Towards rational computational peptide design

Liwei Chang[1,2†], Arup Mondal[1,2†] and Alberto Perez[1,2]*

[1]Department of Chemistry, University of Florida, Gainesville, FL, United States, [2]Quantum Theory Project, University of Florida, Gainesville, FL, United States

Peptides are prevalent in biology, mediating as many as 40% of protein-protein interactions, and involved in other cellular functions such as transport and signaling. Their ability to bind with high specificity make them promising therapeutical agents with intermediate properties between small molecules and large biologics. Beyond their biological role, peptides can be programmed to self-assembly, and they are already being used for functions as diverse as oligonucleotide delivery, tissue regeneration or as drugs. However, the transient nature of their interactions has limited the number of structures and knowledge of binding affinities available—and their flexible nature has limited the success of computational pipelines that predict the structures and affinities of these molecules. Fortunately, recent advances in experimental and computational pipelines are creating new opportunities for this field. We are starting to see promising predictions of complex structures, thermodynamic and kinetic properties. We believe in the following years this will lead to robust rational peptide design pipelines with success similar to those applied for small molecule drug discovery.

Introduction

Computational modeling is routinely used in the early stages of the drug discovery process to identify molecules that might bind with high affinity to a particular protein receptor. Three key aspects that contribute to the success of computational tools are: 1) the availability of small molecule virtual libraries (e.g., (Irwin and Shoichet, 2005; Shivanyuk et al., 2007; Bento et al., 2014)), 2) the efficiency of docking software to identify candidates from the virtual libraries (Murugan et al., 2022), and 3) the accuracy of physics-based approaches such as alchemical free energy perturbation to determine relative and absolute binding affinities amongst candidate molecules binding a receptor (Bhati et al., 2017; Fratev and Sirimulla, 2019; Lee et al., 2020; Limongelli, 2020; Mey et al., 2020). Some of the limitations include the presence of multiple binding modes, plasticity in the receptor, large conformational changes in the binding molecule, highly charged systems, and comparison across different series of compounds (Chodera et al., 2011; Mobley and Klimovich, 2012; Mey et al., 2020). Despite the progress of computational platforms for drug discovery, small molecule drugs typically require a pocket in the receptor protein in order to bind (Laskowski et al., 1996; Liang et al., 1998; Gao and Skolnick, 2013).

Peptides are flexible molecules that can bind specifically to receptors even in the absence of pockets–targets that were once deemed undruggable by small molecule drugs (Balliu and Baltzer, 2017; Wang et al., 2021). Our cells already use peptides as signaling molecules (Cunha et al., 2008; Foster et al., 2019; Dubas et al., 2021), and many protein-protein interactions take place through peptide epitopes (Milroy et al., 2014; Pelay-Gimeno et al., 2015; Wai et al., 2018; Aiyer et al., 2021). Thus, peptides offer the possibility to inhibit interactions present in disease pathways (Pelay-Gimeno et al., 2015). Furthermore, peptides are highly programmable for self-assembly, allowing the generation of functionalized fibers with applications ranging from scaffolds for tissue regeneration (Loo et al., 2015a) to bioink (Loo et al., 2015b). Successful computational pipelines would allow the rational identification of peptides that bind with high affinity to a particular receptor or have specific self-assembly properties. However, peptide's flexible nature, the large sequence space, and presence of multiple weak interactions that stabilize the complex, have stymied the development of peptide discovery pipelines. Achieving success in modeling peptide complexes will require synergy between several fields which we discuss in this perspective (see Figure 1).

