampled conformations made it possible to bias the MD away from visiting already sampled conformations, and thereby encouraging the sampling of previously unsampled states. 4) It should be noted that the ideas in DESP are generalizable and may be used for implementing other forms of biased MD simulations, including targeted and steered MD simulations, and we explore these in our subsequent articles.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Alland, C., Moreews, F., Boens, D., Carpentier, M., Chiusa, S., Lonquety, M., et al. (2005). RPBS: a web resource for structural bioinformatics. *Nucleic Acids Res.* 33, W44. doi:10.1093/nar/gki477

Allison, J. R. (2020). Computational methods for exploring protein conformations. *Biochem. Soc. Trans.* 48 (4), 1707–1724. doi:10.1042/bst20200193

Åqvist, J., Wennerström, P., Nervall, M., Bjelic, S., and Brandsdal, B. O. (2004). Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem. Phys. Lett.* 384 (4–6), 288–294. doi:10.1016/j.cplett.2003.12.039

Babin, V., Roland, C., and Sagui, C. (2008). Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* 128 (13), 134101. doi:10.1063/1.2844595

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28 (1), 235–242. doi:10.1093/nar/28.1.235

Bonati, L., Zhang, Y. Y., and Parrinello, M. (2019). Neural networks-based variationally enhanced sampling. *Proc. Natl. Acad. Sci. U.S.A.* 116 (36), 17641–17647. doi:10.1073/pnas.1907975116

Bussi, G., and Laio, A. (2020). Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* 2, 1–13. doi:10.1038/s42254-020-0153-0

Carlson, H. A., and McCammon, J. A. (2000). Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* 57 (2), 213–218.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The amber biomolecular simulation programs. *J. Comput. Chem.* 26 (16), 1668–1688. doi:10.1002/jcc.20290

Chen, W., and Ferguson, A. L. (2018). Molecular enhanced sampling with autoencoders: on-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* 39 (25), 2079–2102. doi:10.1002/jcc.25520

Chiti, F., and Dobson, C. M. (2017). Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu. Rev. Biochem.* 86, 27–68. doi:10.1146/annurev-biochem-061516-045115

Chow, K. H., and Ferguson, D. M. (1995). Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Comput. Phys. Commun.* 91 (1–3), 283–289. doi:10.1016/0010-4655(95)00059-o

Council, National Research, and others (1989). Opportunities in biology. National academies. Available at: https://www.nap.edu/read/742/chapter/3 (Accessed February 22, 2021).

Cukier, R. I. (2015). Dihedral angle entropy measures for intrinsically disordered proteins. *J. Phys. Chem. B* 119 (9), 3621–3634. doi:10.1021/jp5102412

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98 (12), 10089–10092. doi:10.1063/1.464397

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., et al. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35, W522. doi:10.1093/nar/gkm276

Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Beauchamp, K. A., Wang, L.-P., et al. (2017). OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *Plos Comput. Biol.* 13 (7), e1005659. doi:10.1371/journal.pcbi.1005659

Guo, T., Noble, W., and Hanger, D. P. (2017). Roles of tau protein in health and disease. *Acta Neuropathol.* 133 (5), 665–704. doi:10.1007/s00401-017-1707-9

Hamelberg, D., Mongan, J., and McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120 (24), 11919–11929. doi:10.1063/1.1755656

Harada, R., and Kitao, A. (2013). Parallel cascade selection molecular dynamics to generate conformational transition pathway. *J. Chem. Phys.* 139, 035103. doi:10.1063/1.4813023

Hartl, F. U. (2017). Protein misfolding diseases. *Annu. Rev. Biochem.* 86, 21–26. doi:10.1146/annurev-biochem-061516-044518

He, Y., Chen, Y., Alexander, P. A., Bryan, P. N., and Orban, J. (2012). Mutational tipping points for switching protein folds and functions. *Structure* 20 (2), 283–291. doi:10.1016/j.str.2011.11.018

Hénin, J., and Chipot, C. (2004). Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* 121 (7), 2904–2914. doi:10.1063/1.1773132

Ittisoponpisan, S., Khanna, T., Alhuzimi, E., David, A., and Sternberg, M. J. E. (2019). Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? *J. Mol. Biol.* 431 (11), 2197–2212. doi:10.1016/j.jmb.2019.04.009

Jurrus, E., Engel, D. F., Star, K. B., Monson, K., Brandi, J., Felberg, L. E., et al. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* 27 (1), 112–128. doi:10.1002/pro.3280

Karplus, M., and Kuriyan, J. (2005). Molecular dynamics and protein function. *Proc. Natl. Acad. Sci.* 102 (19), 6679–6685. doi:10.1073/pnas.0408930102

Kingma, D. P., and Ba, J. (2014). A method for stochastic optimization. ArXiv [Preprint]. Available at: http://arXiv:1412.6980 (Accessed December 22, 2014).

Laio, A., and Gervasio, F. L. (2008). Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* 71 (12), 126601. doi:10.1088/0034-4885/71/12/126601

Laskowski, R. A., Stephenson, J. D., Sillitoe, I., Orengo, C. A., and Thornton, J. M. (2020). VarSite: disease variants and protein structure. *Protein Sci.* 29 (1), 111–119. doi:10.1002/pro.3746

Lee, Y., Basith, S., and Choi, S. (2018). Recent advances in structure-based drug design targeting class A G protein-coupled receptors utilizing crystal structures and computational simulations. *J. Med. Chem.* 61 (1), 1–46. doi:10.1021/acs.jmedchem.6b01453

Lemke, T., and Peter, C. (2019). EncoderMap: dimensionality reduction and generation of molecule conformations. *J. Chem. Theor. Comput.* 15 (2), 1209–1215. doi:10.1021/acs.jctc.8b00975

Li, P., and Merz, K. M., Jr (2014). Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J. Chem. Theor. Comput.* 10 (1), 289–297. doi:10.1021/ct400751u

Lin, S.-M., Lin, S.-C., Hsu, J.-N., Chang, C.-K., Chien, C.-M., Wang, Y.-S., et al. (2020). Structure-based stabilization of non-native protein-protein interactions of coronavirus nucleocapsid proteins in antiviral drug design. *J. Med. Chem.* 63 (6), 3131–3141. doi:10.1021/acs.jmedchem.9b01913

Maier, J. A., Martinez, C., Kasavajhala, K., WickstromHauser, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theor. Comput.* 11 (8), 3696–3713. doi:10.1021/acs.jctc.5b00255

McCafferty, C. L., and Sergeev, Y. V. (2016). Silico mapping of protein unfolding mutations for inherited disease. *Sci. Rep.* 6 (1), 1–12. doi:10.1038/srep37298

McCarty, J., and Parrinello, M. (2017). A variational conformational dynamics approach to the selection of collective variables in metadynamics. *J. Chem. Phys.* 147 (20), 204109. doi:10.1063/1.4998598

Miao, Y., Feher, V. A., and McCammon, J. A. (2015). Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J. Chem. Theor. Comput.* 11 (8), 3584–3595. doi:10.1021/acs.jctc.5b00436

Moritsugu, K., Terada, T., and Kidera, A. (2012). Disorder-to-order transition of an intrinsically disordered region of sortase revealed by multiscale enhanced sampling. *J. Am. Chem. Soc.* 134 (16), 7094–7101. doi:10.1021/ja3008402

Mu, Y., Nguyen, P. H., and Stock, G. (2005). Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins* 58 (1), 45–52. doi:10.1002/prot.20310

Ostermeir, K., and Zacharias, M. (2014). Hamiltonian replica-exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles. *J. Comput. Chem.* 35 (2), 150–158. doi:10.1002/jcc.23476

Pastor, R. W., Brooks, B. R., and Szabo, A. (1988). An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* 65 (6), 1409–1419. doi:10.1080/00268978800101881

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: an imperative style, high-performance deep learning library. ArXiv [Preprint]. Available at: http://arXiv:1912.01703 (Accessed December 3, 2019).

Pawełand, S., and Caflisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. *Curr. Opin. Struct. Biol.* 48, 93–102. doi:10.1016/j.sbi.2017.10.010

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., et al. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91 (1), 1–41. doi:10.1016/0010-4655(95)00041-d

Reddi, S. J., Kale, S., and Kumar, S. (2019). On the convergence of Adam and beyond. ArXiv [Preprint]. Available at: http://arXiv:1904.09237 (Accessed April 19, 2019).