Infrastructure similar to small molecule discovery pipelines are already in place for peptides, with an overall lower success rate. Instead of small molecule virtual libraries, bioinformatic approaches are typically used to derive peptide libraries based on known information from the receptor and known binding partners, reducing the active sequence space from millions (20^n, where n is the length of the peptide) to thousands of sequences. But, dealing with peptide flexibility reduces the efficiency of both docking tools and scoring functions to predict accurate structures (Rentzsch and Renard, 2015; Wang et al., 2016; Ciemny et al., 2018). Surprisingly, a recent implementation of AlphaFold (AF) for peptide docking showed an unprecedented success–despite being trained for a different task (protein structure prediction) (Jumper et al., 2021; Ko and Lee, 2021; Tsaban et al., 2022). Combining search strategies and template-based strategies, PatchMAN has recently surpassed even the successes from AF under certain scenarios, leading the way into the structural characterization of
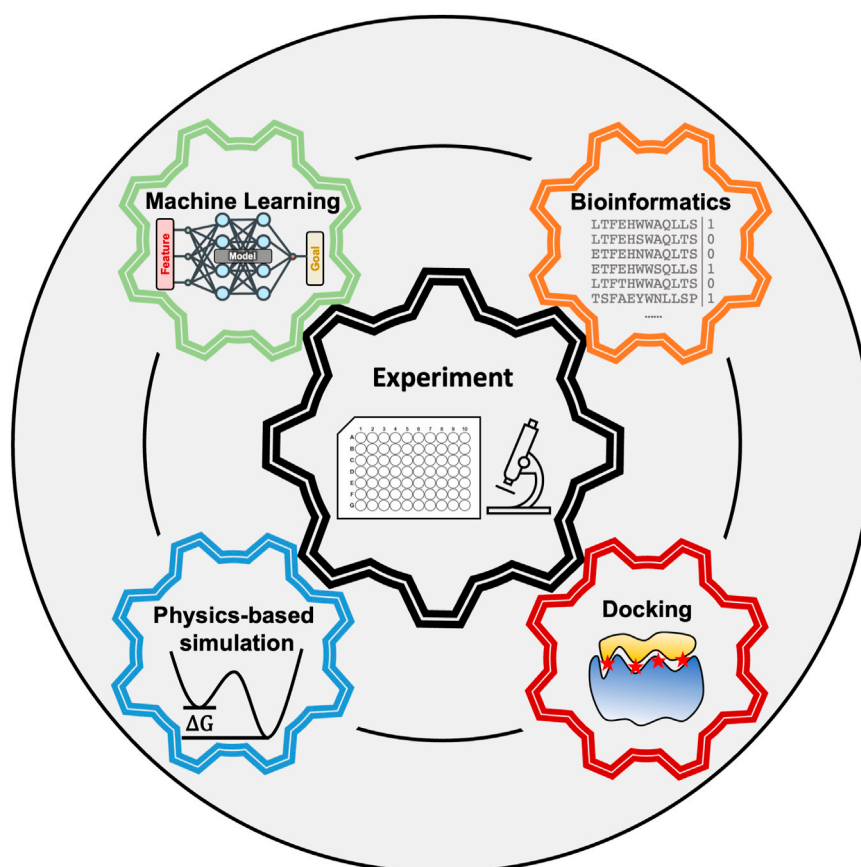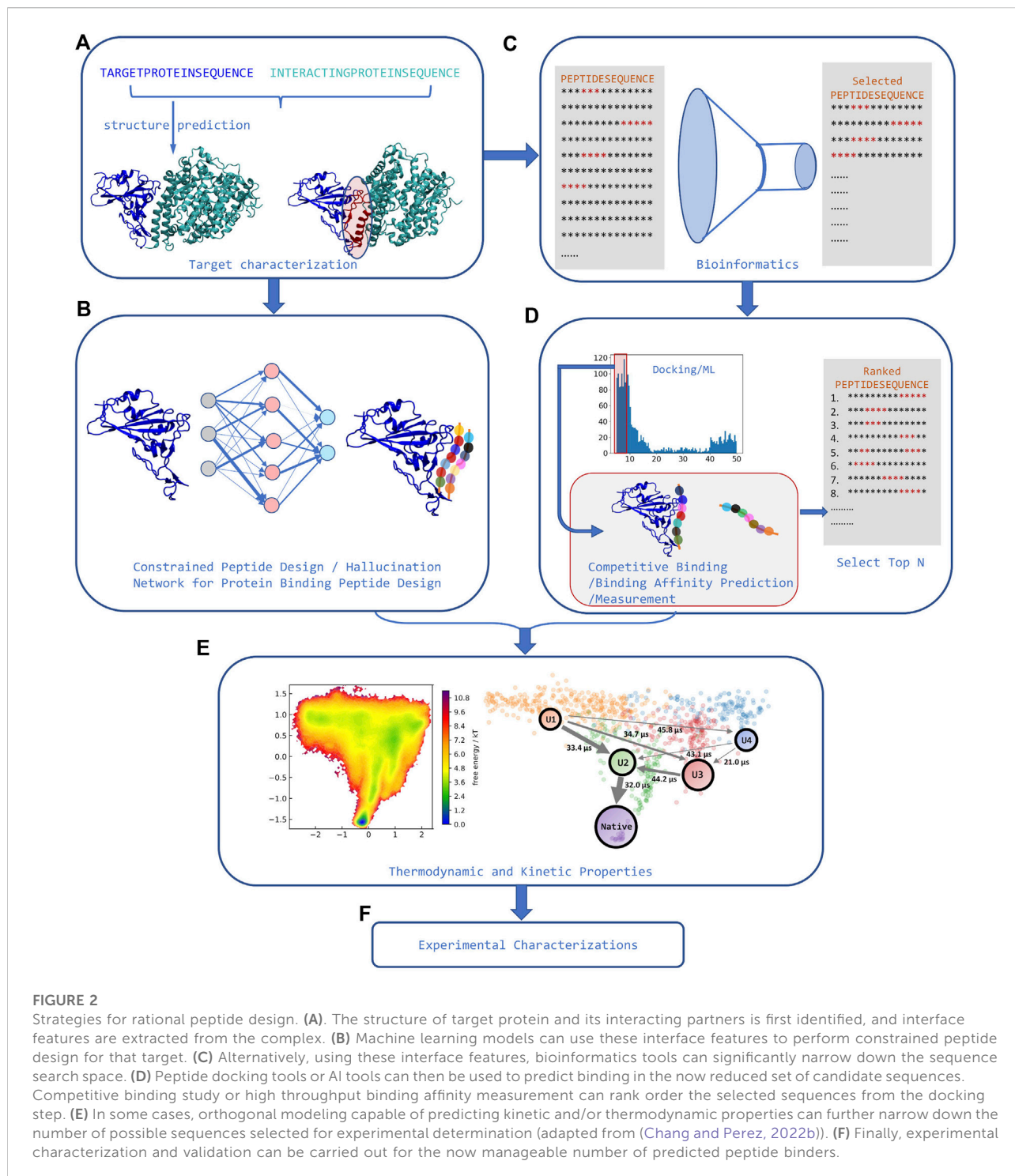

FIGURE 1
Peptide modeling requires synergy between multiple computational techniques and experiments.

FIGURE 2
Strategies for rational peptide design. **(A)**. The structure of target protein and its interacting partners is first identified, and interface features are extracted from the complex. **(B)** Machine learning models can use these interface features to perform constrained peptide design for that target. **(C)** Alternatively, using these interface features, bioinformatics tools can significantly narrow down the sequence search space. **(D)** Peptide docking tools or AI tools can then be used to predict binding in the now reduced set of candidate sequences. Competitive binding study or high throughput binding affinity measurement can rank order the selected sequences from the docking step. **(E)** In some cases, orthogonal modeling capable of predicting kinetic and/or thermodynamic properties can further narrow down the number of possible sequences selected for experimental determination (adapted from (Chang and Perez, 2022b)). **(F)** Finally, experimental characterization and validation can be carried out for the now manageable number of predicted peptide binders.

previously unknown peptide-protein interactions (Khramushin et al., 2022).

The last step needed for a pipeline that can identify high affinity peptide binders is a way to rank-order peptides by binding affinity that overcomes limitations in traditional scoring functions (see Figure 2). A recent assay using

competitive binding with AF has shown the potential to differentiate between weak and strong peptide binders under certain conditions (Chang and Perez, 2022a). Some of the limitations observed in the assay are similar to those found when using AF for docking and stem from the fact that AF was only parametrized for predicting protein structures, rather

than binding–thus AF has an implicit bias against bound peptides. A different approach using an extra layer on top of AF trained on binding affinity data for the Major Histocompatibility Complex (MHC type I and II) shows the ability to differentiate active binders from inactive peptides (Motmaen et al., 2022). Finally, some physics-based approaches try to establish methodologies capable of capturing both the flexibility and binding affinity based on known binding modes (Morrone et al., 2017; Liu et al., 2020; Mondal et al., 2022). However, they are still more computationally demanding and there are no good benchmark sets showing transferability, convergence, and associated errors.

Based on our current understanding of the field we expect significant changes coming from better training of AF like strategies to the problem of peptide-protein complexes. There are already instances of this, and the community has developed efforts for facilitating the use (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) (Mirdita et al., 2022) and training (https://github.com/aqlaboratory/openfold) of these technologies. However, these are all focused on the structure prediction. What will we need to predict binding affinities or binding mechanisms? Approaches such as the one presented for MHC are only possible when enough existing data on binding affinities is known and are thus not directly transferable to all systems. Similar to the Protein Structure Initiative (Montelione, 2012), will the collection and creation of databases of binding affinities for multiple systems (e.g., similar to those curated for transcription factor protein-DNA interactions (Chiu et al., 2019; Fornes et al., 2019)) lead to new possibilities for the virtual screening of peptide libraries based on artificial intelligence? And what is the role of physics-based approaches? Worldwide competition events such as CAPRI (Janin et al., 2003) and CASP(Moult, 2005) have spurred development, independent assessment and standardized benchmark sets for the community–and they play an important role for the development of similar model challenges in related fields (e.g. the CyroEM model challenge) (Lawson et al., 2021).

Key areas of synergy for modeling peptide behavior

Increasing the size of peptide-protein databases

Machine learning has already shown promising results for predicting peptide-protein interactions despite being trained for a different task–training for peptide binding will require increasing our structural and energetic understanding of these interactions. However, the transient nature of peptide-protein interactions has made it challenging to accumulate large datasets, both in terms of structures and binding free energies. Both are needed to develop computational software that predicts high affinity peptide binders (Cunningham et al., 2020; Motmaen et al., 2022). Two advances will play key roles to increase database knowledge to feed machine learning databases: 1) curating protein-protein structural databases to identify peptide epitopes and their interaction patterns, as recently shown by PatchMAN (Khramushin et al., 2022) and others (Peterson et al., 2017; Aderinwale et al., 2020), 2) establishment of high throughput/high sensitivity techniques for determining binding affinities (Nguyen et al., 2019).

The protein data bank (Berman et al., 2002) has an overrepresentation of structures amenable to the techniques needed to solve their structures. Thus, stable protein monomers are overrepresented with respect to multimeric structures. Furthermore, many intrinsically disordered proteins (IDP) contain short linear motifs (SLiMs) that bind proteins transiently through peptide epitopes 3–12 amino acids long (Krystkowiak and Davey, 2017; Ivarsson and Jemth, 2019). These interactions are hard to characterize both experimentally and computationally, but their prevalence (estimated to be up to 100,000 SLiMs) and biological importance makes their characterization especially important (Tompa et al., 2014). As many peptide sequences will be able to bind in sites were SLiMs bind, it will be more important to characterize not only binding affinities (in terms of K_d) but also the specificity of the binding region.

A similar issue of binding affinity/specificities can be seen in determining where in a genome a transcription binding protein will bind. This community has developed high-throughput techniques to identify binding affinities both *in vivo* and *in vitro* which are summarized in position binding motifs (PBMs) and compiled in databases such as TRANSFAC(Matys et al., 2006) or JASPAR(Fornes et al., 2019). We expect to see expansion of peptide databases to the levels seen in the protein-DNA field. Many high-throughput techniques for detecting binding affinities (Tonikian et al., 2008; Ivarsson et al., 2014; Jensen et al., 2018; Parker et al., 2019) are limited by one or more of the following: low binding affinities, aggregation, rapid equilibrium, low accuracy in estimating purity and/or concentration of the peptides, qualitative (presence of binding) instead of quantitative (K_d), and fast dissociation rates (Nguyen et al., 2019)—thus, resulting in high uncertainties. Generally, higher accuracy data requires lower throughput methods such as isothermal titration calorimetry or surface plasmon resonance (Abraham et al., 2005), limiting the amount of available data. More recently, developments such as microfluidic-based approaches (Vincentelli et al., 2015; Nguyen et al., 2019; Hein et al., 2020) enable a faster complementary method to high-throughput methods and will play a key role in increasing our understanding of peptide-protein binding affinities.

Bioinformatics

While some high-throughput methods use combinatoric peptides libraries, bioinformatic tools use knowledge of the system to create libraries that are more likely to contain peptides that bind a particular target (Turk and Cantley, 2003). Thanks to advances in sequencing and metagenomic approaches, there are now extensive sequence libraries (Bateman et al., 2017; Breitwieser et al., 2017) and efficient tools to process them (Finn et al., 2011; Remmert et al., 2012; Potter et al., 2018). Furthermore, their importance for both computational and experimental researchers has led to effective pipelines without computational expertise requirement (Joshi and Blankenberg, 2022). Thus, bioinformatic approaches take the role of narrowing sequence space.

Machine learning

With the rapid evolution of this field, any prediction of how exactly this field will evolve and which databases will be needed for training will likely be outdated in a few weeks. A promising trend highlighted by the application of AlphaFold to problems of protein-protein and peptide-protein structure predictions is the potential for transferability across macromolecular interactions. Recently, AlphaFold principles have been used to derive a protein-nucleic acid structure prediction software (Baek et al., 2022). Future AI should be aiming to understand macromolecular interactions independently of the type of molecule: rather than having an AI for proteins and a different for small molecules or nucleic acids we need integrated AI to handle all aspects of molecular recognition–transferable to other biomolecules like peptoids, modified proteins, nucleic acids and capable of interacting with small molecues. A second area of development for AI will be the identification and correction of biases arising from training datasets. For example, the training for finding folded states creates a bias to favor bound structures over unbound ones–predicting some protein-peptide complexes even when the peptide should not bind. Finally, the prediction of structures and binding affinities seem to be on two independent tracks at the moment–some tools are good for sampling bound states and others for rank-ordering them by binding affinity (Cunningham et al., 2020; Chang and Perez, 2022a; Motmaen et al., 2022). While development in affinity prediction has lagged behind the structure prediction problem, there is a recent increase in the number of methods available, some of them now available as webservers with promising accuracy (Romero-Molina et al., 2022). It is feasible to think that as binding affinity databases increase in size and accuracy, these two independent pieces will be trained together. In this direction, machine learning based approaches that predict peptide-protein interactions and peptide

binding residues (Lei et al., 2021) can help increase or build peptide-protein databases.

Physics-based approaches

The promise of these approaches has been their potential to capture bound states, as well as the thermodynamic and kinetic properties connecting the different states. This promise has been compromised by inaccuracies in physics models (force fields), efficient sampling strategies, and computational cost (Durrant and McCammon, 2011; Petrov and Zagrovic, 2014). At this point it seems unfeasible that physics-based approaches will be able to match the speed and accuracy of structure predictions coming from machine learning in the near-future. However, once the structure is known, it might still play a role in determining thermodynamic (e.g., K_d) and kinetic properties (e.g., k_{on} and k_{off}) (Zuckerman and Chong, 2016; Paul et al., 2021, 2017; Dickson et al., 2017; Zimmerman et al., 2018; Wang and Miao, 2020).

With an increasing number of methods that capture thermodynamics, the question is whether or not they will become computationally feasible and routine to match the success of FEP calculations in small molecules. These types of methods could address some of the limitations that are prevalent in small chemical changes (single point mutations) and explicitly account for posttranslational modifications (Edwards et al., 2021). Furthermore, the development and maturity of platforms to analyze ensembles for kinetic properties (e.g., weighted ensemble methods, and Markov state models) are leading to pioneering works capturing not only the thermodynamics but also kinetics of macromolecular interactions (Zuckerman and Chong, 2016; Zhou et al., 2017). Currently dissociation rates seem to have the largest uncertainties (by a few orders of magnitude), which can be in part explained by force fields that tend to favor compact structures. As these methods mature and force field development efforts continue to emphasize capturing folded, IDP, bound and unbound behavior, we expect the accuracy to improve.

Design principles

Recent protein design work is already leveraging machine learning to hallucinate novel folds based on the idealized version of proteins learnt by the algorithms (Anishchenko et al., 2021; Wang et al., 2022). Furthermore, the promise of constraint design based on a given backbone structure is promising for the design of proteins and peptides that bind a particular target (Moffat et al., 2021; Anand et al., 2022). This has already led to the design of mini proteins through the combination of computation (Rosetta) and experiments (Cao et al., 2022). Recent approaches started to take physical property such as solubility

of designed peptides into consideration (Kosugi and Ohue, 2022). With the maturity of this field at the hands of a few expert groups, comes the possibility to design molecules with an increasing range of properties, such as peptide macrocycles that are not only able to adopt a particular structure, but also are able to cross membranes.

Peptide self-assembly

Peptides have programmable self-assembly properties, while maintaining the ability to incorporate functional motifs recognized by other macromolecules. Fortunately, due to these properties, peptides are increasingly used for biomaterial design applications (e.g., functionalized hydrogel scaffolds that induce cellular behavior) (Loo et al., 2015a). Despite its promise, atomistic tools to predict the structures, properties, accessibility of functional sites and contributing to the development of materials have lagged their experimental counterpart (Rauscher and Pomès, 2017; Smadbeck et al., 2014). We believe two independent directions will benefit from recent approaches: 1) the identification of new functional motifs to queue cellular behavior; 2) the ability to predict structures and physical properties of self-assembling peptides to complement experimental efforts.

Post translational modifications

The recent Sars-COV2 pandemic highlighted the need to account for post translational modifications (PTM). Yet, as modelers we are often faced with the challenge of knowing which and where these modifications occur, lack of parameters or datasets (machine learning) to model these modifications. This issue becomes accentuated for the transient interactions of interest in peptide-protein interactions. However, the community has shown a fast response and adaptation to model these systems. We believe going forward the community will collect more data on such PTM that will improve modeling efforts. For example, many of the peptide-protein experimental binding affinity assays already incorporate the ability to introduce such PTMs (Nguyen et al., 2019)—and computational pipelines (Khoury et al., 2013; Croitoru et al., 2021) are adapting to facilitate incorporating such modifications to systems of interest.

Discussion

We are excited by the role that recent advances in a variety of areas can have in the field of peptide-protein structure and affinity prediction. Peptides can be of interest as single molecule therapeutics or as programmable molecules that

aggregate to specific patterns creating scaffolds (Perez et al., 2021). In both scenarios a functional motif will allow these peptides to target and inhibit or enhance interactions and pathways. Peptides have some intrinsic limitations like fast degradation and difficulty crossing barriers that might make them precursors of other molecules such as modified peptides (Szabó et al., 2022), mini-proteins (Cao et al., 2022) and small molecule peptidomimetics (Rubin et al., 2018; Perez, 2021). However, predicting the initial structure and identifying high affinity binders will be critical to develop these peptide derivatives. As the field moves forward, identifying novel binding motifs and the sequence/structure relationships that allows these peptides to bind will in turn increase our understanding of binding plasticity, the role of multiple binding modes in binding affinities for more accurate predictions.

A particular field of interest where we see methods expanding their presence is in material bioengineering (Pérez et al., 2013). Here, peptide epitopes embedded in fibers that make up the extra cellular matrix (ECM) bind different integrin proteins on the cellular membrane which trigger different cascade of events that govern cellular behavior (such as adhesion, growth, or migration) (Collier et al., 2010; Wade and Burdick, 2012; Loo et al., 2015a; Hellmund and Koksch, 2019). We are aware of under 100 linear motifs that bind integrins, where both the sequence and conformation are important. However, it is likely that more motifs can bind in these sites, offering opportunities for designed biomaterials. The second challenge such materials must meet is the ability to self-assemble in order to mimic the ECM–with different biomechanical properties in different tissues (e.g., softer for tissues like the brain, and stiffer for tissues like bone or muscles) (Collier et al., 2010). Computational tools in this area are scarce (Smadbeck et al., 2014; Rauscher and Pomès, 2017), but increase in computational power, force fields that better represent the balance between folded/unfolded and new sampling strategies make us optimistic in this area as well. Starting from peptides that are experimentally known to self-assemble, the first goal is to robustly and reproducible generate atomistic models of the resulting scaffold. Analysis of these scaffolds will allow us to distinguish mechanical properties like stiffness and accessibility of the functional motifs. A second requirement will be to identify pipelines that are sensitive to sequence properties: many experimental designs show the evolution of the peptide self-assembling motif, from initial sequences that barely oligomerized to successful scaffolds (Papapostolou et al., 2007). With self-assembling materials small errors in force field or machine learning preferences get amplified throughout the scaffold. Thus, a higher level of sensitivity is required. Finally, these models should be tested for functionality: are the binding motifs readily available for interaction with desired proteins (e.g., integrins that control cellular behavior). Can these models explain some of the

behavior observed when mixing two successful binding motifs that have different functionality (e.g., adhesion and growth signals) (Liu et al., 2015).

Conclusion

The rapid pace of advances in AI, coupled with increased computational power, access to larger and more curated databases and improvements in other computational modeling pipelines is changing many aspects of structural biology pipelines. We believe this presents an opportunity to advance our understanding of systems involving peptide molecules, where advances both experimental and computational have faced the challenges of working with such flexible molecules. Peptides already have multiple applications ranging from drugs, to nanodevices, or tissue regeneration to name a few. Developing robust computational pipelines to design and control the properties of these molecules will allow a rapid increase in the number of applications and uses of these molecules.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LC and AM prepared figures, revised and discussed the manuscript. AP conceptualized the work and wrote the first draft.

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 DK declared a past co-authorship with the author AP.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abraham, T., Lewis, R. N. A. H., Hodges, R. S., and McElhaney, R. N. (2005). Isothermal titration calorimetry studies of the binding of a rationally designed analogue of the antimicrobial peptide gramicidin S to phospholipid bilayer membranes. *Biochemistry* 44, 2103–2112. doi:10.1021/bi048077d

Aderinwale, T., Christoffer, C. W., Sarkar, D., Alnabati, E., and Kihara, D. (2020). Computational structure modeling for diverse categories of macromolecular interactions. *Curr. Opin. Struct. Biol.* 64, 1–8. doi:10.1016/j.sbi.2020.05.017

Aiyer, S., Swapna, G. V. T., Ma, L.-C., Liu, G., Hao, J., Chalmers, G., et al. (2021). A common binding motif in the ET domain of BRD3 forms polymorphic structural interfaces with host and viral proteins. *Structure* 29, 886–898.e6. e6. doi:10.1016/j.str.2021.01.010

Anand, N., Eguchi, R., Mathews, I. I., Perez, C. P., Derry, A., Altman, R. B., et al. (2022). Protein sequence design with a learned potential. *Nat. Commun.* 13, 746. doi:10.1038/s41467-022-28313-9

Anishchenko, I., Pellock, S. J., Chidyausiku, T. M., Ramelot, T. A., Ovchinnikov, S., Hao, J., et al. (2021). De novo protein design by deep network hallucination. *Nature* 600, 547–552. doi:10.1038/s41586-021-04184-w

Baek, M., McHugh, R., Anishchenko, I., Baker, D., and DiMaio, F. (2022). *Accurate prediction of nucleic acid and protein-nucleic acid complexes using RoseTTAFoldNA.* doi:10.1101/2022.09.09.507333

Balliu, A., and Baltzer, L. (2017). Exploring non-obvious hydrophobic binding pockets on protein surfaces: Increasing affinities in peptide–protein interactions. *Chembiochem* 18, 1396–1407. doi:10.1002/cbic.201700048

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., et al. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169. doi:10.1093/nar/gkw1099

Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., et al. (2014). The ChEMBL bioactivity database: An update. *Nucleic Acids Res.* 42, D1083–D1090. doi:10.1093/nar/gkt1031

Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., et al. (2002). The protein data bank. *Acta Crystallogr. D. Biol. Crystallogr.* 58, 899–907. doi:10.1107/s0907444902003451

Bhati, A. P., Wan, S., Wright, D. W., and Coveney, P. V. (2017). Rapid, accurate, precise, and reliable relative free energy prediction using ensemble based thermodynamic integration. *J. Chem. Theory Comput.* 13, 210–222. doi:10.1021/acs.jctc.6b00979

Breitwieser, F. P., Lu, J., and Salzberg, S. L. (2017). A review of methods and databases for metagenomic classification and assembly. *Brief. Bioinform.* 20, 1125–1136. doi:10.1093/bib/bbx120

Cao, L., Coventry, B., Goreshnik, I., Huang, B., Park, J. S., Jude, K. M., et al. (2022). Design of protein-binding proteins from the target structure alone. *Nature* 605, 551–560. doi:10.1038/s41586-022-04654-9

Chang, L., and Perez, A. (2022a). AlphaFold encodes the principles to identify high affinity peptide binders. *Biorxiv* 18, 484931. doi:10.1101/2022.03.18.484931

Chang, L., and Perez, A. (2022b). Deciphering the folding mechanism of proteins G and L and their mutants. *J. Am. Chem. Soc.* 144, 14668–14677. doi:10.1021/jacs.2c04488

Chiu, T.-P., Xin, B., Markarian, N., Wang, Y., and Rohs, R. (2019). TFBSshape: An expanded motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Res.* 48, D246–D255. doi:10.1093/nar/gkz970

Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K., and Pande, V. S. (2011). Alchemical free energy methods for drug discovery: Progress and challenges. *Curr. Opin. Struct. Biol.* 21, 150–160. doi:10.1016/j.sbi.2011.01.011

Ciemny, M., Kurcinski, M., Kamel, K., Kolinski, A., Alam, N., Schueler-Furman, O., et al. (2018). Protein–peptide docking: Opportunities and challenges. *Drug Discov. Today* 23, 1530–1537. doi:10.1016/j.drudis.2018.05.006

Collier, J. H., Rudra, J. S., Gasiorowski, J. Z., and Jung, J. P. (2010). Multi-component extracellular matrices based on peptide self-assembly. *Chem. Soc. Rev.* 39, 3413. doi:10.1039/b914337h

Croitoru, A., Park, S.-J., Kumar, A., Lee, J., Im, W., MacKerell, A. D., et al. (2021). Additive CHARMM36 force field for nonstandard amino acids. *J. Chem. Theory Comput.* 17, 3554–3570. doi:10.1021/acs.jctc.1c00254

Cunha, F. M., Berti, D. A., Ferreira, Z. S., Klitzke, C. F., Markus, R. P., and Ferro, E. S. (2008). Intracellular peptides as natural regulators of cell signaling. *J. Biol. Chem.* 283, 24448–24459. doi:10.1074/jbc.m801252200

Cunningham, J. M., Koytiger, G., Sorger, P. K., and AlQuraishi, M. (2020). Biophysical prediction of protein–peptide interactions and signaling networks using machine learning. *Nat. Methods* 17, 175–183. doi:10.1038/s41592-019-0687-1

Dickson, A., Tiwary, P., and Vashisth, H. (2017). Kinetics of ligand binding through advanced computational approaches: A review. *Curr. Top. Med. Chem.* 17, 2626–2641. doi:10.2174/1568026617666170414142908

Dubas, E., Żur, I., Moravčíková, J., Fodor, J., Krzewska, M., Surówka, E., et al. (2021). Proteins, small peptides and other signaling molecules identified as inconspicuous but possibly important players in microspores reprogramming toward embryogenesis. *Front. Sustain. Food Syst.* 5, 745865. doi:10.3389/fsufs.2021.745865

Durrant, J. D., and McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol.* 9, 71. doi:10.1186/1741-7007-9-71

Edwards, T., Foloppe, N., Harris, S. A., and Wells, G. (2021). The future of biomolecular simulation in the pharmaceutical industry: What we can learn from aerodynamics modelling and weather prediction. Part 1. Understanding the physical and computational complexity of *in silico* drug design. *Acta Crystallogr. D. Struct. Biol.* 77, 1348–1356. doi:10.1107/s2059798321009712

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi:10.1093/nar/gkr367

Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., et al. (2019). Jaspar 2020: Update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 48, D87–D92. doi:10.1093/nar/gkz1001

Foster, S. R., Hauser, A. S., Vedel, L., Strachan, R. T., Huang, X.-P., Gavin, A. C., et al. (2019). Discovery of human signaling systems: Pairing peptides to G protein-coupled receptors. *Cell* 179, 895–908.e21. e21. doi:10.1016/j.cell.2019.10.010

Fratev, F., and Sirimulla, S. (2019). An improved free energy perturbation FEP+ sampling protocol for flexible ligand-binding domains. *Sci. Rep.* 9, 16829. doi:10.1038/s41598-019-53133-1

Gao, M., and Skolnick, J. (2013). A comprehensive survey of small-molecule binding pockets in proteins. *PLoS Comput. Biol.* 9, e1003302. doi:10.1371/journal.pcbi.1003302

Hein, J., Nguyen, H., Cyert, M., and Fordyce, P. (2020). Protocol for peptide synthesis on spectrally encoded beads for MRBLE-pep assays. *Bio. Protoc.* 10, e3669. doi:10.21769/bioprotoc.3669

Hellmund, K. S., and Koksch, B. (2019). Self-assembling peptides as extracellular matrix mimics to influence stem cell's fate. *Front. Chem.* 7, 172. doi:10.3389/fchem.2019.00172

Irwin, J. J., and Shoichet, B. K. (2005). Zinc - a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182. doi:10.1021/ci049714+

Ivarsson, Y., Arnold, R., McLaughlin, M., Nim, S., Joshi, R., Ray, D., et al. (2014). Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2542–2547. doi:10.1073/pnas.1312296111

Ivarsson, Y., and Jemth, P. (2019). Affinity and specificity of motif-based protein–protein interactions. *Curr. Opin. Struct. Biol.* 54, 26–33. doi:10.1016/j.sbi.2018.09.009

Janin, J., Henrick, K., Moult, J., Eyck, L. T., Sternberg, M. J. E., Vajda, S., et al. (2003). Capri: A critical assessment of PRedicted interactions. *Proteins.* 52, 2–9. doi:10.1002/prot.10381

Jensen, K. K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J. A., Yan, Z., et al. (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* 154, 394–406. doi:10.1111/imm.12889

Joshi, J., and Blankenberg, D. (2022). Pdaug: A galaxy based toolset for peptide library analysis, visualization, and machine learning modeling. *Bmc Bioinforma.* 23, 197. doi:10.1186/s12859-022-04727-6

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2

Khoury, G. A., Thompson, J. P., Smadbeck, J., Kieslich, C. A., and Floudas, C. A. (2013). Forcefield_PTM: *Ab initio* charge and AMBER forcefield parameters for frequently occurring post-translational modifications. *J. Chem. Theory Comput.* 9, 5653–5674. doi:10.1021/ct400556v

Khramushin, A., Ben-Aharon, Z., Tsaban, T., Varga, J. K., Avraham, O., and Schueler-Furman, O. (2022). Matching protein surface structural patches for high-resolution blind peptide docking. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2121153119. doi:10.1073/pnas.2121153119

Ko, J., and Lee, J. (2021). Can AlphaFold2 predict protein-peptide complex structures accurately? *bioRxiv* 27, 453972. doi:10.1101/2021.07.27.453972

Kosugi, T., and Ohue, M. (2022). Solubility-aware protein binding peptide design using AlphaFold. *Biomedicines* 10, 1626. doi:10.3390/biomedicines10071626

Krystkowiak, I., and Davey, N. E. (2017). SLiMSearch: A framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. *Nucleic Acids Res.* 45, W464–W469. doi:10.1093/nar/gkx238

Laskowski, R. A., Luscombe, N. M., Swindells, M. B., and Thornton, J. M. (1996). Protein clefts in molecular recognition and function. *Protein Sci.* 5, 2438–2452. doi:10.1002/pro.5560051206

Lawson, C. L., Kryshtafovych, A., Adams, P. D., Afonine, P. V., Baker, M. L., Barad, B. A., et al. (2021). Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. *Nat. Methods* 18, 156–164. doi:10.1038/s41592-020-01051-w

Lee, T.-S., Allen, B. K., Giese, T. J., Guo, Z., Li, P., Lin, C., et al. (2020). Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.* 60, 5595–5623. doi:10.1021/acs.jcim.0c00613

Lei, Y., Li, Shuya, Liu, Z., Wan, F., Tian, T., Li, Shao, et al. (2021). A deep-learning framework for multi-level peptide–protein interaction prediction. *Nat. Commun.* 12, 5465. doi:10.1038/s41467-021-25772-4

Liang, J., Woodward, C., and Edelsbrunner, H. (1998). Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* 7, 1884–1897. doi:10.1002/pro.5560070905

Limongelli, V. (2020). Ligand binding free energy and kinetics calculation in 2020. *WIREs Comput. Mol. Sci.* 10. doi:10.1002/wcms.1455

Liu, C., Brini, E., Perez, A., and Dill, K. A. (2020). Computing ligands bound to proteins using MELD-accelerated MD. *J. Chem. Theory Comput.* 16, 6377–6382. doi:10.1021/acs.jctc.0c00543

Liu, J., Liu, S., Chen, Y., Zhao, X., Lu, Y., and Cheng, J. (2015). Functionalized self-assembling peptide improves INS-1 β-cell function and proliferation via the integrin/FAK/ERK/cyclin pathway. *Int. J. Nanomedicine* 10, 3519–3531. doi:10.2147/ijn.s80502

Loo, Y., Goktas, M., Tekinay, A. B., Guler, M. O., Hauser, C. A. E., and Mitraki, A. (2015a). Self-assembled proteins and peptides as scaffolds for tissue regeneration. *Adv. Healthc. Mat.* 4, 2557–2586. doi:10.1002/adhm.201500402

Loo, Y., Lakshmanan, A., Ni, M., Toh, L. L., Wang, S., and Hauser, C. A. E. (2015b). Peptide bioink: Self-assembling nanofibrous scaffolds for three-dimensional organotypic cultures. *Nano Lett.* 15, 6919–6925. doi:10.1021/acs.nanolett.5b02859

Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., et al. (2006). TRANSFAC(R) and its module TRANSCompel(R): Transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* 34, D108–D110. –D110. doi:10.1093/nar/gkj143

Mey, A. S. J. S., Allen, B. K., Macdonald, H. E. B., Chodera, J. D., Hahn, D. F., Kuhn, M., et al. (2020). Best practices for alchemical free energy calculations [article v1.0]. *Living J. comput. Mol. Sci.* 2, 18378. doi:10.33011/livecoms.2.1.18378

Milroy, L.-G., Grossmann, T. N., Hennig, S., Brunsveld, L., and Ottmann, C. (2014). Modulators of protein–protein interactions. *Chem. Rev.* 114, 4695–4748. doi:10.1021/cr400698c

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi:10.1038/s41592-022-01488-1

Mobley, D. L., and Klimovich, P. V. (2012). Perspective: Alchemical free energy calculations for drug discovery. *J. Chem. Phys.* 137, 230901. doi:10.1063/1.4769292

Moffat, L., Greener, J. G., and Jones, D. T. (2021). Using AlphaFold for rapid and accurate fixed backbone protein design. *Biorxiv* 24, 457549. doi:10.1101/2021.08.24.457549

Mondal, A., Swapna, G. V. T., Hao, J., Ma, L., Roth, M. J., Montelione, G. T., et al. (2022). Structure determination of protein-peptide complexes from NMR chemical shift data using MELD. *Biorxiv* 31, 474671. doi:10.1101/2021.12.31.474671

Montelione, G. T. (2012). The protein structure initiative: Achievements and visions for the future. *F1000 Biol. Rep.* 4, 7. doi:10.3410/b4-7

Morrone, J. A., Perez, A., MacCallum, J., and Dill, K. A. (2017). Computed binding of peptides to proteins with MELD-accelerated molecular dynamics. *J. Chem. Theory Comput.* 13, 870–876. doi:10.1021/acs.jctc.6b00977

Motmaen, A., Dauparas, J., Baek, M., Abedi, M. H., Baker, D., and Bradley, P. (2022). Peptide binding specificity prediction using fine-tuned protein structure prediction networks. *Biorxiv* 12, 499365. doi:10.1101/2022.07.12.499365

Moult, J. (2005). A decade of CASP: Progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* 15, 285–289. doi:10.1016/j.sbi.2005.05.011

Murugan, N. A., Podobas, A., Gadioli, D., Vitali, E., Palermo, G., and Markidis, S. (2022). A review on parallel virtual screening softwares for high-performance computers. *Pharm. (Basel).* 15, 63. doi:10.3390/ph15010063

Nguyen, H. Q., Roy, J., Harink, B., Damle, N. P., Latorraca, N. R., Baxter, B. C., et al. (2019). Quantitative mapping of protein-peptide affinity landscapes using spectrally encoded beads. *Elife* 8, e40499. doi:10.7554/elife.40499

Papapostolou, D., Smith, A. M., Atkins, E. D. T., Oliver, S. J., Ryadnov, M. G., Serpell, L. C., et al. (2007). Engineering nanoscale order into a designed protein fiber. *Proc. Natl. Acad. Sci. U. S. A.* 104, 10853–10858. doi:10.1073/pnas.0700801104

Parker, B. W., Goncz, E. J., Krist, D. T., Statsyuk, A. V., Nesvizhskii, A. I., and Weiss, E. L. (2019). Mapping low-affinity/high-specificity peptide–protein interactions using ligand-footprinting mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 116, 21001–21011. doi:10.1073/pnas.1819533116

Paul, A., Samantray, S., Anteghini, M., Khaled, M., and Strodel, B. (2021). Thermodynamics and kinetics of the amyloid-β peptide revealed by Markov state models based on MD data in agreement with experiment. *Chem. Sci.* 12, 6652–6669. doi:10.1039/d0sc04657d

Paul, F., Wehmeyer, C., Abualrous, E. T., Wu, H., Crabtree, M. D., Schöneberg, J., et al. (2017). Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nat. Commun.* 8, 1095. doi:10.1038/s41467-017-01163-6

Pelay-Gimeno, M., Glas, A., Koch, O., and Grossmann, T. N. (2015). Structure-based design of inhibitors of protein–protein interactions: Mimicking peptide binding epitopes. *Angew. Chem. Int. Ed.* 54, 8896–8927. doi:10.1002/anie.201412070

Perez, J. J. (2021). Exploiting knowledge on structure–activity relationships for designing peptidomimetics of endogenous peptides. *Biomedicines* 9, 651. doi:10.3390/biomedicines9060651

Perez, J. J., Perez, R. A., and Perez, A. (2021). Computational modeling as a tool to investigate PPI: From drug design to tissue engineering. *Front. Mol. Biosci.* 8 (681617), 1–20. doi:10.3389/fmolb.2021.681617

Pérez, R. A., Won, J.-E., Knowles, J. C., and Kim, H.-W. (2013). Naturally and synthetic smart composite biomaterials for tissue regeneration. *Adv. Drug Deliv. Rev.* 65, 471–496. doi:10.1016/j.addr.2012.03.009

Peterson, L. X., Roy, A., Christoffer, C., Terashi, G., and Kihara, D. (2017). Modeling disordered protein interactions from biophysical principles. *PLoS Comput. Biol.* 13, e1005485. doi:10.1371/journal.pcbi.1005485

Petrov, D., and Zagrovic, B. (2014). Are current atomistic force fields accurate enough to study proteins in crowded environments? *PLoS Comput. Biol.* 10, e1003638. doi:10.1371/journal.pcbi.1003638

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi:10.1093/nar/gky448

Rauscher, S., and Pomès, R. (2017). The liquid structure of elastin. *Elife* 6, e26526. doi:10.7554/elife.26526

Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2012). HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* 9, 173–175. doi:10.1038/nmeth.1818

Rentzsch, R., and Renard, B. Y. (2015). Docking small peptides remains a great challenge: An assessment using AutoDock vina. *Brief. Bioinform.* 16, 1045–1056. doi:10.1093/bib/bbv008

Romero-Molina, S., Ruiz-Blanco, Y. B., Mieres-Perez, J., Harms, M., Münch, J., Ehrmann, M., et al. (2022). PPI-affinity: A web tool for the prediction and optimization of protein–peptide and protein–protein binding affinity. *J. Proteome Res.* 21, 1829–1841. doi:10.1021/acs.jproteome.2c00020

Rubin, S. J. S., Tal-Gan, Y., Gilon, C., and Qvit, N. (2018). Conversion of protein active regions into peptidomimetic therapeutic leads using backbone cyclization and cycloscan – how to do it yourself. *Curr. Top. Med. Chem.* 18, 556–565. doi:10.2174/1568026618666180518094322

Shivanyuk, A., Ryabukhin, S., Tolmachev, A., Bogolyubsky, A., Mykytenko, D., Chupryna, A., et al. (2007). Enamine real database: Making chemical diversity real. *Chem. Today* 25, 58–59.

Smadbeck, J., Chan, K. H., Khoury, G. A., Xue, B., Robinson, R. C., Hauser, C. A. E., et al. (2014). De novo design and experimental characterization of ultrashort self-associating peptides. *PLoS Comput. Biol.* 10, e1003718. doi:10.1371/journal.pcbi.1003718

Szabó, I., Yousef, M., Soltész, D., Bató, C., Mező, G., and Bánóczi, Z. (2022). Redesigning of cell-penetrating peptides to improve their efficacy as a drug delivery system. *Pharmaceutics* 14, 907. doi:10.3390/pharmaceutics14050907

Tompa, P., Davey, N. E., Gibson, T. J., and Babu, M. M. (2014). A million peptide motifs for the molecular biologist. *Mol. Cell* 55, 161–169. doi:10.1016/j.molcel.2014.05.032

Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J.-H., Reva, B., et al. (2008). A specificity map for the PDZ domain family. *PLoS Biol.* 6, e239. doi:10.1371/journal.pbio.0060239

Tsaban, T., Varga, J. K., Avraham, O., Ben-Aharon, Z., Khramushin, A., and Schueler-Furman, O. (2022). Harnessing protein folding neural networks for peptide–protein docking. *Nat. Commun.* 13, 176. doi:10.1038/s41467-021-27838-9

Turk, B. E., and Cantley, L. C. (2003). Peptide libraries: At the crossroads of proteomics and bioinformatics. *Curr. Opin. Chem. Biol.* 7, 84–90. doi:10.1016/s1367-5931(02)00004-2

Vincentelli, R., Luck, K., Poirson, J., Polanowska, J., Abdat, J., Blémont, M., et al. (2015). Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat. Methods* 12, 787–793. doi:10.1038/nmeth.3438

Wade, R. J., and Burdick, J. A. (2012). Engineering ECM signals into biomaterials. *Mat. TodayKidlingt.* 15, 454–459. doi:10.1016/s1369-7021(12)70197-9

Wai, D. C. C., Szyszka, T. N., Campbell, A. E., Kwong, C., Wilkinson-White, L. E., Silva, A. P. G., et al. (2018). The BRD3 ET domain recognizes a short peptide motif through a mechanism that is conserved across chromatin remodelers and transcriptional regulators. *J. Biol. Chem.* 293, 7160–7175. doi:10.1074/jbc.ra117.000678

Wang, J., Lisanza, S., Juergens, D., Tischer, D., Watson, J. L., Castro, K. M., et al. (2022). Scaffolding protein functional sites using deep learning. *Science* 377, 387–394. doi:10.1126/science.abn2100

Wang, J., and Miao, Y. (2020). Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *J. Chem. Phys.* 153, 154109. doi:10.1063/5.0021399

Wang, X., Ni, D., Liu, Y., and Lu, S. (2021). Rational design of peptide-based inhibitors disrupting protein-protein interactions. *Front. Chem.* 9, 682675. doi:10.3389/fchem.2021.682675

Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., et al. (2016). Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: The prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* 18, 12964–12975. doi:10.1039/c6cp01555g

Zhou, G., Pantelopulos, G. A., Mukherjee, S., and Voelz, V. A. (2017). Bridging microscopic and macroscopic mechanisms of p53-MDM2 binding with kinetic network models. *Biophys. J.* 113, 785–793. doi:10.1016/j.bpj.2017.07.009

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

Zuckerman, D. M., and Chong, L. T. (2016). Weighted ensemble simulation: Review of methodology, applications, and software. *Annu. Rev. Biophys.* 46, 43–57. doi:10.1146/annurev-biophys-070816-033834