Ribeiro, J. M. L., Bravo, P., Wang, Y., and Tiwary, P. (2018). Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* 149 (7), 72301. doi:10.1063/1.5025487

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23 (3), 327–341. doi:10.1016/0021-9991(77)90098-5

Salawu, E. O. (2020). Enhanced sampling of nucleic acids' structures using deep-learning-derived biasing forces. *IEEE Symp. Ser. Comput. Intel.* 11, 1648–1654. doi:10.1109/SSCI47803.2020.9308559

Salawu, E. O. (2018b). In silico study reveals how E64 approaches, binds to, and inhibits falcipain-2 of *Plasmodium falciparum* that causes malaria in humans. *Sci. Rep.* 8 (1), 16380. doi:10.1038/s41598-018-34622-1

Salawu, E. O. (2016). Random forests secondary structure assignment for coarse-grained and all-atom protein systems. *Cogent Biol.* 2 (1), 1214061. doi:10.1080/23312025.2016.1214061

Salawu, E. O. (2018a). The impairment of TorsinA's binding to and interactions with its activator: an atomistic molecular dynamics study of primary dystonia. *Front. Mol. Biosci.* 5, 64. doi:10.3389/fmolb.2018.00064

Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013a). An overview of the Amber biomolecular simulation package. *Wires Comput. Mol. Sci.* 3 (2), 198–210. doi:10.1002/wcms.1121

Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013b). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle Mesh Ewald. *J. Chem. Theor. Comput.* 9 (9), 3878–3888. doi:10.1021/ct400314y

Salvador, P. (2014). "Dependencies of J-couplings upon dihedral angles on proteins," in *Annual reports on NMR spectroscopy*. Amsterdam, Netherlands: Elsevier, 185–227.

Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., et al. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* 51 (7), 91–97. doi:10.1145/1364782.1364802

Shaw, D. E., Dror, R. O., Salmon, J. K., Grossman, J. P., Mackenzie, K. M., Joseph, A., et al. (2009). Millisecond-scale molecular dynamics simulations on anton. *Proc. Conf. High Perform. Comput. Netw. Storage Anal.* 65, 1–11. doi:10.1145/1654059.1654126

Sidky, H., Chen, W., and Ferguson, A. L. (2020). Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. *Mol. Phys.* 118 (5), e1737742. doi:10.1080/00268976.2020.1737742

Sinko, W., Miao, Y., de Oliveira, C. A. F., and McCammon, J. A. (2013). Population based reweighting of scaled molecular dynamics. *J. Phys. Chem. B* 117 (42), 12759–12768. doi:10.1021/jp401587e

Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314 (1–2), 141–151. doi:10.1016/s0009-2614(99)01123-9

Tramutola, A., Lanzillotta, C., Perluigi, M., and Butterfield, D. A. (2017). Oxidative stress, protein modification and Alzheimer disease. *Brain Res. Bull.* 133, 88–96. doi:10.1016/j.brainresbull.2016.06.005

Tribello, G. A., Ceriotti, M., and Parrinello, M. (2010). A self-learning algorithm for biased molecular dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 107 (41), 17509–17514. doi:10.1073/pnas.1011511107

Wang, X., Song, K., Li, L., and Chen, L. (2018). Structure-based drug design strategies and challenges. *Curr. Top. Med. Chem.* 18 (12), 998–1006. doi:10.2174/1568026618666180813152921

Wehmeyer, C., and Noé, F. (2018). Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. *J.Chem. Phys.* 148 (24), 241703. doi:10.1063/1.5011399

Yang, L.-Q., Sang, P., Tao, Y., Fu, Y.-X., Zhang, K.-Q., Xie, Y.-H., et al. (2014). Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *J. Biomol. Struc. Dyn.* 32 (3), 372–393. doi:10.1080/07391102.2013.770372

Conflict of Interest: Author EOS is employed by the company Amazon Web Services (AWS). There are no other authors. The work and substantial parts of the manuscript were completed as an independent research scientist before joining AWS.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership