

Computational drug discovery for emerging viral infections

Edited by

Sinosh Skariyachan, Kumar K. M. and
Arli Aditya Parikesit

Published in

Frontiers in Microbiology



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-4049-7
DOI 10.3389/978-2-8325-4049-7

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

Computational drug discovery for emerging viral infections

Topic editors

Sinosh Skariyachan — Department of Microbiology, St. Pius X College Rajapuram, India

Kumar K. M. — Pondicherry University, India

Akli Aditya Parikesit — Indonesia International Institute for Life-Sciences (i3L), Indonesia

Citation

Skariyachan, S., Kumar, K. M., Parikesit, A. A., eds. (2023). *Computational drug discovery for emerging viral infections*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4049-7

Table of contents

- 05 **Editorial: Computational drug discovery for emerging viral infections**
Sinosh Skariyachan, Kumar Kalavathi Murugan and Arli Aditya Parikesit
- 08 **Network pharmacology and molecular docking-based investigations of *Kochia* Fructus's active phytomolecules, molecular targets, and pathways in treating COVID-19**
Shakeel Ahmad Khan and Terence Kin Wah Lee
- 26 **Mechanism of N-0385 blocking SARS-CoV-2 to treat COVID-19 based on molecular docking and molecular dynamics**
Jun-Feng Cao, Xingyu Yang, Li Xiong, Mei Wu, Shengyan Chen, Chenyang Xiong, Peiyong He, Yonghua Zong, Lixin Zhang, Hongjiao Fu, Yue Qi, Xiran Ying, Dengxin Liu, Xiaosong Hu and Xiao Zhang
- 42 **Identification of natural compounds as SARS-CoV-2 inhibitors *via* molecular docking and molecular dynamic simulation**
Tiantian Han, Ziqing Luo, Lichun Ji, Peng Wu, Geng Li, Xiaohong Liu and Yanni Lai
- 59 **Discovering common pathogenic processes between COVID-19 and HFRS by integrating RNA-seq differential expression analysis with machine learning**
Fatima Noor, Usman Ali Ashfaq, Abu Bakar, Waqar ul Haq, Khaled S. Allemailem, Basmah F. Alharbi, Wafa Abdullah I. Al-Megrin and Muhammad Tahir ul Qamar
- 71 **Immunogenic cell death-led discovery of COVID-19 biomarkers and inflammatory infiltrates**
Jianzhen Zhuo, Ke Wang, Zijun Shi and Chunlei Yuan
- 83 **Machine learning on large scale perturbation screens for SARS-CoV-2 host factors identifies β -catenin/CBP inhibitor PRI-724 as a potent antiviral**
Maximilian A. Kelch, Antonella Vera-Guapi, Thomas Beder, Marcus Oswald, Alicia Hiemisch, Nina Beil, Piotr Wajda, Sandra Ciesek, Holger Erfle, Tuna Toptan and Rainer Koenig
- 101 ***In silico* generation of novel ligands for the inhibition of SARS-CoV-2 main protease (3CL^{pro}) using deep learning**
Prejwal Prabhakaran, Ananda Vardhan Hebbani, Soumya V. Menon, Biswaranjan Paital, Sneha Murmu, Sunil Kumar, Mahender Kumar Singh, Dipak Kumar Sahoo and Padma Priya Dharmavaram Desai
- 117 **Target specific inhibition of West Nile virus envelope glycoprotein and methyltransferase using phytocompounds: an *in silico* strategy leveraging molecular docking and dynamics simulation**
Shopnil Akash, Imren Bayıl, Md. Anisur Rahman, Nobendu Mukerjee, Swastika Maitra, Md. Rezaul Islam, Sanchaita Rajkhowa, Arabinda Ghosh, Sami A. Al-Hussain, Magdi E. A. Zaki, Vikash Jaiswal, Sanjit Sah, Joshuan J. Barboza and Ranjit Sah

- 137 **Development of a new drug candidate for the inhibition of Lassa virus glycoprotein and nucleoprotein by modification of evodiamine as promising therapeutic agents**
Shopnil Akash, Javiera Baeza, Sajjat Mahmood, Nobendu Mukerjee, Vetriselvan Subramaniyan, Md. Rezaul Islam, Gaurav Gupta, Vinibha Rajakumari, Suresh V. Chinni, Gobinath Ramachawolran, Fayez M. Saleh, Ghadeer M. Albadrani, Amany A. Sayed and Mohamed M. Abdel-Daim
- 153 **Novel computational and drug design strategies for inhibition of monkeypox virus and *Babesia microti*: molecular docking, molecular dynamic simulation and drug design approach by natural compounds**
Shopnil Akash, Showkat Ahmad Mir, Sajjat Mahmood, Saddam Hossain, Md. Rezaul Islam, Nobendu Mukerjee, Binata Nayak, Hiba-Allah Nafidi, Yousef A. Bin Jardan, Amare Mekonnen and Mohammed Bourhia
- 175 **Design, evaluation, and immune simulation of potentially universal multi-epitope mpox vaccine candidate: focus on DNA vaccine**
Nino Rcheulishvili, Jiawei Mao, Dimitri Papukashvili, Shunping Feng, Cong Liu, Xingyun Wang, Yunjiao He and Peng George Wang
- 191 **The use of Xuanbai Chengqi decoction on monkeypox disease through the estrone-target AR interaction**
Yanqi Jiao, Chengcheng Shi and Yao Sun



OPEN ACCESS

EDITED AND REVIEWED BY

Anna Kramvis,
University of the Witwatersrand, South Africa

*CORRESPONDENCE

Sinosh Skariyachan
✉ sinoshmicro@stpius.ac.in;
✉ sinoshskariya@gmail.com

RECEIVED 23 October 2023

ACCEPTED 08 November 2023

PUBLISHED 20 November 2023

CITATION

Skariyachan S, Kalavathi Murugan K and
Parikesit AA (2023) Editorial: Computational
drug discovery for emerging viral infections.
Front. Microbiol. 14:1326731.
doi: 10.3389/fmicb.2023.1326731

COPYRIGHT

© 2023 Skariyachan, Kalavathi Murugan and
Parikesit. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Editorial: Computational drug discovery for emerging viral infections

Sinosh Skariyachan^{1*}, Kumar Kalavathi Murugan² and
Arli Aditya Parikesit³

¹Department of Microbiology, St. Pius X College Rajapuram, Kasaragod, Kerala, India, ²Department of Bioinformatics, Pondicherry University, Puducherry, Tamil Nadu, India, ³Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life-Sciences (i3L), Jakarta, Indonesia

KEYWORDS

emerging viral infections, drug discovery, data science, computational biology, bioinformatics, integrated omics, novel insights, cutting-edge research

Editorial on the Research Topic

Computational drug discovery for emerging viral infections

The emerging and reemerging viral infections mainly SARS-CoV2, Nipah, Zika, Lassa, Monkeypox (mpox), West Nile, and Ebola, are significant concerns for public health in recent the past with high mortality and morbidity throughout the world. Though vaccines have been developed for some of these infections, the lack of promising lead and drug candidates against many of these emerging viruses is a present concern (Morens and Fauci, 2020). Screening of potential molecular targets and unexplored lead molecules from natural sources and repurposing of drug molecules are recent trends in drug discovery. Computational Biology and Bioinformatics potentially contribute to the modern drug discovery protocol (Zhang et al., 2022). The integration of data science, machine learning, and artificial intelligence with computational biology offers great scope in screening lead candidates with ideal drug likeliness, pharmacokinetic and toxicities properties for drug development (Rai et al., 2023). The computational models not only provide insights into the experimental validation, but also reduce the cost, time and other complexities associated with traditional drug discovery and development. Hence, with the advent of molecular docking and dynamics method, computational research is becoming more important than ever for drug development against viral infections (Karplus and McCammon, 2002; Shoichet et al., 2002).

This Research Topic aimed to provide a recent outline and update on various aspects of drug discovery for emerging viral infections that utilize data science, computational biology, bioinformatics, and integrated omics sciences. The major goal of this Research Topic is to prioritize cutting-edge *in silico* drug discovery approaches that can be scaled up to the experimental level and contribute to modern drug discovery against emerging viral infections.

The Research Topic was introduced in May 2022, within the next 6 months we received 23 submissions, of which 12 full-length research articles were accepted and 11 manuscripts were rejected. We have contacted 100% of the potential authors and 100 authors contributed to the Research Topic.

As of the second week of October 2023, 19K total views and downloads were reported. The accepted manuscript contributes substantial breakthroughs in the scope and applications of computational drug discovery and predictive models in future-generation drug discovery to emerging viral infections. Though the Research Topic welcomed the submissions related to emerging viral infections, most of the articles were focused on the role of computational biology and other *in silico* approaches against SARS-CoV-2.

Khan and Lee comprehensively reported the network pharmacological and molecular docking-based studies of using the phytochemicals of *Kochia fructus* toward the potential molecular targets for the lead discovery against SARS-COV2. The molecular interaction and binding stability of small molecule N-0385 toward TMPRSS2, ACE2, and DPP4 receptors of SARS-CoV-2 were reported (Cao et al.). Han et al. also identified five potential natural compounds against the S protein of Lambda and Delta mutants of SARS-CoV-2. Similarly, with the help of structure-based virtual screening and deep learning approaches, potential ligands were screened toward the main proteases of SARS-CoV-2 (Prabhakaran et al.). The common pathogenic processes between SARS-COV2 and Hemorrhagic fever with renal syndrome (HFRS) are well illustrated by the integration of RNA-seq differential expression analysis and machine learning approaches (Noor et al.). Further, machine learning approaches were used for the large-scale screening for SARS-CoV-2 host factories and prioritized PRI-724 as a potent antiviral against β -catenin/CBP (Kelch et al.). The specific role of immunogenic cell death (ICD) and related genes in SARS-COV2 was analyzed by Zhuo et al. they suggested that CASP1, CD4, and EIF2AK3 are probably the diagnostic genes of SARS-COV2 and that are correlated with immune activity.

In addition to SARS-COV2, the topic collection also highlighted the role of various computational biology approaches and models toward the drug discovery against Lassa virus, West Nile virus, Monkeypox virus and tick-borne pathogens such as human *Babesia microti*. Rcheulishvili et al. designed a universal multiepitope DNA vaccine candidate against mpox and its efficiency was evaluated by immunoinformatic approaches. Several natural plant-based molecules were subjected to computational screening and potential lead molecules were identified as the probable molecular targets of *Babesia microti* and monkeypox virus (Akash, Ahmad Mir et al.). The scope of Xuanbai Chengqi decoction (XBCQD) on monkeypox is illustrated by several computational biology resources by Jioa et al. suggested that the estrone-target AR interaction model is probably one of the best models against monkeypox drug discovery. The modified evodiamine derivatives are probably promising lead candidate against the glycoprotein and nucleoprotein targets of the Lassa virus (Akash, Baeza et al.). Akash, Bayil et al. also performed the interaction modeling of envelope glycoprotein and methyltransferase of West Nile virus and phytocompounds using several *in silico* approaches.

The major focus of the Research Topics is *Computational drug discovery for emerging viral infections*, the authors comprehensively

illustrated the scope and applications of various predictive models in drug screening and new lead discovery against various emerging infections such as SARS-CoV2, Monkeypox and Lassa Viral infections, the computational models require extensive studies and validation in experimental level. Computational drug discovery certainly reduces the time, complexities, and initial investments associated with conventional vaccine design and drug discovery. The predictive models provide substantial breakthroughs and a foundation to scale up the hypothetical aspects into the experimental level and can scale up the approach into the industrial and healthcare sectors.

Author contributions

SS: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Validation, Writing—original draft. KK: Investigation, Project administration, Supervision, Visualization, Writing—review & editing. AP: Investigation, Project administration, Resources, Supervision, Writing—review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank all our potential reviewers, researchers and scientists who contributed to the Research Topics. We also thank all the members of the Board of the Frontiers Publishing group for their timely updates, guidelines, and continuous and constant support throughout.

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.

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

- Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 35, 646–652. doi: 10.1038/nsb0902-646
- Morens, D. M., and Fauci, A. S. (2020). Emerging pandemic diseases: how we got to SARS-COV2. *Cell* 182, 1077–1092. doi: 10.1016/j.cell.2020.08.021
- Rai, M., Singh, A. V., Paudel, N., Kanase, A., Falletta, E., Kerkar, P., et al. (2023). Herbal concoction unveiled: a computational analysis of phytochemicals' pharmacokinetic and toxicological profiles using novel approach methodologies (NAMs). *Curr. Res. Toxicol.* 5:100118. doi: 10.1016/j.crttox.2023.100118
- Shoichet, B. K., McGovern, S. L., Wei, B., and Irwin, J. J. (2002). Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* 6, 439–446. doi: 10.1016/S1367-5931(02)00339-3
- Zhang, Y., Luo, M., Wu, P., Wu, S., Lee, T. Y., and Bai, C. (2022). Application of computational biology and artificial intelligence in drug design. *Int. J. Mol. Sci.* 23:13568. doi: 10.3390/ijms232113568



OPEN ACCESS

EDITED BY

Arli Aditya Parikesit,
Indonesia International Institute
for Life-Sciences (i3L), Indonesia

REVIEWED BY

Arif Ansori,
Airlangga University, Indonesia
Yunus Effendi,
Leibniz University Hannover, Germany

*CORRESPONDENCE

Shakeel Ahmad Khan
terence.kw.lee@polyu.edu.hk
Terence Kin Wah Lee
shakilahmad56@gmail.com

SPECIALTY SECTION

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

RECEIVED 18 June 2022

ACCEPTED 19 July 2022

PUBLISHED 05 August 2022

CITATION

Khan SA and Lee TKW (2022) Network
pharmacology and molecular
docking-based investigations
of Kochiae Fructus's active
phytomolecules, molecular targets,
and pathways in treating COVID-19.
Front. Microbiol. 13:972576.
doi: 10.3389/fmicb.2022.972576

COPYRIGHT

© 2022 Khan and Lee. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Network pharmacology and molecular docking-based investigations of Kochiae Fructus's active phytomolecules, molecular targets, and pathways in treating COVID-19

Shakeel Ahmad Khan^{1*} and Terence Kin Wah Lee^{1,2*}

¹Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China, ²State Key Laboratory of Chemical Biology and Drug Discovery, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China

COVID-19 disease is caused by SARS-CoV-2. Hyper-inflammation mediated by proinflammatory cytokines is humans' primary etiology of SARS-CoV-2 infection. Kochiae Fructus is widely used in China as traditional Chinese medicine (TCM) to treat inflammatory diseases. Due to its anti-inflammatory properties, we hypothesized that Kochiae Fructus would be a promising therapeutic agent for COVID-19. The active phytomolecules, targets, and molecular pathways of Kochiae Fructus in treating COVID-19 have not been explored yet. Network pharmacology analysis was performed to determine the active phytomolecules, molecular targets, and pathways of Kochiae Fructus. The phytomolecules in Kochiae Fructus were retrieved from the Traditional Chinese Medicine Systems Pharmacology (TCMSP) database, and their potential targets were predicted with the SwissTargetPrediction webserver. COVID-19-related targets were recovered from the GeneCards database. Intersecting targets were determined with the VENN tool. The Protein-protein interaction (PPI) and Molecular Complex Detection (MCODE) network analyses were constructed using the Cytoscape software. Using the DAVID tool, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed on the intersecting targets. AutoDock Vina (version 1.2.0.) was used for molecular docking analysis. Six active phytomolecules and 165 their potential targets, 1,745 COVID-19-related targets, and 34 intersecting targets were identified. Network analysis determined 13 anti-COVID-19 core targets and three key active phytomolecules (Oleanolic acid, 9E,12Z-octadecadienoic acid, and 11,14-eicosadienoic acid). Three key pathways (pathways in cancer, the TNF signaling pathway, and lipid and atherosclerosis) and the top six anti-COVID-19 core targets (IL-6, PPARG, MAPK3, PTGS2, ICAM1, and MAPK1) were determined to be involved in the treatment of COVID-19 with active phytomolecules of Kochiae Fructus. Molecular docking analysis revealed that

three key active phytomolecules of *Kochiae Fructus* had a regulatory effect on the identified anti-COVID-19 core targets. Hence, these findings offer a foundation for developing anti-COVID-19 drugs based on phytomolecules of *Kochiae Fructus*.

KEYWORDS

Kochiae Fructus, COVID-19, network pharmacology, molecular docking, molecular pathways

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) initially emerged in December 2019 in Wuhan, China, and then spread fast across the country, causing extensive respiratory infections (Rizwan et al., 2020). Till now, we have remained unsuccessful in halting the transmission of SARS-CoV-2 and its subsequent disease, “Corona Virus Disease 2019” (COVID-19). They are posing a continuous threat to human health and are claiming the deaths of numerous people over the globe daily (Tang et al., 2022). One of the most significant impediments in treating COVID-19 is the production of proinflammatory cytokines throughout the disease’s progression, referred to as cytokine release syndrome (CRS). CRS has been identified as humans’ primary etiology of SARS- and MERS-CoV infection (Channappanavar and Perlman, 2017; Mehta et al., 2020; Niu et al., 2021). COVID-19 patients have been observed to have systemic hyper-inflammation, also known as macrophage activation syndrome or cytokine storm (McGonagle et al., 2020). Reports demonstrate that a biomarker for CRS is an increase in pro-inflammatory cytokines such as IL-6. In patients with COVID-19, IL-6-induced CRS has been observed (Fehr et al., 2017; Ruan et al., 2020a; Zhang X. et al., 2020). Studies have reported that Tocilizumab, an IL-6 receptor antagonist, can be effective in suppressing cytokine storms in patients with COVID-19. However, early evidence from randomized clinical trials remains inconclusive (No author list, 2020; Niu et al., 2021). The only anti-COVID-19 therapies available are anti-SARS-CoV-2 vaccines such as BNT162b2, CoronaVac, ChAdOx1 nCoV-19, and mRNA-1273 which have been administered to people all over the globe (Tang et al., 2022). The vaccination program of anti-SARS-CoV-2 vaccines has successfully prevented thousands of deaths from COVID-19. Despite their therapeutic efficacy, these vaccines are reported to induce severe complications, including hypersensitivity myocarditis, Bell’s palsy, vaccine-induced immune thrombocytopenia, and thrombosis (VITT) (NICE, 2021; Kounis et al., 2022; Wan et al., 2022). As a corollary, developing more effective therapeutics with a

reduced risk of adverse effects against COVID-19 remains a critical demand.

In this instance, traditional Chinese medicine (TCM) may be a viable alternative therapy for COVID-19. TCM has been used for thousands of years to treat a variety of diseases. Recent research has shown that TCM is quite beneficial in the treatment of COVID-19 (Liu et al., 2020; Ma Q. et al., 2020; Yang et al., 2020). The phytochemicals in TCM, such as polyphenols, flavonoids, alkaloids, and terpenoids, have been shown to effectively suppress pro-inflammatory cytokines (Zhou H. X. et al., 2020; El-Missiry et al., 2021; Alzaabi et al., 2022; Ge et al., 2022). Moreover, compared to conventional therapeutic drugs, the potential of TCM to treat just the symptoms of a disease while causing no damage to healthy cells makes it a more tempting choice for drug design and development.

Kochiae Fructus is the dried fruit of *Kochia scoparia* (L.) Schrad., has been used in TCM for over 2000 years as a topical and edible drug (Choi et al., 2014; Luo et al., 2021). *Kochiae Fructus* was initially reported in “Shennong Ben Cao Jing” as a “top grade” medicinal herb (Matsuda et al., 1999; Zou et al., 2021). *Kochiae Fructus*’s traditional efficacy was also reported in ancient medicinal books such as the Herbal Canon, Compendium of Materia Medica, etc., (Xiao et al., 2018; Zou et al., 2021). *Kochiae Fructus* has traditionally been used to treat back pain, malignant sores, skin itching, eczema, abnormal leucorrhea, male impotence, frequent urination, and eye conditions (Choi et al., 2014; Xiao et al., 2018; Luo et al., 2021; Zou et al., 2021). More than 150 phytochemicals have been identified in *Kochiae Fructus* that are characterized as heterocyclics, essential oils, organic acids, amino acids, carbohydrates (primarily mono- and disaccharides), flavonoids, and triterpenoids (Zou et al., 2021). Pharmacological studies demonstrated that extracts and components of *Kochiae Fructus* have anti-cancer, anti-microbial, anti-gastric mucosal damage, anti-edema, anti-nociceptive, anti-allergic, anti-inflammatory, anti-dermatitic, anti-pruritic, anti-itching, and hypoglycemic effects (Zou et al., 2021). Due to the presence of phytochemicals (polyphenols, flavonoids, terpenoids, etc.) and their anti-inflammatory properties, we hypothesized that *Kochiae Fructus*

would be a promising therapeutic agent for COVID-19. The active molecules and mechanisms of action of *Kochiae Fructus* in treating COVID-19 have not been reported yet. Thus, we have provided several justifications for our hypothesis about the effectiveness of *Kochiae Fructus* in treating COVID-19.

Network pharmacology and bioinformatics are emerging interdisciplinary fields involved in drug research and development. They utilize artificial intelligence as well as big data to identify active drug molecules and their molecular pathways (Xia, 2017). According to TCM's holistic approach, network pharmacology intends to investigate the effectiveness of pharmaceuticals holistically, shifting research away from conventional one drug, one target mode, and toward a developing one drug, network targets paradigm (Zhou Z. et al., 2020). In the current study, we, therefore, employed different bioinformatics tools, including network pharmacology and molecular docking approaches, to predict biologically active molecules, molecular targets, as well as molecular pathways implicated in *Kochiae Fructus*'s anti-COVID-19 effects (Figure 1).

Materials and methods

Collection of *Kochiae Fructus*'s active molecules

We have retrieved the active phytochemicals of *Kochiae Fructus* utilizing the TCMSP (Version 2.3¹) (Ru et al., 2014). The TCMSP database includes the relationships between each component's absorption, distribution, metabolism, and excretion (ADME) characteristics. Oral bioavailability "OB" is a term that refers to "the rate and degree to which an active component or active moiety is absorbed from a therapeutic product that becomes accessible at the targeted site." Drug Likeness "DL" is a qualitative paradigm for drug design that incorporates the ADME qualities of ingredients and established medications. Thus, the therapeutic phytochemicals of *Kochiae Fructus* that satisfied the requirements of both $OB \geq 20\%$ and $DL \geq 0.10$ were selected for further analysis (Chen et al., 2001; Ru et al., 2014; Ma X. et al., 2020; Wang et al., 2021).

Screening of COVID-19-related targets

The COVID-19-related targets were retrieved from the GeneCards® : The Human Gene Database² by searching the keywords "Novel Coronavirus Pneumonia" and "Novel Coronavirus," (Rebhan et al., 1997). After combining the targets

of both keywords, the duplication of targets was removed using Venny 2.1 online database³ (Oliveros, 2007).

SwissTargetPrediction of *Kochiae Fructus*'s active phytochemicals

By using the SwissTargetPrediction database,⁴ the potential targets of *Kochiae Fructus*'s active phytochemicals were acquired (Daina et al., 2019). The only potential targets with a probability score greater than 0 were selected for further analysis.

Identification of intersecting targets

The intersection targets between the potential targets of the *Kochiae Fructus*'s active phytochemicals and the COVID-19-related targets were identified using the Venny 2.1 online database (see text footnote 3) (Oliveros, 2007). These identified intersecting targets were recognized as potential anti-COVID-19 key targets.

Protein-protein interaction analysis

The potential anti-COVID-19 key targets (identified in section "Identification of intersecting targets") were then assessed for protein-protein interaction (PPI) analysis using the STRING database (version 11.5, accessed on April 8, 2022⁵) at a confidence score > 0.4 and a species limited to "Homo sapiens" (von Mering et al., 2003; Chu et al., 2020). The string PPI analysis results were then uploaded to Cytoscape software (version 3.9.0, Boston, MA, the United States, accessed on April 8, 2022) to identify the potential anti-COVID-19 core targets (Lopes et al., 2010).

Molecular Complex Detection analysis

The important modules in the PPI network of 34 potential anti-COVID-19 key targets were then determined by employing the Molecular Complex Detection (MCODE) plug-in in Cytoscape software (accessed on April 8, 2022) (Lopes et al., 2010). The conditions for MCODE analysis were Find clusters: in the whole network, Degree cutoff: 2, Node score cutoff: 0.2, K-core: 2, and Maximum Depth: 100 (Cao et al., 2021).

1 <https://old.tcmsp-e.com/tcmsp.php>, accessed on April 8, 2022

2 <https://www.genecards.org/>, accessed on April 8, 2022

3 <https://bioinfogp.cnb.csic.es/tools/venny/>, accessed on April 8, 2022

4 <http://www.swisstargetprediction.ch/>, accessed on April 8, 2022

5 <https://string-db.org/>

was performed on 34 potential anti-COVID-19 key targets (identified in section “Identification of intersecting targets”) using DAVID (Version 6.8)⁶ (Huang et al., 2008; Zhang et al., 2021). The GO terms were categorized into three types: cellular component (CC), biological process (BP), and molecular function (MF). By uploading the data to the Bioinformatics platform,⁷ the top 10 GO analysis data (BP, CC, and MF) and top 30 KEGG pathways were further exhibited in the form of an enrichment dot bubble (Weishengxin, 2022). The classical hypergeometric test was used to determine statistical significance. The adjusted p -value < 0.05 was utilized as the significant threshold in our investigation after utilizing the Benjamini–Hochberg method to control the false discovery rate (FDR) for multiple hypothesis testing (Chu et al., 2020).

Molecular docking

Briefly, two-dimensional (2D) structures of *Kochiae Fructus*’s key active phytomolecules were obtained from the NCBI PubChem⁸ online database in Spatial Data File (SDF) format and three-dimensional (3D) structures were designed utilizing BIOVIA Discovery Studio Visualizer 2021 and saved in PDB file format (BIOVIA DS, 2016; NIH, 2022). Protein crystal structures of top-six potential anti-COVID-19 core targets were obtained in PDB format from the Protein Data Bank⁹ (Berman et al., 2000; RCSB PDB, 2022). The ligands and water molecules from the protein crystal structure complexes were retrieved using BIOVIA Discovery Studio Visualizer 2021 software. Following that, a grid of each protein was constructed using it, polar hydrogen was added, and the resulting protein was saved in PDB file format (BIOVIA DS, 2016). The PDB-formatted proteins were then uploaded to AutoDock Vina (version 1.2.0), and the Kollman and Gasteiger partial charges were applied. The key active phytomolecules of *Kochiae Fructus* were then transferred to AutoDock Vina in PDB format. Proteins and key active phytomolecules were converted to pdbqt format, and AutoDock Vina scripts were written for molecular docking (Trott and Olson, 2010). Finally, the acquired docked complexes were visualized using the software BIOVIA Discovery Studio Visualizer 2021 to assess the binding capacity of the key active phytomolecules and potential anti-COVID-19 core targets (BIOVIA DS, 2016). Binding energy less than zero shows that a ligand may spontaneously bind to the receptor. It is generally accepted that the lower the energy score of the ligand-receptor binding configuration, the

greater the likelihood of binding (Trott and Olson, 2010; Chu et al., 2020).

Results

Active compounds screening

From the TCMSP database, 19 phytomolecules from *Kochiae Fructus* were retrieved. In addition, screening for active phytomolecules based on $OB \geq 20\%$ and $DL \geq 0.10$ resulted in identifying six active phytomolecules in *Kochiae Fructus*. Table 1 lists each of *Kochiae Fructus*’s six active phytomolecules.

Potential targets of *Kochiae Fructus*’s active phytomolecules

Using the SwissTargetPrediction online database, the potential targets of *Kochiae Fructus*’s active phytomolecules were identified. With a probability score > 0 , 459 potential targets of six active phytomolecules were identified. 165 potential targets were chosen for further investigation after redundancies were eliminated.

COVID-19-related targets

By searching the GeneCards database with the terms “Novel Coronavirus Pneumonia” and “Novel Coronavirus,” 1745 COVID-19-related targets were identified.

Intersecting targets analysis

Using the VENNY 2.1.0 online tool, the intersecting targets between the potential targets of *Kochiae Fructus*’s active phytomolecules and the COVID-19-related targets were analyzed, and 34 intersecting targets were determined as potential anti-COVID-19 key targets (Figure 2).

Protein-protein interaction network construction and analysis

The PPI network was established by uploading 34 potential anti-COVID-19 key targets into the STRING database version 11.5, as illustrated in Figure 3A. The result reveals that the network has 34 nodes and 115 edges. Correspondingly, the average node degree, local clustering coefficient, expected number of edges, and PPI enrichment p -values were 6.76, 0.667, 32, and $p < 0.0000001$.

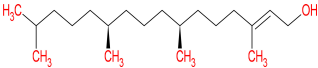
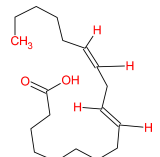
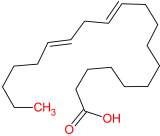
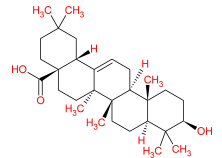
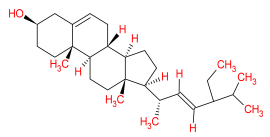
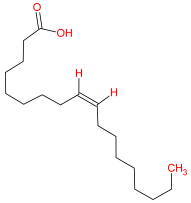
⁶ <https://david.ncifcrf.gov/>, accessed on April 8, 2022

⁷ <http://www.bioinformatics.com.cn/>, accessed on April 8, 2022

⁸ <https://pubchem.nlm.nih.gov/>, accessed on April 9, 2022

⁹ <https://www.rcsb.org/>, accessed on April 9, 2022

TABLE 1 List of *Kochiae Fructus*'s active phytomolecules based on OB \geq 20% and DL \geq 0.10.

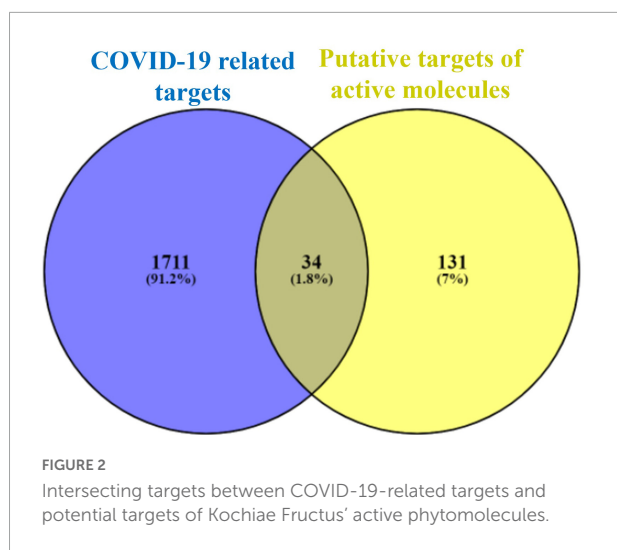
Phytomolecules ID	Phytomolecules name	Phytomolecules structure	Molecular weight	OB (%)	DL
MOL001442	Phytol		296.6	33.82	0.13
MOL002038	9E,12Z-octadecadienoic acid		280.5	41.9	0.14
MOL002211	11,14-eicosadienoic acid		308.56	39.99	0.2
MOL000263	Oleanolic acid		456.78	29.02	0.76
MOL000449	Stigmasterol		412.77	43.83	0.76
MOL000675	Oleic acid		282.52	33.13	0.14

In addition, the PPI findings in a simple text data format (.tsv) file were then imported into the Cytoscape 3.9.0 software to visualize and analyze the PPI network. As seen in **Figure 3B**, the PPI network consisted of 32 nodes (after removing two unconnected nodes) and 115 edges, with an average shortest path length of 2.022 between all node pairings. The network's density, diameter, and radius were 0.232, 4, and 3, respectively. The average number of neighbors, clustering coefficient, network heterogeneity, and network centralization were 7.188, 0.61, 0.75, and 0.475, respectively. As a PPI network node's degree lowers, its color changes from red to yellow. The 13 nodes satisfying the degree centrality (DC) > average value (7.187) requirement were retrieved and identified as anti-COVID-19 core targets (**Figure 3C**). The 13 anti-COVID-19 core targets sorted by DC are shown as a bar graph in **Figure 4**. The top six anti-COVID-19 core targets, IL-6, PPARG, MAPK3, PTGS2, ICAM1, and MAPK1, were selected for molecular docking analyses with key active phytomolecules of *Kochiae Fructus* identified in section

"Network of anti-COVID-19 targets and the *Kochiae Fructus*'s active phytomolecules."

Molecular Complex Detection analysis

The MCODE analysis identified two cluster networks inside the PPI network of 34 anti-COVID-19 key targets, as seen in **Figure 5**. Cluster network 1 consisted of 7 nodes and 17 edges, with a score of 5.667. The components of cluster network 1 were IL6, AR, PPARG, MAPK3, MDM2, PTGS2, and TLR9. In cluster network 1, IL6, PPARG, MAPK3, and PTGS2 have extensive interconnections with multiple targets. Cluster network 2, on the other hand, has 10 nodes and 22 edges with a score of 4.889. PRKCA, PRKCB, PRKCCE, PRKCQ, MAPK1, NOS2, PLA2G4A, MAPK14, KEAP1, and ICAM1 were all part of Cluster Network 2. PRKCA, PRKCB, and MAPK1 are the highly interconnected targets in cluster network 2. Moreover, both cluster networks have found the



top six anti-COVID-19 core targets. Cluster network 1 had four of them (IL6, PPARG, MAPK3, and PTGS2). Two of them (ICAM1 and MAPK1) were found in cluster network 2; thus, both the PPI network and MCODE analyses yielded consistent findings.

Network of anti-COVID-19 targets and the *Kochiae Fructus*'s active phytomolecules

As shown in **Figure 6A**, the network between the 34 potential anti-COVID-19 key targets and the active phytomolecules of *Kochiae Fructus* was established using Cytoscape software 3.9.0. It consists of 40 nodes and 86 edges. Each edge represents the interaction between active phytomolecules and potential anti-COVID-19 key targets. In addition, the network's diameter and radius were, respectively, four and two. A node's degree reflects the number of edges connecting it to other network nodes. As the degree of a node rises, its color changes from yellow (lowest degree) to red (highest degree).

In addition, the hub network between the 13 potential anti-COVID-19 core targets and the active phytomolecules of *Kochiae Fructus* was constructed using the Cytoscape software version 3.9.0, as shown in **Figure 6B**. All six active phytomolecules of *Kochiae Fructus* have been shown to interact with thirteen potential anti-COVID-19 core targets. The six active phytomolecules of *Kochiae Fructus* with their degree values in the hub network are presented as a bar graph in **Figure 7**. MOL002038 (9E,12Z-octadecadienoic acid), MOL002211 (11,14-eicosadienoic acid), and MOL000263 (oleanolic acid) are, respectively, the top three active phytomolecules that interact with nine, seven, and five anti-COVID-19 core targets ($DC \geq 5$). Furthermore, the

network findings revealed that single key active phytomolecules may interact with multiple anti-COVID-19 core targets and that multiple key active phytomolecules can interact synergistically with a single anti-COVID-19 core target. Thus, our results verified the combinatorial interaction mechanism between numerous anti-COVID-19 core targets and multiple key active phytomolecules in *Kochiae Fructus*.

Gene ontology enrichment analysis

The 34 potential anti-COVID-19 key targets were subjected to enrichment analysis of GO terms. **Figure 8** depicts the top 10 enrichment terms of MF, CC, and BP for 34 potential anti-COVID-19 key targets of *Kochiae Fructus*. According to GO enrichment analysis, the gene targets associated with BP are embroiled in platelet activation, peptidyl-serine phosphorylation, inflammatory response, positive regulation of gene expression, intracellular transduction, positive regulation of transcription from RNA, polymerase II promoter, apoptotic process, etc. Gene targets in CC are primarily found in the cytoplasm, plasma membrane, nucleoplasm, cytosol, nucleus, endoplasmic reticulum, etc. GO enrichment analysis indicates further that protein serine/threonine kinase activity, calcium-dependent protein kinase activity, protein kinase C activity, enzyme binding, integrin binding, MAPK kinase activity, etc., dominate the enriched MF ontology.

Kyoto Encyclopedia of Genes and Genomes analysis

Figure 9 depicts the thirty most significant KEGG pathways of the thirty-four potential anti-COVID-19 key targets. The KEGG pathway enrichment study suggested that *Kochiae Fructus* anti-COVID-19 key targets may be implicated in the AGE-RAGE signaling pathway in diabetic complications, the VEGF signaling pathway, cancer pathways, neurodegeneration-multiple diseases pathways, inflammatory mediator regulation of TRP channels, GnRH signaling pathway, Fc epsilon RI signaling pathway, Sphingolipid signaling pathway, TNF signaling pathway, HIF-1 signaling pathway, etc. These pathways may contribute substantially to the underlying mechanisms of *Kochiae Fructus* in alleviating COVID-19.

Network between top six anti-COVID-19 core targets and thirty enriched KEGG pathways

A network was established connecting the top six anti-COVID-19 core targets and their related pathways to determine

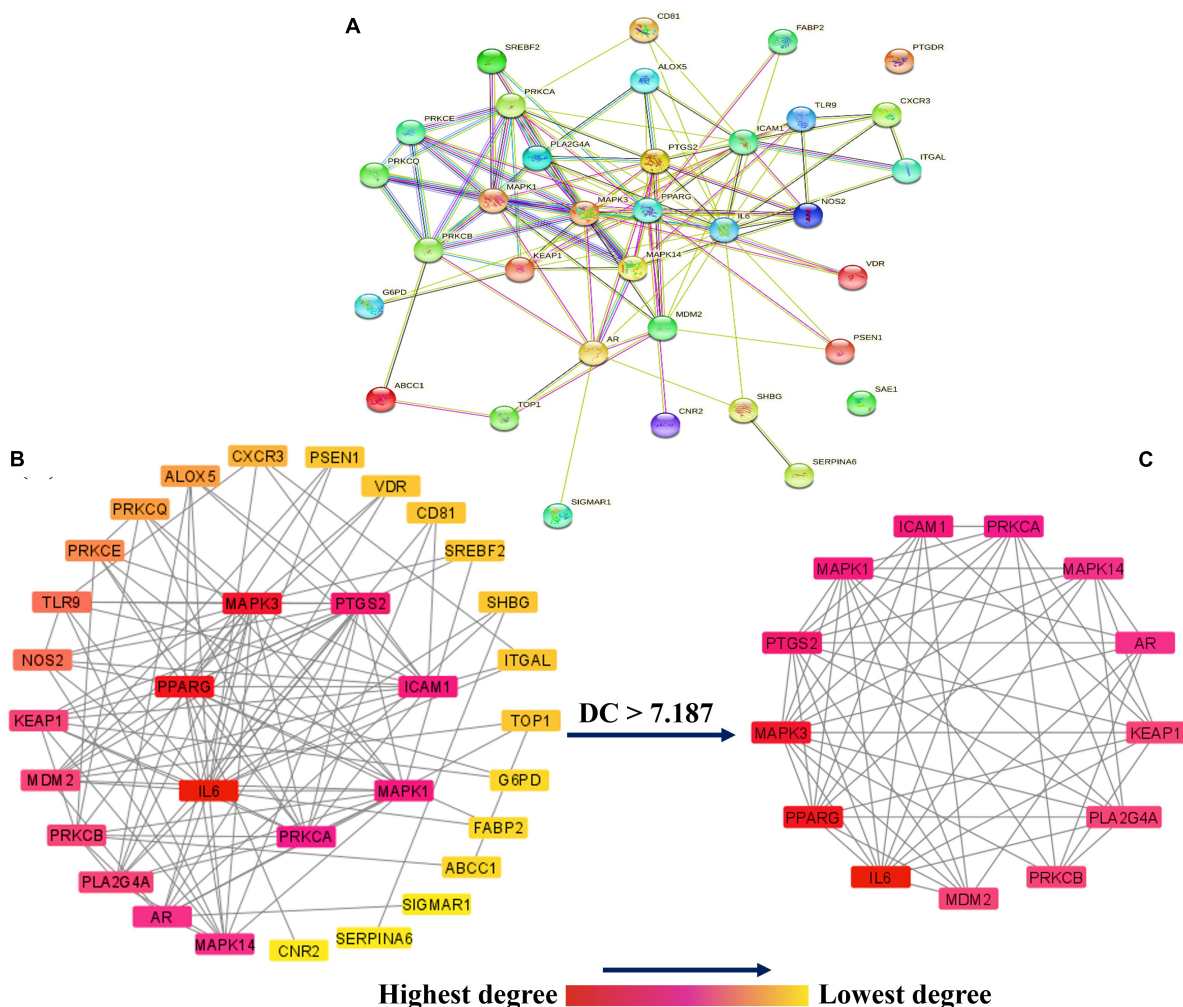


FIGURE 3

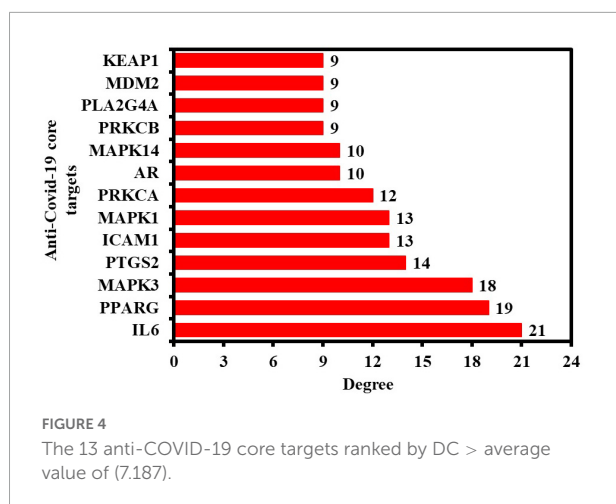
(A) The PPI network of 34 potential anti-COVID-19 key targets constructed by employing STRING. The PPI network of (B) 34 potential anti-COVID-19 key targets, and (C) 13 potential anti-COVID-19 core targets constructed by using the Cytoscape software. The color of each node changes from red (highest) to yellow (lowest) as its degree decreases. DC stands for degree centrality.

the key pathways involved in the anti-COVID-19 effects of *Kochia fructus*' active phytochemicals. The network findings revealed that five anti-COVID-19 core targets followed the pathways in cancer (IL6, MAPK1, PPARG, PTGS2, and MAPK3) (degree = 5), TNF signaling pathway (IL6, MAPK1, PTGS2, ICAM1, and MAPK3) (degree = 5), and Lipid and atherosclerosis (IL6, MAPK1, PPARG, ICAM1, and MAPK3) (degree = 5) (Figure 10).

Furthermore, the pathways in the network were sorted by $DC > \text{average value (3.03)}$ to identify the key pathways. As illustrated in Figure 11, seven significant pathways were determined and shown in a bar graph with their degree values in the network. By regulating the expression of anti-COVID-19 core targets, these seven pathways may contribute considerably to the anti-COVID-19 effects of the active phytochemicals in *Kochia fructus*.

Molecular docking studies

The top six anti-COVID-19 core targets, IL-6, PPARG, MAPK3, PTGS2, ICAM1, and MAPK1, were molecularly docked with the three key active phytochemicals of *Kochia fructus*. Table 2 displays the outcomes of molecular docking in terms of docking scores. Figures 12A–L, 13A–L and Supplementary Figures 1A–L present in 2D and 3D the docked complexes of key active phytochemicals and anti-COVID-19 core targets with the lowest docking scores. The docking score represents an active phytochemicals (ligands) affinity for anti-COVID-19 core targets (receptor). The lower the docking score, the better the ligand's interaction with the receptor (Trott and Olson, 2010; Khan and Lee, 2022). Docking scores < -4.25 , < -5.0 , and < -7.0 , respectively, correspond to the existence, good, and strong docking activity between the key



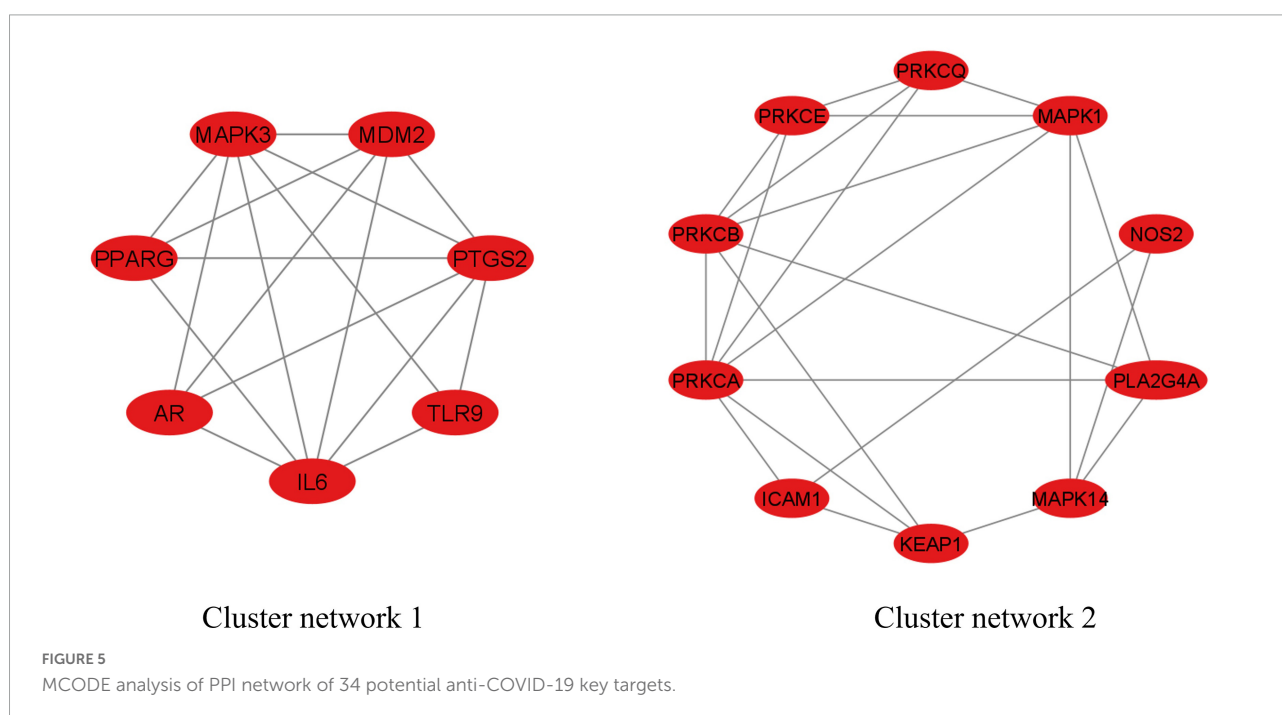
active phytomolecules and the anti-COVID-19 core targets (Xie et al., 2021).

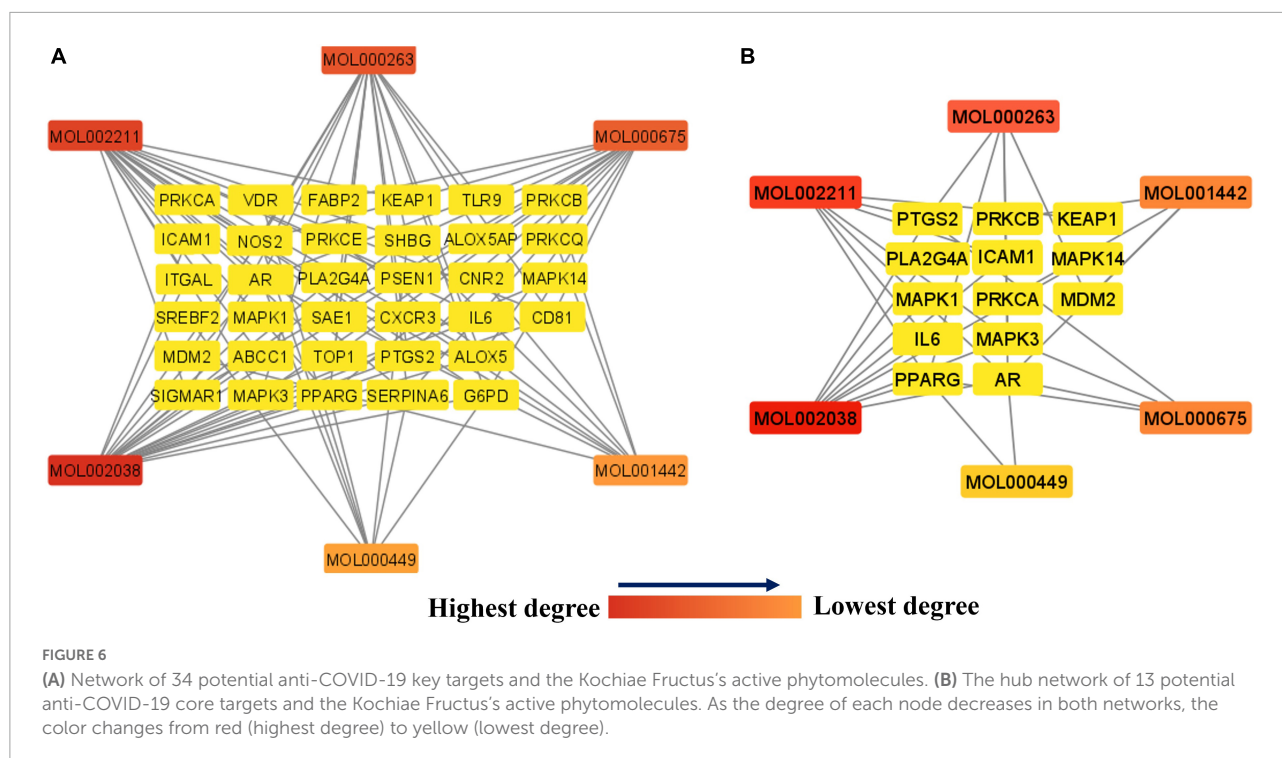
The results demonstrate that the MOL000263 (Oleanolic acid) exhibited a strong binding affinity with IL-6 by presenting a docking score of -6.1 (Table 2) and interacted with amino acid residues (ARG182 and LEU178) through conventional hydrogen bonding interaction (Figure 12A). MOL002038 (9E,12Z-octadecadienoic acid) and MOL002211 (11,14-eicosadienoic acid) demonstrated strong binding affinity with PPARG by exhibiting docking scores (-6.1 and -6.0) respectively, compared to MOL000263 (oleanolic acid), which presented a -1.0 -docking score and interaction with amino acid residue (HIS449) via pi-sigma bond interaction (Figure 12C).

The docking activity of all three key active phytomolecules with MAPK3 ranged from strong to excellent. MOL000263 (Oleanolic acid) provided the best docking score (-8.0) and was coupled with amino acid residue (ASP128) through conventional hydrogen bonding interaction (Figure 12E). Compared to the other two key active phytomolecules, MOL000263 (Oleanolic acid) coupled with PTGS2's amino acid residues (ASN144) through conventional hydrogen bond interaction and exhibited a docking score of -4.9 (Figure 12G). All three active phytomolecules displayed a high affinity for binding with MAPK1. MOL000263 (Oleanolic acid) coupled with MAPK1's amino acid residues (LYS151 and SER153) through conventional hydrogen bonding and carbon-hydrogen bonding, respectively, to achieve the best docking score (-8.0) (Figure 12I). The docking scores of MOL002038 (9E,12Z-octadecadienoic acid), MOL002211 (11,14-eicosadienoic acid), and MOL000263 (Oleanolic acid) with ICAM1 were -4.7 , -4.8 , and -5.9 , respectively. These results suggested that the key active phytomolecules of *Kochia fructus* seemed to have a regulatory impact on the anti-COVID-19 core targets investigated, including IL-6, PPARG, MAPK3, PTGS2, ICAM1, and MAPK1. In addition, the findings of molecular docking and network pharmacology screening were consistent, indicating the validity of network pharmacology in this research.

Discussion

This work investigated *Kochia fructus*'s active phytomolecules and molecular pathways for COVID-19

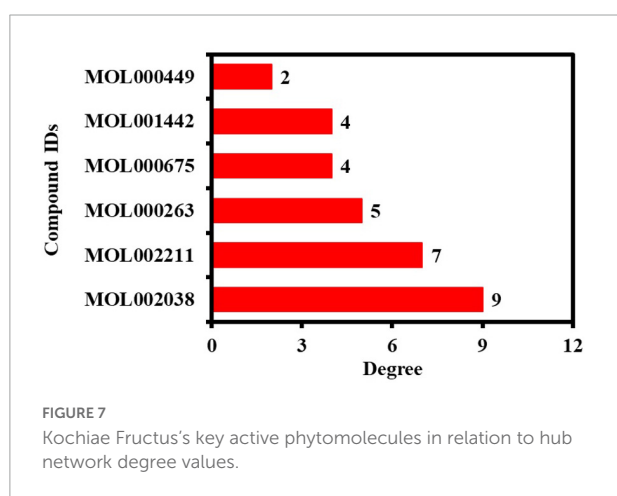




therapy. Based on $OB \geq 20\%$ and $DL \geq 0.10$, six compounds were identified as active phytomolecules in *Kochiae Fructus* (Table 1). MOL002038 (9E,12Z-octadecadienoic acid), MOL002211 (11,14-eicosadienoic acid), and MOL000263 (oleanolic acid) are the top three key active phytomolecules interacting with more than four anti-COVID-19 core targets. In addition, the network analysis demonstrated that a single key active phytomolecule may interact with many anti-COVID-19 core targets and that several key active phytomolecules can interact synergistically with a single anti-COVID-19 core target. Consequently, our findings revealed a combinatorial

interaction between the numerous anti-COVID-19 core targets and multiple key active phytomolecules in *Kochiae Fructus* for COVID-19 therapy.

The PPI network analysis reveals that multiple targets, including IL-6, PPARG, MAPK3, PTGS2, ICAM1, MAPK1, etc., are involved in the anti-COVID-19 effects of *Kochiae Fructus*' active phytomolecules (Figures 3, 4). In COVID-19 patients, IL6 was markedly overexpressed and attributed to inflammation. CRS mediated by IL-6 is prevalent among COVID-19 patients and is accountable for their severe acute respiratory distress. By suppressing IL6 with tocilizumab, Xu et al. showed a reduction in CRS and a quick improvement in symptoms in patients (Ruan et al., 2020b; Wang et al., 2020; Xu et al., 2020). Reports indicate that overactivation/upregulation of MAPK (MAK1 and MAPK3) mediates the production of inflammatory cytokines such as IL-1 β , IL6, IL10, TNF- α , IL4, and INF- γ (de Souza et al., 2014; Grimes and Grimes, 2020; Oh et al., 2021) and these cytokines induce CRS. Furthermore, MAPK overactivation has been linked to thrombotic events and vascular endothelial infections in critically sick COVID-19 patients (Grimes and Grimes, 2020; Varga et al., 2020; Zhang Y. et al., 2020; Zhou F. et al., 2020). Furthermore, owing to MAPK overactivation, alveolar tissues are injured, resulting in reduced ventilation, acute lung injury, and acute respiratory distress syndrome (Goel et al., 2021). Consequently, targeting IL6, MAPK1, and MAPK3 may be a viable alternative therapeutic approach for treating CRS and COVID-19. On the other hand, allergic responses are caused by soluble ICAM1 in nasal



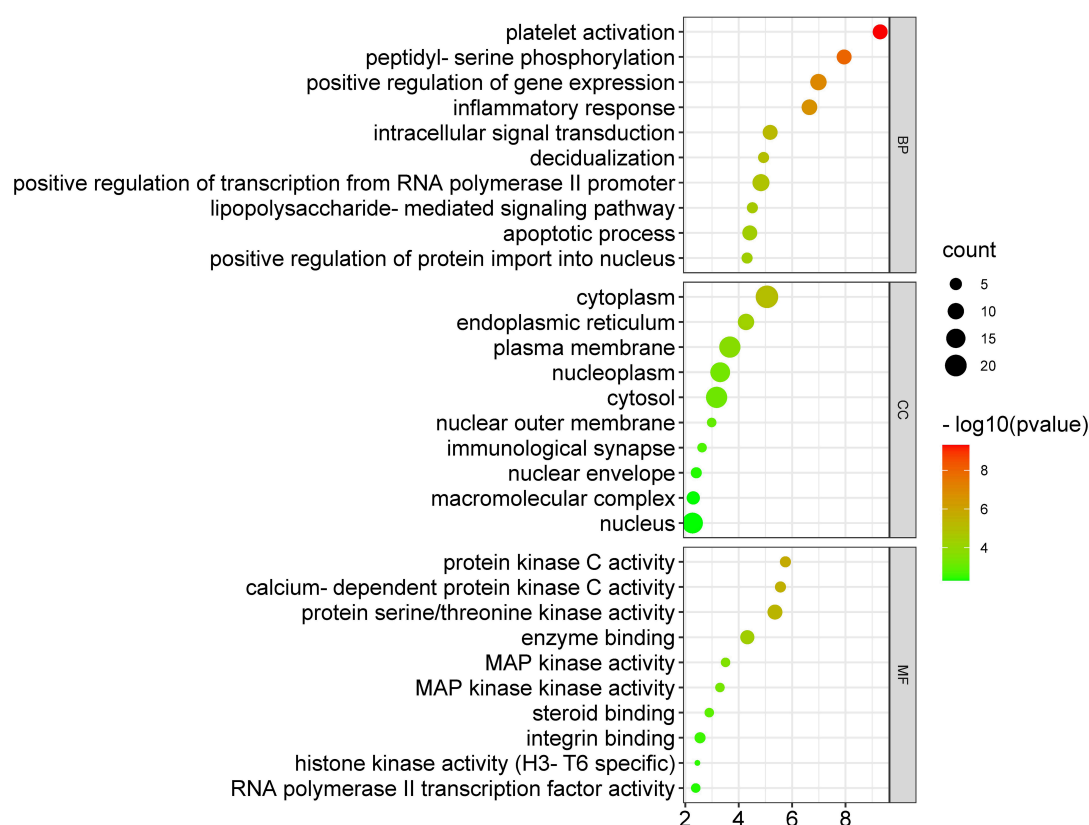


FIGURE 8

GO enrichment analysis of 34 potential anti-COVID-19 key targets.

epithelial cells. ICAM1 is found in a significant proportion of bronchial asthma patients. ICAM1 offers a suitable atmosphere for the coronavirus to infiltrate and live inside the human nose (Winther et al., 2002; Sharma et al., 2021). Nagashima et al. (2020) reported that COVID-19 endothelial cells overexpressed ICAM1 relative to controls, which might attract leukocytes (endotheliitis) and send intracellular signals leading to a pro-inflammatory state. A proinflammatory condition causes systemic endothelial dysfunction and endothelial cell death. The adhesion molecules' chronic inflammatory signaling contributes to thrombosis (Nagashima et al., 2020). PTGS2 (COX2) is involved in controlling the homeostasis of the organism. Inflammatory stimuli modulate PTGS2 expression (Ricciotti and Fitzgerald, 2011). Chen et al. reported that SARS-CoV-2 infection increased the expression of PTGS2 in animal systems and human cell cultures (Chen et al., 2021). Inhibiting ICAM1 and PTGS2 with inhibitors can be a potential therapeutic strategy in treating COVID-19. PPARs, especially PPARγ, is responsible for regulating glucose and lipid metabolism. Additionally, they control homeostasis and inhibit the production of certain proinflammatory cytokines. However, alterations in their expression (downregulation) result in the inflammation involved with COVID-19 etiology

(Ciavarella et al., 2020). Desterke et al. (2020) reported the suppression of PPARγ in severe COVID-19 and proposed that it contributes to monocyte/macrophage-mediated inflammatory storm. Thus, activating PPARγ with an agonist might be a potential COVID-19 therapeutic modality.

According to GO enrichment analysis, *Kochia fructus* active phytochemicals might be implicated in affecting COVID-19 targets associated with multiple BP, including platelet activation, peptidyl-serine phosphorylation, inflammatory response, positive regulation of gene expression, intracellular signal transduction, positive regulation of transcription from RNA, polymerase II promoter, apoptotic process, etc. Platelet hyperactivation has been observed in COVID-19 patients. Platelets secrete procoagulants in critically ill individuals. They influence immune responses by interacting with other immune cells, resulting in severe thromboinflammation. Targeting these pathways may reduce immune response (Jevtic and Nazy, 2022; Uzun et al., 2022). Patients with COVID-19 have been shown to have elevated levels of peptidyl-serine phosphorylation (Ong et al., 2021; Chatterjee and Thakur, 2022). The inflammatory response is caused by an overactivation of the innate and adaptive immune systems and the invasion of SARS-CoV-2, which invades and activates numerous immune cells

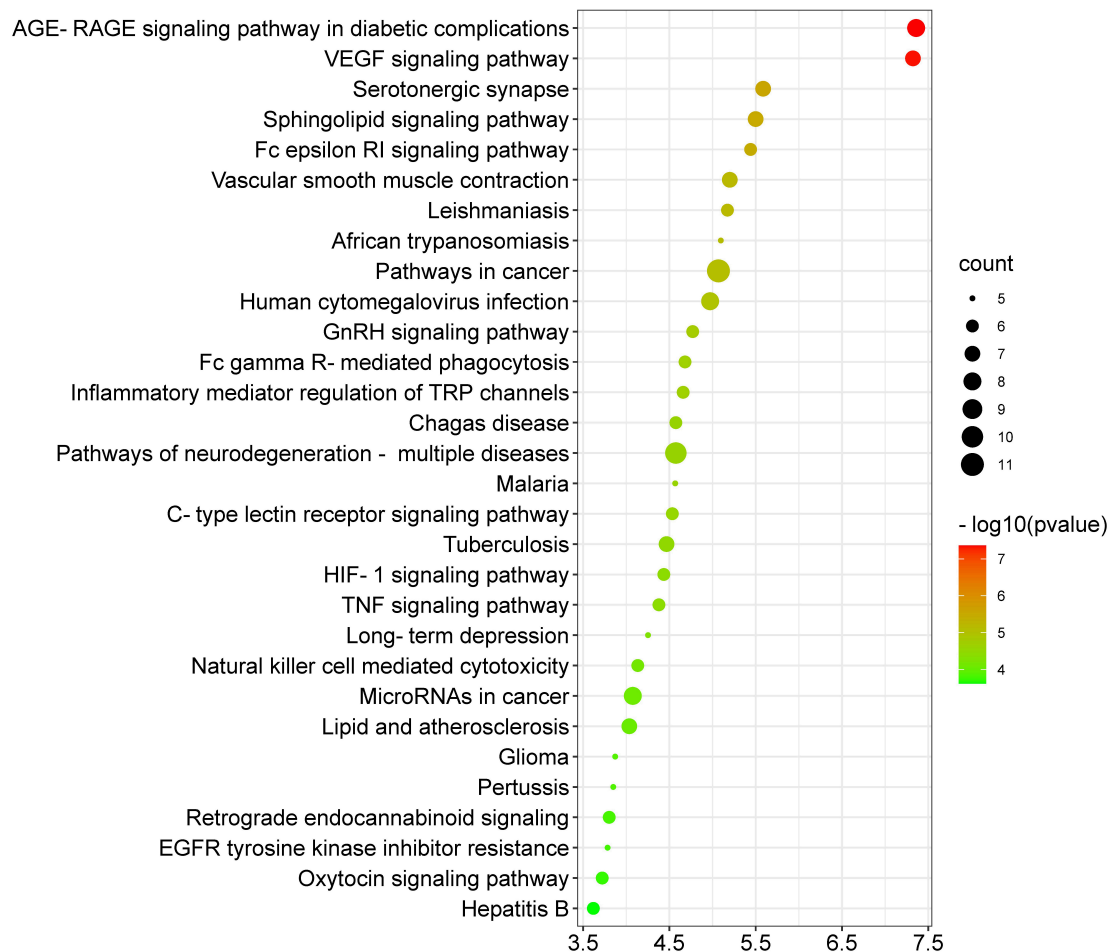


FIGURE 9

The top thirty enriched KEGG pathways of 34 potential anti-COVID-19 key targets.

(Tan et al., 2021). Multiple intracellular transductions, such as TNF-alpha, mTOR, NF- κ B, etc., are implicated in the pathogenesis of COVID-19 (Farahani et al., 2022). *Kochia* Fructus' active phytochemicals impacting COVID-19 targets associated with numerous CC are predominantly located in the cytoplasm, plasma membrane, nucleoplasm, cytosol, nucleus, endoplasmic reticulum, etc. Moreover, these COVID-19 targets perform multiple MF such as protein serine/threonine kinase activity, calcium-dependent protein kinase activity, protein kinase C activity, enzyme binding, integrin binding, MAPK kinase activity, etc.

The KEGG pathway enrichment investigation revealed that the molecular pathways driving the anti-COVID-19 effects of *Kochia* Fructus' active phytochemicals might be involved in the AGE-RAGE signaling pathway in diabetic complications, the VEGF signaling pathway, cancer pathways, neurodegeneration-multiple diseases pathways, inflammatory mediator regulation of TRP channels, GnRH signaling pathway, Fc epsilon RI signaling pathway, Sphingolipid signaling

pathway, TNF signaling pathway, HIF-1 signaling pathway, etc. Moreover, the network analysis of the top six anti-COVID-19 core targets and thirty enriched KEGG pathways demonstrated that seven pathways, pathways in cancer, TNF signaling pathway, lipid and atherosclerosis, AGE-RAGE signaling pathway in diabetic complications, human cytomegalovirus infection, pathways of neurodegeneration-multiple diseases, and C-type lectin receptor signaling pathways, are the key pathways ranked by $DC > \text{average value}$ (3.03). Among them, pathways in cancer and the TNF signaling pathway are the core pathways followed by five anti-COVID-19 core targets. The abnormal expression of targets (IL6, MAPK1, PPARG, PTGS2, and MAPK3) leads to cell proliferation and a variety of cancer types. Regulating pathways in cancer may modulate their expression. The pathophysiology of COVID-19 is also attributed to the expression/activation of these targets. In COVID-19 treatment, their expression may also be regulated by targeting cancer-related pathways. For instance, Hayashi et al. demonstrated that

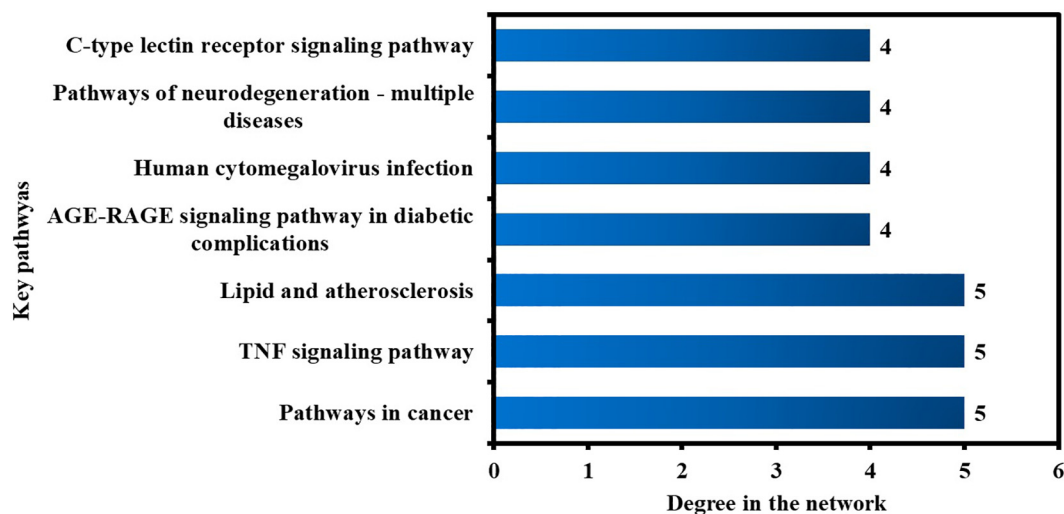
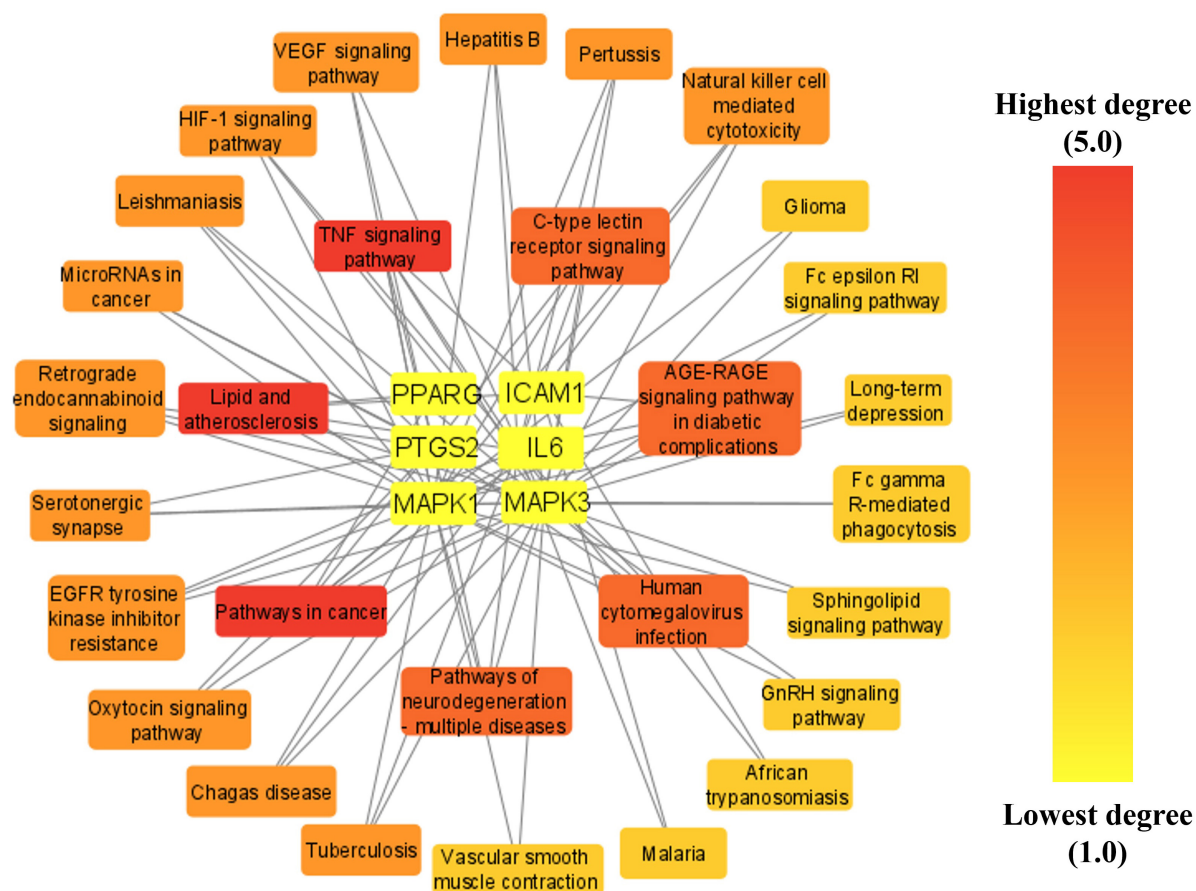
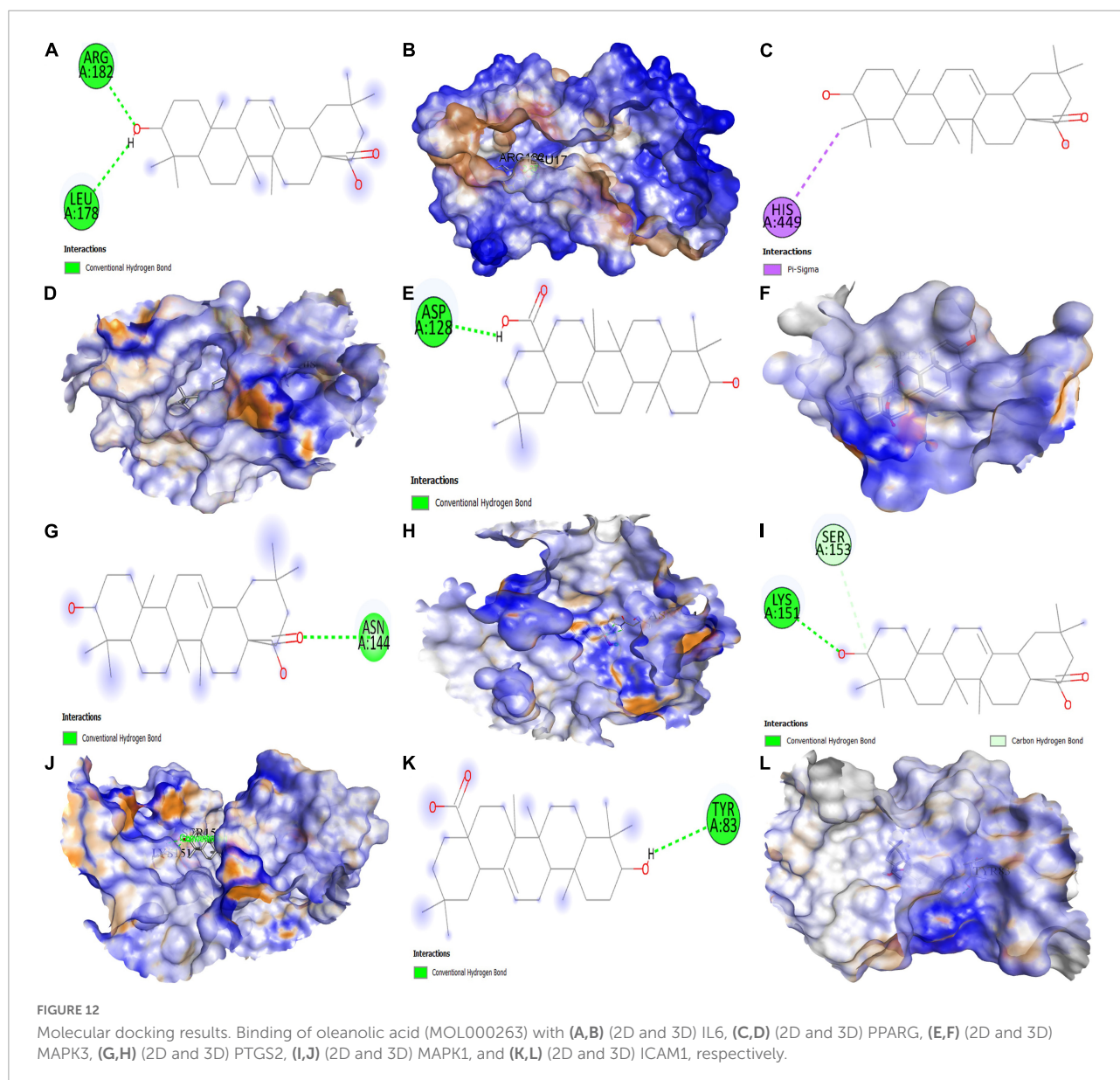


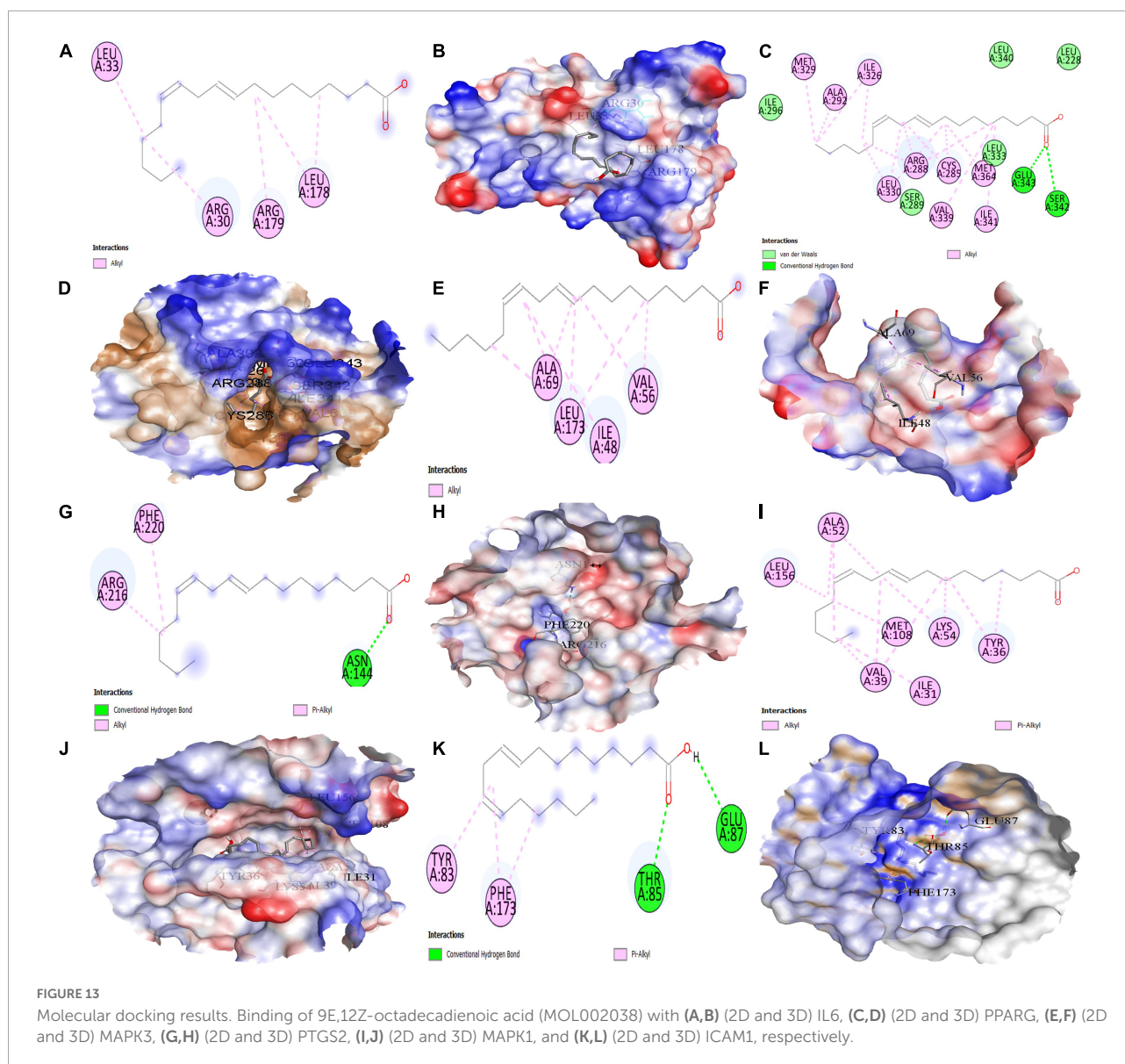
TABLE 2 Presents the binding affinities of *Kochia fructus*'s key active phytomolecules with the top six anti-COVID-19 core targets.

Phytomolecule IDs	Phytomolecule names	Binding affinities (kcal/mol)					
		IL6	PPARG	MAPK3	PTGS2	MAPK1	ICAM1
MOL000263	Oleanolic acid	-6.1	-1.0	-8.0	-4.9	-9.3	-5.9
MOL002038	9E,12Z-octadecadienoic acid	-4.5	-6.1	-5.5	-2.9	-5.9	-4.7
MOL002211	11,14-eicosadienoic acid	-4.1	-6.0	-5.6	-3.1	-6.3	-4.8



antineoplastic drugs inhibit MAPK and prevent SARS-CoV-2 replication (Hayashi and Konishi, 2021). Reports show that an overactive TNF signaling pathway has been identified as a key contributor to the development and severity of COVID-19. Its overactivation is induced by various factors (innate

immune activation, cytokines release, high glucose level, etc.), leading to the onset of inflammation and the eventual failure of multiple organs (heart failure, insulin resistance, lung damage, blood clotting, etc.) (Ablamunits and Lepsy, 2022). Hence, these pathways may play a significant role in the underlying



mechanisms of *Kochiae Fructus*' active phytomolecules in COVID-19 alleviation.

Conclusion

In this research, we have successfully determined the active phytomolecules and molecular pathways of *Kochiae Fructus* for COVID-19 therapy. Thirteen anti-COVID-19 core targets and three key active phytomolecules have been determined. This research also showed that, out of thirteen targets, the top six (IL-6, PPARG, MAPK3, PTGS2, ICAM1, and MAPK1) are likely to be implicated in the anti-COVID-19 effects of *Kochiae Fructus*' active phytomolecules. The underlying mechanisms through which *Kochiae Fructus*'

active phytomolecules demonstrate anti-COVID-19 effects are the inhibition/regulation of the multiple BP (platelet activation, peptidyl-serine phosphorylation, inflammatory response, positive regulation of gene expression, intracellular transduction, positive regulation of transcription from RNA, polymerase II promoter, apoptotic process, etc.). We identified three key pathways (pathways in cancer, the TNF signaling pathway, and lipid and atherosclerosis) involved in the treatment of COVID-19 with active phytomolecules of *Kochiae Fructus*. Consequently, our results demonstrated a synergistic impact between the multiple anti-COVID-19 core targets, multiple molecular pathways, and key active phytomolecules in *Kochiae Fructus* for treating COVID-19. The findings of molecular docking revealed that the key active phytomolecules of *Kochiae Fructus* had a regulatory effect

on the anti-COVID-19 core targets. Furthermore, molecular docking and network pharmacology screening results were consistent, demonstrating network pharmacology's validity in this research. Hence, these findings offer a foundation for developing anti-COVID-19 drugs further in the future based on phytomolecules of *Kochia fructus*.

Data availability statement

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

Author contributions

SK and TL: conceptualization, validation, investigation, and writing—review and editing and original draft preparation. and editing. SK: methodology, software, formal analysis, data curation, and visualization. TL: supervision, project administration, and funding acquisition. Both authors have read and agreed to the published version of the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (81871935) and the PolyU Distinguished Postdoctoral Fellowship Scheme.

References

- Ablamunits, V., and Lepsy, C. (2022). Blocking TNF signaling may save lives in COVID-19 infection. *Mol. Biol. Rep.* 49, 2303–2309. doi: 10.1007/S11033-022-07166-X
- Alzaabi, M. M., Hamdy, R., Ashmawy, N. S., Hamoda, A. M., Alkhayat, F., Khademi, N. N., et al. (2022). Flavonoids are promising safe therapy against COVID-19. *Phytochem. Rev.* 21, 291–312. doi: 10.1007/S11101-021-09759-Z
- 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, 235–242. doi: 10.1093/NAR/28.1.235
- BIOVIA DS (2016). BIOVIA discovery studio visualizer. 20:779.
- Cao, F., Cheng, Y. S., Yu, L., Xu, Y. Y., and Wang, Y. (2021). Bioinformatics analysis of differentially expressed genes and protein–protein interaction networks associated with functional pathways in Ulcerative Colitis. *Med. Sci. Monitor* 27:e927917. doi: 10.12659/MSM.927917
- Channappanavar, R., and Perlman, S. (2017). Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. *Semin. Immunopathol.* 39, 529–539. doi: 10.1007/S00281-017-0629-X
- Chatterjee, B., and Thakur, S. S. (2022). SARS-CoV-2 infection triggers phosphorylation: potential target for Anti-COVID-19 therapeutics. *Front. Immunol.* 13:829474 doi: 10.3389/fimmu.2022.829
- Chen, J. S., Alfajaro, M. M., Chow, R. D., Wei, J., Filler, R. B., Eisenbarth, S. C., et al. (2021). Nonsteroidal anti-inflammatory drugs dampen the cytokine and antibody response to SARS-CoV-2 infection. *J. Virol.* 95:e00014-21. doi: 10.1128/JVI.00014-21/SUPPL_FILE/JVI.00014-21-S0002.PDF
- Chen, M. L., Shah, V., Patnaik, R., Adams, W., Hussain, A., Conner, D., et al. (2001). Bioavailability and bioequivalence: An FDA regulatory overview. *Pharm. Res.* 8, 1645–1650. doi: 10.1023/A:1013319408893
- Choi, Y. Y., Kim, M. H., Lee, J. Y., Hong, J., Kim, S. H., and Yang, W. M. (2014). Topical application of *Kochia scoparia* inhibits the development of contact dermatitis in mice. *J. Ethnopharmacol.* 154, 380–385. doi: 10.1016/J.JEP.2014.04.009
- Chu, M., Tsang, M. S. M., He, R., Lam, C. W. K., Quan, Z. B., and Wong, C. K. (2020). The active compounds and therapeutic mechanisms of pentaherbs formula for oral and topical treatment of atopic dermatitis based on network pharmacology. *Plants* 9:1166. doi: 10.3390/PLANTS9091166
- Ciavarella, C., Motta, I., Valente, S., and Pasquinelli, G. (2020). Pharmacological (or Synthetic) and nutritional agonists of PPAR- γ as candidates for Cytokine storm modulation in COVID-19 disease. *Molecule* 25:2076. doi: 10.3390/MOLECULES25092076
- Daina, A., Michielin, O., and Zoete, V. (2019). SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res.* 47, W357–W364. doi: 10.1093/NAR/GKZ382
- de Souza, A. P., Vale, V. L. C., Silva, M. D. C., Araújo, I. B. D. O., Trindade, S. C., Moura-Costa, L. F., et al. (2014). MAPK involvement in cytokine production

Acknowledgments

We thank PolyU Distinguished Postdoctoral Fellowship Scheme at The Hong Kong Polytechnic University, Hong Kong, and the award indicated above for helping to make this research successful.

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.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.972576/full#supplementary-material>

- in response to *Corynebacterium pseudotuberculosis* infection. *BMC Microbiol.* 14:230. doi: 10.1186/s12866-014-0230-6
- Desterke, C., Turhan, A. G., Bennaceur-Griscelli, A., and Griscelli, F. (2020). PPAR γ cistrome repression during activation of lung monocyte-macrophages in severe COVID-19. *iScience* 23:101611. doi: 10.1016/j.isci.2020.101611
- El-Missiry, M. A., Fekri, A., Kesar, L. A., and Othman, A. I. (2021). Polyphenols are potential nutritional adjuvants for targeting COVID-19. *Phytother. Res.* 35, 2879–2889. doi: 10.1002/PTR.6992
- Farahani, M., Niknam, Z., Mohammadi Amirabad, L., Amiri-Dashatan, N., Koushki, M., Nemati, M., et al. (2022). Molecular pathways involved in COVID-19 and potential pathway-based therapeutic targets. *Biomed. Pharmacother.* 145:112420. doi: 10.1016/j.biopha.2021.112420
- Fehr, A. R., Channappanavar, R., and Perlman, S. (2017). Middle east respiratory syndrome: emergence of a pathogenic human coronavirus. *Ann. Rev. Med.* 68, 387–399. doi: 10.1146/ANNUREV-MED-051215-031152
- Ge, J., Liu, Z., Zhong, Z., Wang, L., Zhuo, X., Li, J., et al. (2022). Natural terpenoids with anti-inflammatory activities: Potential leads for anti-inflammatory drug discovery. *Bioorg. Chem.* 124:105817. doi: 10.1016/j.bioorg.2022.105817
- Goel, S., Saheb Sharif-Askari, F., Saheb Sharif, Askari, N., Madkhana, B., Alwaa, A. M., et al. (2021). SARS-CoV-2 switches 'on' MAPK and NF κ B signaling via the reduction of nuclear DUSP1 and DUSP5 expression. *Front. Pharmacol.* 12:631879. doi: 10.3389/fphar.2021.631879
- Grimes, J. M., and Grimes, K. A. (2020). p38 MAPK inhibition: A promising therapeutic approach for COVID-19. *J. Mol. Cell. Cardiol.* 144, 63–65. doi: 10.1016/j.yjmcc.2020.05.007
- Hayashi, T., and Konishi, I. (2021). Cancer therapy with decreased SARS-CoV-2 infection rates in cancer patients. *Br. J. Cancer* 126, 521–522. doi: 10.1038/s41416-021-01685-3
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4:14, 44–57. doi: 10.1038/nprot.2008.211
- Jevtic, S. D., and Nazy, I. (2022). The COVID complex: A review of platelet activation and immune complexes in COVID-19. *Front. Immunol.* 13:807934. doi: 10.3389/FIMMU.2022.807934
- Khan, S. A., and Lee, T. K. W. (2022). Network-pharmacology-based study on active phytochemicals and molecular mechanism of *Cnidium monnieri* in treating Hepatocellular Carcinoma. *Int. J. Mol. Sci.* 23:5400.
- Kounis, N. G., Mplani, V., Konari, I., and Velissaris, D. (2022). Hypersensitivity myocarditis and COVID-19 vaccines. *Kardiol. Polska (Polish Heart J.)* 80, 109–110. doi: 10.33963/KP.A2021.0166
- Liu, M., Gao, Y., Yuan, Y., Yang, K., Shi, S., Zhang, J., et al. (2020). Efficacy and safety of integrated traditional Chinese and western medicine for corona virus disease 2019 (COVID-19): a systematic review and meta-analysis. *Pharmacol. Res.* 158:104896. doi: 10.1016/j.phrs.2020.104896
- Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q., Bader, G. D., et al. (2010). Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 26, 2347–2348. doi: 10.1093/BIOINFORMATICS/BTQ430
- Luo, M., Zeng, B., Wang, H., Yang, Z., Peng, Y., Zhang, Y., et al. (2021). *Kochia scoparia* Saponin Momordin Ic Modulates HaCaT Cell Proliferation and Apoptosis via the Wnt/ β -catenin pathway. *Evid. Based Compl. Altern. Med.* 2021:5522164. doi: 10.1155/2021/5522164
- Ma, Q., Pan, W., Li, R., Liu, B., Li, C., Xie, Y., et al. (2020). Liu Shen capsule shows antiviral and anti-inflammatory abilities against novel coronavirus SARS-CoV-2 via suppression of NF- κ B signaling pathway. *Pharmacol. Res.* 158:104850. doi: 10.1016/j.phrs.2020.104850
- Ma, X., Wu, J., Liu, C., Li, J., Dong, S., Zhang, X., et al. (2020). Deciphering of key pharmacological pathways of Poria Cocos intervention in breast cancer based on integrated pharmacological method. *Evid. Based Complementary Altern. Med.* 2020:4931531. doi: 10.1155/2020/4931531
- Matsuda, H., Li, Y., Yamahara, J., and Yoshikawa, M. (1999). Inhibition of gastric emptying by Triterpene Saponin, Momordin Ic, in Mice: Roles of Blood Glucose, Capsaicin-sensitive sensory nerves, and central nervous system. *J. Pharmacol. Exp. Ther.* 289, 729–734.
- McGonagle, D., Sharif, K., O'Regan, A., and Bridgewood, C. (2020). The role of Cytokines including Interleukin-6 in COVID-19 induced Pneumonia and Macrophage activation syndrome-like disease. *Autoimmun. Rev.* 19:102537. doi: 10.1016/j.autrev.2020.102537
- Mehta, P., McAuley, D. F., Brown, M., Sanchez, E., Tattersall, R. S., and Manson, J. J. (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* 395, 1033–1034. doi: 10.1016/S0140-6736(20)30628-0
- Nagashima, S., Mendes, M. C., Camargo Martins, A. P., Borges, N. H., Godoy, T. M., Miggiolaro, A. F. R. D. S., et al. (2020). Endothelial dysfunction and thrombosis in patients with COVID-19 - Brief report. *Arterioscler. Thromb. Vasc. Biol.* 40, 2404–2407. doi: 10.1161/ATVBAHA.120.314860
- NICE (2021). Overview | COVID-19 Rapid Guideline: Vaccine-Induced Immune Thrombocytopenia and Thrombosis (VITT) | Guidance | NICE. Available online at: <https://www.nice.org.uk/guidance/ng200> (Accessed May 4, 2022).
- NIH. (2022). PubChem. Available online at: <https://pubchem.ncbi.nlm.nih.gov/> (Accessed April 9, 2022).
- Niu, W. H., Wu, F., Cao, W. Y., Wu, Z. G., Chao, Y. C., and Liang, C. (2021). Network pharmacology for the identification of phytochemicals in traditional Chinese medicine for COVID-19 that may regulate interleukin-6. *Biosci. Rep.* 41:20202583. doi: 10.1042/BSR20202583
- No author list. (2020). Cytokine Storm Drugs Move from CAR T to COVID-19. *Cancer Discov.* 10:OF8. doi: 10.1158/2159-8290.CD-ND2020-008/336334/P/CYTOKINE-STORM-DRUGS-MOVE-FROM-CAR-T-TO-COVID
- Oh, K. K., Adnan, M., and Cho, D. H. (2021). Network pharmacology approach to decipher signaling pathways associated with target proteins of NSAIDs against COVID-19. *Sci. Rep.* 11:9606. doi: 10.1038/s41598-021-88313-5
- Oliveros, J. C. (2007). Venny 2.1.0. Available online at: <https://bioinfo.cnb.csic.es/tools/venny/> (accessed April 8, 2022).
- Ong, E. Z., Kalimuddin, S., Chia, W. C., Ooi, S. H., Koh, C. W., Tan, H. C., et al. (2021). Temporal dynamics of the host molecular responses underlying severe COVID-19 progression and disease resolution. *EBioMedicine* 65:103262. doi: 10.1016/j.ebiom.2021.103262/ATTACHMENT/FBA55CEE-5852-43BB-AB4D-E398AFCF801D/MMC1.DOCX
- RCSB PDB (2022). RCSB PDB: Search. Available online at: <https://www.rcsb.org/search> (Accessed April 9, 2022).
- Rebhan, M., Chalifa-Caspi, V., Prilusky, J., and Lancet, D. (1997). GeneCards: integrating information about genes, proteins and diseases. *Trends Gen.* 13:163. doi: 10.1016/S0168-9525(97)01103-7
- Ricciotti, E., and Fitzgerald, G. A. (2011). Prostaglandins and inflammation. *Arterioscler. Thromb. Vasc. Biol.* 31, 986–1000. doi: 10.1161/ATVBAHA.110.207449
- Rizwan, K., Rasheed, T., Khan, S. A., Bilal, M., and Mahmood, T. (2020). Current perspective on diagnosis, epidemiological assessment, prevention strategies, and potential therapeutic interventions for severe acute respiratory infections caused by 2019 novel coronavirus (SARS-CoV-2). *Hum. Vacc. Immunother.* 16, 3001–3010. doi: 10.1080/21645515.2020.1794684
- Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., et al. (2014). TCMSP: A database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminform.* 6, 1–6. doi: 10.1186/1758-2946-6-13/FIGURES/2
- Ruan, Q., Yang, K., Wang, W., Jiang, L., and Song, J. (2020a). Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med.* 46, 846–848. doi: 10.1007/S00134-020-05991-X/FIGURES/1
- Ruan, Q., Yang, K., Wang, W., Jiang, L., and Song, J. (2020b). Correction to: Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med.* 46, 1294–1297. doi: 10.1007/S00134-020-06028-Z
- Sharma, P., Pandey, A. K., and Bhattacharyya, D. K. (2021). Determining crucial genes associated with COVID-19 based on COPD Findings^{*,**}. *Comp. Biol. Med.* 128:104126. doi: 10.1016/j.compbiomed.2020.104126
- Tan, L. Y., Komarasamy, T. V., and Rmt Balasubramaniam, V. (2021). Hyperinflammatory immune response and COVID-19: A double edged sword. *Front. Immunol.* 12:742941. doi: 10.3389/FIMMU.2021.742941
- Tang, Y., Li, X., Yuan, Y., Zhang, H., Zou, Y., Xu, Z., et al. (2022). Network pharmacology-based predictions of active components and pharmacological mechanisms of *Artemisia annua* L. for the treatment of the novel Corona virus disease 2019 (COVID-19). *BMC Complementary Med. Ther.* 22:3523. doi: 10.1186/S12906-022-03523-2/FIGURES/11
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comp. Chem.* 31, 455–461. doi: 10.1002/JCC.21334
- Uzun, G., Pelzl, L., Singh, A., and Bakchoul, T. (2022). Immune-mediated platelet activation in COVID-19 and vaccine-induced immune thrombotic thrombocytopenia. *Front. Immunol.* 13:837629. doi: 10.3389/FIMMU.2022.837629
- Varga, Z., Flammer, A. J., Steiger, P., Haberecker, M., Andermatt, R., Zinkernagel, A. S., et al. (2020). Endothelial cell infection and endothelitis in COVID-19. *Lancet* 395, 1417–1418. doi: 10.1016/S0140-6736(20)30937-5

- von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 31, 258–261. doi: 10.1093/NAR/GKG034
- Wan, E. Y. F., Chui, C. S. L., Lai, F. T. T., Chan, E. W. Y., Li, X., Yan, V. K. C., et al. (2022). Bell's palsy following vaccination with mRNA (BNT162b2) and inactivated (CoronaVac) SARS-CoV-2 vaccines: a case series and nested case-control study. *Lancet Infect. Dis.* 22, 64–72. doi: 10.1016/S1473-3099(21)00451-5
- Wang, C. R., Chen, H. W., Li, Y., Zhou, M. Y., Wong, V. K. W., Jiang, Z. H., et al. (2021). Network pharmacology exploration reveals anti-apoptosis as a common therapeutic mechanism for Non-Alcoholic fatty liver disease treated with blueberry leaf polyphenols. *Nutrients* 13:4060. doi: 10.3390/NU13114060
- Wang, Z., Yang, B., Li, Q., Wen, L., and Zhang, R. (2020). Clinical Features of 69 Cases With Coronavirus Disease 2019 in Wuhan, China. *Clin. Infect. Dis.* 71, 769–777. doi: 10.1093/CID/CIAA272
- Weishengxin (2022). *Weishengxin - Data Analysis and Visualization Experts Around You*. Available online at: <http://www.bioinformatics.com.cn/> (accessed April 8, 2022).
- Winther, B., Arruda, E., Witek, T. J., Marlin, S. D., Tsianco, M. M., Innes, D. J., et al. (2002). Expression of ICAM-1 in Nasal Epithelium and Levels of Soluble ICAM-1 in Nasal Lavage fluid during human experimental Rhinovirus infection. *Arch. Otolaryngol. Head Neck Surg.* 128, 131–136. doi: 10.1001/ARCHOTOL.128.2.131
- Xia, X. (2017). Bioinformatics and drug discovery. *Curr. Top. Med. Chem.* 17, 1709–1726. doi: 10.2174/1568026617666161116143440
- Xiao, Z., Xiao, S., Zhang, Y., Pan, T., and Ouyang, B. (2018). The anti-inflammatory effect of Fructus Kochiae on Allergic contact dermatitis rats via pERK1/2/TLR4/NF- κ B pathway activation. *Evid. Based Complementary Altern. Med.* 2018:1096920. doi: 10.1155/2018/1096920
- Xie, Y., Zhou, K., Wang, Y., Yang, S., Liu, S., Wang, X., et al. (2021). Revealing the Mechanism of *Astragali radix* against cancer-related fatigue by network pharmacology and molecular docking. *Evid. Based Complementary Altern. Med. eCAM* 2021:7075920. doi: 10.1155/2021/7075920
- Xu, X., Han, M., Li, T., Sun, W., Wang, D., Fu, B., et al. (2020). Effective treatment of severe COVID-19 patients with tocilizumab. *Proc. Natl. Acad. Sci. U.S.A.* 117, 10970–10975. doi: 10.1073/PNAS.2005615117
- Yang, R., Liu, H., Bai, C., Wang, Y., Zhang, X., Guo, R., et al. (2020). Chemical composition and pharmacological mechanism of Qingfei Paidu Decoction and Ma Xing Shi Gan decoction against coronavirus disease 2019 (COVID-19): In silico and experimental study. *Pharmacol. Res.* 157:104820. doi: 10.1016/J.PHRS.2020.104820
- Zhang, X., Tan, Y., Ling, Y., Lu, G., Liu, F., Yi, Z., et al. (2020). Viral and host factors related to the clinical outcome of COVID-19. *Nature* 583, 437–440. doi: 10.1038/s41586-020-2355-0
- Zhang, Y., Xiao, M., Zhang, S., Xia, P., Cao, W., Jiang, W., et al. (2020). Coagulopathy and Antiphospholipid Antibodies in patients with Covid-19. *N. Engl. J. Med.* 382:e38. doi: 10.1056/NEJMC2007575
- Zhang, Y. Z., Yang, J. Y., Wu, R. X., Fang, C., Lu, H., Li, H. C., et al. (2021). Network Pharmacology-based identification of key mechanisms of Xihuang Pill in the treatment of triple-negative breast cancer stem cells. *Front. Pharmacol.* 12:2582. doi: 10.3389/FPHAR.2021.714628/BIBTEX
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395, 1054–1062. doi: 10.1016/S0140-6736(20)30566-3
- Zhou, H. X., Li, R. F., Wang, Y. F., Shen, L. H., Cai, L. H., Weng, Y. C., et al. (2020). Total alkaloids from *Alstonia scholaris* inhibit influenza A virus replication and lung immunopathology by regulating the innate immune response. *Phytomedicine* 77:153272. doi: 10.1016/J.PHYMED.2020.153272
- Zhou, Z., Chen, B., Chen, S., Lin, M., Chen, Y., Jin, S., et al. (2020). Applications of network pharmacology in traditional chinese medicine research. *Evid. Based Complementary Altern. Med.* 2020:1646905. doi: 10.1155/2020/1646905
- Zou, W., Tang, Z., Long, Y., Xiao, Z., Ouyang, B., and Liu, M. (2021). Kochiae Fructus, the fruit of common potherb *Kochia scoparia* (L.) Schrad: A review on phytochemistry, pharmacology, toxicology, quality control, and pharmacokinetics. *Evid. Based Complementary Altern. Med.* 2021:5382684. doi: 10.1155/2021/5382684



OPEN ACCESS

EDITED BY

Sinosh Skariyachan,
St. Pius X College, Rajapuram, India

REVIEWED BY

Ahmed A. Al-Karmalawy,
Horus University, Egypt
Chandrabose Selvaraj,
Alagappa University,
India
Mithun Rudrapal,
Rasiklal M. Dhariwal Institute of
Pharmaceutical Education and Research,
India

*CORRESPONDENCE

Xiaosong Hu
183652467@qq.com
Xiao Zhang
954073462@qq.com

[†]These authors have contributed equally to
this work and share first authorship

SPECIALTY SECTION

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

RECEIVED 08 August 2022

ACCEPTED 26 September 2022

PUBLISHED 18 October 2022

CITATION

Cao J-F, Yang X, Xiong L, Wu M, Chen S,
Xiong C, He P, Zong Y, Zhang L, Fu H, Qi Y,
Ying X, Liu D, Hu X and Zhang X (2022)
Mechanism of N-0385 blocking SARS-
CoV-2 to treat COVID-19 based on
molecular docking and molecular
dynamics.
Front. Microbiol. 13:1013911.
doi: 10.3389/fmicb.2022.1013911

COPYRIGHT

© 2022 Cao, Yang, Xiong, Wu, Chen, Xiong,
He, Zong, Zhang, Fu, Qi, Ying, Liu, Hu and
Zhang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Mechanism of N-0385 blocking SARS-CoV-2 to treat COVID-19 based on molecular docking and molecular dynamics

Jun-Feng Cao^{1†}, Xingyu Yang^{1†}, Li Xiong¹, Mei Wu¹,
Shengyan Chen¹, Chenyang Xiong¹, Peiyong He¹,
Yonghua Zong², Lixin Zhang³, Hongjiao Fu¹, Yue Qi¹,
Xiran Ying¹, Dengxin Liu¹, Xiaosong Hu^{4*} and Xiao Zhang^{4*}

¹Clinical Medicine, Chengdu Medical College, Chengdu, China, ²University of Tibetan Medicine, Lhasa, China, ³Yunnan Academy of Forestry Sciences, Kunming, Yunnan, China, ⁴Chengdu Medical College of Basic Medical Sciences, Chengdu, China

Purpose: 2019 Coronavirus disease (COVID-19) has caused millions of confirmed cases and deaths worldwide. TMPRSS2-mediated hydrolysis and maturation of spike protein is essential for SARS-CoV-2 infection *in vivo*. The latest research found that a TMPRSS2 inhibitor called N-0385 could effectively prevent the infection of the SARS-CoV-2 and its variants. However, it is not clear about the mechanism of N-0385 treatment COVID-19. Therefore, this study used computer simulations to investigate the mechanism of N-0385 treatment COVID-19 by impeding SARS-CoV-2 infection.

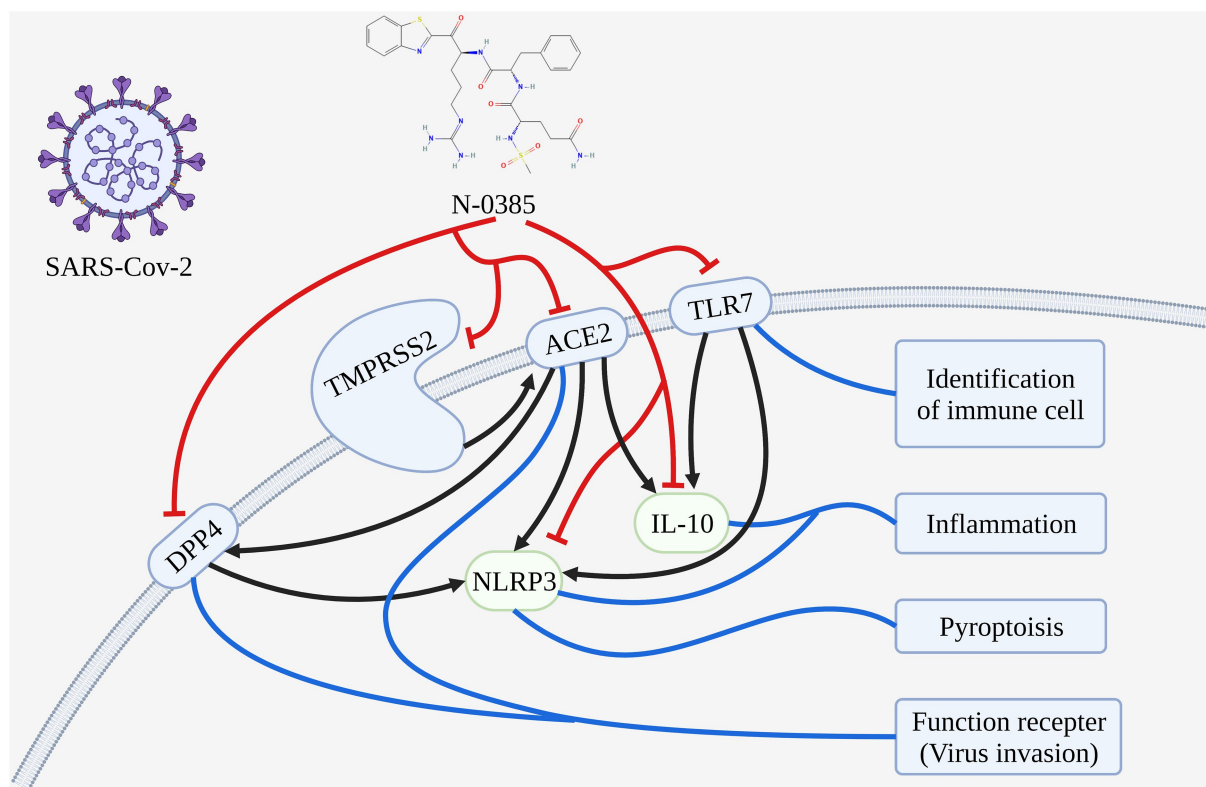
Methods: The GeneCards database was used to search disease gene targets, core targets were analyzed by PPI, GO and KEGG. Molecular docking and molecular dynamics were used to validate and analyze the binding stability of small molecule N-0385 to target proteins. The supercomputer platform was used to simulate and analyze the number of hydrogen bonds, binding free energy, stability of protein targets at the residue level, radius of gyration and solvent accessible surface area.

Results: There were 4,600 COVID-19 gene targets from GeneCards database. PPI, GO and KEGG analysis indicated that signaling pathways of immune response and inflammation played crucial roles in COVID-19. Molecular docking showed that N-0385 could block SARS-CoV-2 infection and treat COVID-19 by acting on ACE2, TMPRSS2 and NLRP3. Molecular dynamics was used to demonstrate that the small molecule N-0385 could form very stable bindings with TMPRSS2 and TLR7.

Conclusion: The mechanism of N-0385 treatment COVID-19 was investigated by molecular docking and molecular dynamics simulation. We speculated that N-0385 may not only inhibit SARS-CoV-2 invasion directly by acting on TMPRSS2, ACE2 and DPP4, but also inhibit the immune recognition process and inflammatory response by regulating TLR7, NLRP3 and IL-10 to prevent SARS-CoV-2 invasion. Therefore, these results suggested that N-0385 may act through multiple targets to reduce SARS-CoV-2 infection and damage caused by inflammatory responses.

KEYWORDS

COVID-19, N-0385, bioinformatics analysis, molecular docking, molecular dynamics



GRAPHICAL ABSTRACT

The mechanisms analysis of N-0385 blocking SARS-CoV-2 infection in the treatment of COVID-19.

Introduction

Abbreviations: ACE, Angiotensin-converting enzyme; ADFR, AutoDockFR; AMBER, Aamber biomolecular simulation programs; ARDS, acute respiratory distress syndrome; BP, biological process; caspase-1, cysteine aspartase 1; CC, cell component; CCR9, C–C trendy factor receptor 9; COVID-19, 2019 coronavirus disease; DAVID, database for annotation, visualization and integrated discovery; DPP4, dipeptidyl peptidase-4; EGFR, epidermal growth factor receptor; GO, gene ontology; HUVEC, human umbilical vein endothelial cells; IC50, inhibitory concentration; IFN, interferon; IL-10, interleukin 10; IL-18, interleukin 18; IL-1 β , interleukin 1 β ; INS, insulin; KEGG, Kyoto Encyclopedia of Genes and Genomes; MD, molecular dynamics; MERS-CoV, middle east respiratory syndrome coronavirus; MF, molecular function.; MMGBSA, molecular mechanics-generalized born surface area; NLRP3, nucleotide-binding oligomerization domain; PAMPs, pathogen-associated molecular patterns; PME, particle-Mesh-Ewald; PMS2, misty fixing protein 2; PPI, protein–protein Interactions; RBD, receptor-binding domain; RMSD, root mean square deviation; RMSF, root mean square fluctuation; Rog, radius of gyration; S, spike-in; SA, surface area; SASA, solvent accessible surface area; ssRNA, single strand RNA; TLR7, toll-like receptor 7; TMPRSS2, transmembrane serine protease 2; VdW, Van der Waals force.

COVID-19 is a malignant infectious disease caused by SARS-CoV-2 infection, it is still an ongoing life-threatening event. SARS-CoV-2 belongs to the genus beta coronavirus (Perلمان and Netland, 2009; Su et al., 2016; Rai et al., 2021). It is found that SARS-CoV-2 is transmitted person to person mainly through inhalation or contact with infectable droplets, with an incubation period of 2–14 days (Lin et al., 2020; Liu et al., 2020; Rohit et al., 2020). The SARS-CoV-2 spike protein binding to ACE2 can promote viral infection of cells (Khateeb et al., 2021). The clinical symptoms of COVID-19 include respiratory syndrome (such as: shortness of breath, cough, dyspnea, fever and viral pneumonia; Cheng et al., 2020; Huang et al., 2020). Studies have found that exacerbation of COVID-19 disease was associated with secondary systemic, inflammatory responses and cytokine storms. Critically ill patients may progress to ARDS, septic shock, multi-organ failure and ultimately death (Rai et al., 2021). Current research has found that there are many key protein targets that may play a key role in COVID-19. Transmembrane Serine Protease 2

(TMPRSS2) is a serine protease family protein, and TMPRSS2 can facilitate virus (such as: SARS-CoV-2, HCoV-229E and MERS-CoV) entry into host cells through proteolytic cleavage and activation of viral envelope glycoproteins (Koch et al., 2021). Angiotensin Converting Enzyme 2 (ACE2) is a metalloproteinase, and the main physiological role of ACE2 is to regulate vasoconstriction and blood pressure (Lan et al., 2020). Some studies have made certain that ACE2 is the receptor for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; Beyerstedt et al., 2021). Toll Like Receptor 7 (TLR7) is a member of the Toll-like receptor family. TLR7 plays a key role in recognizing SARS-CoV-2 and initiating the development of an early antiviral immune response (Solanich et al., 2021). Dipeptidyl Peptidase 4 (DPP4) is a serine exopeptidase that selectively degrades a variety of substrates (including incretin hormones, growth factors and cytokines; Nargis and Chakrabarti, 2018). Study has been reported that there may be a tight interaction between the COVID-19 spike glycoprotein and DPP4 (Vankadari and Wilce, 2020). Interleukin 10 (IL-10) is a key anti-inflammatory mediator capable of protecting the host from damage caused by over-activated inflammatory responses (Wang et al., 2019; Saraiva et al., 2020). It has been demonstrated that there is a dramatic increase in IL-10 expression in the COVID-19 cytokine storm, which is thought to be a fundamental function for suppressing inflammation (Lu et al., 2021). NLR Family Pyrin Domain Containing 3 (NLRP3) promotes inflammasome formation, and NLRP3 activates MAPK and NF- κ B signaling cascades to regulate innate and adaptive immune systems (Zhen and Zhang, 2019).

In current clinical treatment, clinical drugs and vaccines remain the mainstay of prevention and treatment of COVID-19, but vaccine approaches have little efficacy against new variants of the new crown. Current clinical drugs include antivirals (Ritonavir, Lopinavir and Raltegravir), antimalarials (Hydroxychloroquine and Chloroquine) and anti-inflammatory corticosteroids (Dexamethasone and Prednisolone; Khan and Al-Balushi, 2021). However, the effectiveness of current clinical therapeutic agents is not satisfactory. Although clinical improvement has been reported in patients with neoconjunctivitis after the use of Lopinavir/Ritonavir (Lim et al., 2020). Li showed little benefit of lopinavir/ritonavir in improving clinical outcomes in patients hospitalized with mild/moderate COVID-19 (Li et al., 2020), and these drugs triggered a minority of gastrointestinal adverse events. Although chloroquine and hydroxychloroquine have been effective in clinical treatment as anti-malarial drugs, they are inherently toxic, especially to the eyes and heart (Gautret et al., 2020; Wang et al., 2020). Monupivir has shown significant benefit in the treatment of mild SARS-CoV-2, but the therapeutic role of monupivir with moderate to severe COVID-19 is unclear (Singh et al., 2021). An *in vitro* study showed that atazanavir inhibited the replication of SARS-CoV-2 more than lopinavir (Fintelman-Rodrigues et al., 2020). In addition, it has been shown that plasma or clean monoclonal antibodies developed

from fully recovered COVID-19 patients can be offered as therapeutic agents to new COVID-19 patients (Ashour et al., 2022). Long-term administration of chloroquine and hydroxychloroquine can lead to irreversible harm (such as: retinopathy and cardiomyopathy; Motarjemizadeh et al., 2015; Chatre et al., 2018). The role of anti-inflammatory corticosteroids in the treatment of serious infections has been in controversy, and there are limited data on the treatment of COVID-19. Although drug repurposing can be used as a contingency plan in the treatment of diseases, the results of therapies used to treat patients with COVID-19 are highly controversial due to the different study populations (Al-Karmalawy et al., 2021). Therefore, there is a lack of drugs that can effectively hinder viral infection with fewer side effects in the clinical treatment of COVID-19.

N-0385 is a Transmembrane Serine Protease 2 (TMPRSS2) targeted peptidomimetic compound. N-0385 is also known as Ms-QFR-Kbt. The molecular weight of N-0385 is 644.77, and the molecular formula of N-0385 is $C_{28}H_{36}N_8O_6S_2$ (Shapira et al., 2022). TMPRSS2 is a key protease for the entry of SARS-CoV-2 into cells, and TMPRSS2 is able to cleave the SARS-CoV-2 spike protein to initiate virus invasion and infection of cells (Hoffmann et al., 2020; Zhang et al., 2020; Kishimoto et al., 2021). Study found that N-0385 could effectively inhibit SARS-CoV-2 infection in human lung epithelial cells at the half-maximal inhibitory concentration (IC₅₀) of 12.3 ± 1.9 nM. Interestingly, N-0385 was able to target TMPRSS2 to effectively inhibit SARS-CoV-2 infection in Calu-3 cells, these results indicated that N-0385 was a highly efficient inhibitor of TMPRSS2 and could block SARS-CoV-2 entry into cells. Surprisingly, when K18-hACE2 mice were infected with SARS-CoV-2B.1.617. N-0385 treatment completely prevented SARS-CoV-2-induced mortality and significantly prevented weight loss, pulmonary pathology and viral infection (Shapira et al., 2022). This finding deduces that N-0385 can serve as an antiviral agent against pan-mutated viruses. Therefore, N-0385 will provide a proven protocol for the early treatment of COVID-19 and the responding to emerging variant SARS-CoV-2.

However, the research on N-0385 is just beginning, and more in-depth studies on N-0385 are lacking. Therefore, this study investigates the mechanism of N-0385 treatment COVID-19 through molecular docking and molecular dynamics. Bioinformatics were used to screen for COVID-19 core targets. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Protein-Protein Interactions (PPI) were used to analyze gene targets and explore their mechanisms of action and potential pathways. The molecular system motions were used to simulate the result of complex binding from cellular level to chemical moiety level. Molecular docking was used to determine the affinity of the small molecule N-0385 and protein targets. Molecular dynamics was used to simulate the stability of the binding complex. Therefore, the mechanism of N-0385 treatment COVID-19 in this study will facilitate research and clinical applications related to N-0385.

Materials and methods

Acquisition and screening disease gene targets

In this study, we used “COVID-19” and “SARS-CoV-2” as keywords to obtain disease gene targets from the GeneCards database. The relevance score ≥ 5 was used as a threshold to screen COVID-19 related genes targets from GeneCards database, the relevance score is a comprehensive evaluation of the association of genes with the studied diseases.

Protein–protein interaction network construction

The STRING database was used to analyze protein–protein interactions of COVID-19 and construct the PPI network. We imported COVID-19 related protein targets obtained from GeneCards database into Cytoscape 3.7.1 and STRING database for analysis (Rastelli et al., 2010). The network topology parameters were analyzed by Cytoscape 3.7.1, and the core protein targets were filtered according to the criteria of node degree value and median center value greater than the average value.

Gene target enrichment analysis

We used the related genes targets in the DAVID database for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. GO enrichment was used to obtain biological information about gene targets in biological processes (BP), cellular components (CC) and molecular functions (MF). KEGG pathway enrichment was performed by enriching the signaling pathways involved in the related gene targets. Gene targets screening was performed at $p < 0.05$ to analyze the main signaling pathways and biological processes of COVID-19. Omicshare tool platform was used to visualize the results of GO enrichment and KEGG enrichment (Cao et al., 2022b).

Molecular docking and validation of molecular docking

Molecular docking was used to study the molecular affinity of the small molecule N-0385 to the relevant protein targets of COVID-19. The protein crystal structures were downloaded from the PDB database, and 3D structures of small molecules were downloaded from the PUBCHEM database. We performed the molecular docking work by employing AutoDock Vina 1.1.2 software (Kim et al., 2016). Prior to docking, PyMol 2.5 was used to process all receptor proteins (including removal of water molecules, salt ions and small molecules). ADFRsuite 1.0 was

used to convert all processed small molecules and receptor proteins into the PDBQT format required for docking with AutoDock Vina 1.1.2 (Ravindranath et al., 2015). The docked conformation with the highest molecular docking score was considered to be the binding conformation for subsequent molecular dynamics simulations. We analyzed and compared the binding site poses, chemical bond lengths and chemical bond angles of the original crystal ligand to the protein by re-docking the original crystal ligand and the protein using the original crystal ligand of the protein target as a positive reference. Finally, the consistency of the binding mode can indicate the correctness of the molecular docking scheme. And the value of molecular docking score was greater than 6.5 among the target proteins and the small molecule N-0385, it was considered to be tightly bound (Cao J. et al., 2022).

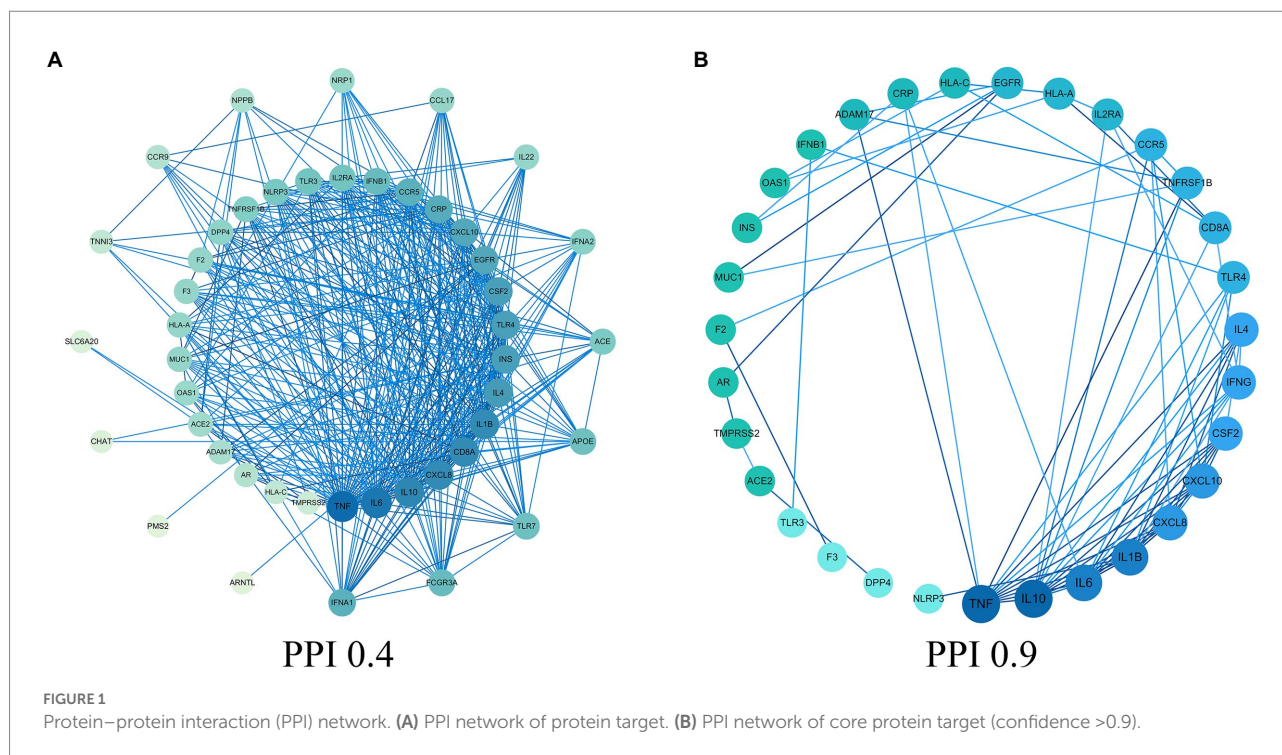
Analysis of proteins localized in the cell membrane and screening of core protein targets

Analysis of current research results on small molecule N-0385, we found that the small molecule N-0385 may act mainly by acting on key proteins (such as: TMPRSS2) in the cellular cytosol thus preventing SARS-CoV-2 from invading cells to reduce the infection (Shapira et al., 2022). Therefore, this study will focus on the small molecule N-0385 and the proteins expressed on the cell membrane involved in SARS-CoV-2 infected cells. We imported COVID-19 related targets obtained from the GeneCards database into the COMPARTMENT database. The COMPARTMENT database is a database for describing and analyzing protein localization at subcellular level. Proteins localized at the cell membrane with Z-score > 4 were considered highly expressed at the cell membrane. The protein targets that bound tightly to the small molecule N-0385 and the protein targets that were localized at the cell membrane were intersected. Therefore, we obtained core protein targets that bind tightly to the small molecule N-0385 and may be involved in SARS-CoV-2 infected cells.

Molecular dynamics

In this study, molecular dynamics simulations were further validated the stability of the complex formed by the core target protein and the small molecule N-0385, which was screened by combining bioinformatics analysis and molecular docking results.

Molecular dynamics (MD) relies on Newtonian mechanics to simulate the motion of molecular systems, and the stability of the complex is studied by taking samples from a complex composed of different states of the molecular system and calculating macroscopic properties such as the thermodynamic quantities of the system (Skariyachan, 2022). We used AMBER 18 software for all-atom molecular dynamics simulations (Maier et al., 2015). The ligands were



parameterized using a combination of AmberTools18 and ACPYPE51 protocols. The energy of all proteins was minimized by using the most rapid descent method and the conjugate gradient method, and the coordinates and energies of the system are saved every 10 ps. Finally, molecular dynamics simulations were performed for 100 ns for each system under periodic boundary conditions (Larini et al., 2007; Skariyachan et al., 2021). Molecular dynamics simulations included protein-ligand complex root mean square deviation (RMSD), root mean square fluctuation (RMSF), solvent accessible surface area (SASA) and radius of gyration (Rog).

Molecular Mechanics/Poisson Boltzmann (Generalized Born) Surface Area binding free energy calculation

The Molecular Mechanics/Poisson Boltzmann (Generalized Born) Surface Area (MM-GBSA) method was used to calculate the free energy of binding between proteins and ligands (Rastelli et al., 2010). We used 100 ns molecular dynamics simulations for the calculation. The calculation equation is as follows:

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - (\Delta G_{\text{receptor}} + \Delta G_{\text{ligand}}) = \Delta E_{\text{internal}} + \Delta E_{\text{VDW}} + \Delta E_{\text{elec}} + \Delta G_{\text{GB}} + \Delta G_{\text{SA}}$$

The non-polar solvation free energy (ΔG_{GA}) was calculated based on solvent accessible surface area (SASA) and product

of surface tension (γ), $\Delta G_{\text{GA}} = 0.0072 \times \text{SASA}$ (Cao et al., 2022a).

Results

Acquisition of disease related genes targets and construction of protein interaction network

In this study, there were 4,600 COVID-19 gene targets. We obtained COVID-19 related gene targets from the GeneCards database based on the relevance score. The relevance score ≥ 5 was considered as COVID-19 related gene targets, and 51 COVID-19 related gene targets were obtained. The protein interaction network of COVID-19 was constructed by using the STRING database, shown in Figure 1A. The 31 core protein targets (such as: TNF, IL-10, IL6, etc.) were obtained by increasing the confidence score ≥ 0.9 . Bioinformatic analysis showed that inflammatory factors (such as: CSF2, IFNG, and CXCL8) occupied the majority of the core protein interaction network, and this is followed by growth factors that affect cell growth and receptor proteins involved in SARS-CoV-2 infected cells. Moreover, the factors involved in the regulation of acute inflammation (such as: IL-1 β , IL-10 and IL6) were very closely linked to other targets suggesting that acute inflammation and viral infection played a major role in COVID-19, and the core protein interaction network was

reconstructed by using the 31 core protein targets, shown in [Figure 1B](#).

Gene ontology and KEGG enrichment analysis

The 51 COVID-19 related gene targets were imported into DAVID database for enrichment analysis. At $p < 0.05$, Gene Ontology (GO) enrichment analysis yielded 256 GO, including 215 biological processes (BP), 22 cellular components (CC) and 19 molecular function (MF). The results showed that biological processes were correlated with cell growth and production of inflammatory factors, mainly involving cellular response to lipopolysaccharide, immune response and positive regulation of gene expression. In cellular component, extracellular region, extracellular space, cell surface accounted for a larger proportion. In molecular functions, cytokine activity, protein binding and growth factor activity were relatively high, shown in [Figures 2A–F](#). KEGG pathway analysis yielded 75 pathways, KEGG enrichment analysis revealed signaling pathways involved in infection and immune response, mainly coronavirus disease—COVID-19, cytokine–cytokine receptor interaction, influenza A and other signaling pathways, shown in [Figures 2G,H](#).

Screening and analysis of key protein targets

The binding energy scores of 51 protein targets (such as: ACE2, INS and TLR3) and the small molecule drug N-0385 were obtained by molecular docking simulations, the molecular docking results are shown in [Table 1](#). The value of molecular docking score was greater than 6.5 among the target proteins and the small molecule N-0385, it was considered to be tightly bound. We obtained 31 protein targets (such as: PMS2, NLRP3 and DPP4) that could form tight bindings to the small molecule N-0385. And we obtained the distribution of the corresponding protein target in the cell by COMPARTMENT database. Proteins localized at the cell membrane with Z-score > 4 were considered highly expressed at the cell membrane. We obtained 40 protein targets (such as: EGFR, ACE2 and TMPRSS2) that were highly expressed in cell membranes. The protein targets that are highly expressed in the cell membrane and those that bind tightly to the small molecule N-0385 were intersected by Venny. The 23 intersecting protein targets (such as: CCR9, ACE2 and TMPRSS2) were those that bound tightly to the small molecule N-0385 and were highly expressed in the cell membrane, shown in [Figure 3](#). Since the results so far have shown that only proteins with both of these characteristics are likely to be the targets of N-0385. We further obtained the scores of these proteins from the GeneCards database, the relevance score ≥ 12 was considered as key protein target. The 6 key protein targets (including: ACE2,

TLR7, TMPRSS2, IL-10, NLRP3 and DPP4) were screened by relevance score.

Molecular docking

The results indicated that N-0385/ACE2 was mainly maintained by hydrogen bonding and hydrophobic interactions. The small molecule N-0385 interacted with Ser-409, Arg-518, Thr-371, Pro-346, His-345 and Arg-273 on ACE2 protein by hydrogen bonding and with Phe-274, Trp-271, Asp-269, Asn-149 and Tyr-515 by hydrophobic interactions. And N-0385 also interacted with Glu-406 by ionic bonding, shown in [Figure 4A](#). The small molecule N-0385 interacted with Tyr-662, Tyr-666, Arg-125, His-126, Arg-358, Tyr-547 and Ser-209 on DPP4 protein by hydrogen bonding and with Lys-554 by hydrophobic interactions. And N-0385 interacted with Tyr-547 and Phe-357 on DPP4 protein by pi-pi interactions and with Tyr-666 by cationic-pi interactions, shown in [Figure 4B](#). In N-0385/IL-10, the small molecule N-0385 interacted with Leu-23, Leu-26, Tyr-72, Leu-98, Leu-65, Ile-69, Met-68, Phe-56 and Leu52 on IL-10 protein by hydrophobic interactions, shown in [Figure 4C](#). The binding of N-0385/NLRP3 indicated that N-0385 interacted with Ser-626, Glu-624, Ala-228, Tyr-632, Arg-578 and Thr-439 on NLRP3 protein by hydrogen bonding and with Leu-628, Arg-351, Phe-410, Ile-411, Tyr-632 and Ile-574 by hydrophobic interactions. And N-0385 also interacted with Glu-629 by ionic bonding, shown in [Figure 4D](#). The binding of N-0385/TLR7 indicated that N-0385 interacted with Val-38, Thr-801 and His-800 on TLR7 protein by hydrogen bonding and with Thr-804, Leu-808 and Ile-826 by hydrophobic interactions. And N-0385 also interacted with Glu-802 by ionic bonding, shown in [Figure 4E](#). The binding of N-0385/TMPRSS2 indicated that N-0385 interacted with Thr-393, Gln-438, Ser-441, His-296, Ser-436 and Gly-464 on TMPRSS2 protein by hydrogen bonding and with Val-280 and Gln-438 by hydrophobic interactions. And N-0385 also interacted with Asp-435 by ionic bonding, shown in [Figure 4F](#).

Molecular dynamics results

The root mean square deviation (RMSD) of molecular dynamics simulations can reflect the motion of the complexes. N-0385/DPP4, N-0385/TMPRSS2, N-0385/ACE2 and N-0385/NLRP3 reached convergence at the beginning of the simulation, and they all preserved stable fluctuations in the subsequent simulations. Although N-0385/TLR7 and N-0385/IL-10 fluctuated sharply at the beginning of the simulation, they both gradually enter a stable state in the later part of the simulation. The overall stability was ranked from high to low as N-0385/DPP4, N-0385/TMPRSS2, N-0385/NLRP3, N-0385/ACE2, N-0385/TLR7 and N-0385/IL-10. The results are shown in [Figure 5](#).

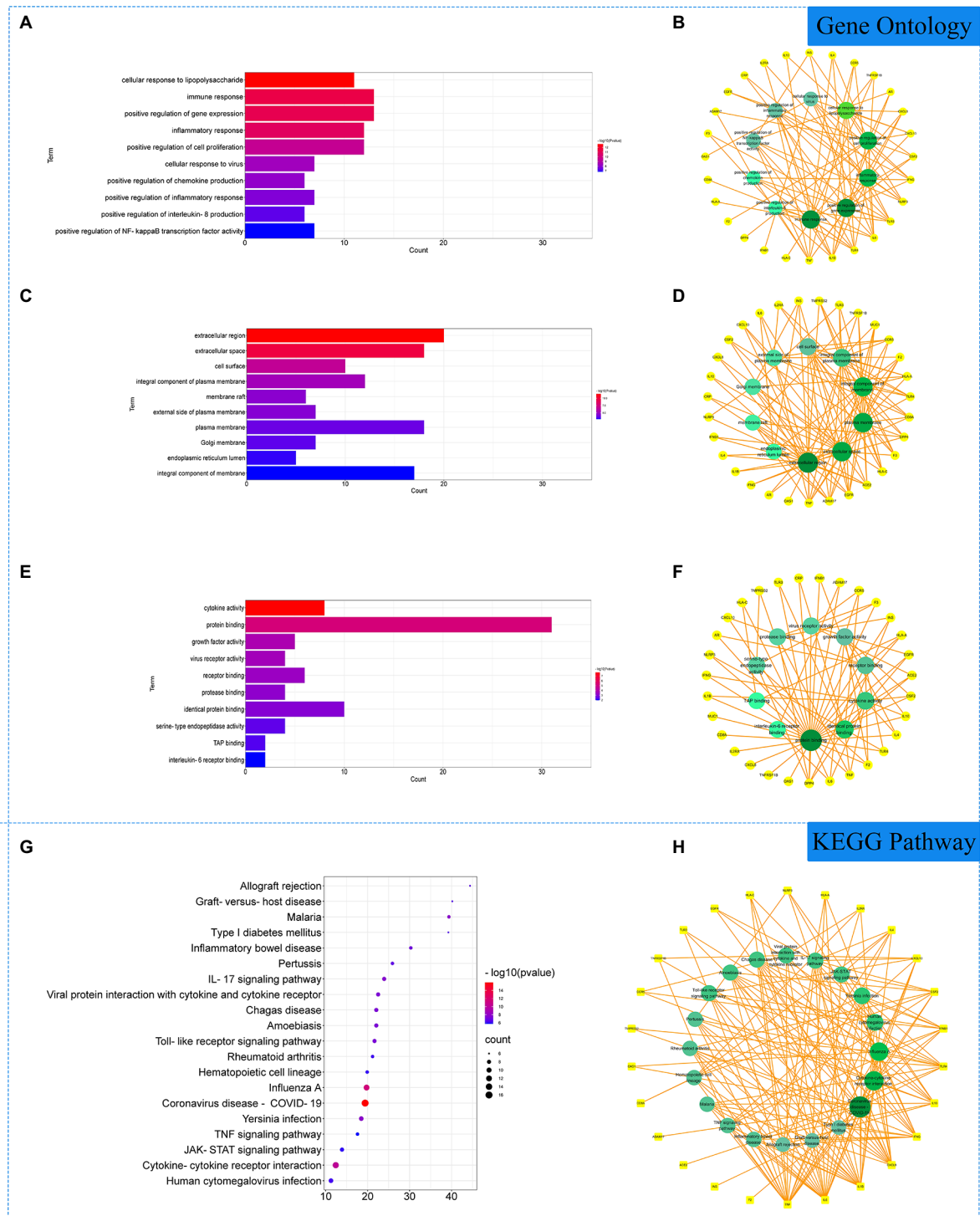


FIGURE 2

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of related genes. **(A)** The top 10 terms in biological processes (BP) were greatly enriched. **(B)** The subnetwork displayed the top 10 BP terms and related genes. **(C)** The top 10 terms in cellular components (CC) were greatly enriched. **(D)** The subnetwork displayed the top 10 CC terms and related genes. **(E)** The top 10 terms in molecular function (MF) were greatly enriched. **(F)** The subnetwork displayed the top 10 MF terms and related genes. **(G)** The top 20 KEGG pathways were showed. **(H)** The subnetworks displayed the top 20 KEGG pathways.

TABLE 1 Results of molecular docking scores.

Target_name	Ligand_name	Docking_score	Target_name	Ligand_name	Docking_score
PMS2	N-0385	−9.6	IFNA2	N-0385	−6.8
MUC1	N-0385	−8.9	INS	N-0385	−6.7
ACE	N-0385	−8.7	IFNG	N-0385	−6.7
CCR5	N-0385	−8.6	FCGR3A	N-0385	−6.7
NLRP3	N-0385	−8.6	ABO	N-0385	−6.6
DPP4	N-0385	−8.4	TLR4	N-0385	−6.5
TAMM41	N-0385	−8.3	IFNA1	N-0385	−6.5
ADAM17	N-0385	−8.2	CD8A	N-0385	−6.5
F2	N-0385	−8.1	CRP	N-0385	−6.5
IL2RA	N-0385	−8.1	IFNB1	N-0385	−6.4
HLA-C	N-0385	−8.1	APOE	N-0385	−6.3
SLC6A20	N-0385	−8	TLR3	N-0385	−6.2
AR	N-0385	−7.9	IL6	N-0385	−6.2
NRP1	N-0385	−7.8	IL-1β	N-0385	−6.1
TNFRSF1B	N-0385	−7.8	CXCL10	N-0385	−6.1
EGFR	N-0385	−7.7	NPPB	N-0385	−6
TLR7	N-0385	−7.7	TNNI3	N-0385	−5.9
ARNTL	N-0385	−7.6	IL22	N-0385	−5.9
HLA-A	N-0385	−7.6	CSF2	N-0385	−5.8
CHAT	N-0385	−7.6	F3	N-0385	−5.8
CCR9	N-0385	−7.4	TNF	N-0385	−5.7
ACE2	N-0385	−7.3	CD99	N-0385	−5.7
IL-10	N-0385	−7.2	IL4	N-0385	−5.5
INPP5E	N-0385	−7.2	CCL17	N-0385	−5.4
TMPRSS2	N-0385	−6.8	CXCL8	N-0385	−5.3
OAS1	N-0385	−6.8			

Results of combining free energy

We calculated the binding energy based on the trajectory of molecular dynamics simulations by the MM-GBSA method, and the binding energy can more accurately reflect the binding mode of small molecules and target proteins. The results of combining free energy of N-0385/ACE2, N-0385/DPP4, N-0385/IL-10, N-0385/NLRP3, N-0385/TLR7 and N-0385/TMPRSS2 were -12.37 ± 1.68 kcal/mol, -25.83 ± 2.35 kcal/mol, -27.75 ± 2.35 kcal/mol, -18.46 ± 1.84 kcal/mol, -42.73 ± 3.21 kcal/mol and -30.99 ± 2.37 kcal/mol. The results showed that the small molecule N-0385 and the corresponding proteins possessed some strong binding affinity. The binding free energies of N-0385/TLR7 and N-0385/TMPRSS2 were high, and their binding free energy values were less than -30.0 kcal/mol. The binding energies of these complexes were mainly contributed by Van der Waals energy and electrostatic energy. The experimental results are shown in Table 2.

Hydrogen bond analysis

Hydrogen bonding is one of the strongest non-covalent binding interactions, and hydrogen bonding is an important basis

for the formation of stable binding small molecule ligands and protein targets. The simulation results showed that all the complexes had a high number of hydrogen bonds during the simulation, and the number of hydrogen bonds was basically maintained above 2. N-0385/ACE2, N-0385/DPP4, N-0385/IL-10 and N-0385/TLR7 even had more than 4 hydrogen bonds in the late stage of simulation, suggesting that hydrogen bonds were very important for the formation of these complexes. The number of hydrogen bonds in N-0385/TMPRSS2 has been maintained at 5–6, and there is an abnormal fluctuation at the beginning of the simulation of N-0385/TMPRSS2. This abnormal fluctuation may be due to a change in the binding state of the receptor protein and the ligand small molecule or the effect of the protein's own peptide chain. The results are shown in Figure 6.

Analysis of root mean square fluctuations

Root mean square fluctuations (RMSF) can respond to the flexibility of the protein during molecular dynamics simulation. In general, the flexibility of the protein decreases after drug binding to stabilize the protein. ACE2, DPP4, NLRP3, TLR7 and TMPRSS2 proteins still had low RMSF fluctuations within 3 Å

after N-0385 binding, indicating that these proteins have low flexibility and close binding of small molecules. Although the RMSF of the receptor protein IL-10 and the ligand small molecule

N-0385 showed large fluctuations after binding, the RMSF was low at the rest of the site except for the two ends of the protein, indicating that the core structure of the protein has good rigidity. The results are shown in Figure 7.

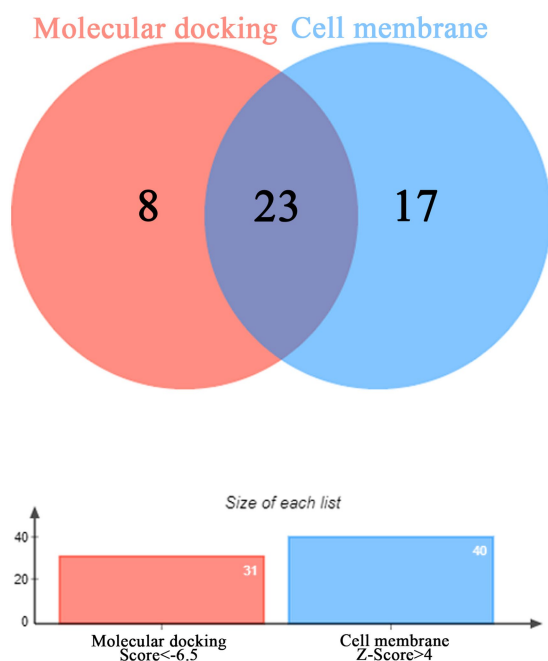


FIGURE 3
The Venny diagram of intersection targets. The intersection of protein targets highly expressed in cell membranes and protein targets tightly bound to the small molecule N-0385.

Analysis of radius of gyration

The radius of gyration (RoG) can reflect the degree of compactness of the complex. The RoG simulation results showed that the convergence from the system were N-0385/NLRP3, N-0385/DPP4, N-0385/TMPRSS2, N-0385/ACE2, N-0385/TLR7 and N-0385/IL-10 from the largest to the smallest. The results are shown in Figure 8.

Analysis of solvent accessible surface area

The solvent accessible surface area (SASA) indicates the area where the complex can come in contact with the aqueous solution. In addition, the fluctuation of SASA reflects the exposure of the protein surface and the changes that occur in the buried area. Because the systems of the six groups of complexes analyzed in this study are different, the SASA values of different complexes have no comparative value or significance. However, the SASA fluctuations of all complexes were stable indicating that the complexes were able to form tightly bound. The results are shown in Figure 9.

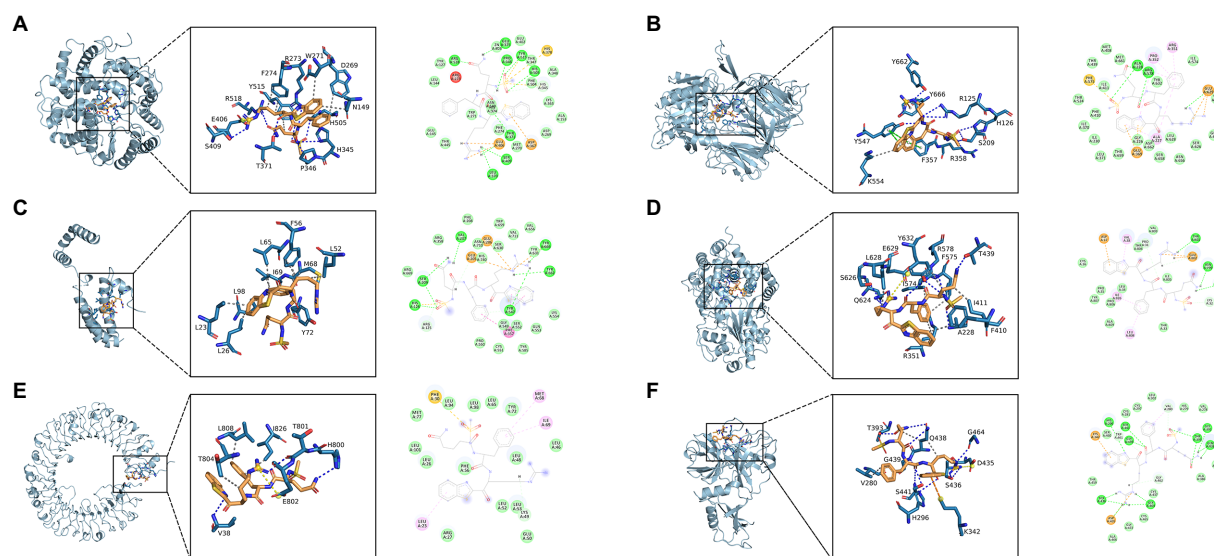


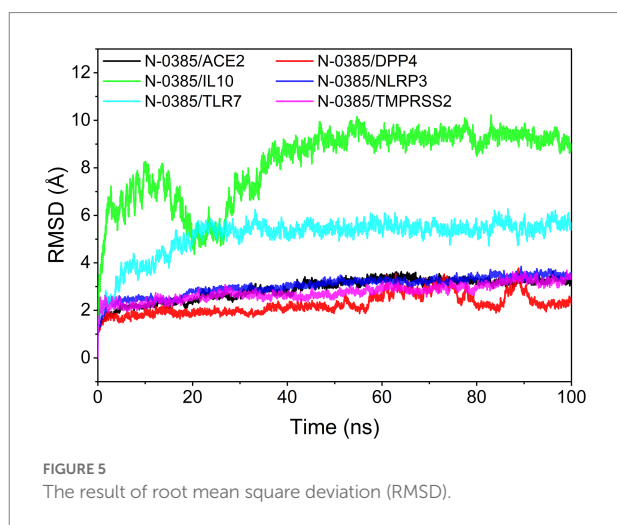
FIGURE 4
Molecular docking of small molecule N-0385 and protein targets. (A) N-0385/ACE2, (B) N-0385/DPP4, (C) N-0385/IL-10, (D) N-0385/NLRP3, (E) N-0385/TLR7, (F) N-0385/TMPRSS2.

Discussion

In this study, we investigate the mechanism of N-0385 treatment COVID-19 by molecular docking and molecular dynamics simulation. We speculated that N-0385 may not only inhibit SARS-CoV-2 invasion directly by acting on TMPRSS2, ACE2 and DPP4, but also inhibit the immune recognition process and inflammatory response by regulating TLR7, NLRP3 and IL-10 to prevent SARS-CoV-2 invasion.

Firstly, we conjectured that N-0385 may block hydrolysis and cleavage of SARS-CoV-2 protein by directly acting on TMPRSS2 and ACE2, thereby preventing viral entry into cells. Secondly, N-0385 may reduce viral infection of cells and regulate that *via* metabolism through PDD4, thereby reducing body damage. Finally, N-0385 may reduce the inflammatory response and block the immune recognition process by regulating TLR7, NLRP3 and IL-10, thereby reducing cellular damage and alleviating the disease.

Therefore, these results indicated that N-0385 may act through multiple targets to reduce viral infection and reduce damage consisting of inflammatory responses.



Analysis of molecular docking and molecular dynamics

Molecular docking can be identified by spatial matching of the drug small molecule N-0385 and the protein macromolecule *in vivo* to each other, and molecular docking can predict their interactions, binding modes and affinities. Molecular dynamics is a powerful tool to analyze structural and dynamic information of biological macromolecular systems. By simulating the microstructural mechanisms and dynamic trajectory behavior of molecules, physicochemical data of molecular systems can be obtained. Molecular dynamics can be used to reveal the microstructural changes of drug small molecule-protein macromolecule interaction systems. Therefore, the results of molecular docking and molecular dynamics can be used to explore the mechanism of N-0385 as it might be used in the treatment of COVID-19.

In this study, the MMGBSA method was used to calculate the binding free energy that is able to reflect the binding stability of ligand small molecules to receptor proteins. The binding free energy results showed that N-0385/TMPRSS2 was -30.99 ± 2.37 kcal/mol, and that N-0385 binds to TMPRSS2 mainly through hydrogen bonding. N-0385/TMPRSS2 converged at the beginning of the RMSD simulation and remained stable in the subsequent simulations. The binding free energy results showed that N-0385/ACE2 and N-0385/DPP4 were -12.37 ± 1.68 kcal/mol and -25.83 ± 2.35 kcal/mol. For N-0385/ACE2, the energy analysis showed that hydrophobic interactions were the main contributing energy. And N-0385 was mainly bound to DPP4 through hydrogen bonding. ACE2 and DPP4 still had low RMSF fluctuations within 3 Å after N-0385 binding, indicating that these proteins were low in flexibility and tightly bound to small molecules. The binding free energy results showed -42.73 ± 3.21 kcal/mol, -18.46 ± 1.84 kcal/mol and -27.75 ± 2.35 kcal/mol for N-0385/TLR7, N-0385/NLRP3 and N-0385/IL-10. The binding energies of these complexes were mainly contributed by Van der Waals and electrostatic energies. N-0385/IL-10 and N-0385/TLR7 formed more than four hydrogen bonds at the late stage of the

TABLE 2 Binding free energies and energy components predicted by MM/GBSA (kcal/mol).

System name	ΔE_{vdw}	ΔE_{elec}	ΔG_{GB}	ΔG_{SA}	ΔG_{bind}
N-0385/ACE2	-44.78 ± 4.28	-442.35 ± 14.52	481.69 ± 16.90	-6.93 ± 0.44	-12.37 ± 1.68
N-0385/DPP4	-42.69 ± 2.84	-382.48 ± 15.82	405.95 ± 14.17	-6.61 ± 0.34	-25.83 ± 2.35
N-0385/IL-10	-44.44 ± 3.10	-60.49 ± 5.82	83.55 ± 4.12	-6.37 ± 0.24	-27.75 ± 2.35
N-0385/NLRP3	-43.90 ± 5.35	-98.43 ± 15.41	131.09 ± 11.66	-7.21 ± 0.45	-18.46 ± 1.84
N-0385/TLR7	-42.49 ± 2.49	-140.09 ± 10.69	146.94 ± 11.98	-7.08 ± 0.64	-42.73 ± 3.21
N-0385/TMPRSS2	-53.79 ± 3.80	-103.12 ± 13.55	133.01 ± 12.84	-7.10 ± 0.25	-30.99 ± 2.37

ΔE_{vdw} : Van der Waals energy.

ΔE_{elec} : electrostatic energy.

ΔG_{GB} : electrostatic contribution to solvation.

ΔG_{SA} : non-polar contribution to solvation.

ΔG_{bind} : binding free energy.

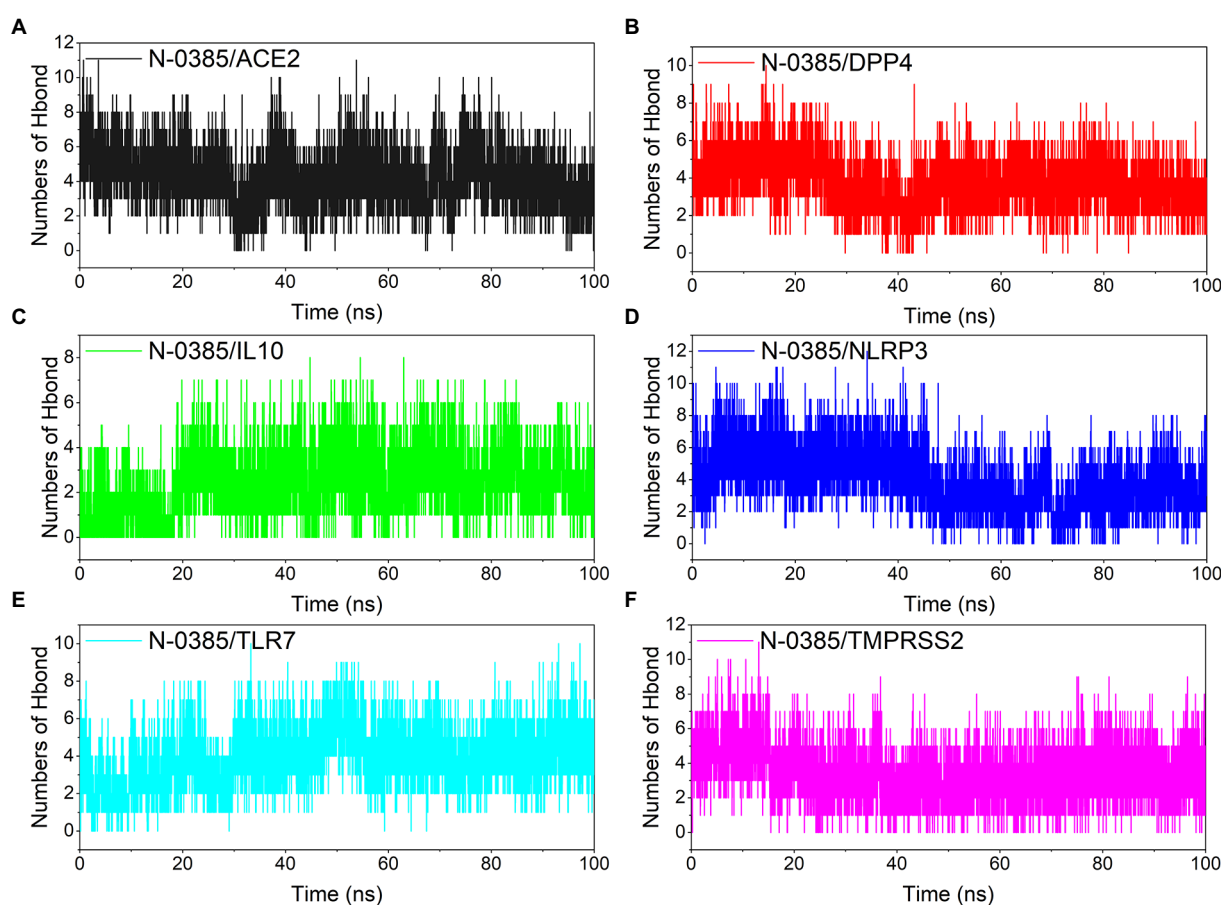


FIGURE 6

Changes in the number of hydrogen bonds among small molecule ligands and protein receptors in complex system simulations. (A) N-0385/ACE2, (B) N-0385/DPP4, (C) N-0385/IL-10, (D) N-0385/NLRP3, (E) N-0385/TLR7, (F) N-0385/TMPRSS2.

simulation, indicating that hydrogen bonds were crucial for the formation of these complexes.

Analysis of the potential mechanism of N-0385 to block SARS-CoV-2 infection of cells

N-0385 may inhibit the invasion of SARS-CoV-2 directly by acting on TMPRSS2, ACE2 and DPP4. N-0385 may block the hydrolysis and cleavage of SARS-CoV-2 protein by acting directly on TMPRSS2 and ACE2, thus preventing the virus from entering the cells, and N-0385 may reduce SARS-CoV-2 infection of cells through DPP4 and regulate organism metabolism, thus reducing organism damage.

Bioinformatics analysis showed that transmembrane serine protease 2 (TMPRSS2) facilitated virus entry into host cells through proteolytic cleavage and activation of viral envelope glycoproteins. KEGG signaling pathway analysis indicated that TMPRSS2 was involved in the mitochondrial immune response to SARS-CoV-2. GO analysis showed that TMPRSS2 was associated with serine-type endopeptidase activity and scavenger

receptor activity. Angiotensin-converting enzyme 2 (ACE2) is a functional receptor for the spike glycoprotein of SARS-CoV-2, and plays an important role in the regulation of cardiovascular and renal function and fertility. KEGG signaling pathway analysis included peptide hormone metabolism. GO analysis included metalloproteinase activity and peptide binding. Dipeptidyl peptidase 4 (DPP4) is involved in insulin metabolism and immune regulation. And DPP4 has been shown to be a functional receptor for Middle East respiratory syndrome coronavirus (MERS-CoV). KEGG signaling pathway analysis showed that DPP4 was involved in the metabolism of peptide hormones and the regulation of intestinal insulin. GO analysis suggested that DPP4 was associated with protein homodimerization activity and signaling receptor binding. Protein interaction network analysis showed that TMPRSS2, ACE2 and DPP4 were closely associated with viral invasion targets.

The SARS-CoV-2 spike protein, transmembrane protease serine 2 (TMPRSS2) and human receptor angiotensin-converting enzyme 2 (ACE2) are the major host-pathogen determinants affecting infection (Zhou P. et al., 2020; Senapati et al., 2021). The spike protein is located on the viral outer membrane and the

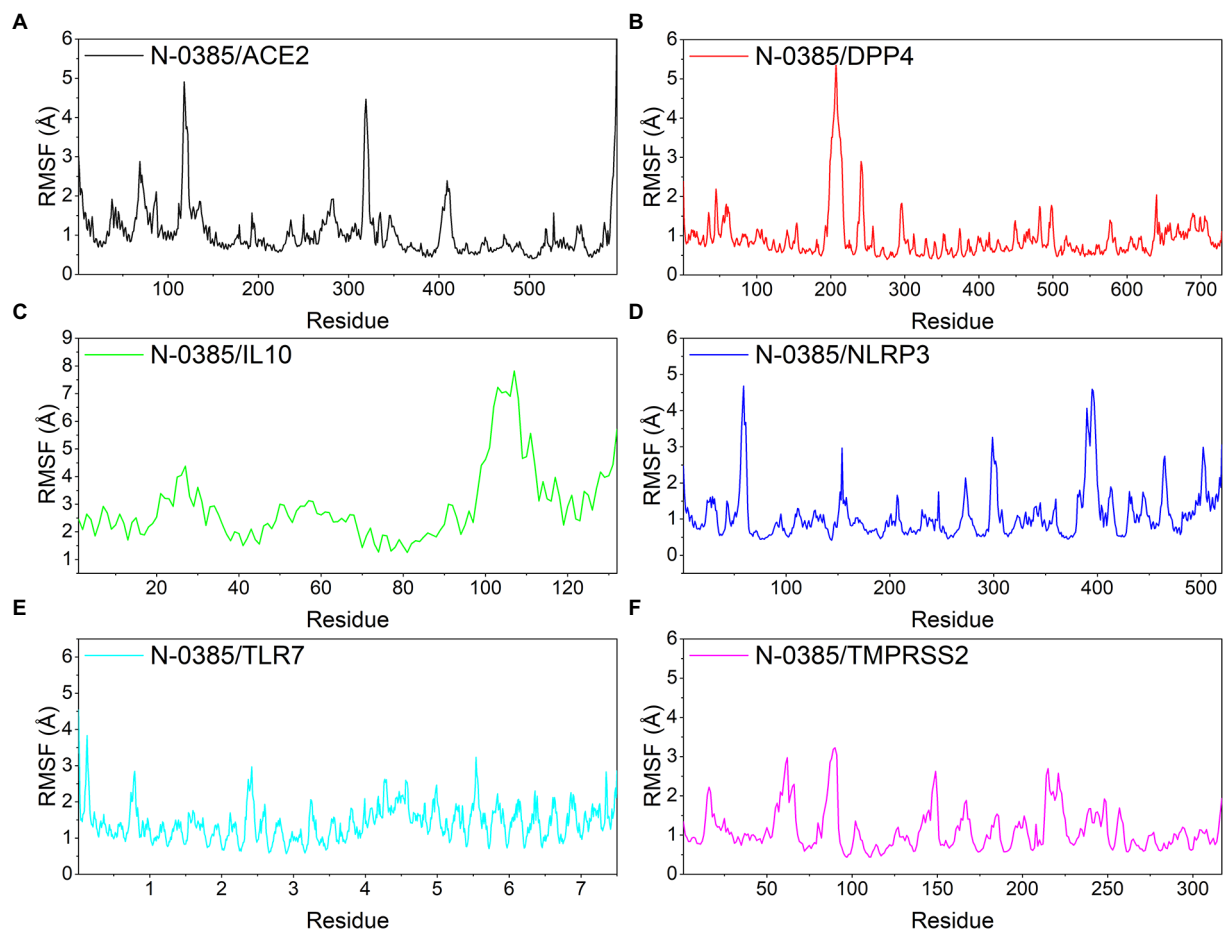


FIGURE 7
Changes in the stability of protein targets at the residue level. (A) N-0385/ACE2, (B) N-0385/DPP4, (C) N-0385/IL-10, (D) N-0385/NLRP3, (E) N-0385/TLR7, (F) N-0385/TMPRSS2.

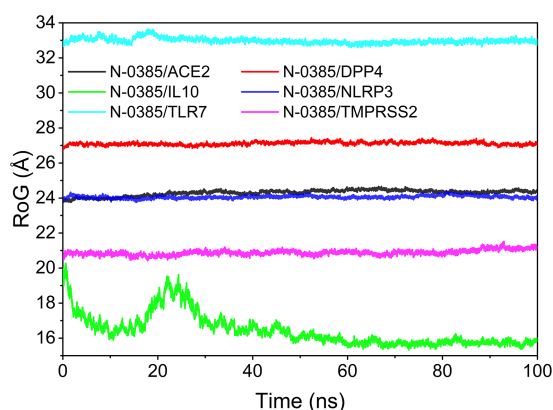
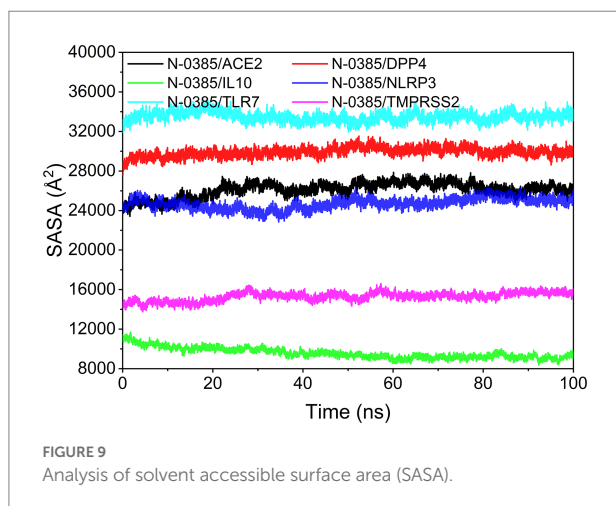


FIGURE 8
The result of radius of gyration (RoG).

protein has two major functional subunits (a long N-terminal S1 subunit and a relatively short C-terminal S2 subunit). Receptor-Binding Domain (RBD) is located in the S1 subunit of the spiked

protein (Li et al., 2005), and TMPRSS2 is located on the surface of type II alveolar cells (Shulla et al., 2011). ACE2 is located on the surface of type II alveolar cells, and it is involved in the regulation of signaling pathways of the renin-angiotensin system and integrin signaling (Shulla et al., 2011; Beacon et al., 2021). In addition, ACE2 can act as a carboxypeptidase, removing a single amino acid from the C-terminus of the substrate (Turner and Hooper, 2002). Study was shown that amino acid-born mutations in the binding sites of spike protein, TMPRSS2 and ACE2 alter the affinity of the proteins, which may affect the structural stability of the complexes (Hussain et al., 2020; MacGowan and Barton, 2020; Yan et al., 2020; Beacon et al., 2021).

The spike protein of the SARS-CoV-2 enters human cells by binding to TMPRSS2 and ACE2. Two distinct modes of cell entry exist for SARS-CoV-2, and the two entry modes differ in the second cleavage. The entry of SARS-CoV-2 into cells occurs by two spike protein cleavages (Senapati et al., 2021). The first occurrence of cleavage takes place when the spike protein and ACE2 are bound together, there are 20 ACE2 residues interacting with 17 residues from the spike protein RBD, triggering a partial conformational rearrangement in the spike



protein and increased sensitivity to hydrolytic digestion of the protein at the junction of the S1 and S2 subunits of the spike protein. This is followed by exposure of the S2' cleavage site in the S2 subunit (Simmons et al., 2005; Li et al., 2006; Jackson et al., 2022). The second cleavage occurs at the exposed S2' cleavage site, where multiple arginine-rich sites (Arg667 and Arg797) are recognized and cleaved by two different protein hydrolases, releasing the S2' subunit. If the target cell TMPRSS2 is not adequately expressed, or if the viral-ACE2 complex does not encounter TMPRSS2, the viral-ACE2 complex is internalized into the endolysosome by histone proteases *via* reticulon-mediated endocytosis to perform hydrolytic cleavage. This cleavage has to occur at the cell surface under acidic conditions and the presence of TMPRSS2 (Jackson et al., 2022). TMPRSS2 can enzymatically cleave a string of hydrophobic amino acids exposed at a site on the S2 subunit of the spiked protein, thereby allowing the virus to rapidly embed into the host cell membrane. Therefore, viruses mediated by TMPRSS2-mediated endocytosis can enter the cell more rapidly (Hoffmann et al., 2020). After cleavage, the fusion peptide is released by intracellular histone protease L15, 16, initiating fusion pore formation (Matsuyama et al., 2010; Glowacka et al., 2011). Subsequently, the unfolded spike protein folds up to fuse the viral outer membrane with the cell membrane thus ensuring that the viral gene can access the cytoplasm.

DPP4 is a serine ectopeptidase, also known as CD26. DPP4 can cleave X-proline dipeptides from the N-terminal end of peptides, and it is expressed in a variety of epithelial and endothelial cells of the systemic vasculature, kidney, lung, small intestine and heart (Solerte et al., 2020a,b). DPP4 is involved in glucose and insulin metabolism and immune regulation. Studies have shown that Middle East respiratory syndrome coronavirus uses DPP4 as its functional receptor (Bassendine et al., 2020). Interestingly, it was demonstrated that there might be a tight interaction between the S1 domain loop of the COVID-19 spike glycoprotein and the CD26 surface (Vankadari and Wilce, 2020). Therefore, like ACE2, DPP4 may act as a SARS-CoV-2 co-receptor into cells (Solerte et al., 2020a,b). Recently, Nádasdi et al. (2022)

found that reduced circulating DPP4 activity is associated with severe COVID-19 disease and is a strong prognostic biomarker of COVID-19 mortality. DPP4 inhibitors are commonly used to treat type 2 diabetes. Interestingly, many studies have shown that some DPP4 inhibitors (such as: selegiline and emetine) may have a therapeutic effect on neocoronary, but clinical trials are still needed to confirm their effects (Mikhael et al., 2022; Zhang et al., 2022). In addition, many data show that diabetes is also a factor in causing severe symptoms of COVID-19 (Zhou F. et al., 2020; Godeau et al., 2021). Thus DPP4 may be a potential target for the treatment or prevention of COVID-19 patients with concomitant type 2 diabetes (Iacobellis, 2020).

N-0385 hinders SARS-CoV-2 invasion by inhibiting recognition processes and inflammatory responses

N-0385 may reduce the inflammatory response and block the immune recognition process by regulating TLR7, NLRP3 and IL-10, thereby reducing cellular damage and alleviating disease.

Bioinformatics analysis suggested that toll like receptor 7 (TLR7) plays an important role in pathogen recognition and innate immune activation, and it recognizes pathogen-associated molecular patterns (PAMPs) expressed on infectious agents and mediates the production of cytokines necessary for effective immunity. KEGG signaling pathway analysis suggested that TLR7 was involved in dendritic cell developmental lineage pathways and in the interaction between immune cells and microRNAs in the tumor microenvironment. GO analysis indicated that NLRP3 was involved in transmembrane signaling receptor activity and double-stranded RNA binding. NLR Family Pyrin Domain Containing 3 (NLRP3) plays a role in the regulation of inflammation, immune response and apoptosis, and NLRP3 can also induce cell death. KEGG signaling pathway analysis included protein metabolism and SARS-CoV-2 activation of the NLRP3 inflammasome. GO analysis included peptidoglycan binding. Interleukin 10 (IL-10) has a potent anti-inflammatory function to which it plays major role in immunomodulatory cytokine, and it can limit excessive tissue destruction caused by inflammation. KEGG signaling pathway analysis suggested that TLR7 was involved in dendritic cell developmental lineage pathways and MIF-mediated glucocorticoid regulation. GO analysis indicated that NLRP3 was involved in cytokine activity and interleukin 10 receptor binding. Protein interaction network analysis showed that TLR7, NLRP3 and IL-10 were closely associated with inflammatory responses and immune regulation.

TLR7 may be involved in SARS-CoV-2 genome recognition through recognition of ssRNA and synthetic oligonucleotides (Khanmohammadi and Rezaei, 2021; Yuan et al., 2021). TLR7 is considered a key cellular sensor of SARS-CoV-2 encoded ssRNA, and it is involved in host resistance and disease pathogenesis of COVID-19 (Salvi et al., 2021). Interestingly, Asano et al. (2021) found that X-linked recessive

TLR7 deficiency was a highly exogenous genetic cause of critical COVID-19 pneumonia, and TLR7 and pDC were essential for protective type I IFN immunity against SARS-CoV-2 in the respiratory tract. Some of research have found that the immune response to SARS-CoV-1 *via* TLR7 activation may trigger a cytokine storm (Tang et al., 2016; de Groot and Bontrop, 2020).

IL-10 is a very important class of anti-inflammatory mediators that protects the host from overreacting to pathogens and microbiota, and it can play an irreplaceable role in sterile wound healing, autoimmunity, cancer and homeostasis (Fiorentino et al., 1989; Moore et al., 1990; Saraiva et al., 2020). IL-10 produces a wide range of effects of anti-inflammatory activity by targeting a variety of cells. IL-10 acts mainly through macrophages, and IL-10 triggers a dramatic immunosuppressive response mainly by inhibiting the transcription of cytokines, chemokines, MHCII co-stimulatory and adhesion molecules (Bogdan et al., 1991; Moore et al., 2001; Shouval et al., 2014; Zigmond et al., 2014). Therefore, IL-10 not only regulates the local cytokine microenvironment, but also restricts antigen presentation preventing T cell responses and the spread of inflammation. Studies have found that IL-10 increases ACE2 mRNA expression in lung-derived Calu-3 cells and human umbilical vein endothelial cells (HUVEC), thereby providing cardiopulmonary protection in COVID-19 using a negative feedback mechanism that inhibits inflammation. At the same time, IL-10 promotes viral binding and entry using organismal protective mechanisms to kill pathogens, thereby reducing inflammatory cytokine production and protection from further tissue damage (Albini et al., 2021; Tabares-Guevara et al., 2021).

The NLRP3 inflammasome consists of NLRP3, apoptosis-associated speckle-like protein (ASC) and cysteine aspartase 1 (caspase-1). Activated inflammasomes contribute to the release of mature cytokines, thereby facilitating the development of an innate immune response. It has been shown that the interaction of ACE2 receptors with SARS-CoV-2 spike proteins causes NLRP3 inflammasomes in cells, which may lead to cellular scorching if NLRP3 is over activated (Ratajczak et al., 2020, 2021). NLRP3 inflammasomes trigger inflammatory immune responses *via* intracellular caspase-1, which leads to their ability to release the potent pro-inflammatory cytokines interleukin 1 β (IL-1 β) and interleukin 18 (IL-18; Franchi et al., 2012; Swanson et al., 2019; Ratajczak et al., 2020; Strollo and Pozzilli, 2020).

Conclusion

In this study, we investigate the mechanism of therapeutical effect of N-0385 in COVID-19 by molecular docking and molecular dynamics. We speculated that N-0385 may not only directly inhibit SARS-CoV-2 invasion by acting on TMPRSS2,

ACE2 and DPP4, but also inhibit the immune recognition process and inflammatory response by regulating TLR7, NLRP3 and IL-10, thus preventing SARS-CoV-2 invasion. N-0385 may have a promising usage in the treatment of COVID-19, for which it blocks SARS-CoV-2 infection *via* multiple targets within inflammatory response. This study renders theoretical basis and new research ideas for N-0385 in the treatment of COVID-19.

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

J-FC, XYY, LX, MW, YZ, and HF contributed to the conception of the study. J-FC, XZ, LX, SC, LZ, CX, PH, YQ, XY, and DL contributed significantly to analysis and manuscript preparation. J-FC, XYY, MW, and SC performed the data analyses and wrote the manuscript. XZ, J-FC, and XH helped to perform the analysis with constructive discussions. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by “Project of Tibetan Medicine Regional Cooperative Innovation Center (no. 2019xtcx006).”

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.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1013911/full#supplementary-material>

References

- Albini, A., Calabrone, L., Carlini, V., Benedetto, N., Lombardo, M., Bruno, A., et al. (2021). Preliminary evidence for IL-10-induced ACE2 mRNA expression in lung-derived and endothelial cells: implications for SARS-CoV-2 ARDS pathogenesis. *Front. Immunol.* 12:718136. doi: 10.3389/fimmu.2021.718136
- Al-Karmalawy, A. A., Soltane, R., Abo Elmaaty, A., Tantawy, M. A., Antar, S. A., Yahya, G., et al. (2021). Coronavirus disease (COVID-19) control between drug repurposing and vaccination: a comprehensive overview. *Vaccines (Basel)* 9:1317. doi: 10.3390/vaccines9111317
- Asano, T., Boisson, B., Onodi, F., Matuozzo, D., Moncada-Velez, M., Maglorius Renkilaraj, M. R. L., et al. (2021). X-linked recessive TLR7 deficiency in ~1% of men under 60 years old with life-threatening COVID-19. *Sci. Immunol.* 6:eab4348. doi: 10.1126/sciimmunol.abl4348
- Ashour, N. A., Abo Elmaaty, A., Sarhan, A. A., Elkadeed, E. B., Moussa, A. M., Erfan, I. A., et al. (2022). A systematic review of the global intervention for SARS-CoV-2 combating: from drugs repurposing to Molnupiravir approval. *Drug Des. Devel. Ther.* 16, 685–715. doi: 10.2147/DDDT.S354841
- Bassendine, M. F., Bridge, S. H., McCaughan, G. W., and Gorrell, M. D. (2020). COVID-19 and comorbidities: a role for dipeptidyl peptidase 4 (DPP4) in disease severity? *J. Diabetes* 12, 649–658. doi: 10.1111/1753-0407.13052
- Beacon, T. H., Delcuve, G. P., and Davie, J. R. (2021). Epigenetic regulation of ACE2, the receptor of the SARS-CoV-2 virus(1). *Genome* 64, 386–399. doi: 10.1139/gen-2020-0124
- Beyerstedt, S., Casaro, E. B., and Rangel, É. B. (2021). COVID-19: angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection. *Eur. J. Clin. Microbiol. Infect. Dis.* 40, 905–919. doi: 10.1007/s10096-020-04138-6
- Bogdan, C., Vodovotz, Y., and Nathan, C. (1991). Macrophage deactivation by interleukin 10. *J. Exp. Med.* 174, 1549–1555. doi: 10.1084/jem.174.6.1549
- Cao, J.-F., Gong, Y., Wu, M., Yang, X., Xiong, L., Chen, S., et al. (2022a). Exploring the mechanism of action of licorice in the treatment of COVID-19 through bioinformatics analysis and molecular dynamics simulation. *Front. Pharmacol.* 13:1003310. doi: 10.3389/fphar.2022.1003310
- Cao, J., Li, L., Xiong, L., Wang, C., Chen, Y., and Zhang, X. (2022). Research on the mechanism of berberine in the treatment of COVID-19 pneumonia pulmonary fibrosis using network pharmacology and molecular docking. *Phytomed. Plus* 2:100252. doi: 10.1016/j.phyplu.2022.100252
- Cao, J.-F., Yang, X., Xiong, L., Wu, M., Chen, S., Xu, H., et al. (2022b). Exploring the mechanism of action of dapansutride in the treatment of gouty arthritis based on molecular docking and molecular dynamics. *Front. Physiol.* 13:990469. doi: 10.3389/fphys.2022.990469
- Chatre, C., Roubille, F., Vernhet, H., Jorgensen, C., and Pers, Y. M. (2018). Cardiac complications attributed to chloroquine and hydroxychloroquine: a systematic review of the literature. *Drug Saf.* 41, 919–931. doi: 10.1007/s40264-018-0689-4
- Cheng, Y., Luo, R., Wang, K., Zhang, M., Wang, Z., Dong, L., et al. (2020). Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int.* 97, 829–838. doi: 10.1016/j.kint.2020.03.005
- de Groot, N. G., and Bontrop, R. E. (2020). COVID-19 pandemic: is a gender-defined dosage effect responsible for the high mortality rate among males? *Immunogenetics* 72, 275–277. doi: 10.1007/s00251-020-01165-7
- Fintelman-Rodrigues, N., Sacramento, C. Q., Ribeiro Lima, C., Souza da Silva, F., Ferreira, A. C., Mattos, M., et al. (2020). Atazanavir, alone or in combination with ritonavir, inhibits SARS-CoV-2 replication and Proinflammatory cytokine production. *Antimicrob. Agents Chemother.* 64:e00825-20. doi: 10.1128/AAC.00825-20
- Fiorentino, D. F., Bond, M. W., and Mosmann, T. R. (1989). Two types of mouse T helper cell. IV. Th2 clones secrete a factor that inhibits cytokine production by Th1 clones. *J. Exp. Med.* 170, 2081–2095. doi: 10.1084/jem.170.6.2081
- Franchi, L., Muñoz-Planillo, R., and Núñez, G. (2012). Sensing and reacting to microbes through the inflammasomes. *Nat. Immunol.* 13, 325–332. doi: 10.1038/ni.2231
- Gautret, P., Lagier, J. C., Parola, P., Hoang, V. T., Meddeb, L., Mailhe, M., et al. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int. J. Antimicrob. Agents* 56:105949. doi: 10.1016/j.ijantimicag.2020.105949
- Glowacka, I., Bertram, S., Müller, M. A., Allen, P., Soilleux, E., Pfefferle, S., et al. (2011). Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J. Virol.* 85, 4122–4134. doi: 10.1128/jvi.02232-10
- Godeau, D., Petit, A., Richard, I., Roquelaure, Y., and Descatha, A. (2021). Return-to-work, disabilities and occupational health in the age of COVID-19. *Scand. J. Work Environ. Health* 47, 408–409. doi: 10.5271/sjweh.3960
- 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. *Cells* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi: 10.1016/s0140-6736(20)30183-5
- Hussain, M., Jabeen, N., Raza, F., Shabbir, S., Baig, A. A., Amanullah, A., et al. (2020). Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *J. Med. Virol.* 92, 1580–1586. doi: 10.1002/jmv.25832
- Iacobellis, G. (2020). COVID-19 and diabetes: can DPP4 inhibition play a role? *Diabetes Res. Clin. Pract.* 162:108125. doi: 10.1016/j.diabres.2020.108125
- Jackson, C. B., Farzan, M., Chen, B., and Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* 23, 3–20. doi: 10.1038/s41580-021-00418-x
- Khan, S. A., and Al-Balushi, K. (2021). Combating COVID-19: the role of drug repurposing and medicinal plants. *J. Infect. Public Health* 14, 495–503. doi: 10.1016/j.jiph.2020.10.012
- Khanmohammadi, S., and Rezaei, N. (2021). Role of toll-like receptors in the pathogenesis of COVID-19. *J. Med. Virol.* 93, 2735–2739. doi: 10.1002/jmv.26826
- Khateeb, J., Li, Y., and Zhang, H. (2021). Emerging SARS-CoV-2 variants of concern and potential intervention approaches. *Crit. Care* 25:244. doi: 10.1186/s13054-021-03662-x
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213. doi: 10.1093/nar/gkv951
- Kishimoto, M., Uemura, K., Sanaki, T., Sato, A., Hall, W. W., Kariwa, H., et al. (2021). TMPRSS2 and TMPRSS13 activate the SARS-CoV-2 spike protein. *Viruses* 13:384. doi: 10.3390/v13030384
- Koch, J., Uckele, Z. M., Doldan, P., Stanifer, M., Boulant, S., and Lozach, P. Y. (2021). TMPRSS2 expression dictates the entry route used by SARS-CoV-2 to infect host cells. *EMBO J.* 40:e107821. doi: 10.15252/embj.2021107821
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5
- 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
- Li, F., Berardi, M., Li, W., Farzan, M., Dormitzer, P. R., and Harrison, S. C. (2006). Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *J. Virol.* 80, 6794–6800. doi: 10.1128/jvi.02744-05
- Li, F., Li, W., Farzan, M., and Harrison, S. C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309, 1864–1868. doi: 10.1126/science.1116480
- Li, Y., Xie, Z., Lin, W., Cai, W., Wen, C., Guan, Y., et al. (2020). Efficacy and safety of Lopinavir/Ritonavir or Arbidol in adult patients with mild/moderate COVID-19: An exploratory randomized controlled trial. *Med.* 1, 105–113.e4. doi: 10.1016/j.medj.2020.04.001
- Lim, J., Jeon, S., Shin, H. Y., Kim, M. J., Seong, Y. M., Lee, W. J., et al. (2020). The Author's response: Case of the index patient who caused tertiary transmission of corona virus disease 2019 in Korea: The application of Lopinavir/Ritonavir for the treatment of COVID-19 Pneumonia monitored by quantitative RT-PCR. *J. Korean Med. Sci.* 35:e89–113.e4. doi: 10.3346/jkms.2020.35.e89
- Lin, Q., Zhao, S., Gao, D., Lou, Y., Yang, S., Musa, S. S., et al. (2020). A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int. J. Infect. Dis.* 93, 211–216. doi: 10.1016/j.ijid.2020.02.058
- Liu, Y., Gayle, A. A., Wilder-Smith, A., and Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* 27:taaa021. doi: 10.1093/jtm/taaa021
- Lu, L., Zhang, H., Dauphars, D. J., and He, Y. W. (2021). A potential role of interleukin 10 in COVID-19 pathogenesis. *Trends Immunol.* 42, 3–5. doi: 10.1016/j.it.2020.10.012
- MacGowan, S. A., and Barton, G. J. (2020). Missense variants in ACE2 are predicted to encourage and inhibit interaction with SARS-CoV-2 spike and contribute to genetic risk in COVID-19. *bioRxiv* [Epub ahead of preprint], doi: 10.1101/2020.05.03.074781 %J
- 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

- Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., and Taguchi, F. (2010). Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J. Virol.* 84, 12658–12664. doi: 10.1128/jvi.01542-10
- Mikhael, E. M., Ong, S. C., and Sheikh Ghadzi, S. M. (2022). Efficacy and safety of Sitagliptin in the treatment of COVID-19. *J. Pharm. Pract.* 8971900221102119: 089719002211021. doi: 10.1177/08971900221102119
- Moore, K. W., de Waal Malefyt, R., Coffman, R. L., and O'Garra, A. (2001). Interleukin-10 and the interleukin-10 receptor. *Annu. Rev. Immunol.* 19, 683–765. doi: 10.1146/annurev.immunol.19.1.683
- Moore, K. W., Vieira, P., Fiorentino, D. F., Trounstein, M. L., Khan, T. A., and Mosmann, T. R. (1990). Homology of cytokine synthesis inhibitory factor (IL-10) to the Epstein-Barr virus gene BCRF1. *Science* 248, 1230–1234. doi: 10.1126/science.2161559
- Motarmjemzadeh, Q., Aidenloo, N. S., and Abbaszadeh, M. (2015). Detection of hydroxychloroquine retinal toxicity by automated Perimetry in 60 rheumatoid arthritis patients with Normal Fundoscopic findings. *Glob J Health Sci* 8, 59–64. doi: 10.5539/gjhs.v8n3p59
- Nádasdi, Á., Sinkovits, G., Bobek, I., Lakatos, B., Föhréc, Z., Prohászka, Z. Z., et al. (2022). Decreased circulating dipeptidyl peptidase-4 enzyme activity is prognostic for severe outcomes in COVID-19 inpatients. *Biomark. Med* 16, 317–330. doi: 10.2217/bmm-2021-0717
- Nargis, T., and Chakrabarti, P. (2018). Significance of circulatory DPP4 activity in metabolic diseases. *IUBMB Life* 70, 112–119. doi: 10.1002/iub.1709
- Perlman, S., and Netland, J. (2009). Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.* 7, 439–450. doi: 10.1038/nrmicro2147
- Rai, P., Kumar, B. K., Deekshit, V. K., Karunasagar, I., and Karunasagar, I. (2021). Detection technologies and recent developments in the diagnosis of COVID-19 infection. *Appl. Microbiol. Biotechnol.* 105, 441–455. doi: 10.1007/s00253-020-11061-5
- Rastelli, G., Del Rio, A., Degliesposti, G., and Sgobba, M. (2010). Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* 31, NA–810. doi: 10.1002/jcc.21372
- Ratajczak, M. Z., Bujko, K., Ciechanowicz, A., Sielatycka, K., Cymer, M., Marlicz, W., et al. (2021). SARS-CoV-2 entry receptor ACE2 is expressed on very small CD45(–) precursors of hematopoietic and endothelial cells and in response to virus spike protein activates the Nlrp3 Inflammasome. *Stem Cell Rev. Rep.* 17, 266–277. doi: 10.1007/s12015-020-10010-z
- Ratajczak, M. Z., Bujko, K., Cymer, M., Thapa, A., Adamiak, M., Ratajczak, J., et al. (2020). The Nlrp3 inflammasome as a "rising star" in studies of normal and malignant hematopoiesis. *Leukemia* 34, 1512–1523. doi: 10.1038/s41375-020-0827-8
- Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J., and Sanner, M. F. (2015). AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS Comput. Biol.* 11:e1004586. doi: 10.1371/journal.pcbi.1004586
- Rohit, A., Rajasekaran, S., Karunasagar, I., and Karunasagar, I. (2020). Fate of respiratory droplets in tropical vs temperate environments and implications for SARS-CoV-2 transmission. *Med. Hypotheses* 144:109958. doi: 10.1016/j.mehy.2020.109958
- Salvi, V., Nguyen, H. O., Sozio, F., Schioppa, T., Gaudenzi, C., Laffranchi, M., et al. (2021). SARS-CoV-2-associated ssRNAs activate inflammation and immunity via TLR7/8. *JCI. Insight* 6:e150542. doi: 10.1172/jci.insight.150542
- Saraiva, M., Vieira, P., and O'Garra, A. (2020). Biology and therapeutic potential of interleukin-10. *J. Exp. Med.* 217:e20190418. doi: 10.1084/jem.20190418
- Senapati, S., Banerjee, P., Bhagavathula, S., Kushwaha, P. P., and Kumar, S. (2021). Contributions of human ACE2 and TMPRSS2 in determining host-pathogen interaction of COVID-19. *J. Genet.* 100:12. doi: 10.1007/s12041-021-01262-w
- Shapira, T., Monreal, I. A., Dion, S. P., Buchholz, D. W., Imbiakha, B., Olmstead, A. D., et al. (2022). A TMPRSS2 inhibitor acts as a pan-SARS-CoV-2 prophylactic and therapeutic. *Nature* 605, 340–348. doi: 10.1038/s41586-022-04661-w
- Shouval, D. S., Biswas, A., Goettl, J. A., McCann, K., Conaway, E., Redhu, N. S., et al. (2014). Interleukin-10 receptor signaling in innate immune cells regulates mucosal immune tolerance and anti-inflammatory macrophage function. *Immunity* 40, 706–719. doi: 10.1016/j.immuni.2014.03.011
- Shulla, A., Heald-Sargent, T., Subramanya, G., Zhao, J., Perlman, S., and Gallagher, T. (2011). A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J. Virol.* 85, 873–882. doi: 10.1128/jvi.02062-10
- Simmons, G., Gosalia, D. N., Rennekamp, A. J., Reeves, J. D., Diamond, S. L., and Bates, P. (2005). Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11876–11881. doi: 10.1073/pnas.0505577102
- Singh, A. K., Singh, A., Singh, R., and Misra, A. (2021). Molnupiravir in COVID-19: a systematic review of literature. *Diabetes Metab. Syndr.* 15:102329. doi: 10.1016/j.dsx.2021.102329
- Skariyachan, S. (2022). Scope of computational biology and bioinformatics toward the discovery of potential therapeutic agents against viral diseases. *Futur. Virol.* 17, 257–260. doi: 10.2217/fvl-2021-0281
- Skariyachan, S., Gopal, D., Deshpande, D., Joshi, A., Uttarkar, A., and Niranjan, V. (2021). Carbon fullerene and nanotube are probable binders to multiple targets of SARS-CoV-2: insights from computational modeling and molecular dynamic simulation studies. *Infect. Genet. Evol.* 96:105155. doi: 10.1016/j.meegid.2021.105155
- Solanich, X., Vargas-Parra, G., van der Made, C. I., Simons, A., Schuurs-Hoeijmakers, J., Antolí, A., et al. (2021). Genetic screening for TLR7 variants in Young and previously healthy men with severe COVID-19. *Front. Immunol.* 12:719115. doi: 10.3389/fimmu.2021.719115
- Solerte, S. B., Di Sabatino, A., Galli, M., and Fiorina, P. (2020a). Dipeptidyl peptidase-4 (DPP4) inhibition in COVID-19. *Acta Diabetol.* 57, 779–783. doi: 10.1007/s00592-020-01539-z
- Solerte, S. B., Di Sabatino, A., Galli, M., and Fiorina, P. (2020b). Dipeptidyl peptidase-4 (DPP4) inhibition in COVID-19. *Acta Diabetol.* 57, 779–783. doi: 10.1007/s00592-020-01539-z
- Strollo, R., and Pozzilli, P. (2020). DPP4 inhibition: preventing SARS-CoV-2 infection and/or progression of COVID-19? *Diabetes Metab. Res. Rev.* 36:e3330. doi: 10.1002/dmrr.3330
- Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C. K., Zhou, J., et al. (2016). Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 24, 490–502. doi: 10.1016/j.tim.2016.03.003
- Swanson, K. V., Deng, M., and Ting, J. P. (2019). The NLRP3 inflammasome: molecular activation and regulation to therapeutics. *Nat. Rev. Immunol.* 19, 477–489. doi: 10.1038/s41577-019-0165-0
- Tabares-Guevara, J. H., Jaramillo, J. C., Ospina-Quintero, L., Piedrahíta-Ochoa, C. A., García-Valencia, N., Bautista-Erazo, D. E., et al. (2021). IL-10-dependent amelioration of chronic inflammatory disease by microdose subcutaneous delivery of a prototypic Immunoregulatory small molecule. *Front. Immunol.* 12:708955. doi: 10.3389/fimmu.2021.708955
- Tang, F. S., Van Ly, D., Spann, K., Reading, P. C., Burgess, J. K., Hartl, D., et al. (2016). Differential neutrophil activation in viral infections: enhanced TLR-7/8-mediated CXCL8 release in asthma. *Respirology* 21, 172–179. doi: 10.1111/resp.12657
- Turner, A. J., and Hooper, N. M. (2002). The angiotensin-converting enzyme gene family: genomics and pharmacology. *Trends Pharmacol. Sci.* 23, 177–183. doi: 10.1016/s0165-6147(00)01994-5
- Vankadari, N., and Wilce, J. A. (2020). Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. *Emerg. Microbes. Infect.* 9, 601–604. doi: 10.1080/22221751.2020.1739565
- Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., et al. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* 30, 269–271. doi: 10.1038/s41422-020-0282-0
- Wang, X., Wong, K., Ouyang, W., and Rutz, S. (2019). Targeting IL-10 family cytokines for the treatment of human diseases. *Cold Spring Harb. Perspect. Biol.* 11:a028548. doi: 10.1101/cshperspect.a028548
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444–1448. doi: 10.1126/science.abb2762
- Yuan, Q., Zhou, Q., Ren, J., Wang, G., Yin, C., Shang, D., et al. (2021). WGCNA identification of TLR7 as a novel diagnostic biomarker, progression and prognostic indicator, and immunotherapeutic target for stomach adenocarcinoma. *Cancer Med.* 10, 4004–4016. doi: 10.1002/cam4.3946
- Zhang, K., Wang, K., Zhang, C., Teng, X., Li, D., and Chen, M. (2022). Exploring the potential mechanism of emetine against coronavirus disease 2019 combined with lung adenocarcinoma: bioinformatics and molecular simulation analyses. *BMC Cancer* 22:687. doi: 10.1186/s12885-022-09763-2
- Zhang, C., Zhang, Y., Zhang, S., Wang, Z., Sun, S., Liu, M., et al. (2020). Intracellular autoactivation of TMPRSS11A, an airway epithelial transmembrane serine protease. *J. Biol. Chem.* 295, 12686–12696. doi: 10.1074/jbc.RA120.014525
- Zhen, Y., and Zhang, H. (2019). NLRP3 Inflammasome and inflammatory bowel disease. *Front. Immunol.* 10:276. doi: 10.3389/fimmu.2019.00276
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395, 1054–1062. doi: 10.1016/s0140-6736(20)30566-3
- Zigmond, E., Bernshtein, B., Friedlander, G., Walker, C. R., Yona, S., Kim, K. W., et al. (2014). Macrophage-restricted interleukin-10 receptor deficiency, but not IL-10 deficiency, causes severe spontaneous colitis. *Immunity* 40, 720–733. doi: 10.1016/j.immuni.2014.03.012



OPEN ACCESS

EDITED BY

Arli Aditya Parikesit,
Indonesia International Institute for
Life-Sciences (i3L), Indonesia

REVIEWED BY

Surjit Singh,
Sister Nivedita University,
India
Zhen Luo,
Jinan University,
China
Viol Kharisma,
Airlangga University,
Indonesia

*CORRESPONDENCE

Geng Li
✉ lg@gzucm.edu.cn
Xiaohong Liu
✉ drlxh@foxmail.com
Yanni Lai
✉ laianni@gzucm.edu.cn

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

RECEIVED 10 November 2022

ACCEPTED 29 December 2022

PUBLISHED 01 February 2023

CITATION

Han T, Luo Z, Ji L, Wu P, Li G, Liu X and
Lai Y (2023) Identification of natural compounds
as SARS-CoV-2 inhibitors *via* molecular docking
and molecular dynamic simulation.
Front. Microbiol. 13:1095068.
doi: 10.3389/fmicb.2022.1095068

COPYRIGHT

© 2023 Han, Luo, Ji, Wu, Li, Liu and Lai. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Identification of natural compounds as SARS-CoV-2 inhibitors *via* molecular docking and molecular dynamic simulation

Tiantian Han^{1†}, Ziqing Luo^{2†}, Lichun Ji³, Peng Wu¹, Geng Li^{2*},
Xiaohong Liu^{1*} and Yanni Lai^{4*}

¹The First Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, China, ²Laboratory Animal Center, Guangzhou University of Chinese Medicine, Guangzhou, China, ³The Third Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, China, ⁴School of Basic Medical Sciences, Guangzhou University of Chinese Medicine, Guangzhou, China

Background: Base mutations increase the contagiousness and transmissibility of the Delta and Lambda strains and lead to the severity of the COVID-19 pandemic. Molecular docking and molecular dynamics (MD) simulations are frequently used for drug discovery and relocation. Small molecular compounds from Chinese herbs have an inhibitory effect on the virus. Therefore, this study used computational simulations to investigate the effects of small molecular compounds on the spike (S) protein and the binding between them and angiotensin-converting enzyme 2 (ACE2) receptors.

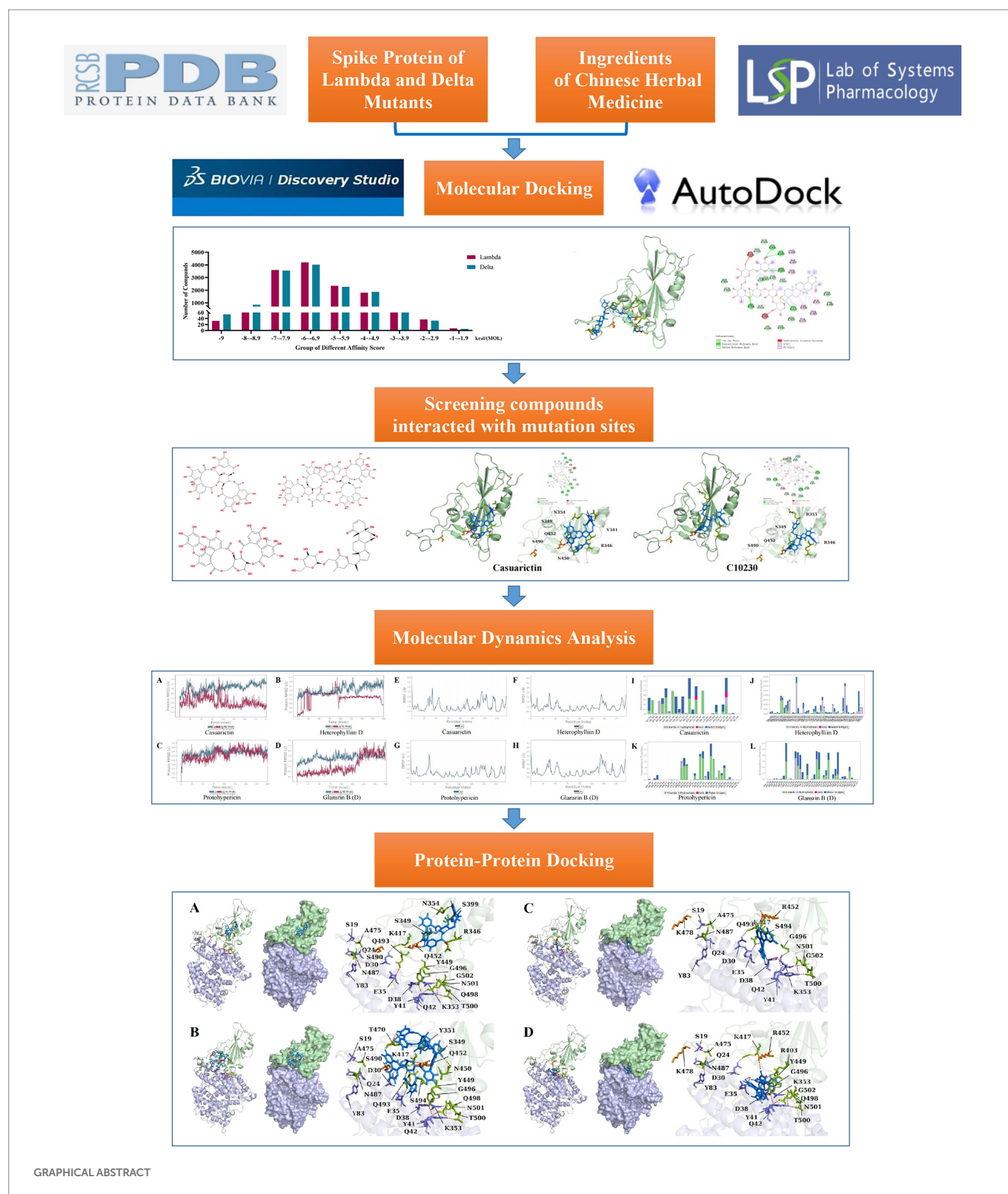
Methods: In this study, molecular docking, MD simulation, and protein–protein analysis were used to explore the medicinal target inhibition of Chinese herbal medicinal plant chemicals on SARS-CoV-2. 12,978 phytochemicals were screened against S proteins of SARS-CoV-2 Lambda and Delta mutants.

Results: Molecular docking showed that 65.61% and 65.28% of the compounds had the relatively stable binding ability to the S protein of Lambda and Delta mutants (docking score ≤ -6). The top five compounds with binding energy with Lambda and Delta mutants were clematichinenoside AR2 (−9.7), atratoglucoside, b (−9.5), physalin b (−9.5), atratoglucoside, a (−9.4), Ochnaflavone (−9.3) and neo-przewaquinone a (−10), Wikstrosin (−9.7), xilingsaponin A (−9.6), ardisianoside G (−9.6), and 23-epi-26-deoxyactein (−9.6), respectively. Four compounds (Casuarictin, Heterophyllin D, Protohypericin, and Glansrin B) could interact with S protein mutation sites of Lambda and Delta mutants, respectively, and MD simulation results showed that four plant chemicals and spike protein have good energy stable complex formation ability. In addition, protein–protein docking was carried out to evaluate the changes in ACE2 binding ability caused by the formation of four plant chemicals and S protein complexes. The analysis showed that the binding of four plant chemicals to the S protein could reduce the stability of the binding to ACE2, thereby reducing the replication ability of the virus.

Conclusion: To sum up, the study concluded that four phytochemicals (Casuarictin, Heterophyllin D, Protohypericin, and Glansrin B) had significant effects on the binding sites of the SARS-CoV-2S protein. This study needs further *in vitro* and *in vivo* experimental validation of these major phytochemicals to assess their potential anti-SARS-CoV-2.

KEYWORDS

molecular docking, molecular dynamics simulation, binding energy, SARS-CoV-2 inhibitors, Chinese herbal medicine



Abbreviations: ACE2, Angiotensin-converting enzyme 2; MD, Molecular dynamics; S, Spike; RBD, Receptor binding domain; VOC, Variety of concern; NAb, Neutralizing antibodies; TCMSP, Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform; Leu, L, leucine; Gln, Q, glutamine; F, F, Fnylalanine; Ser, S, serine; Arg, R, Arginine; Thr, T, Threonine; Lys, K, Lysine; CTD, C-terminal domain; RMSD, Root Mean Square Deviation; RMSF, Root Mean Square Fluctuation; VOI, Variety of interest.

1. Introduction

COVID-19, caused by the Novel Coronavirus, remains a worldwide pandemic, with more than 628 million cumulative confirmed cases and more than 6.57 million cumulative deaths reported globally as of November 3, 2022. More and more studies have shown that COVID-19 can not only cause lung damage but

also attack the liver (Bangash et al., 2020), kidneys (Balawender et al., 2022), heart (Madjid et al., 2020), reproductive system (Fan et al., 2020), and nervous system (Chuang et al., 2021). Patients infected with COVID-19 will show symptoms such as fever, exertion, dry cough, loss of smell and taste, dyspnea, and severely infected patients will have respiratory and circulatory failure and even multi-organ failure, which will lead to death (Tsang et al., 2021). In addition, the latest research shows that the long-term symptoms caused by the COVID-19 infection cause patients to develop muscle pain, tiredness, fear, depression, and other psychological disorders and increase the risk of cardiovascular disease (Hanson et al., 2022). Although coronaviruses have proofreading mechanisms to maintain their long genomes and have a relatively low mutation rate, different variants have emerged with severe economic and social impacts (Robson et al., 2020).

The SARS-CoV-2 virus infects *via* the engagement of human ACE2 by the virus receptor binding domain (RBD) of its S protein (Lan et al., 2020). The dominant strain currently circulating is the Omicron variant, the Delta variant, and the Lambda variant were the dominant strain worldwide before the Omicron variant, which mutations occurring in the RBD region (Pulliam et al., 2022). Delta has two mutations within its RBD, L452R, and T478K. The Lambda variant, once a major strain in Argentina and Chile, has two mutations in its RBD, L452Q, and F490S (Romero et al., 2021). Compared with the wild cohort, the Delta variant spreads faster, has a shorter incubation period, higher viral load, and longer viral clearance time, and elderly patients infected with Delta variety of concerns (VOCs) are more likely to develop critical illness (Wang et al., 2021). In addition to being highly transmissible, the Lambda variant is more infectious and resistant to neutralizing antibodies (NAb), resulting in a decrease in the protective effect of all currently approved anti-SARS-CoV-2 vaccines (Kimura and Kosugi et al., 2022). The Omicron variant is currently circulating; although the transmissibility and immune escape are both high, the pathogenicity has “substantially decreased.” Moreover, the probability of developing new crowns after infection with the Omicron variant has decreased by about 20%–50% compared with the Delta variant (Nyberg et al., 2022; Xie et al., 2022). At present, although vaccines can provide a high level of prevention of hospitalization and reduce mortality, vaccines cannot prevent new coronavirus infection or reinfection, and there are currently no effective anti-new coronavirus drugs (Nyberg et al., 2022).

Studies have shown that traditional Chinese medicine has played a particular role in the treatment of COVID-19, suggesting that antiviral and even anti-variant drug candidates can be found in various Chinese herbal medicines (Ren et al., 2020; Runfeng et al., 2020). Drug screening in the preclinical and clinical stages is costly and time-consuming (Walls et al., 2020). Virtual screening by computer can provide rapid, considerable, and new testable hypotheses for drug repositioning (Cheng et al., 2017).

In this study, the protein structures of Delta and Lambda variants were constructed and, respectively, docked with 12,978 small molecule compounds, which were verified and extracted from Chinese herbal medicine downloaded from the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP). The compound with the highest docking score was screened out, and then MD was simulated. The protein–protein docking was carried out to understand its mode of action. This study is expected to provide a reference for screening anti-virus and even anti-mutant strains.

2. Materials and methods

2.1. Protein structure and evaluation

Download the S protein and ACE2 binding structure file (6LZG) from the PDB database.¹ The mutation sites of the Delta and Lambda strain of the binding domain of the S protein were confirmed. Using PyMOL 2.1 to carry out virtual mutation function, 452 amino acids of S protein RBD were mutated from leucine (Leu, L) to glutamine (Gln, Q) and 490 amino acids were mutated from Pnylalanine (F, F) to serine (Ser, S) to form Lambda strain. The 452 amino acid of RBD of S protein was mutated from L to arginine (Arg, R), and the 478 amino acid was mutated from threonine (Thr, T) to lysine (Lys, K) to form Delta strain. The protein was added hydrogen by Autodock Tool 1.5.6 and formed a PDBQT file. The rationality of protein conformations was evaluated by Ramachandran plot using UCLA-DOE's SAVES server v 6.0.²

2.2. Compound structure and optimization

The three-dimensional conformations of the 13,144 small molecule compounds from 500 Chinese herbs were downloaded from the TCMSP database.³ The compounds were minimized by assigning force field MM2 by using ChemBio3D Ultra 13.0, and the optimized structure was prepared.

2.3. Virtual screening and molecular docking

Autodock Vina 1.1.2 was used in silicon docking. The docking was carried out, and the binding pocket covered the mutation base with the following parameters –42.602, 33.01, and 9.399 for the X, Y, and Z axes, respectively. These coordinates represent the binding site area covering the C-terminal domain (CTD) of the Sprotein of SARS-CoV-2. The x, y, and z length of the grid box is 50. The affinity score was analyzed by sectional statistics, and GraphPad Prism 9.0 was used for visualization.

2.4. Molecular dynamics simulation

MD simulations of protein and compound complexes were performed by Desmond v2020. OPLS3e was selected as the molecular field for MD simulation. TIP3 water model was used in the MD system. Neutralize the system charge by adding ions. The energy minimization of the entire system is achieved by using the OPLS3e force field since it is a full atomic-type force field. The geometry of water molecules, bond lengths, and bond angles of heavy atoms are constrained by the SHAKE algorithm. A continuous system is simulated by applying periodic boundary conditions. Long-range static electricity is maintained by the particle mesh Ewald method. An NPT method harness at 300 K and 1.0 bar was used to balance the system. Berendsen coupling

¹ <https://www.rcsb.org/>

² <https://saves.mbi.ucla.edu/>

³ <https://old.tcm-sp-e.com/tcm-sp.php>

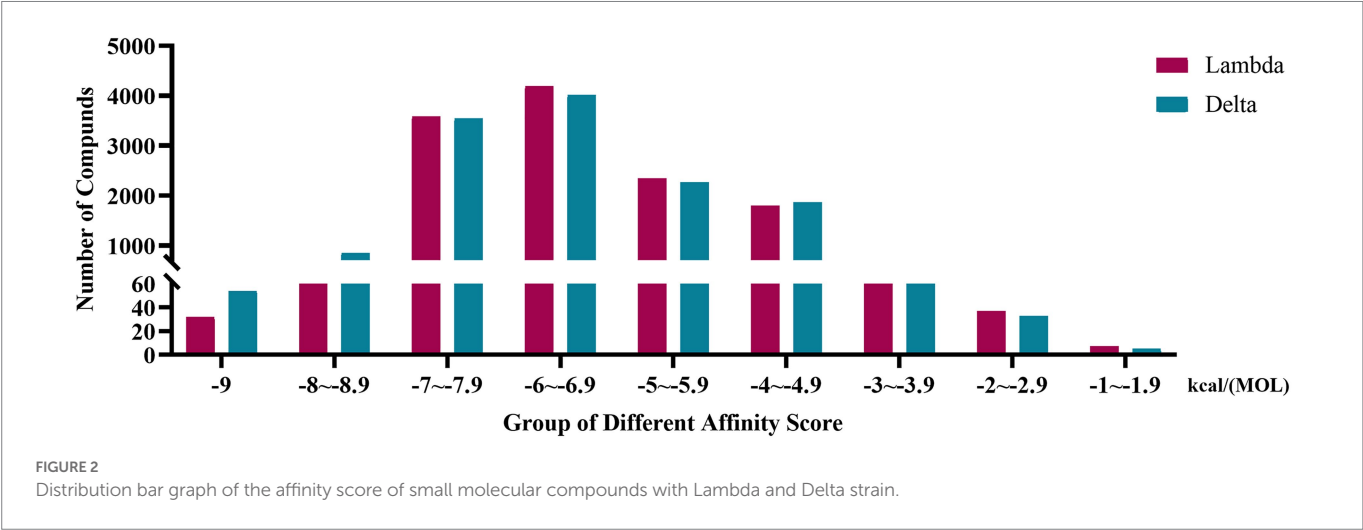
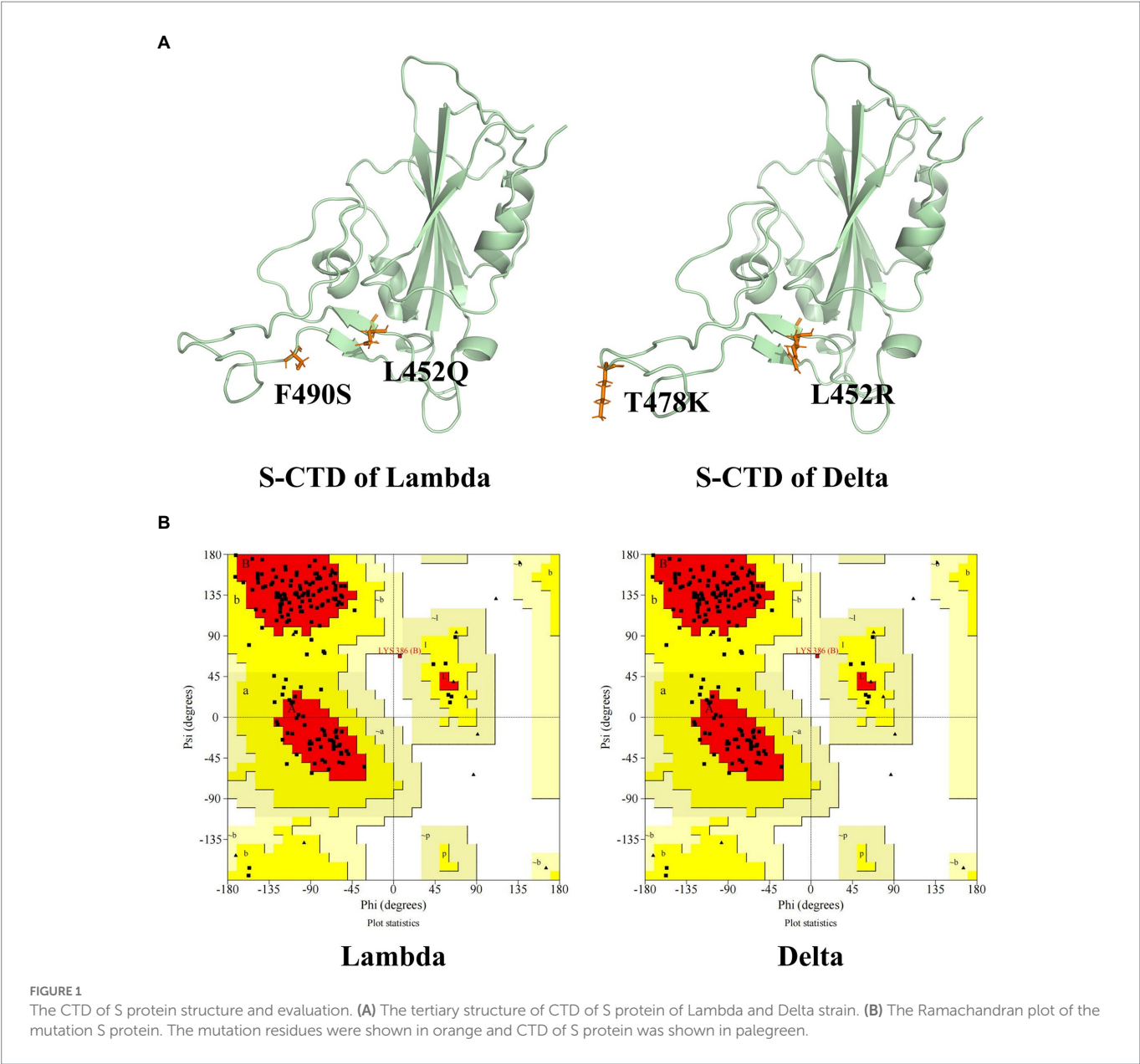


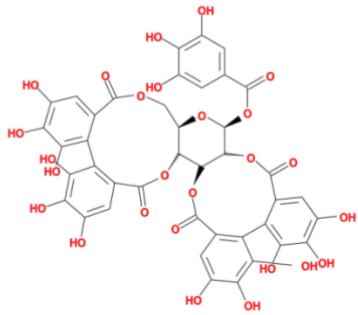
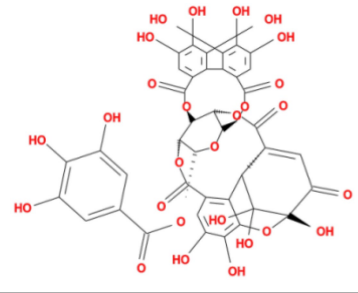
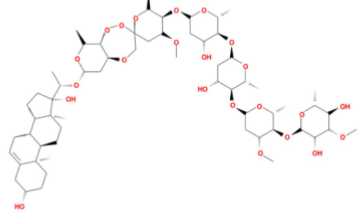
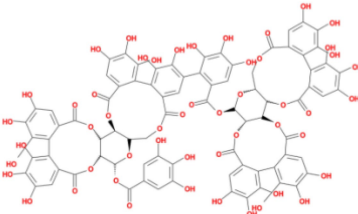
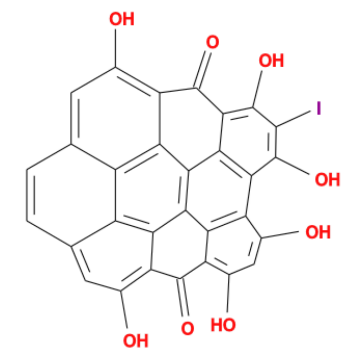
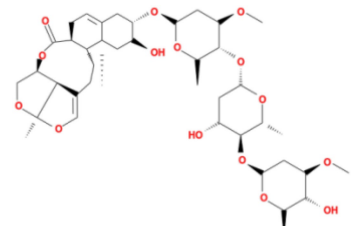
TABLE 1 Molecular docking results of Lambda S-CTD and compounds (top 20).

Rank	Affinity (kcal/MOL)	MOL_ID	molecule_name
1	−9.7	MOL000753	clematichinenoside AR2
2	−9.5	MOL006876	atrato-glucoside, b
3	−9.5	MOL007238	physalin b
4	−9.4	MOL006874	atrato-glucoside, a
5	−9.3	MOL003009	Ochnaflavone
6	−9.3	MOL003351	cyclopseudo-hypericin
7	−9.3	MOL005678	periplocoside J
8	−9.3	MOL006757	Bryonolic acid
9	−9.3	MOL007353	solamargine
10	−9.2	MOL003278	salaspermic acid
11	−9.2	MOL005671	periplocoside C
12	−9.2	MOL009468	β1-solamargine
13	−9.2	MOL009472	26-O-β-D-glucopyranosyl-nuatigenin-3-O-α-L-rhamnopyranosyl(1→2)-β-D-glucopyranoside
14	−9.2	MOL011100	bisindigotin
15	−9.1	MOL002067	hypericin
16	−9.1	MOL002659	kihadanin A
17	−9.1	MOL004500	Markogenin-3-O-beta-D-glucopyranosyl-(1→2)-beta-D-galactopyranoside
18	−9.1	MOL005459	Diosgenin-3-O-beta-D-glucopyranoside
19	−9.1	MOL006855	solamargine
20	−9.1	MOL006895	glucoside, d

TABLE 2 Molecular docking results of Delta S-CTD and compounds (top 20).

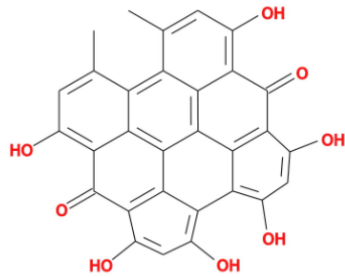
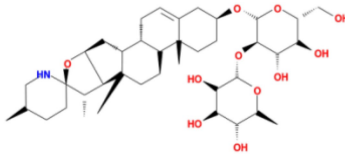
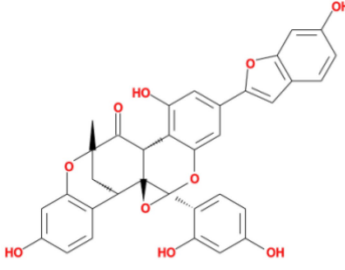
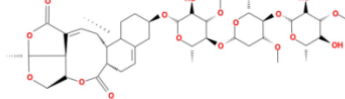
Rank	Affinity (kcal/MOL)	MOL_ID	molecule_name
1	−10	MOL007062	neo-przewaquinone a
2	−9.7	MOL011124	Wikstrosin
3	−9.6	MOL004521	xilingsaponin A
4	−9.6	MOL011030	ardisianoside G
5	−9.6	MOL011990	23-epi-26-deoxyactein
6	−9.6	MOL012727	mulberrofurin K
7	−9.5	MOL004509	Timosaponin A III
8	−9.5	MOL007238	physalin b
9	−9.5	MOL009474	26-O-β-D-glucopyranosyl-nuatigenin-3-O-α-L-rhamnopyranosyl(1→2)-o-[β-D-glucopyranosyl(1→4)]-β-D-glucopyranoside
10	−9.4	MOL004444	Ziebeimine
11	−9.4	MOL005626	cynapanoside B
12	−9.4	MOL006757	Bryonolic acid
13	−9.4	MOL006874	atrato-glucoside, a
14	−9.4	MOL006876	atrato-glucoside, b
15	−9.4	MOL011540	withanolide D
16	−9.4	MOL012929	glucoside C
17	−9.3	MOL003351	cyclopseudo-hypericin
18	−9.3	MOL005678	periplocoside J
19	−9.3	MOL008582	trillin
20	−9.3	MOL011486	Datuarmeteloside B

TABLE 3 Molecular docking results of Lambda S-CTD mutant bases and compounds.

Rank	Affinity (kcal/MOL)	MOL_ID	Molecule name	Structure
1	−9.1	MOL010726	Casuarictin	
2	−9.0	MOL002505	C10230	
3	−8.9	MOL005678	Periplocoside J	
4	−8.1	MOL009110	Heterophylliin D	
5	−7.9	MOL003351	Cyclopseudo-hypericin	
6	−7.9	MOL006895	Glucoside,d	

(Continued)

TABLE 3 (Continued)

Rank	Affinity (kcal/MOL)	MOL_ID	Molecule name	Structure
7	−7.8	MOL002067	Hypericin	
8	−7.6	MOL009468	β1-solamargine	
9	−7.5	MOL012732	Mulberrofuran Q	
10	−7.2	MOL006876	Atratoglucoside,b	

algorithm is used to couple temperature–pressure parameters. At a later stage of preparation of the system, 100 ns was run at a time step of 1.2 fs. The track was recorded every 100 ps, recording a total of 1,000 frames. The Root Mean Square Deviation (RMSD) of backbone atoms was calculated, and a graphical analysis was performed to understand the nature of protein–ligand interactions. The Root Mean Square Fluctuation (RMSF) of each residue was calculated to understand the major conformational changes of the residue between the initial state and the kinetic state.

2.5. Protein–protein docking

Protein structure (6LZG) was obtained from the PDB database (see text footnote 1, respectively) and processed by PyMOL 2.1 software, including the removal of water and ions, protonation, the addition of missing atoms, and completion of chemical groups. HDock SERVER⁴ is used to set the protein as a rigid state. The range of Receptor protein was set as 19:A, 24:A, 35:A, 38:A, 42:A, 353:A, and 30:A. The range of Delta and Lambda S protein was set as 487:B, 475:B, 417:B, 493:B, 502:B, 496:B, 449:B, and 498:B. 100 conformations were generated in silicon docking, and the best conformation was selected by scoring in ascending order.

⁴ <http://hdock.phys.hust.edu.cn/>

3. Results

3.1. Evaluation of protein structure rationality

Ramachandran Plot was used to assess the rationality of the mutation S protein (Morris et al., 1992). As can be seen in Figure 1, the total number of residues of CTD of S protein is 195. The residues in most favored regions take part of 82.1%, and residues in additional allowed regions were 17.3%. The residues in disallowed regions were 0.6% which was <5%, indicating the rationality of the structure of the S protein of the mutation strain.

3.2. Optimization results of compound structure

There were 13,144 compounds in the TCMSP database, and 12,978 compounds were optimized after removing the compounds with unreasonable structures 166 compounds could not be optimized.

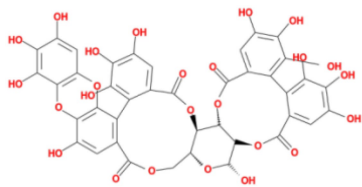
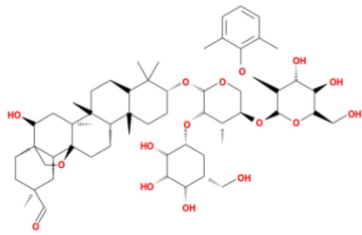
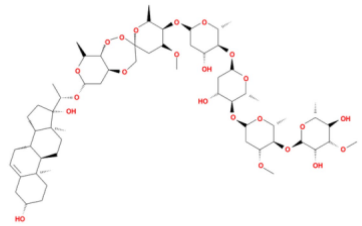
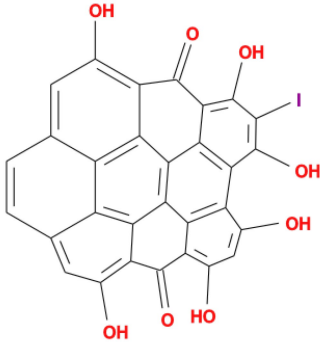
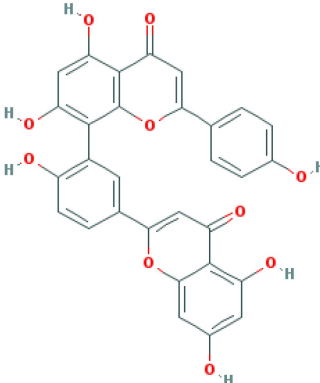
3.3. Molecular docking analysis

Affinity score was used to evaluate the result of virtual screening. The lower the affinity score, the more stable the confirmation. The docking results of 12,978 compounds with Lambda S-CTD and Delta S-CTD were ranked from lowest to highest according to affinity scores.

The affinity scores were normally distributed. Most compounds were concentrated in a fraction of -4 to -7.9 kcal/(MOL). Generally, the conformations with affinity scores below -8 were considered reliable. There were 734 (5.66%) Lambda conformation and 910 (7.01%) Delta conformations below -8 (Figure 2), indicating those compounds which consisted of conformation had an interactive relationship with CTD of S protein.

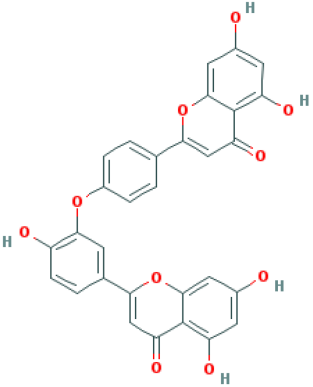
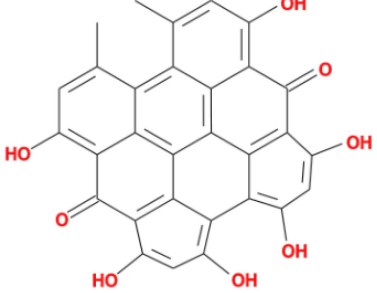
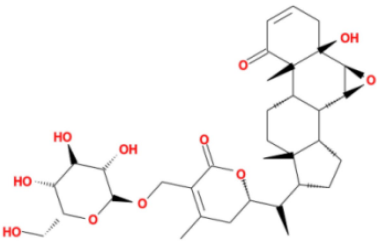
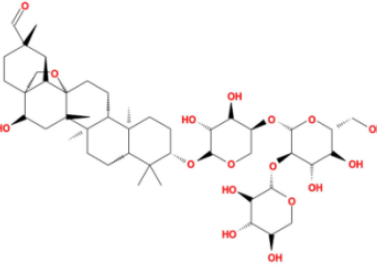
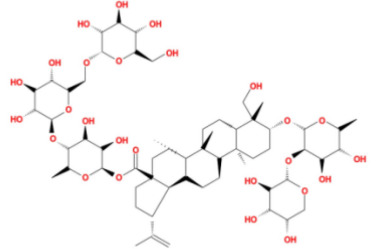
The top 20 compounds were shown in Tables 1, 2. Among the top 100 results, not all compounds bind to the mutation base. By checking the 9 conformations manually of the top 100 results, the conformation of compounds bound to the mutation base was selected and listed in Tables 3, 4 according to affinity score from lowest to highest. The 2D structure of compounds was shown in Tables 3, 4.

TABLE 4 Molecular docking results of Delta S-CTD mutant bases and compounds.

Rank	Affinity (kcal/MOL)	MOL_ID	Molecule name	Structure
1	-8.9	MOL009107	Glansrin B	
2	-8.9	MOL008305	Ardisiacrispin A	
3	-8.6	MOL005678	Periplocoside J	
4	-8.4	MOL003351	Cyclopseudo-hypericin	
5	-8.4	MOL002037	Amentoflavone	

(Continued)

TABLE 4 (Continued)

Rank	Affinity (kcal/MOL)	MOL_ID	Molecule name	Structure
6	−8.4	MOL003009	Ochnaflavone	
7	−8.3	MOL002067	Hypericin	
8	−8.2	MOL003350	Protohypericin	
9	−8.2	MOL011486	Datuarmeteloside B	
10	−8.2	MOL010968	Primulanin	

It showed that Casuarictin, C10230, Cyclopseudo-hypericin, Glucoside,d, Hypericin, β 1-solamargine, Mulberrofuran Q, and Atratoglucoside, bind to Q452 of lambda strain, while Periplocoside J binds to S490. Heterophylliin D binds to both Q452 and S490 through van der Waals and conventional hydrogen bonds (Figure 3, Table 5). Glansrin B, Ardisiacrispin A, Periplocoside J, Cyclopseudo-hypericin,

Amentoflavone, Ochnaflavone, Hypericin, Datuarmeteloside B, and Primulanin only bind to R452 of delta strain, while Protohypericin binds to R452 and K478, respectively, in a different conformation. Protohypericin binds to R452 through conventional hydrogen bond, and it binds to K478 through Pi-cation, Pi-Alkyl and conventional hydrogen bond (Figure 4, Table 6).

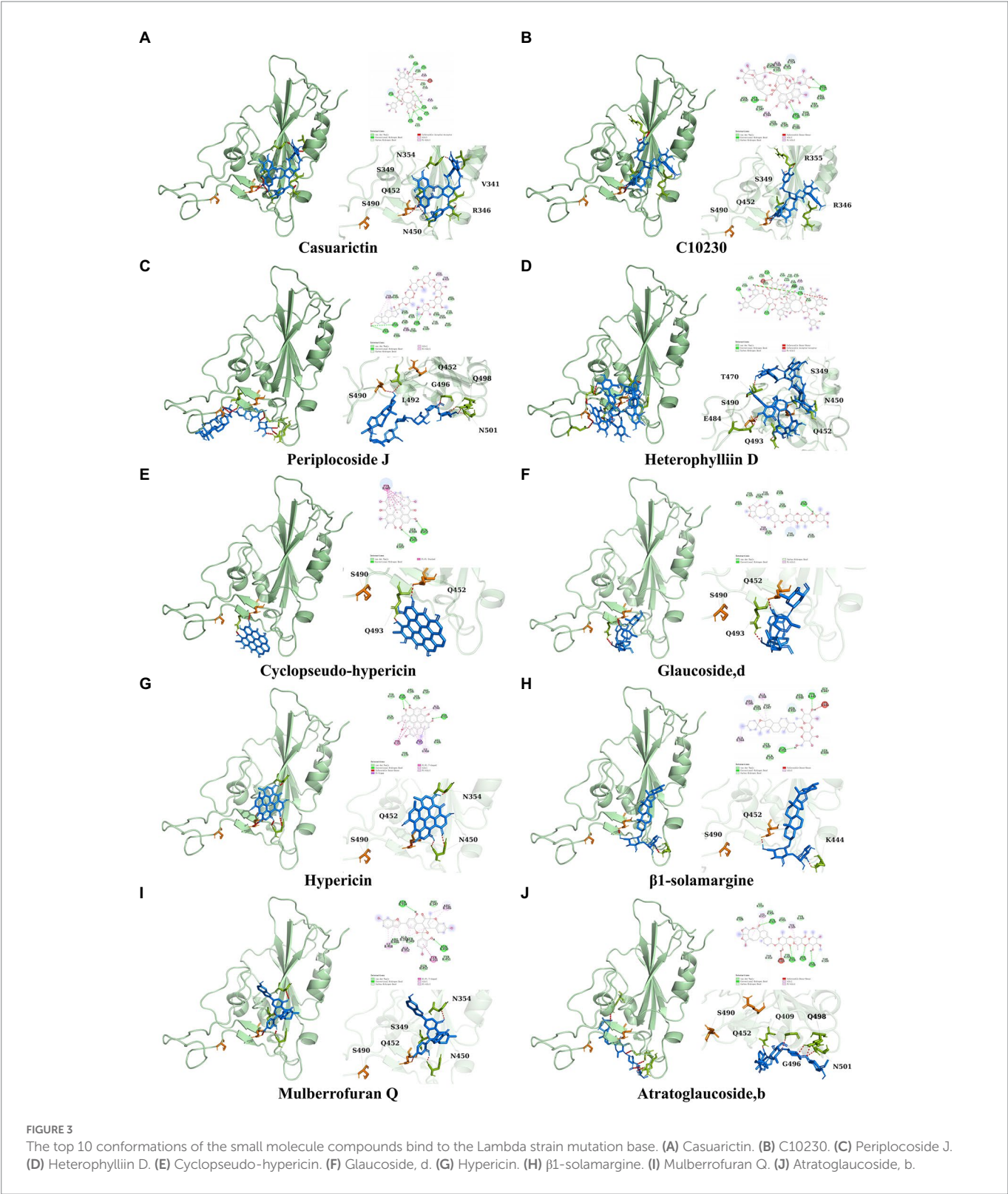


TABLE 5 The residues for the chemical bonds between compounds and spike proteins of Lambda strain.

Compounds	Van de waals	Conventional hydrogen bond	Unfavorable acceptor-acceptor	Carbon hydrogen bond	Pi-Alky1
Casuarictin	F357, Y451, N354, K356, I468	E340, V341, R346, S349, Y351, N450, Q452	S399	A348	A344, A352
Heterophylliin D	P491, E471, I472, S490, S494, Y451, Y449, F347, A348, S349, Y351, S469, I468	Q493, Q452, E484, T470	L492	N450	A352

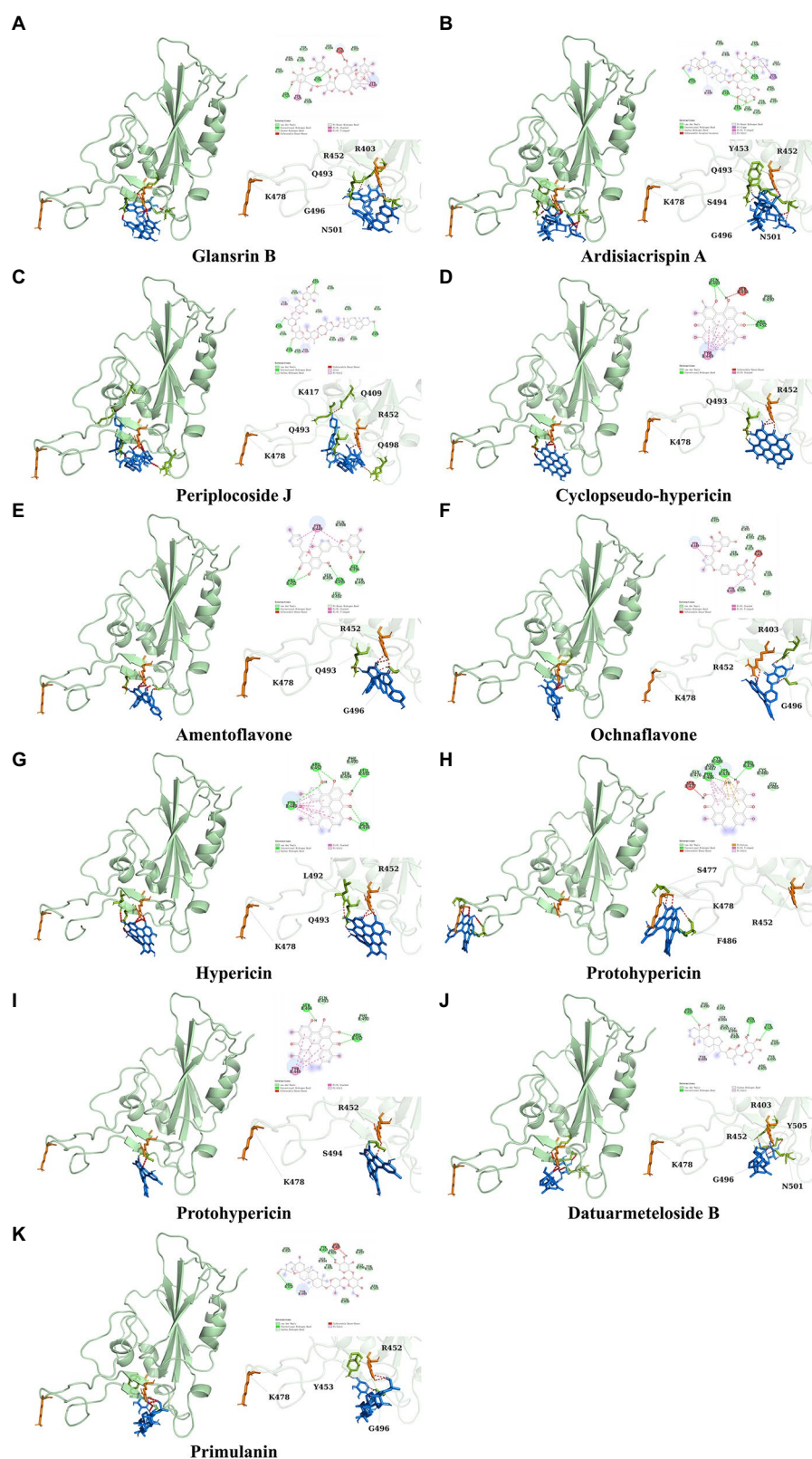


FIGURE 4

The top 10 conformations of the small molecule compounds bind to the Delta strain mutation base. (A) Glansrin B. (B) Ardisiacrispin A. (C) Periplocoside J. (D) Cyclopseudo-hypericin. (E) Amentoflavone. (F) Ochnaflavone. (G) Hypericin. (H, I) Protohypericin. (J) Datuarmeteloside B. (K) Primulanin.

3.4. Molecular dynamics analysis

Molecular docking cannot fully consider the flexibility of protein structure. To further elucidate the critical interactions between small

molecule compounds and S protein, MD simulation was performed on the optimal molecular docking model in this study. The model with the highest score and the model with the most mutation sites were considered to be the best model. They were Casuarictin, Heterophyllin

D and S protein of Lambda strain, Protohypericin, and Glansrin B and S protein of Delta strain, respectively.

In MD trajectory analysis, root means square deviation (RMSD) and root mean square fluctuation (RMSF) is the most frequently used index. The larger RMSD is, the more unstable the conformation is. As seen from [Figures 5A–D](#), the small molecule fluctuated at the beginning and gradually tend to be stable in the movement process. In the MD simulation of 200ns, the RMSD between the four compounds and the S protein reached a relative equilibrium state. It revealed that these four compounds approach the appropriate position of the S protein, thus promoting the stability between them and indicating that the small molecule was well combined with the protein surface. RMSF can be used to observe the allosteric of local sites in the simulation process. The larger the RMSF is, the more pronounced the conformational change of residue is. 0.2 was used to be the cut-off value. As shown in [Figures 5E–H](#), in the binding between Casuarictin and S protein, residues 30–40, 140–145, 150–156 were Highly fluctuating. In the binding between Heterophyllin D and S protein, residues 35–40 and 140–145 were highly fluctuating. In the binding between Protohypericin and S protein, residues 35–40 were highly fluctuating, while the residues 145–150 were highly fluctuating in the binding between Glansrin B and S protein.

Protein-ligand interactions can be monitored throughout the simulation. These interactions can be classified and summarized by type, as shown in [Figures 5I–L](#). Protein-ligand interactions was divided into four types. They were hydrogen bonds, hydrophobic interaction, ions, and water bridges. The binding of Casuarictin with S of Lambda strain and the binding of Protohypericin and Glansrin B with S-CTD of Delta strain were mainly connected with H-bonds. The binding of Heterophyllin D with S of Lambda strain displayed with hydrophobic interaction.

Combing the result of RMSD, RMSF, and Protein-ligand interactions, Casuarictin and Heterophyllin D bind to the S-CTD of the Lambda strain and showed reliable stability. Protohypericin and Glansrin B bind to S-CTD of Delta strain and displayed reliable stability.

3.5. Protein–protein docking analysis

ACE2 receptors of the host provide a binding site for the S protein of SARS-CoV-2, which indicates that the destruction of the binding between these two proteins provides the potentiality for COVID-19 treatment. To explore the effect of compound candidates during the recognition and fusion between S protein and ACE2, protein–protein docking was carried out. Binding Energy (kcal/mol) was used to indicate the stability between two proteins. The lower the score, the stronger the bond. As shown in [Table 7](#), the predicted binding energy of the Lambda strain and Delta strain were -349 kcal/mol and -355 kcal/mol, respectively, which were higher than the standard strain.

Lambda and Delta strains are more contagious than standard strains, as seen clinically, and reflected in the binding energy of protein–protein docking results mentioned above. However, when the S protein binds to small molecule compounds, such as Casuarictin, Heterophyllin D, Protohypericin, and Glansrin B, the binding energy between the S protein and ACE2 receptor was raised, and the stability was shaken. The binding energy between the S protein of Lambda-strain along with Casuarictin and Heterophyllin D were -329 kcal/mol and -344 kcal/mol, which were higher than -349 kcal/mol of S protein binding to ACE2 alone. The binding energy of Glansrin B, along with S protein and ACE2, was significantly raised to -143 kcal/mol from -355 kcal/mol and displayed the best potentiality. The interactions between S protein along with small molecule compounds and ACE2 are shown in [Figure 6](#).

4. Discussion

SARS-CoV-2 continues to infect, hospitalize and kill millions of people worldwide ([Drożdżal et al., 2021](#)). The virus infects human ACE2 through the binding of the virus RBD of its S protein ([Lan et al., 2020](#)). SARS-CoV-2 Delta and Lambda variants are the worldwide epidemic virus strains before the prevalence of the Omicron variant. They have named a variety of concerns (VOC) and a variety of interest (VOI), respectively ([Markarian et al., 2022](#)). In this study, we constructed a protein model of the CTD of the S protein of SARS-CoV-2 Delta and Lambda variants and used molecular docking and MD simulation methods to study the infection mediated by all plant chemicals in the Chinese herbal medicine ingredient database (TCMSP) and SARS-CoV-2. The virus RBD of S protein is the region where SARS-CoV-2 binds to the human ACE2 receptor. The mutation of the RBD region makes SARS-CoV-2 Delta and Lambda variants more beneficial to the combination with ACE2 and immune escape ([Arbeitman et al., 2021](#)). Our research showed that the stable binding of four compounds (Casuarictin, Heterophyllin D, Protohypericin, and Glansrin B) to the mutation sites of SARS-CoV-2 Delta and Lambda variants could weaken the stability of the binding of the RBD region of the virus to the receptor ACE2. Therefore, we speculate that these compounds (Casuarictin, Heterophyllin D, Protohypericin, and Glansrin B) can weaken the spread of SARS-CoV-2 in the population by acting on the RBD domain of the virus.

S protein is a potential fragment that can be used as an antigen in vaccine development ([Zhang et al., 2020](#)). This protein plays a crucial role in the first step of the infection process because it binds to the ACE2 receptor and then enters the host cell ([Petruk et al., 2020](#)). According to the mutation sites of SARS-CoV-2 Delta and Lambda mutants, we constructed reasonable CTD protein models of the S protein. Delta strain was formed by mutating L452R and T478K in the RBD region of

TABLE 6 The residues for the chemical bonds between compounds and spike proteins of Delta strain.

Compounds	Van de waals	Conventional hydrogen bond	Carbon hydrogen bond	Unfavorable donor-donor	Pi-Pi T-shaped	Pi-Pi Stacked
Glansrin B	Y453, Y495, F497, S494, R452, Q498	G496, N501	R403	Q493	Y505	Y449
Protohypericin (Conformation 1)	Q476, N487, C489, G485	P479, C488, K478, F486	–	S477	–	Y449
Protohypericin (Conformation 2)	Q493, F490	S494, R452	–	–	–	Y449

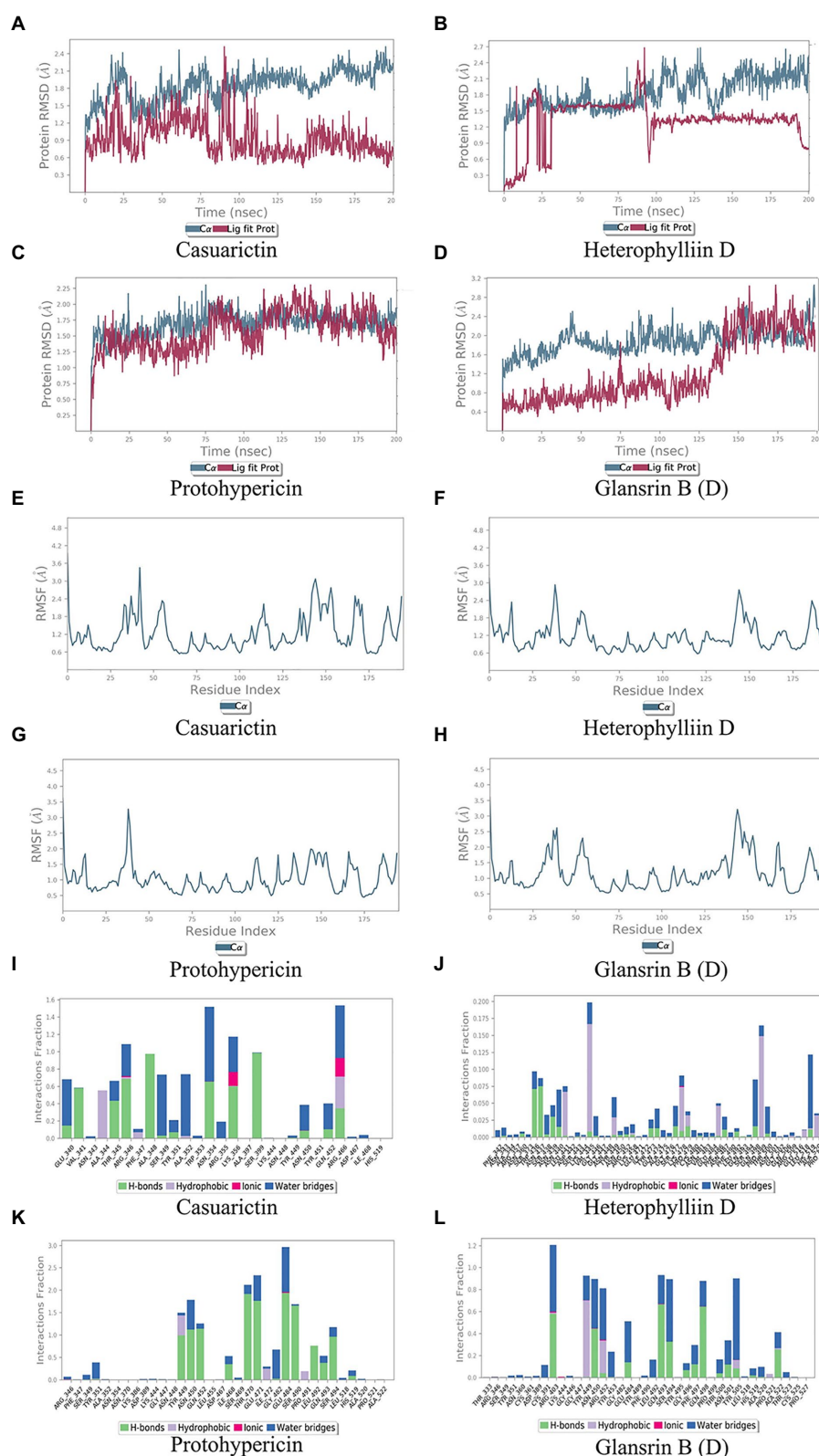


FIGURE 5

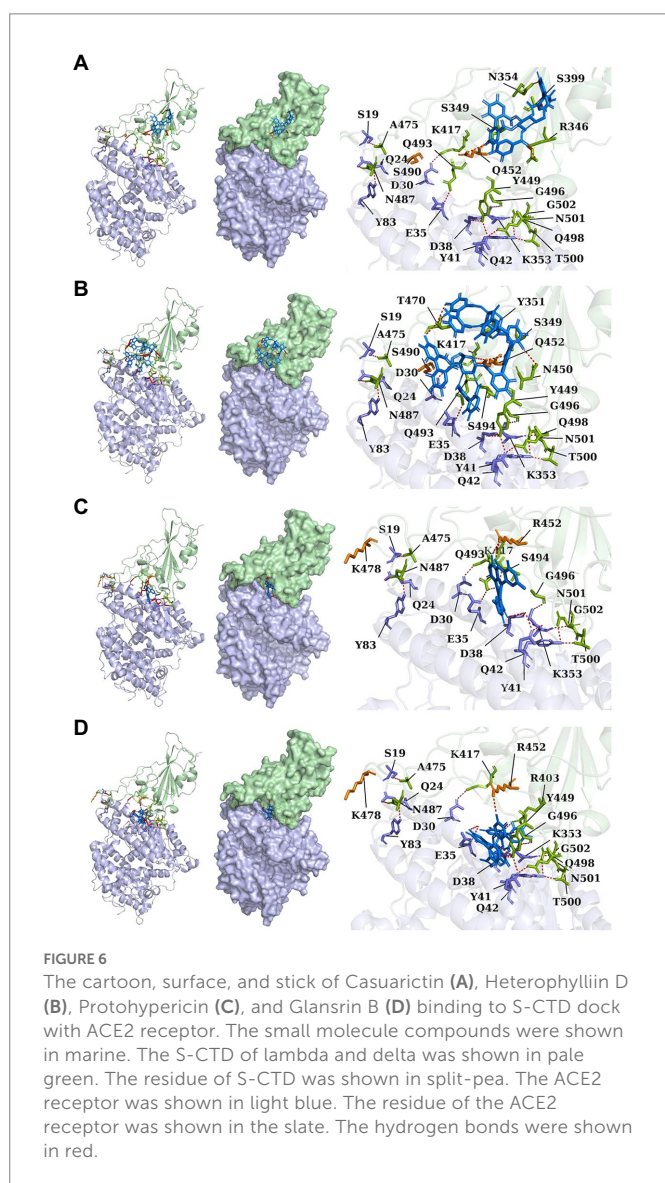
Molecular dynamics analysis of the top 2 conformations of small molecular compounds bind to mutation base. The results of protein RMSD, the RMSF of residue index and interactions fraction of Casuarictin (A, E, I), Heterophyllin D (B, F, J) Protohypericin (C, G, K) and Glansrin B (D, H, L) binds to the S-CTD of Lambda strain and Delta strain.

S protein, while the Lambda strain was formed by mutating L452Q and F490S in the RBD region (Hamill et al., 2022; Wang et al., 2022). In the constructed protein model, we studied the binding ability of 12,978

medicinal natural plant chemicals to SARS-CoV-2S protein. In this study, more than 65% of the compounds in the molecular docking of Lambda and Delta mutants showed a good binding ability to their

TABLE 7 The results of protein–protein docking.

	Binding Energy (kcal/mol)	The residue of ACE2	The residue of S protein
S protein of Standard strain with ACE2	−330	S19, Q24, D30, E35, D38, Y41, Q42, Y83, K353	K417, Y449, A475, N487, Q493, G496, Q498, T500, N501, G502
S protein of Lambda strain with ACE2	−349	S19, Q24, D30, E35, Q42, K353	A475, N487, Q493, K417, G496, G446, G502, Q498
Casuarictin binding to S protein of Lambda strain with ACE2	−329	S19, Q24, D30, E35, D38, Y41, Q42, Y83, K353	K417, Y449, A475, N487, Q493, G496, Q498, T500, N501, G502
Heterophyllin D binding to S protein of Lambda strain with ACE2	−344	S19, Q24, D30, E35, D38, Y41, Q42, Y83, K353	K417, Y449, A475, N487, Q493, G496, Q498, T500, N501
S protein of Delta strain with ACE2	−355	S19, Y83, K31, K353, N330	A475, N487, L492, Q493, G496, G446, G502, Q498, T500
Glansrin B binding to S protein of Delta strain with ACE2	−143	S19, Q24, D30, E35, D38, Y41, Q42, Y83, K353	K417, A475, N487, R403, Q393, R452, Y449, G496, Q498, T500, N501, G502
Protohypericin binding to S protein of Delta strain with ACE2	−340	S19, Q24, D30, E35, D38, Y41, Q42, Y83, K353	K417, R452, A475, N487, Q493, S494, G496, T500, N501, G502



respective RBD regions (docking score ≤ -6), which were 65.61% (8,515/12,978) and 65.28% (8,472/12,978) respectively (Figure 2), indicating that Chinese herbal medicine is a huge database with great potential to find anti-SARS-CoV-2 compounds. The compounds clematichinenoside AR2 (−9.7), atratoglucoside, b (−9.5), physalin b (−9.5), atratoglucoside, a (−9.4), Ochnaflavone (−9.3), cyclopseudo-hypericin (−9.3), periplocoside J (−9.3), Bryonolic acid (−9.3), solamargine (−9.3), and salaspermic acid (−9.2) showed good binding ability with Lambda mutants, while the compounds neo-przewaquinone a (−10), Wikstrosin (−9.7), xilingsaponin A (−9.6), ardisianoside G (−9.6), 23-epi-26-deoxyactein (−9.6), mulberrofuran K (−9.6), Timosaponin A III (−9.5), physalin b (−9.5), 26-O-β-D-glucopyranosyluatigenin-3-O-α-L-rhamnopyranosyl(1→2)-o-[β-D-glucopyranosyl(1→4)]-β-D-glucopyranoside (−9.5), and Ziebeimine (−9.4) showed good binding ability with Delta mutant (Tables 1, 2). Among the screened compounds with high scores, some have been reported to have antiviral activity, such as Hypericin, bryonolic acid, solamargine, and salaspermic acid, and some have anti-inflammatory effects, such as ochnaflavone, mulberrofuran K and timosaponin A-III. Hypericin is reported to have inhibition activity to α Coronavirus by targeting 3CL protease (Zhang et al., 2021). In addition, Hypericin can block the function of HSV-1 alkaline nuclease and inhibit virus replication (Cao et al., 2022). A study reported that bryonolic acid targeted the hotspot residues of SARS-CoV-2 main protease (Mpro and S protein), which has an essential role in mediating the viral replication therefore compounds targeting this key enzyme are expected to block the viral replication and transcription (Alagu Lakshmi et al., 2021). Chou et al. (2012) showed solamargine had potent activity against HBsAg, with an IC₅₀ of 1.57 microM. Salaspermic acid inhibited HIV reverse transcriptase and HIV replication in H9 lymphocyte cells (Chen et al., 1992). Ochnaflavone, a double flavonoid compound, has the activity of resisting *Escherichia coli*, *Staphylococcus aureus*, *Enterococcus faecalis*, and *Pseudomonas aeruginosa* (Makhafola et al., 2012). Mulberrofuran K showed anti-inflammatory activities in lipopolysaccharide (LPS)-stimulated murine macrophages by inhibiting transcriptional activation of nuclear factor-κB (NF-κB) and extracellular-regulated kinases (ERK; Shim et al., 2018). Oral administration of timosaponin A-III at 25–50 mg/

kg significantly inhibited the inflammatory markers in LPS-induced ALI mice, including the lung inflammatory index and the total number of inflammatory cells in the bronchoalveolar lavage fluid (BALF) (Park et al., 2018).

Among the top 100 compounds with good binding energy, we summarized the top 10 compounds that interact with mutant bases (Figures 3, 4, Tables 3, 4). All of these candidate ingredients showed a strong affinity for receptor molecular target sites with high binding energy. The receptor-ligand complex with the high score is stable through non-covalent interactions such as hydrogen bonding, van der Waals, and electrostatic interaction (Chen and Kurgan, 2009). These interactions are indeed prerequisites for biological functions and the success of drug development. Nine compounds (Casuarictin, C10230, Periplocoside J, Cyclopseudo-hypericin, Glucoside, d, Hypericin, β 1-solamargine, Mulberrofuran Q, and Atratoglucoside, b) interacted with one mutation site of Lambda strain, and compound Heterophyllin D interacted with two mutation sites (Figure 3). In addition, compounds Glansrin B, Ardisiacrispin A, Periplocoside J, Cyclopseudo-hypericin, Amentoflavone, Ochnaflavone, Hypericin, Protohypericin, Datuarmeloside B, and Primulanin interacted with a mutant base (L452R) of Delta strain, and compound Protohypericin also interacted with the mutant base T478K in another conformation (Figure 4). Casuarictin showed a robust binding activity with Lambda S-CTD by presenting a docking score of -9.1 and interacted with mutation residues (Q452) through conventional hydrogen bonding interaction (Figure 3A). In contrast, Heterophyllin D interacted with mutation residues (Q452) through conventional hydrogen bonding interaction and mutation residues (SERB490) through van der Waals bonding interaction (-8.1 ; Figure 3D). In addition, Glansrin B demonstrated vigorous binding activity with Delta S-CTD by exhibiting docking scores (-8.9), compared to Protohypericin, which presented a -8.2 -docking score with Delta S-CTD and interaction with amino acid residue (R452) via conventional hydrogen bond interaction (Figures 4A, I).

To further observe the stability of the compounds binding to the mutant sites of Lambda and Delta strains, MD analysis was carried out for Casuarictin, Heterophyllin D, Protohypericin, and Glansrin B. RMSD is a parameter to calculate the distance between protein atoms. The average distance between atoms in target proteins that are unbound and ligand/standard inhibitor bound allows us to evaluate the comparative conformation and stability of proteins (Gupta et al., 2021). Casuarictin and Heterophyllin D bind to the S-CTD of Lambda strain converged at the beginning of the RMSD simulation and remained stable in the subsequent simulations (Figures 5A, B). And Protohypericin and Glansrin B bind to S-CTD of Delta strain also display reliable stability (Figures 5C, D). RMSF is an important parameter used to evaluate the changes of protein atoms in the whole time period from the reference position. This allows us to study the comparison results of target protein fractions (residues) before and after ligand binding. S-CTD of Delta strain and Lambda strain still had low RMSF fluctuations within 3 \AA after Casuarictin and Heterophyllin D, or Protohypericin and Glansrin B binding, respectively, indicating that these proteins were low in flexibility and tightly bound to small molecules (Figures 5E–H). It is noteworthy that the compound casuarictin and the compound Protohypericin have been reported to have antiviral effects. Chandra et al. (2022) reported that casuarictin was identified to bind with the M Pro with the numerically lowest binding energies (-12.2 kcal/mol), with an alliance of five hydrogen bonds with amino acids T199 (3.02 \AA), D197 (3.79 \AA), R131 (2.76 \AA), K137 (2.46 \AA), and L287 (3.53 \AA) and three hydrophobic interactions with amino acids

L287 (5.30 \AA), Leu272 (5.12 \AA), and Y239 (5.20 \AA). Tamura et al. (2010) used the model virus disclosed casuarictin as the HCV invasion inhibitor, which showed 87.4% inhibition rate and 45.1% inhibition rate to E1E2 virus and G* virus with a concentration of $10 \mu\text{M}$, respectively. Casuarictin, which acts as a pure NF- κ B inhibitor, inhibited IL-8 secretion in TNF α -treated human gastric epithelial cells by dampening the NF- κ B signaling (Fumagalli et al., 2016). Protohypericin showed antiviral activity against a normal laboratory HCMV strain, AD-169, with an IC $_{50}$ of $5.7 \mu\text{M}$ (Barnard et al., 1992). In addition, Protohypericin showed an anti-influenza virus activity (3.1 ng/mL) by the HA-assay (Yasuda et al., 2010).

The SARS-CoV-2 S protein, transmembrane protease serine 2 (TMPRSS2), and human receptor ACE2 are the main determinants of host pathogens affecting infection. The amino acid mutation of S protein, TMPRSS2, and ACE2 binding sites changed the protein affinity, which may affect the structural stability of the complex (Beacon et al., 2021; AlGhamdi et al., 2022). Therefore, S glycoprotein and host cell receptor ACE2 are one of the drug targets of SARS-CoV-2. To verify the binding complex of compounds (Casuarictin, Heterophyllin D, Protohypericin, and Glansin B) with S protein will affect the binding ability with human ACE2 receptor, the protein-protein docking method was used to observe the binding stability between small molecule-S protein complex and ACE2 receptor. Binding Energy (kcal/mol) was used to indicate the stability between two proteins. The lower the score, the stronger the bond. Our research indicated that the binding of four compounds (Casuarictin, Heterophyllin D, Protohypericin, and Glansin B) with S protein decreased the binding stability between S protein and ACE2 receptor, with the binding energy of small molecule-S protein complex to ACE2 increases (Table 7). It was worth noting that compared with the Lambda strain (-349 kcal/mol) and Delta strain (-355 kcal/mol), the binding energy of the standard strain and ACE2 (-330 kcal/mol) was higher. This result showed that the combination of the Lambda strain and Delta strain with ACE2 was strengthened, which also explained the high transmissibility of variant strains in clinical practice. Therefore, from this perspective, the combination of compounds (Casuarictin, Heterophyllin D, Protohypericin, and Glansin B) and S protein can reduce the strong infectivity of Lambda and Delta mutants to a certain extent.

5. Conclusion

This study explored the biological activities of 12,978 small molecules in Chinese herbal medicine through molecular docking and MD simulation analysis to prevent and treat COVID-19 infection, especially Lambda and Delta variants. In the docking results, more than 65% of the compounds had a relatively stable binding ability with S proteins. Among them, compounds Casuarictin, Heterophyllin D, Protohypericin, and Glansin B showed possible antagonistic resistance to the mutation sites of Lambda and Delta mutants and had significant binding energy. The MD simulation verified that these four phytochemicals, as strong interaction compounds, stabilized with the minimum deviation from the interaction site within the observed total simulation time. In addition, protein-protein docking between the complexes of these plant chemicals and S proteins and ACE2 receptors was carried out to evaluate their binding stability. The results showed that they could reduce the binding ability of S proteins to ACE2. Therefore, these phytochemicals may be feasible candidate drugs

against SARS-CoV-2. However, extensive preclinical studies are needed to determine their effectiveness as antiviral agents.

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

YL conceived and designed the study. TH conducted the molecular docking and simulation and drafted the manuscript. ZL conducted data analysis and plotted the figures. LJ and PW conducted data analysis and performed literature searches. GL, XL, and YL reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the China Postdoctoral Science Foundation (grant no. 2021M700965), Guangdong Provincial Bureau of Traditional Chinese Medicine Research Foundation (grant no. 20231093), National Natural Science Foundation of China (grant no. 81973814), and Science, Technology and Innovation Commission of Shenzhen Municipality (grant no. JSGG20220226090550002).

References

- Alagu Lakshmi, S., Shafreen, R. M. B., Priya, A., and Shunmugiah, K. P. (2021). Ethnomedicines of Indian origin for combating COVID-19 infection by hampering the viral replication: using structure-based drug discovery approach. *J. Biomol. Struct. Dyn.* 39, 4594–4609. doi: 10.1080/07391102.2020.1778537
- AlGhamdi, N. A., Alsawat, H. S., Borgio, J. F., and AbdulAzeed, S. (2022). Emerging of composition variations of SARS-CoV-2 spike protein and human ACE2 contribute to the level of infection: in silico approaches. *J. Biomol. Struct. Dyn.* 40, 2635–2646. doi: 10.1080/07391102.2020.1841032
- Arbeitman, C. R., Rojas, P., Ojeda-May, P., and Garcia, M. E. (2021). The SARS-CoV-2 spike protein is vulnerable to moderate electric fields. *Nat. Commun.* 12:5407. doi: 10.1038/s41467-021-25478-7
- Balawender, K., Pliszka, A., Krowiak, A., Sito, M., Grabarek, B. O., and Boroń, D. (2022). Does SARS-CoV-2 affect male urogenital system? *Curr. Pharm. Biotechnol.* 23, 1792–1799. doi: 10.2174/1389201023666220307102147
- Bangash, M. N., Patel, J., and Parekh, D. (2020). COVID-19 and the liver: little cause for concern. *Lancet Gastroenterol. Hepatol.* 5, 529–530. doi: 10.1016/S2468-1253(20)30084-4
- Barnard, D. L., Huffman, J. H., Morris, J. L. B., Wood, S. G., Hughes, B. G., and Sidwell, R. W. (1992). Evaluation of the antiviral activity of anthraquinones, anthrones and anthraquinone derivatives against human cytomegalovirus. *Antivir. Res.* 17, 63–77. doi: 10.1016/0166-3542(92)90091-I
- Beacon, T. H., Delcuve, G. P., and Davie, J. R. (2021). Epigenetic regulation of ACE2, the receptor of the SARS-CoV-2 virus(1). *Genome* 64, 386–399. doi: 10.1139/gen-2020-0124
- Cao, K., Zhang, Y., Yao, Q., Peng, Y., Pan, Q., Jiao, Q., et al. (2022). Hypericin blocks the function of HSV-1 alkaline nuclease and suppresses viral replication. *J. Ethnopharmacol.* 296:115524. doi: 10.1016/j.jep.2022.115524
- Chandra, M. A., Malaisamy, A., Eswaran, M., Meyyazhagan, A., Arumugam, V. A., Rengasamy, K. R. R., et al. (2022). Evaluation of clove phytochemicals as potential antiviral drug candidates targeting SARS-CoV-2 Main protease: computational docking, molecular dynamics simulation, and pharmacokinetic profiling. *Front. Mol. Biosci.* 9:918101. doi: 10.3389/fmolb.2022.918101
- Chen, K., and Kurgan, L. (2009). Investigation of atomic level patterns in protein-small ligand interactions. *PLoS One* 4:e4473. doi: 10.1371/journal.pone.0004473
- Chen, K., Shi, Q., Kashiwada, Y., Zhang, D. C., Hu, C. Q., Jin, J. Q., et al. (1992). Anti-aids agents, 6. Salaspermic acid, an anti-HIV principle from *Tripterygium wilfordii*, and the structure-activity correlation with its related compounds. *J. Nat. Prod.* 55, 340–346. doi: 10.1021/np50081a010
- Cheng, F., Hong, H., Yang, S., and Wei, Y. (2017). Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Brief. Bioinform.* 18, 682–697. doi: 10.1093/bib/bbw051
- Chou, S. C., Huang, T. J., Lin, E. H., Huang, C. H., and Chou, C. H. (2012). Antihepatitis B virus constituents of *Solanum elaeagnifolium*. *Nat. Prod. Commun.* 7, 153–156. doi: 10.1177/1934578X1200700205
- Chuang, D. T., Aydemir, S., Magda, P., Thomas, C., and Zarnegar, R. (2021). Neurological manifestations as primary presentation of COVID-19 in hospitalized patients. *Acta Neurol. Scand.* 143, 569–574. doi: 10.1111/ane.13399
- Drożdżal, S., Rosik, J., Lechowicz, K., Machaj, F., Szostak, B., Przybyciński, J., et al. (2021). An update on drugs with therapeutic potential for SARS-CoV-2 (COVID-19) treatment. *Drug Resist. Updat.* 59:100794. doi: 10.1016/j.drup.2021.100794
- Fan, C., Lu, W., Li, K., Ding, Y., and Wang, J. (2020). ACE2 expression in kidney and testis may cause kidney and testis infection in COVID-19 patients. *Front. Med. (Lausanne)* 7:563893. doi: 10.3389/fmed.2020.563893
- Fumagalli, M., Sangiovanni, E., Vrhovsek, U., Piazza, S., Colombo, E., Gasperotti, M., et al. (2016). Strawberry tannins inhibit IL-8 secretion in a cell model of gastric inflammation. *Pharmacol. Res.* 111, 703–712. doi: 10.1016/j.phrs.2016.07.028
- Gupta, S., Singh, A. K., Kushwaha, P. P., Prajapati, K. S., Shuaib, M., Senapati, S., et al. (2021). Identification of potential natural inhibitors of SARS-CoV2 main protease by molecular docking and simulation studies. *J. Biomol. Struct. Dyn.* 39, 4334–4345. doi: 10.1080/07391102.2020.1776157
- Hamill, V., Noll, L., Lu, N., Tsui, W. N. T., Porter, E. P., Gray, M., et al. (2022). Molecular detection of SARS-CoV-2 strains and differentiation of Delta variant strains. *Transbound. Emerg. Dis.* 69, 2879–2889. doi: 10.1111/tbed.14443
- Hanson, S. W., Abbafati, C., Aerts, J. G., Al-Aly, Z., Ashbaugh, C., Ballouz, T., et al. (2022). Estimated global proportions of individuals with persistent fatigue, cognitive, and respiratory symptom clusters following symptomatic COVID-19 in 2020 and 2021. *JAMA* 328, 1604–1615. doi: 10.1001/jama.2022.18931

Acknowledgments

The authors thank the outstanding researchers for their generous and selfless disclosure of the online database involved above in our study.

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.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1095068/full#supplementary-material>

- Kimura, I., Kosugi, Y., Wu, J., Zahradnik, J., Yamasoba, D., Butlertanaka, E. P., et al. (2022). The SARS-CoV-2 lambda variant exhibits enhanced infectivity and immune resistance. *Cell Rep.* 38:110218. doi: 10.1016/j.celrep.2021.110218
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5
- Madjid, M., Safavi-Naeini, P., Solomon, S. D., and Vardeny, O. (2020). Potential effects of coronaviruses on the cardiovascular system: a review. *JAMA Cardiol.* 5, 831–840. doi: 10.1001/jamacardio.2020.1286
- Makhafola, T. J., Samuel, B. B., Elgorashi, E. E., and Eloff, J. N. (2012). Ochnaflavone and ochnaflavone 7-O-methyl ether two antibacterial biflavonoids from *Ochna pretoriensis* (Ochnaceae). *Nat. Prod. Commun.* 7, 1601–1604. doi: 10.1177/1934578X1200701216
- Markarian, N. M., Galli, G., Patel, D., Hemmings, M., Nagpal, P., Berghuis, A. M., et al. (2022). Identifying markers of emerging SARS-CoV-2 variants in patients with secondary immunodeficiency. *Front. Microbiol.* 13:933983. doi: 10.3389/fmicb.2022.933983
- Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins* 12, 345–364. doi: 10.1002/prot.340120407
- Nyberg, T., Ferguson, N. M., Nash, S. G., Webster, H. H., Flaxman, S., Andrews, N., et al. (2022). Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. *Lancet* 399, 1303–1312. doi: 10.1016/S0140-6736(22)00462-7
- Park, B. K., So, K. S., Ko, H. J., Kim, H. J., Kwon, K. S., Kwon, Y. S., et al. (2018). Therapeutic potential of the rhizomes of *Anemarrhena asphodeloides* and timosaponin A-III in an animal model of lipopolysaccharide-induced lung inflammation. *Biomol. Ther. (Seoul)* 26, 553–559. doi: 10.4062/biomolther.2017.249
- Petrak, G., Puthia, M., Petrlova, J., Samsudin, F., Strömdahl, A. C., Cerps, S., et al. (2020). SARS-CoV-2 spike protein binds to bacterial lipopolysaccharide and boosts proinflammatory activity. *J. Mol. Cell Biol.* 12, 916–932. doi: 10.1093/jmcb/mjaa067
- Pulliam, J. R. C., van Schalkwyk, C., Govender, N., von Gottberg, A., Cohen, C., Groome, M. J., et al. (2022). Increased risk of SARS-CoV-2 reinfection associated with emergence of omicron in South Africa. *Science* 376:eabn4947. doi: 10.1126/science.abn4947
- Ren, J. L., Zhang, A. H., and Wang, X. J. (2020). Traditional Chinese medicine for COVID-19 treatment. *Pharmacol. Res.* 155:104743. doi: 10.1016/j.phrs.2020.104743
- Robson, F., Khan, K. S., le, T. K., Paris, C., Demirbag, S., Barfuss, P., et al. (2020). Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol. Cell* 79, 710–727. doi: 10.1016/j.molcel.2020.07.027
- Romero, P. E., Dávila-Barclay, A., Salvatierra, G., González, L., Cuicapuza, D., Solís, L., et al. (2021). The emergence of Sars-CoV-2 variant lambda (C.37) in South America. *Microbiol. Spectr.* 9:e00789-21. doi: 10.1128/Spectrum.00789-21
- Runfeng, L., Yunlong, H., Jicheng, H., Weiqi, P., Qin Hai, M., Yongxia, S., et al. (2020). Lianhuaqingwen exerts antiviral and anti-inflammatory activity against novel coronavirus (SARS-CoV-2). *Pharmacol. Res.* 156:104761. doi: 10.1016/j.phrs.2020.104761
- Shim, S. Y., Sung, S. H., and Lee, M. (2018). Anti-inflammatory activity of mulberrofuran K isolated from the bark of *Morus bombycis*. *Int. Immunopharmacol.* 58, 117–124. doi: 10.1016/j.intimp.2017.11.002
- Tamura, S., Yang, G. M., Yasueda, N., Matsuura, Y., Komoda, Y., and Murakami, N. (2010). Tellimagrandin I, HCV invasion inhibitor from *Rosae Rugosae* Flos. *Bioorg. Med. Chem. Lett.* 20, 1598–1600. doi: 10.1016/j.bmcl.2010.01.084
- Tsang, H. F., Chan, L. W. C., Cho, W. C. S., Yu, A. C. S., Yim, A. K. Y., Chan, A. K. C., et al. (2021). An update on COVID-19 pandemic: the epidemiology, pathogenesis, prevention and treatment strategies. *Expert Rev. Anti-Infect. Ther.* 19, 877–888. doi: 10.1080/14787210.2021.1863146
- 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. *Cells* 181, 281–292.e6. doi: 10.1016/j.cell.2020.02.058
- Wang, Y., Chen, R., Hu, F., Lan, Y., Yang, Z., Zhan, C., et al. (2021). Transmission, viral kinetics and clinical characteristics of the emergent SARS-CoV-2 Delta VOC in Guangzhou, China. *EClinicalMedicine* 40:101129. doi: 10.1016/j.eclinm.2021.101129
- Wang, M., Zhang, L., Li, Q., Wang, B., Liang, Z., Sun, Y., et al. (2022). Reduced sensitivity of the SARS-CoV-2 lambda variant to monoclonal antibodies and neutralizing antibodies induced by infection and vaccination. *Emerg. Microbes Infect.* 11, 18–29. doi: 10.1080/22221751.2021.2008775
- Xie, Y., Xu, E., Bowe, B., and al-Aly, Z. (2022). Long-term cardiovascular outcomes of COVID-19. *Nat. Med.* 28, 583–590. doi: 10.1038/s41591-022-01689-3
- Yasuda, T., Yamaki, M., Iimura, A., Shimotai, Y., Shimizu, K., Noshita, T., et al. (2010). Anti-influenza virus principles from *Muehlenbeckia hastulata*. *J. Nat. Med.* 64, 206–211. doi: 10.1007/s11418-009-0386-9
- Zhang, Y., Chen, H., Zou, M., Oerlemans, R., Shao, C., Ren, Y., et al. (2021). Hypericin inhibit alpha-coronavirus replication by targeting 3CL protease. *Viruses* 13:1825. doi: 10.3390/v13091825
- Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., et al. (2020). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* 11:6013. doi: 10.1038/s41467-020-19808-4



OPEN ACCESS

EDITED BY

Arli Aditya Parikesit,
Indonesia International Institute for Life-
Sciences (i3L), Indonesia

REVIEWED BY

Anand Anbarasu,
VIT University, India
Surjit Singh,
Sister Nivedita University, India

*CORRESPONDENCE

Muhammad Tahir ul Qamar
✉ tahirulqamar@gcuf.edu.pk

SPECIALTY SECTION

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

RECEIVED 28 February 2023

ACCEPTED 29 March 2023

PUBLISHED 05 May 2023

CITATION

Noor F, Ashfaq UA, Bakar A, ul Haq W,
Allemailem KS, Alharbi BF, Al-Megrin WAI and
Tahir ul Qamar M (2023) Discovering common
pathogenic processes between COVID-19 and
HFRS by integrating RNA-seq differential
expression analysis with machine learning.
Front. Microbiol. 14:1175844.
doi: 10.3389/fmicb.2023.1175844

COPYRIGHT

© 2023 Noor, Ashfaq, Bakar, ul Haq,
Allemailem, Alharbi, Al-Megrin and Tahir ul
Qamar. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Discovering common pathogenic processes between COVID-19 and HFRS by integrating RNA-seq differential expression analysis with machine learning

Fatima Noor¹, Usman Ali Ashfaq¹, Abu Bakar², Waqar ul Haq²,
Khaled S. Allemailem³, Basmah F. Alharbi⁴,
Wafa Abdullah I. Al-Megrin⁵ and Muhammad Tahir ul Qamar^{1*}

¹Integrative Omics and Molecular Modeling Laboratory, Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan, ²Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad, Pakistan, ³Department of Medical Laboratories, College of Applied Medical Sciences, Qassim University, Buraydah, Saudi Arabia, ⁴Department of Basic Health Science, College of Applied Medical Sciences, Qassim University, Buraydah, Saudi Arabia, ⁵Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

Zoonotic virus spillover in human hosts including outbreaks of Hantavirus and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) imposes a serious impact on the quality of life of patients. Recent studies provide a shred of evidence that patients with Hantavirus-caused hemorrhagic fever with renal syndrome (HFRS) are at risk of contracting SARS-CoV-2. Both RNA viruses shared a higher degree of clinical features similarity including dry cough, high fever, shortness of breath, and certain reported cases with multiple organ failure. However, there is currently no validated treatment option to tackle this global concern. This study is attributed to the identification of common genes and perturbed pathways by combining differential expression analysis with bioinformatics and machine learning approaches. Initially, the transcriptomic data of hantavirus-infected peripheral blood mononuclear cells (PBMCs) and SARS-CoV-2 infected PBMCs were analyzed through differential gene expression analysis for identification of common differentially expressed genes (DEGs). The functional annotation by enrichment analysis of common genes demonstrated immune and inflammatory response biological processes enriched by DEGs. The protein-protein interaction (PPI) network of DEGs was then constructed and six genes named RAD51, ALDH1A1, UBA52, CUL3, GADD45B, and CDKN1A were identified as the commonly dysregulated hub genes among HFRS and COVID-19. Later, the classification performance of these hub genes were evaluated using Random Forest (RF), Poisson Linear Discriminant Analysis (PLDA), Voom-based Nearest Shrunken Centroids (voomNSC), and Support Vector Machine (SVM) classifiers which demonstrated accuracy >70%, suggesting the biomarker potential of the hub genes. To our knowledge, this is the first study that unveiled biological processes and pathways commonly dysregulated in HFRS and COVID-19, which could be in the next future used for the design of personalized treatment to prevent the linked attacks of COVID-19 and HFRS.

KEYWORDS

hemorrhagic fever with renal syndrome, COVID-19, transcriptomic data, hub genes, genes ontologies, pathways, machine learning

1. Introduction

Hemorrhagic fever with renal syndrome (HFRS) is a major rodent-borne zoonosis caused by different species of hantaviruses including Hantaan virus (HTNV) (Lee et al., 1978), Puumala virus (PUUV) (Sironen et al., 2001), Seoul virus (SEOV) (Lee et al., 1982), and Dobrava-Belgrade virus (DOBV) (Papa, 2012). HFRS is primarily transmitted to humans through contact with rodents, excretions, or their saliva (Noor et al., 2022; Tariq and Kim, 2022). The disease can cause fever, hemorrhage, and kidney failure, and can be fatal in severe cases (Krautkrämer et al., 2013; Garanina et al., 2019). HFRS is a significant public health concern in some parts of the world, particularly in Asia and Europe. Most importantly, some latest studies suggested that individuals with HFRS are at increased risk of severe COVID-19 infection (Noor, 2020). COVID-19 is a highly infectious respiratory disease caused by the SARS-CoV-2 virus that was first identified in December 2019 (Li et al., 2020). This disease has spread rapidly around the world and has caused a global pandemic (Zhu and Cai, 2020). The symptoms of COVID-19 range from mild to severe, and the disease can be fatal, particularly in vulnerable populations such as the elderly and those with pre-existing health conditions (Flaherty et al., 2020). Additionally, the severity of COVID-19 appears to be influenced by various factors, including age, comorbidities, and viral load (Hasanoglu et al., 2021).

Recent studies reported that patients with HFRS who also had COVID-19 had more severe symptoms and longer hospital stays than those with HFRS alone (Singh et al., 2020; Subramaniam et al., 2022). Their study also reported a higher mortality rate among HFRS patients with COVID-19. Geladari et al. (2022) present a case study on patient with dialysis-dependent acute kidney injury due to hantavirus complicated with SARS-CoV-2 infection. Further, Cetin and Sahin (2021) presents a case followed up with the differential diagnosis of COVID-19 during the pandemic and diagnosed with HFRS due to hantavirus. To sum up, different case studies are reported in the literature, but, the absence of proper diagnostic tests hinders the diagnosis of co-infection among infected individuals. Overall, the evidence suggests a strong link between HFRS and COVID-19 comorbidity, particularly in terms of increased severity and mortality. However, the underlying mechanisms behind the increased severity of COVID-19 in individuals with HFRS are not yet fully understood. It is believed that the immune response to hantavirus infection may predispose individuals to a dysregulated immune response to SARS-CoV-2, leading to more severe illness (Wan et al., 2021). Furthermore, this co-infection leads to a more severe disease course, as both viruses can cause respiratory and renal failure, and ultimately their co-infection led to diagnostic challenges, as the symptoms of both diseases can overlap. The bell is ringing slightly thus, it is the need of the hour to develop effective diagnostic tools that can differentiate between single and co-infection and provide effective management for patients with COVID-19 and HFRS co-infection.

The spread of viral infections can be controlled through early detection of co-infections. However, current detection methods such as real-time polymerase chain reaction (RT-PCR) are not only time-consuming but also suffer from limited sensitivity and specificity (Bustin, 2000; Auwul et al., 2021). Moreover, the success of these techniques is highly reliant on skilled manpower, appropriate sample collection, and preparation, all of which pose significant challenges, particularly in developing countries. As a result, there is a pressing

need for more efficient and reliable detection methods that can be easily deployed and implemented in resource-limited settings.

The transcriptomic analysis of COVID-19 and HFRS PBMCs can provide valuable insights in understanding the molecular mechanisms underlying the co-infection as well as the identification of potential biomarkers which could ultimately serve as potential targets for developing a single treatment strategy that could tackle both diseases simultaneously. Other studies have been carried out assessing other potential comorbidities with respect to COVID-19 including chronic kidney disease and diabetes mellitus (Auwul et al., 2021; Rahman et al., 2021). Thus, sensing these opportunities, this study combined transcriptomics and bioinformatics approaches for the identification of common genes and shared pathways among HFRS and COVID-19. Initially, the Differentially Expressed Genes (DEGs) were identified through transcriptomic analysis. The functional enrichment analysis was then performed to analyze the shared pathways for elucidating the immune response to the co-infection, which can lead to a better understanding of disease pathogenesis and potential targets for intervention. Later, a Protein-Protein interaction (PPI) network was constructed for the identification of hub genes from the pool of DEGs. However, the validity of these hub genes needs to be confirmed through rigorous validation processes. Thus, we employed supervised machine learning methods including Random Forest (RF), Poisson Linear Discriminant Analysis (PLDA), Voom-based Nearest Shrunken Centroids (voomNSC), and Support Vector Machine (SVM) to determine the validity of these hub genes. Here, for the first time, we have characterized the biological processes and pathways commonly dysregulated in COVID-19 and HFRS, which could be in the next future used in the designing personalized treatment of COVID-19 patients suffering from HFRS as comorbidity.

2. Materials and methods

2.1. Transcriptomic data acquisition

The collection of disease-related datasets is considered a preliminary step in the RNA-seq data analysis pipeline. The PBMCs transcriptomic datasets of SARS-CoV-2 (COVID-19) and HFRS-causing hantaviruses were collected from Gene Expression Omnibus (GEO) (Clough and Barrett, 2016), a public repository of functional high-throughput experimental data obtained through next-generation sequencing and microarrays. The criteria for disease-related dataset selection were completely based on the fact that all datasets must be from *Homo sapiens*. The dataset must contain transcriptomic data and the transcriptomic data contain no drug treatment. Two gene expression raw counts datasets of COVID-19 were retrieved through accession numbers; GSE160351 and GSE152418. The GSE160351 dataset was submitted by Brunetta et al. (2021) containing a total of 9 samples (three healthy controls and six infected individuals). While GSE152418 dataset was deposited by Arunachalam et al. (2020) comprising of total 34 samples (17 infected samples and 17 healthy samples). On the other hand, the other hantavirus-related dataset were obtained with accession number GSE158712 (Li et al., 2020) which consists of total of 30 samples (3 control and 27 infected samples) (Table 1). The raw sequence data of selected datasets were retrieved from the NCBI SRA toolkit using the prefetch command. The data was downloaded in .sra file which is not an acceptable format for different

TABLE 1 Characteristics of datasets.

Disease	Acc. no#	Source	Total runs	Mock infected/ control	Infected	Instrument
HFRS	GSE158712	PBMCs	30	3	27	Bgiseq-500 18
COVID-19	GSE160351	PBMCs	9	3	6	NextSeq 550
	GSE152418	PBMCs	33	17	16	Illumina NovaSeq 6,000

tools, therefore fastq-dump command was used for converting .sra data to .fastq format.

2.2. Data pre-processing

Depending on the sequencing technology, different strategies are used for processing and analyzing the raw sequencing data. Pre-processing the raw sequences data, such as performing quality control to check the read length, presence of any overrepresented sequences (k-mers), average quality score at each sequenced base, and percentage of GC content is now the most time-consuming step in RNA-seq data analysis. Firstly, the raw binary SRA data was turned into sequencing data. Later, the sequence data quality of each sample was controlled by FastQC (Brown et al., 2017). Further, to reduce the noise level, the obtained sequences were trimmed out for low-quality reads and adaptors by applying Sickle (Criscuolo and Brisse, 2013) Trimmomatic (Sewu et al., 2022), and FASTp (Sewu et al., 2022) tools on raw reads (Chen et al., 2018). Trimming of adapter sequences from raw reads was then performed using Sickle, Trimmomatic, and FASTp for identification of overlap adapters among forward and reverse reads. After trimming, the samples were prepared for further analysis.

2.3. Screening for differentially expressed transcripts and genes

After pre-processing, the raw reads were aligned with the reference genome to figure out which gene a read came from. Mapping and assembly of high-quality reads with reference genomes were performed with “New Tuxedo Suit” (HISAT2/StringTie) (Pertea et al., 2016), using default parameters. HISAT2 is a fast and accurate aligner for RNA-seq reads to a reference genome, and it uses a graph-based approach to account for splice junctions and other complex features in the genome. StringTie, on the other hand, is used for transcript assembly and quantification of gene expression. It takes the aligned RNA-seq reads produced by HISAT2 and assembles them into transcripts, estimates their abundances, and generates a file of gene expression values. Initially, indexing of the “*Homo sapiens*” reference genome and alignment of reads to the “*Homo sapiens*” reference genome was done using HISAT2. The aligned reads were then taken and used for the transcript assembly using the StringTie Tool. After assemblies, samples’ GTF (General Transfer Format) documents and Homo_sapiens.GRCh38.109.gtf. Were merged with the StringTie -merge option and transcript abundances of each sample were estimated with the StringTie-eB option (Goksuluk et al., 2019). Transcripts with variance across samples less than one were removed and then differentially expressed transcripts and genes between healthy individuals and infected individuals were screened using the

stattest function from Ballgown [Version 2.12.0 (Weinstein et al., 2019)] with the getFC=TRUE parameter. The batch effect of two sources of transcriptome data was considered during our analysis. Transcripts and genes with a fold change >1.0 and value of $p < 0.05$ were identified as differentially expressed transcripts and genes. All DEGs, including the genes corresponding to the differentially expressed transcripts and differentially expressed genes screened out by Ballgown, were used for subsequent steps.

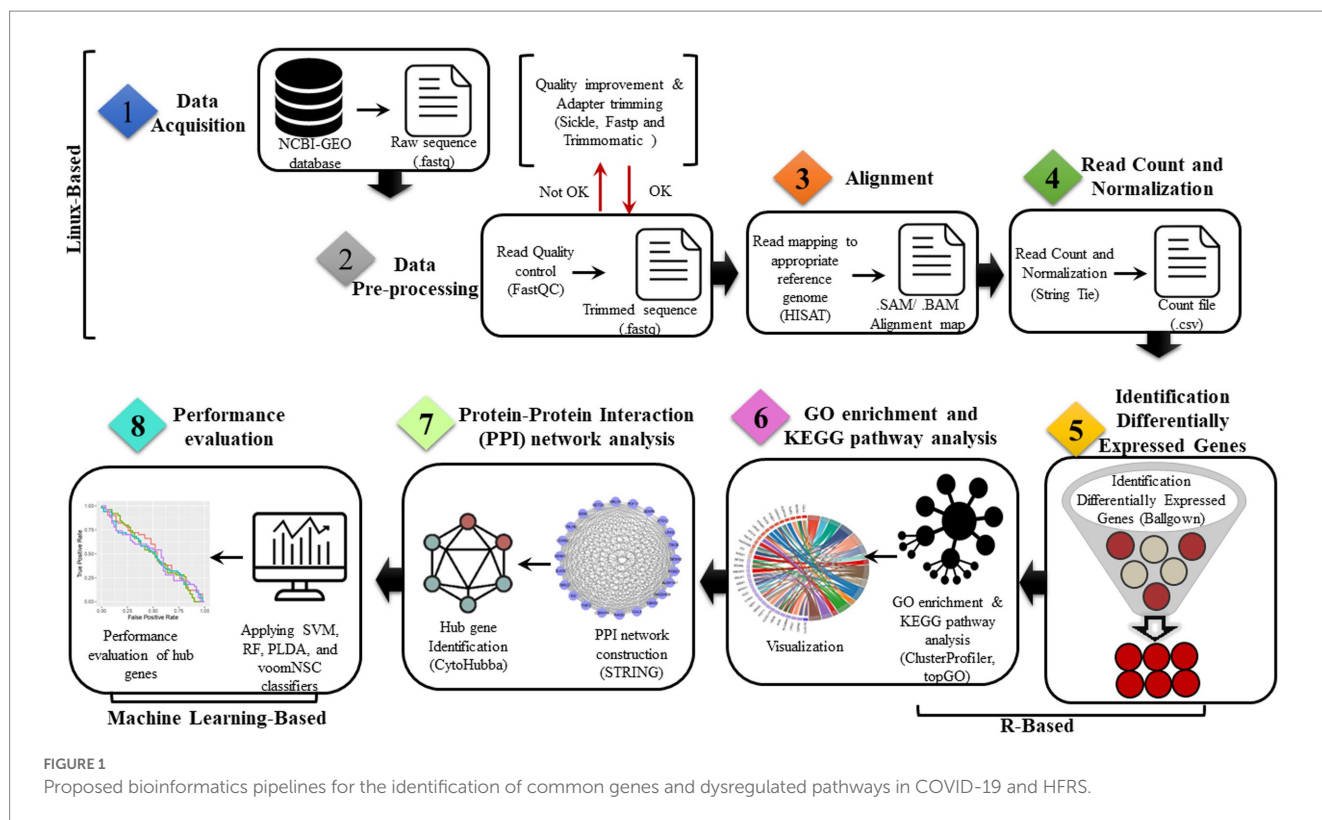
2.4. Identification of common transcriptional signatures and pathways between COVID-19 and HFRS PBMCs

Our study mainly aims to identify common transcriptional signatures, regulators, and pathways between COVID-19 and HFRS. At first, all the DEGs related to COVID-19 and HFRS were obtained and imported into a venn diagram tool¹ to predict overlapped genes for uncovering their common pathogenic processes. Eventually, a group of mutual genes was acquired and considered for further analysis. The overlapped genes were then subjected to ClusterProfiler (Yu et al., 2012) and TopGO (Alexa and Rahnenführer, 2009) packages of R for the identification of common pathways shared among COVID-19 and HFRS. To identify significantly enriched GO terms and KEGG pathways in the analysis, a statistical threshold criterion was applied, with an adjusted p -value of less than 0.05. This criterion was used to select the most relevant and statistically significant GO terms and KEGG pathways. In short, the characterization of common transcriptional signatures, biological processes, molecular function, and pathways dysregulated in COVID-19 and HFRS, could be in the next future used for the design of personalized treatment of COVID-19 patients suffering from HFRS as comorbidity.

2.5. PPI network construction and identification of hub genes

Protein–Protein Interaction (PPI) networks are remarkably significant due to their high versatility, adaptability, and specificity. The functional interactions among overlapped genes with a combined score of more than 0.4 were determined using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (von Mering et al., 2003). Initially, the overlapped genes between COVID-19 and HFRS were then submitted to the STRING database for the construction of the PPI network. The resulting PPI network

1 <http://bioinformatics.psb.ugent.be/webtools/Venn/>



was then subjected to Cytoscape version 3.8 (Shannon et al., 2003) for the identification of hub genes. Hub genes are the highly connected nodes that have a large number of interactions with other genes or proteins. These genes are considered critical components of the PPI network as they play important roles in maintaining the integrity and stability of the network, and they are often associated with key biological processes and pathways. Hub genes are identified by analyzing the topology of the PPI network, using measures such as degree, betweenness, or closeness centrality. In current study, degree methods available in CytoHubba was used for the identification of hub genes.

2.6. Performance evaluation of the hub genes with classification algorithms

To assess the credibility of the identified hub genes, five commonly used classification algorithms—support vector machine (SVM) (Boser et al., 1992) with radial basis kernel function, random forest (RF) (Ho, 1995), Poisson linear discriminant analysis (PLDA) (Witten, 2011), and Voom-based Nearest Shrunken Centroids (voomNSC) (Goksuluk et al., 2019) were employed. Datasets were split into two subsets, training data (70%) and testing data (30%). Then, the classification algorithms were conducted through the MLSeq package in R. The evaluation of model performance is typically subjective and based on a comparison of the model's predictions with the known values of the dependent variable in a given dataset. However, for our study, we have defined the ideal model performance as having metrics' results within the range of 70–90%. On the other hand, a performance exceeding 90–100% may indicate the possibility of overfitting. The DESeq

normalization and VST transformation methods were applied to the count dataset for the SVM and RF classifiers. The HFRS dataset GSE158712 was utilized for the classification analysis, in which the hub genes were incorporated. The classification performance was evaluated using four metrics: accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, and specificity based on the data with hub genes. An overview of the present study is shown in Figure 1.

3. Results

3.1. Screening of DEGs among COVID-19 and HFRS

We obtained two PBMCs gene expression datasets from COVID-19 infected subjects and matched healthy controls (with accession numbers, GSE160351 and GSE152418) for a total of 42 samples, 22 infected individuals and 20 healthy controls. The RNA-seq data analysis of GSE160351 and GSE152418 datasets yielded 1734 DEGs (1,108 upregulated and 626 downregulated). Screening criteria for identification of DEGs were set as value of $p < 0.05$ and $\text{LogFC} > 1.0$. For GSE158712, the selection criteria for DEGs screening were set as value of $p < 0.05$ and $\text{LogFC} > 1.0$. After the screening, 630 DEGs (390 downregulated and 240 upregulated) were identified from GSE158712. The volcano plot of DEGs obtained from GSE160351, GSE152418, and GSE158712 was shown in Figure 2 which provides a pictorial representation of upregulated DEGs, downregulated DEGs, and non-significant genes obtained from each dataset.

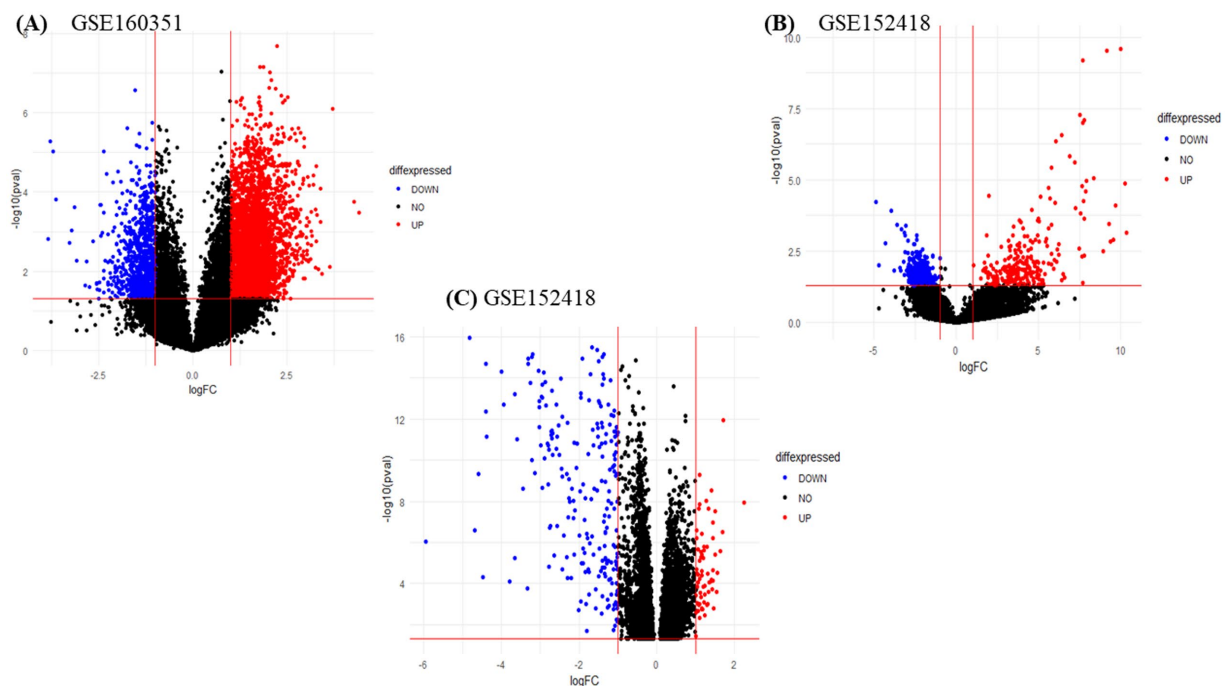


FIGURE 2
Volcano plot (A) GSE160351, (B) GSE152418, and (C) GSE152418.

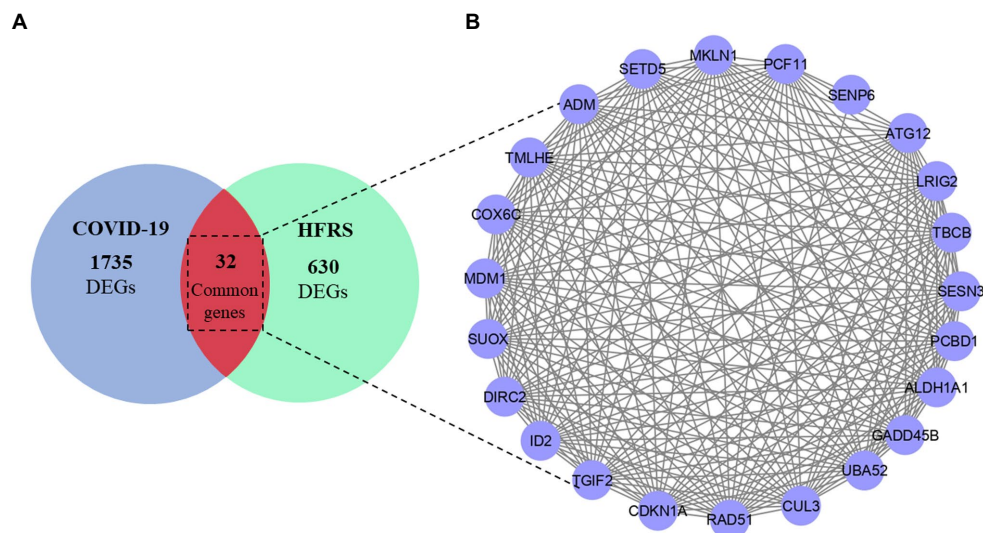


FIGURE 3
(A) Overlapped genes between COVID-19 and HFRS. (B) PPI network of 32 common dysregulated genes among COVID-19 and HFRS.

3.2. Identification of common transcriptional signatures between COVID-19 and HFRS PBMCs

After DEGs identification, a Venn diagram was constructed which indicated 32 common genes between COVID-19 and HFRS. Among them, 17 genes were commonly upregulated and 15

DEGs were commonly downregulated in COVID-19 and HFRS. Overall, the comparative analysis of the transcriptional signatures characterizing COVID-19 and HFRS PBMCs suggests the presence of commonly dysregulated genes between COVID-19 and HFRS (Figure 3 and Supplementary Table S1). After the identification of common genes, a PPI network was constructed for analyzing interaction among 32 commonly dysregulated proteins.

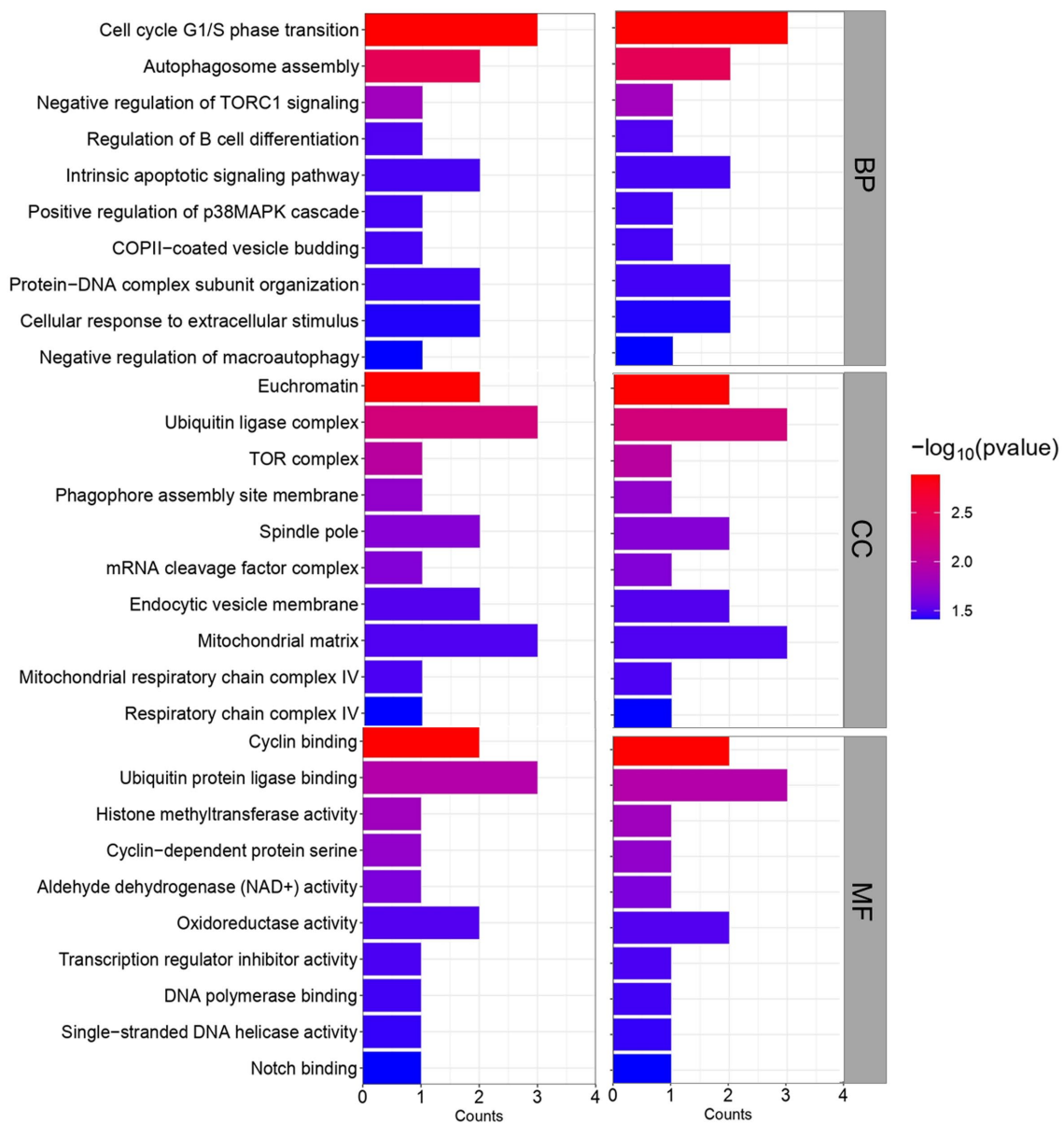


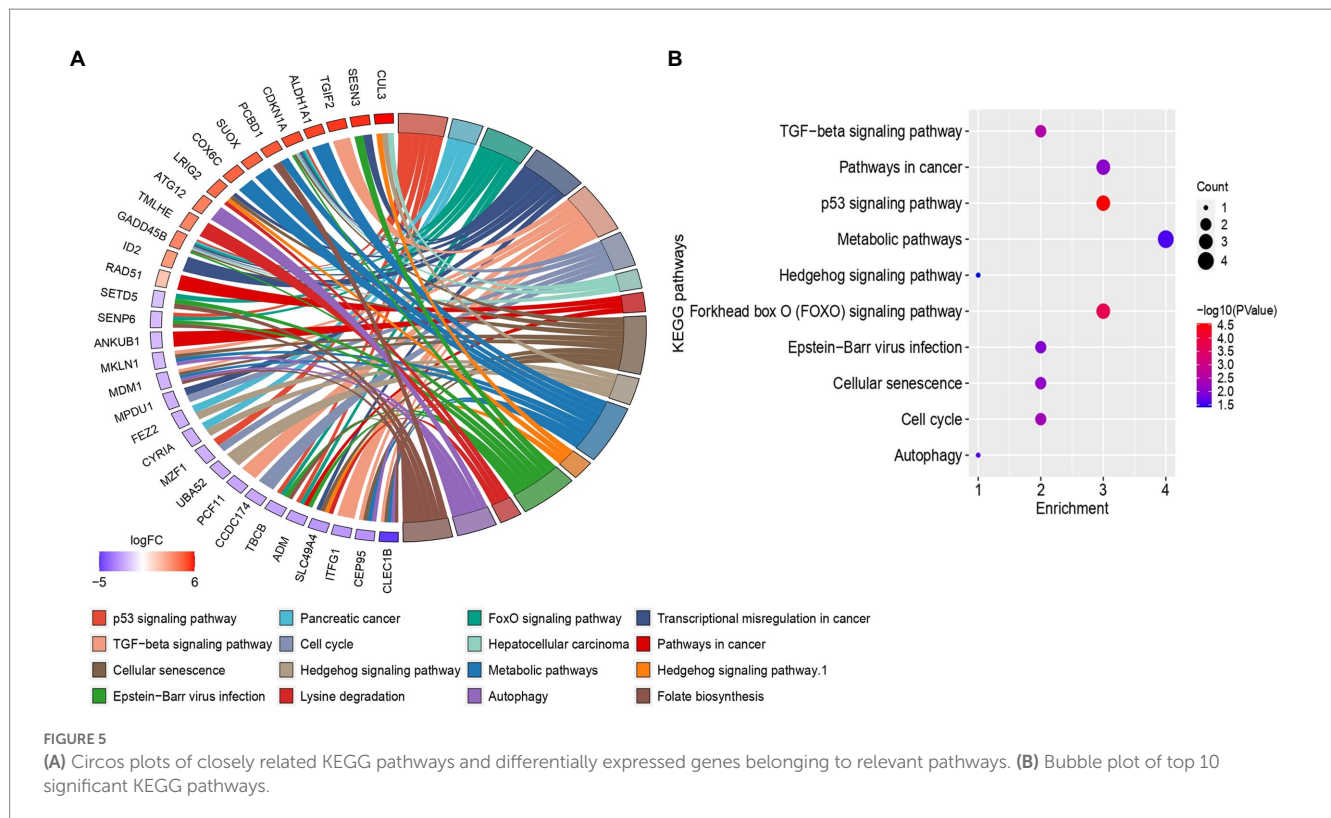
FIGURE 4
GO enrichment analysis of common 32 genes between COVID-19 and HFRS.

3.3. Biological insights of the four-module genes

GO and KEGG pathway analysis was then performed to obtain further biological insight into the commonly dysregulated genes of HFRS and COVID-19. The findings of GO analysis revealed that in terms of BP, the genes were mainly involved in the intrinsic apoptotic signaling pathway, cell cycle G1/S phase transition, autophagosome assembly, negative regulation of TORC1 signaling, regulation of B cell differentiation, COPII-coated vesicle budding, positive regulation of p38MAPK cascade, protein-DNA complex subunit organization, cellular response to extracellular stimulus, and negative regulation of macroautophagy (Figure 4). The most significant cellular components (CC) overlapped genes are enriched in several cell compartments. The significant molecular

function (MF) mainly enriched in the binding-related functions including cyclin binding, ubiquitin protein ligase binding, histone methyltransferase activity, cyclin-dependent protein serine, aldehyde dehydrogenase (NAD⁺) activity, oxidoreductase activity, transcription regulator inhibitor activity, DNA polymerase binding, single-stranded DNA helicase activity, and notch binding.

After GO enrichment analysis, KEGG pathway analysis was performed for providing biological context and insight into the underlying mechanisms of HFRS comorbidity in COVID-19 patients. Through clusterprofiler, only 16 KEGG pathways were identified which fulfill the criteria of value of $p < 0.05$, while those with value of $p > 0.05$ were excluded from the study. The KEGG pathways of overlapped genes are mainly enriched in several pathways such as infection-related pathways, i.e., epstein-barr virus infection, p53 signaling pathway, forkhead box O (FOXO)



signaling pathway, TGF-beta signaling pathway, cell cycle cellular senescence, pathways in cancer, metabolic pathways, autophagy, and hedgehog signaling pathway. The circos plot represented the common genes along with their associated pathway are shown in Figure 5A. Gene involvement in the KEGG pathways was identified by colored connecting lines. Further, the bubble plot of top 10 significant KEGG pathways was presented in Figure 5B.

3.4. Identification of hub genes

The PPI network of common genes was constructed through the STRING database. The PPI network was then imported to Cytoscape for the identification of hub genes. There are 11 topological methods available in cytoHubba. From these methods, degree, MCC, MNC, closeness, and betweenness were selected for hub gene identification. The top 10 genes obtained from the degree, MCC, MNC, closeness, and betweenness were subjected to venn plot (Supplementary Table S2). A total of 6 six genes named GADD45B, UBA52, CDKN1A, RAD51, CUL3, and ALDH1A1 were found to be common in each method (Figure 6). The logFC values and p -values of selected hub genes are presented in Table 2.

3.5. Performance evaluation of the hub genes with a classification algorithm

After the identification of hub genes, different supervised machine-learning classifiers were implemented to evaluate the discriminative performance of predicted hub genes. Four different types of popular

classification algorithms including RF, SVM, PLDA, and voomNSC were executed for computing the performance measure including sensitivity, specificity, and accuracy of hub genes (Table 3). To execute this task, we divided the GSE158712 into test datasets and training datasets. MLSeq takes a matrix of raw counts as the input and performs normalization within-fold so that the normalization of the test fold is performed using coefficients estimated from the training folds. The process of randomly splitting samples into training/test folds, training the models, and then testing performance was repeated 10 times to obtain estimates of model performance. The performance metrics were averaged across the repeated folds. Accuracy explains the overall correctness of classification, and is defined as the proportion of all cases that are correctly classified or diagnosed. Mathematically, accuracy can be expressed as: $\text{Accuracy} = (\text{True Positives} + \text{True Negatives}) / (\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives})$. Sensitivity, also known as recall or true positive rate, measures the proportion of actual positives that are correctly identified by the classification. $\text{Sensitivity} = \text{True Positives} / (\text{True Positives} + \text{False Negatives})$. Specificity, on the other hand, measures the proportion of actual negatives that are correctly identified by the classification or diagnostic test. $\text{Specificity} = \text{True Negatives} / (\text{True Negatives} + \text{False Positives})$.

RF had the highest accuracy of 79.41%, indicating that it was able to correctly classify the samples with a high level of accuracy. The sensitivity of 88.24% suggests that the model was able to correctly identify most of the positive cases (i.e., samples with high expression of the hub genes), while the specificity of 70.59% indicates that it was less successful in correctly identifying the negative cases (i.e., samples with low expression of the hub genes). PLDA had a lower accuracy of 70.59% as compared to RF. The sensitivity of 64.71% suggests that the model was not able to correctly identify many of the positive cases, while the specificity of

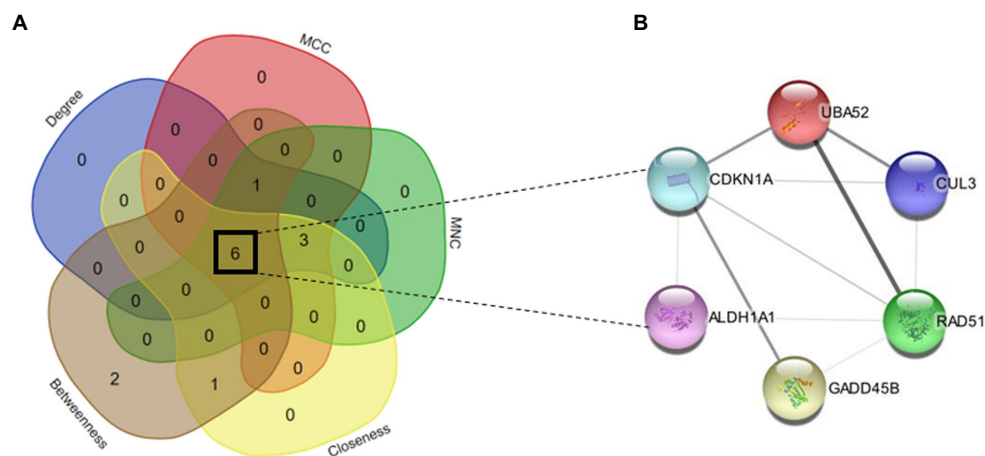


FIGURE 6

(A) Venn map of top hub genes of 5 algorithms (degree, MCC, MNC, closeness, and betweenness). (B) PPI network of six hub genes (GADD45B, UBA52, CDKN1A, RAD51, CUL3, and ALDH1A1).

TABLE 2 Summary of hub genes in COVID-19 and HFRS datasets.

Symbol	Gene name	HFRS datasets		COVID-19 datasets	
		log2FC (>1.0)	P-value (>0.05)	log2FC (>1.0)	P-value (>0.05)
RAD51	RAD51 recombinase	4.776201	0.009451	1.223918	3.50E-07
ALDH1A1	Aldehyde dehydrogenase 1 family member A1	4.29785	0.005062	3.74E-05	2.21E-06
UBA52	Ubiquitin A-52 residue ribosomal protein fusion product 1	3.572751	0.018842	1.071183	0.014559
CUL3	Cullin 3	-4.72487	0.010141	-1.32525	0.000372
GADD45B	Growth arrest and DNA damage inducible beta	-1.85687	0.022523	-1.82173	0.002635
CDKN1A	Cyclin dependent kinase inhibitor 1A	-1.90911	0.03701	-1.92171	0.001758

TABLE 3 Classification performance of six hub genes.

Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)
Random forest (RF)	79.41%	88.24%	70.59%
Poisson linear discriminant analysis (PLDA)	70.59%	64.71%	76.47%
Voom-based Nearest Shrunken Centroids (voomNSC)	77.94%	88.24%	67.65%
Support vector machine (SVM)	76.47%	82.35%	70.59%

76.47% indicates that it was better at identifying the negative cases compared to RF. On the other hand, voomNSC had an accuracy of 77.94%, indicating that it performed better than PLDA but not as well as RF. While the sensitivity was found to be 88.24% suggests that voomNSC correctly identify most of the positive cases, while the specificity of 67.65% indicates that it was less successful in correctly identifying the negative cases compared to both RF and PLDA. Finally, SVM had an accuracy of 76.47%, which is similar to VoomNSC. The sensitivity of 82.35% suggests that it was able to correctly identify most of the positive cases, while the specificity of 70.59% is similar to RF. Recently, Auwul et al. (2021) applied ML algorithms for the performance evaluation of hub genes. Their results indicated that SVM provides greater accuracy of 0.996 followed by RF: 0.955, PLDA: 0.821, and voomDLDA: 0.988. Similarly, we also observed satisfactory sample classification performance (accuracy >0.70) between affected and health control samples for both

datasets. An accuracy value greater than 70% indicates great model performance. The findings of the current study revealed that RF provides high accuracy of 79.41% as compared to SVM (76.47%), voomNSC (77.94%), and PLDA (70.59%) (Figure 7), suggesting the biomarker potential of the hub genes.

4. Discussion

HFRS is a potentially fatal infectious disease with worldwide distribution. PUUV, SEOV, DOBV, and HTNV are the primary causative agents of HFRS (Shid et al., 2022). PUUV, SEOV, DOBV, and HTNV are a member of rodent-borne viruses called hantaviruses that cause life-threatening human diseases in Europe and Asia. COVID-19 is a respiratory disease caused by the novel coronavirus SARS-CoV-2, which

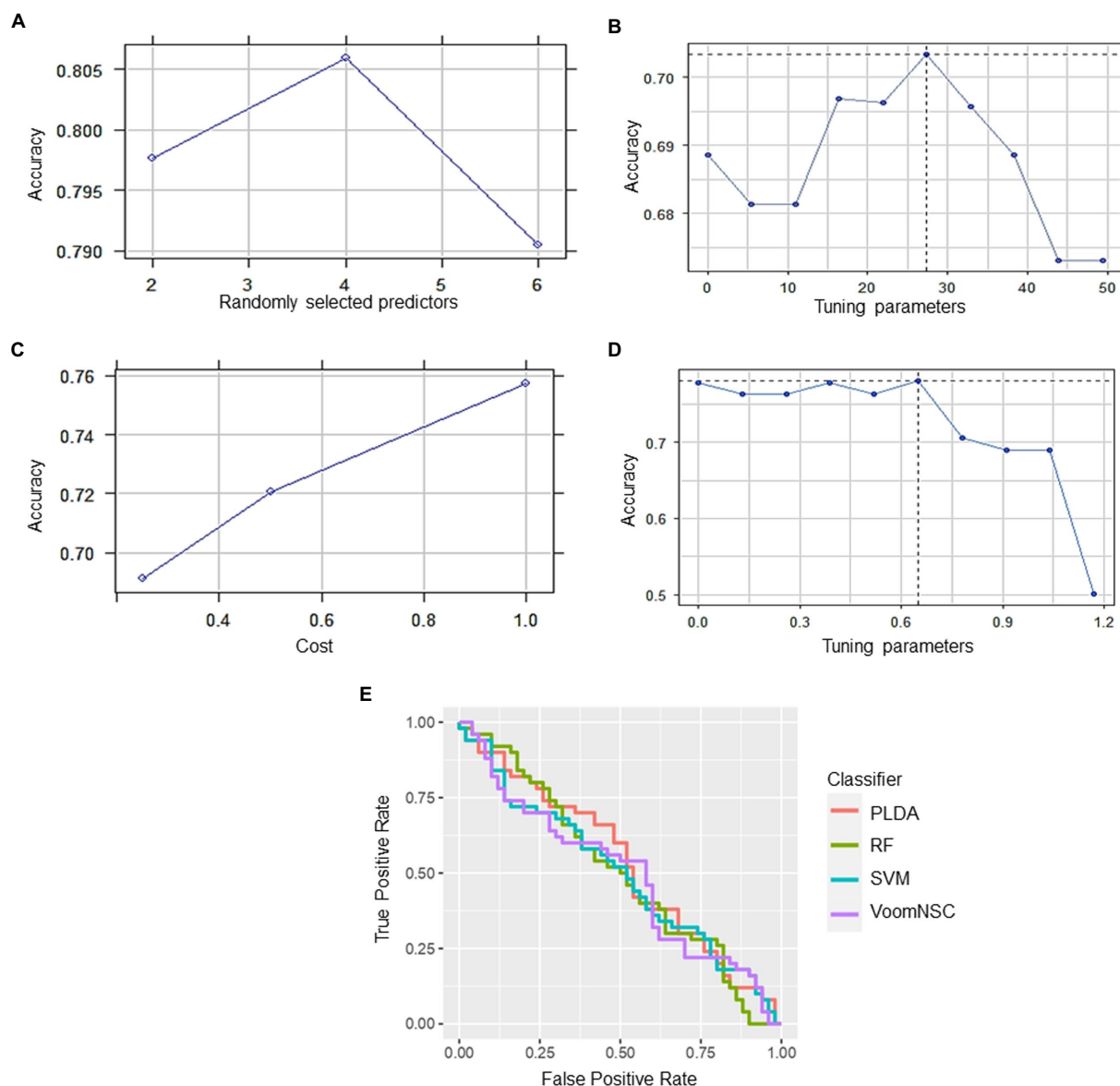


FIGURE 7

Classification performance of four different types of classifiers (A) RF, (B) PLDA, (C) SVM, and (D) voomNSC. (E) Receiver operating curve (ROC) plot of the four classifiers performance based on their accuracy, sensitivity, and specificity.

primarily affects the respiratory system (Yang et al., 2020). While HFRS and COVID-19 have distinct modes of transmission and clinical presentations, there have been reports of co-infections and comorbidities between the two diseases. But the reasons for this are not yet clear, but it is believed immune response to one disease may increase susceptibility or severity of the other such as kidney damage and compromised immune function, which may increase the severity of COVID-19. However, further research is needed to fully understand the nature of this relationship and to develop appropriate treatment strategies.

The presence of comorbidities in individuals co-infected with HFRS and COVID-19 represents a significant challenge that necessitates a solution. It is imperative to develop effective strategies to manage the comorbidities to improve clinical outcomes and reduce the burden of these diseases. Therefore, it is essential to investigate the fundamental genes and pathways involved in coinfection in

order to unravel the molecular associations and mechanisms that are shared by these pathologies. To achieve this, the application of whole-genome transcriptomic analyses has been extensively utilized by researchers to explore autoimmune diseases, cancer, and neurodegenerative disorders, as well as to identify potential pathogenetic mechanisms and novel therapeutic targets. In the current study, integrative bioinformatics approaches were used for the comprehensive analysis of peripheral blood mononuclear cell (PBMC) transcriptomic changes occurring in HFRS and COVID-19. Our study uncovered six genes named RAD51, ALDH1A1, UBA52, CUL3, GADD45B, and CDKN1A were found to be commonly dysregulated among HFRS and COVID-19. Further, machine learning classifiers have gained immense popularity in various bioinformatics tasks. Therefore, we employed different machine learning classification algorithms on hub genes data to determine

their efficacy. Our findings indicate that the classifiers performed satisfactorily with an accuracy of >0.70 for the classification of samples between COVID-19 and HFRS.

Auwul et al. (2021) applied the similar approach to identify the potential drug targets and pathways in COVID-19. Their findings proposed that PLK1, AURKB, AURKA, CDK1, CDC20, KIF11, CCNB1, KIF2C, DTL, and CDC6 were mainly enriched in the inflammatory and immune response, suggesting that these genes are significantly associated with viral infectious diseases. Similarly, Rahman et al. (2021) discovered common pathogenetic processes between COVID-19 and diabetes mellitus by differential gene expression pattern analysis. Their study, for the first time, characterized the biological processes and pathways commonly dysregulated in COVID-19 and diabetes mellitus, which could be in the next future used for the design of personalized treatment of COVID-19 patients suffering from diabetes mellitus as comorbidity. Their study proposed that SARS-CoV-2 could directly determine an impairment of insulin secretion, with consequent disruption of the metabolic control in people already suffering from diabetes mellitus or leading to the development of new-onset diabetes mellitus. Our study explores the possible risk of HFRS after COVID-19 infection by investigating the common molecular mechanisms. By taking advantage of the holistic viewpoint of systems biology, we were able to consider every aspect of both diseases and infer novel hypotheses. Further supplementary studies need to be conducted to clarify the association between COVID-19 and HFRS, as, at the moment, there is little known regarding both of these disease entities. Overall, our analysis highlights various infection-related pathways, i.e., epstein–barr virus infection, p53 signaling pathway, FOXO signaling pathway, TGF- β signaling pathway, cell cycle cellular senescence, pathways in cancer, metabolic pathways, autophagy, and hedgehog signaling pathway which might be the potential links between both COVID-19 and HFRS.

Among six hub genes, RAD51 is a recombinase protein that plays a key role in homologous recombination (Huang et al., 2012). Previous studies demonstrated a potential association between RAD51 and COVID-19. It has been suggested that COVID-19 infection may interfere with the expression and activity of RAD51, thereby impairing DNA repair and increasing the risk of genetic mutations (Biering et al., 2021). Furthermore, studies have also shown that RAD51 may be involved in the inflammatory response to COVID-19 infection (Morenikeji et al., 2021). Elevated levels of inflammation are associated with severe COVID-19, and RAD51 has been shown to modulate inflammatory signaling pathways. In the same vein, RAD51 might play a role in the virus–host interaction by reducing viral replication during hantavirus infection. All these points strengthened the findings that dysregulation of RAD51 expression and activity may contribute to the development of severe HFRS. In conclusion, further research is needed to fully elucidate the mechanisms of RAD51 for understanding the common pathogenic processes between COVID-19 and HFRS, which could pave the way for the development of novel therapeutic strategies. On the other hand, ALDH1A1 has been implicated in the pathogenesis of the disease through its potential involvement in modulating the immune response and influencing the balance between pro-inflammatory and anti-inflammatory cytokines. Furthermore, CDKN1A, also known as p21, is a significant regulator of the cell cycle by controlling the activity of cyclin-dependent kinases (CDKs). In COVID-19, studies have reported that CDKN1A is upregulated in the lung tissue of infected patients, which could lead to the inhibition of

virus replication and decreased inflammation in the host. CDKN1A has been shown to inhibit the replication of some viruses, including herpes simplex virus (HSV) and human immunodeficiency virus (HIV), by blocking the cell cycle progression of infected cells. In HFRS, studies have shown that CDKN1A is upregulated in the kidney tissue of HFRS patients, which may contribute to the development of renal injury (D'Souza, 2022). CDKN1A has been suggested to be involved in the regulation of cellular senescence and apoptosis, which are important processes in the development of renal injury in HFRS. Thus, more future studies are needed to fully understand the role of CDKN1A as well as other hub genes in both COVID-19 and HFRS.

There is limited information available on the translational activity of UBA52 and CUL3 specifically in HFRS. However, both UBA52 and CUL3 are involved in the ubiquitin–proteasome system, which plays an important role in the regulation of various cellular processes, including protein degradation and immune response (Meyer et al., 2020; Jiang et al., 2022). In general, dysregulation of the ubiquitin–proteasome system has been implicated in the pathogenesis of viral infections, including COVID-19, by influencing viral replication and modulating the host immune response (Seyoum, 2023). Therefore, it is possible that UBA52 and CUL3 may also play a role in the pathogenesis of COVID-19 and HFRS coinfection through their involvement in the ubiquitin–proteasome system. In short, there is no direct evidence of interactions between these six hub genes in the context of COVID-19 and HFRS, their involvement in processes related to DNA damage response, inflammation, and immune regulation suggests that they may be part of a larger interactome in their coinfection.

Further, GO and KEGG pathway analysis revealed that the commonly dysregulated genes are mainly involved in infection-related pathways, i.e., epstein–barr virus infection, p53 signaling pathway, forkhead box O (FOXO) signaling pathway, TGF- β signaling pathway, cell cycle, cellular senescence, pathways in cancer, metabolic pathways, autophagy, and hedgehog signaling pathway. These pathways are known to play important roles in various cellular processes, such as cell cycle regulation, DNA damage response, cell proliferation, differentiation, and survival, and are often dysregulated in diseases such as cancer and viral infections. For instance, targeting genes involved in the FOXO signaling pathway may be helpful for treating COVID-19 and HFRS coinfection because it regulates immune responses, oxidative stress, and inflammation. For example, activation of FOXO3a has been shown to enhance the antiviral response by increasing the expression of interferon-stimulated genes and inhibiting viral replication (Wang et al., 2017). Additionally, FOXO3a regulates oxidative stress by increasing the expression of antioxidant enzymes, such as catalase and superoxide dismutase, and inhibiting reactive oxygen species production (Higuchi et al., 2013). In COVID-19 and HFRS, the immune response can be dysregulated, leading to a cytokine storm and inflammation. Activation of FOXO3a can suppress pro-inflammatory cytokines and reduce inflammation, thereby preventing tissue damage and improving clinical outcomes. Moreover, FOXO3a has been shown to promote autophagy which helps in clearing viral infections by degrading and recycling viral components (Wan et al., 2022). Therefore, the identification of genes involved in the FOXO signaling pathway may provide a novel therapeutic approach for COVID-19 and HFRS by regulating immune responses, oxidative stress, and inflammation. Overall, these common pathways may help to better understand the pathogenesis of HFRS and COVID-19, and targeting the linked genes could help

in the development of new therapies to fight against the coinfections. As this field continues to evolve, future studies may benefit from incorporating network-based approaches to predict host-pathogen interactions. Therefore, future studies could incorporate the host-pathogen interactome to gain a deeper understanding of the disease mechanisms and identify additional potential therapeutic targets. For instance, Basu et al. (2022) reported the potential of network-based approaches to predict host-pathogen interactions and identify key host factors involved in viral infections. This approach could be a valuable addition to our computational pipeline in future studies.

To sum up, the integration of transcriptomic with bioinformatics approaches uncovered several pathways and common genes that are involved in both HFRS and COVID-19, suggesting potential similarities in their underlying pathophysiological mechanisms. The identified pathways and genes may serve as potential targets for the development of new therapeutics and which may lend a helping hand in understanding the pathogenesis of both diseases. However, it is important to note that RNA-seq analysis is just one tool in the development of potential therapies for comorbid HFRS and COVID-19. Other factors, such as clinical trials, animal studies, and epidemiological data, also need to be considered before any therapies can be developed and implemented. Additionally, the differences and similarities between HFRS and COVID-19 in terms of genetic factors, clinical manifestations, and disease progression need to be further investigated to develop targeted and effective treatments.

5. Conclusion

Comorbidities are associated with a higher risk of developing severe forms of COVID-19, with a consequent need for mechanical ventilation and an increased death rate. Increased severity of COVID-19 has been observed in patients with HFRS. Following that, the present research aimed to identify gene expression patterns and molecular pathways that were shared between COVID-19 and HFRS. A total of 32 genes were found to be common among COVID-19 and HFRS. From these 32 genes, six named RAD51, ALDH1A1, UBA52, CUL3, GADD45B, and CDKN1A were identified as the hub genes. Later, we demonstrated the classification performance of hub genes with an accuracy greater than 0.70 suggesting the biomarker potential of the hub genes. Further, our study uncovered immune-related pathways that were commonly dysregulated in PBMCs of both COVID-19 and HFRS. In summary, this study provides valuable insights into the molecular mechanisms of HFRS and COVID-19, which may lead to potential therapies for comorbidities. We propose that these events may have important roles in the onset or progression of HFRS. To sum up, our analysis describes possible mechanisms linking COVID-19 and HFRS, elucidating some unknown clues in between. Nonetheless, as this is a thorough computational work,

further case reports and follow-up experiments of COVID-19 patients can corroborate these links.

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

FN collected the data, performed experiments, and write down the first draft of the manuscript. UAA, AB, WQ, KSA, BFA and WAM visualized and validated the results and revised the manuscript. KSA, BFA, WAM and MTQ arranged resources and funding for this study. UAA and MTQ designed and supervised this study, and finalized the manuscript. All authors approved the final version for submission.

Funding

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R39), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

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.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1175844/full#supplementary-material>

References

- Alexa, A., and Rahnenführer, J. (2009). Gene set enrichment analysis with topGO. *Bioconductor Improv.* 27, 1–26.
- Arunachalam, P. S., Wimmers, F., Mok, C. K. P., Perera, R. A., Scott, M., Hagan, T., et al. (2020). Systems biological assessment of immunity to mild versus severe Covid-19 infection in humans. *Science* 369, 1210–1220. doi: 10.1126/science.abc6261
- Auwul, M. R., Rahman, M. R., Gov, E., Shahjaman, M., and Moni, M. A. (2021). Bioinformatics and machine learning approach identifies potential drug targets and pathways in Covid-19. *Brief. Bioinform.* 22 bbab120. doi: 10.1093/bib/bbab120
- Auwul, M. R., Zhang, C., Rahman, M. R., Shahjaman, M., Alyami, S. A., and Moni, M. A. (2021). Network-based transcriptomic analysis identifies the genetic effect

- of Covid-19 to chronic kidney disease patients: A bioinformatics approach. *Saudi J Biol Sci* 28, 5647–5656. doi: 10.1016/j.sjbs.2021.06.015
- Basu, S., Naha, A., Veeraghavan, B., Ramaiah, S., and Anbarasu, A. (2022). *In silico* structure evaluation of Bag3 and elucidating its association with bacterial infections through protein–protein and host–pathogen interaction analysis. *J. Cell. Biochem.* 123, 115–127. doi: 10.1002/jcb.29953
- Biering, S. B., Van Dis, E., Wehri, E., Yamashiro, L. H., Nguyenla, X., Dugast-Darzacq, C., et al. (2021). Screening a library of FDA-approved and bioactive compounds for antiviral activity against SARS-CoV-2. *ACS Infect. Dis.* 7, 2337–2351. doi: 10.1021/acscinfdis.1c00017
- Boser, B. E., Guyon, I. M., and Vapnik, V. N. (eds.). “A training algorithm for optimal margin classifiers.” in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. New York, USA: ACM, (1992).
- Brown, J., Pirrung, M., and McCue, L. A. (2017). FQC dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* 33, 3137–3139. doi: 10.1093/bioinformatics/btx373
- Brunetta, E., Folci, M., Bottazzi, B., De Santis, M., Gritti, G., Protti, A., et al. (2021). Macrophage expression and prognostic significance of the long pentraxin PTX3 in Covid-19. *Nat. Immunol.* 22, 19–24. doi: 10.1038/s41590-020-00832-x
- Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* 25, 169–193. doi: 10.1677/jme.0.0250169
- Cetin, S., and Sahin, A. M. (2021). A case of hantavirus infection detected during the Covid-19 pandemic/Covid-19 pandemisi sirasinda saptanan bir hantavirus infeksiyonu olgusu. *KLİMİK J.* 34, 141–143. doi: 10.36519/kd.2021.3656
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Clough, E., and Barrett, T. (2016). The gene expression omnibus database. *Methods Mol. Biol.* 1418, 93–110. doi: 10.1007/978-1-4939-3578-9_5
- Criscuolo, A., and Brisse, S. (2013). Alientrimmer: A tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* 102, 500–506. doi: 10.1016/j.ygeno.2013.07.011
- D'Souza, M. H. *Biophysical characterisation studies of hantaviral and human long noncoding RNAs*. Lethbridge, AB: University of Lethbridge (2022).
- Flaherty, G. T., Hession, P., Liew, C. H., Lim, B. C. W., Leong, T. K., Lim, V., et al. (2020). Covid-19 in adult patients with pre-existing chronic cardiac, respiratory and metabolic disease: A critical literature review with clinical recommendations. *Trop. Dis. Travel Med. Vaccines* 6, 1–13. doi: 10.1186/s40794-020-00118-y
- Garanina, E., Martynova, E., Davidyuk, Y., Kabwe, E., Ivanov, K., Titova, A., et al. (2019). Cytokine storm combined with humoral immune response defect in fatal hemorrhagic fever with renal syndrome case, Tatarstan, Russia. *Viruses* 11:601. doi: 10.3390/v11070601
- Geladari, V., Sarafidis, P. A., Alexandrou, M.-E., Faitatzidou, D., Nikolaidou, C., Stangou, M., et al. (2022). A patient with dialysis-dependent acute kidney injury due to hantavirus complicated with SARS-CoV-2 infection. *Clin. Nephrol. Case Stud.* 10, 64–70. doi: 10.5414/CNCS110846
- Goksuluk, D., Zararsiz, G., Korkmaz, S., Eldem, V., Zararsiz, G. E., Ozcetin, E., et al. (2019). MLSeq: Machine learning Interface for RNA-sequencing data. *Comput. Methods Prog. Biomed.* 175, 223–231. doi: 10.1016/j.cmpb.2019.04.007
- Hasanoglu, I., Korukluoglu, G., Asilturk, D., Cosgun, Y., Kalem, A. K., Altas, A. B., et al. (2021). Higher viral loads in asymptomatic Covid-19 patients might be the invisible part of the iceberg. *Infection* 49, 117–126. doi: 10.1007/s15010-020-01548-8
- Higuchi, M., Dusting, G. J., Peshavariya, H., Jiang, F., Hsiao, S. T.-F., Chan, E. C., et al. (2013). Differentiation of human adipose-derived stem cells into fat involves reactive oxygen species and Forkhead box O1 mediated upregulation of antioxidant enzymes. *Stem Cells Dev.* 22, 878–888. doi: 10.1089/scd.2012.0306
- Ho, T. K. (ed.). “Random decision forests.” in *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Montreal, Canada, August: IEEE, (1995).
- Huang, F., Mazina, O. M., Zentner, I. J., Cocklin, S., and Mazin, A. V. (2012). Inhibition of homologous recombination in human cells by targeting RAD51 recombinase. *J. Med. Chem.* 55, 3011–3020. doi: 10.1021/jm201173g
- Jiang, Y., Yan, Q., Liu, C.-X., Peng, C.-W., Zheng, W.-J., Zhuang, H.-F., et al. (2022). Insights into potential mechanisms of asthma patients with Covid-19: A study based on the gene expression profiling of Bronchoalveolar lavage fluid. *Comput. Biol. Med.* 146:105601. doi: 10.1016/j.combiomed.2022.105601
- Krautkrämer, E., Zeier, M., and Plyusnin, A. (2013). Hantavirus infection: An emerging infectious disease causing acute renal failure. *Kidney Int.* 83, 23–27. doi: 10.1038/ki.2012.360
- Lee, H. W., Baek, L. J., and Johnson, K. M. (1982). Isolation of Hantaan virus, the etiologic agent of Korean hemorrhagic fever, from wild urban rats. *J. Infect. Dis.* 146, 638–644. doi: 10.1093/infdis/146.5.638
- Lee, H. W., Lee, P. W., and Johnson, K. M. (1978). Isolation of the etiologic agent of Korean hemorrhagic fever. *J. Infect. Dis.* 137, 298–308. doi: 10.1093/infdis/137.3.298
- Li, Y. C., Bai, W. Z., and Hashikawa, T. (2020). The neuroinvasive potential of SARS-CoV-2 may play a role in the respiratory failure of Covid-19 patients. *J. Med. Virol.* 92, 552–555. doi: 10.1002/jmv.25728
- Li, Y., Quan, C., Xing, W., Wang, P., Gao, J., Zhang, Z., et al. (2020). Rapid humoral immune responses are required for recovery from haemorrhagic fever with renal syndrome patients. *Emerg. Microb. Infect.* 9, 2303–2314. doi: 10.1080/22221751.2020.1830717
- Meyer, S. J., Böser, A., Korn, M. A., Koller, C., Bertocci, B., Reimann, L., et al. (2020). Cullin 3 is crucial for pro-B cell proliferation, interacts with CD22, and controls CD22 internalization on B cells. *J. Immunol.* 204, 3360–3374. doi: 10.4049/jimmunol.1900925
- Morenikeji, O. B., Bernard, K., Strutton, E., Wallace, M., and Thomas, B. N. (2021). Evolutionarily conserved long non-coding RNA regulates gene expression in cytokine storm during Covid-19. *Front. Bioeng. Biotechnol.* 8:582953. doi: 10.3389/fbioe.2020.582953
- Noor, R. (2020). General awareness on hantavirus infection: A brief review. *Biomed. Biotechnol. Res. J.* 4:274. doi: 10.4103/bbrj.bbrj_110_20
- Noor, F., Ashfaq, U. A., Asif, M., Adeel, M. M., Alshammari, A., and Alharbi, M. (2022). Comprehensive computational analysis reveals YXXφ [I/L/M/F/V] motif and YXXφ-like tetrapeptides across HFRS causing hantaviruses and their association with viral pathogenesis and host immune regulation. *Front. Immunol.* 13:1031608. doi: 10.3389/fimmu.2022.1031608
- Papa, A. (2012). Dobrava-Belgrade virus: Phylogeny, epidemiology, disease. *Antiviral Res.* 95, 104–117. doi: 10.1016/j.antiviral.2012.05.011
- Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, stringtie and ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095
- Rahman, M. R., Islam, T., Shahjaman, M., Islam, M. R., Lombardo, S. D., Bramanti, P., et al. (2021). Discovering common pathogenetic processes between Covid-19 and diabetes mellitus by differential gene expression pattern analysis. *Brief. Bioinform.* 22:bbab262. doi: 10.1093/bib/bbab262
- Sewe, S. O., Silva, G., Sicut, P., Seal, S. E., and Visendi, P. (2022). “Trimming and validation of Illumina short reads using trimmomatic, trinity assembly, and assessment of RNA-seq data,” in *Plant bioinformatics: Methods and protocols*. ed. D. Edwards (Humana, New York: Springer), 211–232.
- Seyoum, T. F. (2023). The role of ubiquitin-proteasome system in the pathogenesis of severe acute respiratory syndrome coronavirus-2 disease. *Int. J. Inflam.* 2023:6698069. doi: 10.1155/2023/6698069
- 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* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shid, S. J., Sapate, R. B., Raut, I. D., Dhadde, G. S., Mali, H. S., Joshi, S., et al. (2022). Hantavirus infection: An overview. *Asian J. Pharm. Res.* 12, 150–154. doi: 10.5271/12231-5691.2022.00023
- Singh, H., Kaur, H., and Medhi, B. (2020). Novel therapeutic approaches toward Hantaan virus and its clinical features' similarity with Covid-19. *Indian J. Pharmacol.* 52, 347–355. doi: 10.4103/ijp.ijp_1001_20
- Sironen, T., Vaheri, A., and Plyusnin, A. (2001). Molecular evolution of Puumala hantavirus. *J. Virol.* 75, 11803–11810. doi: 10.1128/JVI.75.23.11803-11810.2001
- Subramaniam, A., Pilcher, D., Tiruvoipati, R., Wilson, J., Mitchell, H., Xu, D., et al. (2022). Timely goals of care documentation in patients with frailty in the Covid-19 era: A retrospective multi-site study. *Intern. Med. J.* 52, 935–943. doi: 10.1111/imj.15671
- Tariq, M., and Kim, D.-M. (2022). Hemorrhagic fever with renal syndrome: Literature review, epidemiology, clinical picture and pathogenesis. *Infect. Chemother.* 54:1. doi: 10.3947/ic.2021.0148
- von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). String: A database of predicted functional associations between proteins. *Nucleic Acids Res.* 31, 258–261. doi: 10.1093/nar/gkg034
- Wan, Y. I., Bien, Z., Apea, V. J., Orkin, C. M., Dhairyawan, R., Kirwan, C. J., et al. (2021). Acute kidney injury in Covid-19: Multicentre prospective analysis of registry data. *Clin. Kidney J.* 14, 2356–2364. doi: 10.1093/cj/sfab071
- Wan, S. W., Lee, Y. R., Ho, T. S., and Chang, C. P. (2022). Regulation of innate immune signaling pathways by autophagy in dengue virus infection. *TUBMB Life* 74, 170–179. doi: 10.1002/iub.2554
- Wang, W., Xu, L., Su, J., Peppelenbosch, M. P., and Pan, Q. (2017). Transcriptional regulation of antiviral interferon-stimulated genes. *Trends Microbiol.* 25, 573–584. doi: 10.1016/j.tim.2017.01.001
- Weinstein, D. J., Allen, S. E., Lau, M. C., Erasmus, M., Asalone, K. C., Walters-Conte, K., et al. (2019). The genome of a subterranean nematode reveals adaptations to heat. *Nat. Commun.* 10:5268. doi: 10.1038/s41467-019-13245-8
- Witten, DM. *Classification and clustering of sequencing data using a poisson model*. *Ann. Appl. Stat.* (2011) 5: 2493–2518. doi: 10.1214/11-AOAS493
- Yang, L., Liu, S., Liu, J., Zhang, Z., Wan, X., Huang, B., et al. (2020). Covid-19: Immunopathogenesis and immunotherapeutics. *Signal Transduct. Target. Ther.* 5:128. doi: 10.1038/s41392-020-00243-2
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zhu, J., and Cai, Y. (2020). Engaging the communities in Wuhan, China during the Covid-19 outbreak. *Glob. Health Res. Policy* 5, 1–4. doi: 10.1186/s41256-020-00162-3



OPEN ACCESS

EDITED BY

Arli Aditya Parikesit,
Indonesia International Institute for Life-
Sciences (i3L), Indonesia

REVIEWED BY

Arif Ansori,
Airlangga University, Indonesia
Viol Kharisma,
Airlangga University, Indonesia

*CORRESPONDENCE

Chunlei Yuan
✉ 13802699640@139.com

RECEIVED 21 March 2023

ACCEPTED 18 April 2023

PUBLISHED 09 May 2023

CITATION

Zhuo J, Wang K, Shi Z and Yuan C (2023)
Immunogenic cell death-led discovery of
COVID-19 biomarkers and inflammatory
infiltrates.
Front. Microbiol. 14:1191004.
doi: 10.3389/fmicb.2023.1191004

COPYRIGHT

© 2023 Zhuo, Wang, Shi and Yuan. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Immunogenic cell death-led discovery of COVID-19 biomarkers and inflammatory infiltrates

Jianzhen Zhuo^{1,2}, Ke Wang², Zijun Shi³ and Chunlei Yuan^{1,2*}

¹Guangdong Medical University, Dongguan, Guangdong, China, ²Clinical Laboratory, Boai Hospital of Zhongshan Affiliated to Southern Medical University, Zhongshan, China, ³Reproductive Medical Center, Boai Hospital of Zhongshan Affiliated to Southern Medical University, Zhongshan, China

Immunogenic cell death (ICD) serves a critical role in regulating cell death adequate to activate an adaptive immune response, and it is associated with various inflammation-related diseases. However, the specific role of ICD-related genes in COVID-19 remains unclear. We acquired COVID-19-related information from the GEO database and a total of 14 ICD-related differentially expressed genes (DEGs) were identified. These ICD-related DEGs were closely associated with inflammation and immune activity. Afterward, CASP1, CD4, and EIF2AK3 among the 14 DEGs were selected as feature genes based on LASSO, Random Forest, and SVM-RFE algorithms, which had reliable diagnostic abilities. Moreover, functional enrichment analysis indicated that these feature genes may have a potential role in COVID-19 by being involved in the regulation of immune response and metabolism. Further CIBERSORT analysis demonstrated that the variations in the immune microenvironment of COVID-19 patients may be correlated with CASP1, CD4, and EIF2AK3. Additionally, 33 drugs targeting 3 feature genes had been identified, and the ceRNA network demonstrated a complicated regulative association based on these feature genes. Our work identified that CASP1, CD4, and EIF2AK3 were diagnostic genes of COVID-19 and correlated with immune activity. This study presents a reliable diagnostic signature and offers an overview to investigate the mechanism of COVID-19.

KEYWORDS

COVID-19, immunogenic cell death, biomarker, diagnosis, machine learning

Introduction

Coronavirus infectious disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a public crisis worldwide since 2019 (Zhong et al., 2020; Kai-Wang To et al., 2021). Due to the mutable nature of SARS-CoV-2, numerous countries are suffering multi-dominant viral infections (Araf et al., 2022; Hadj, 2022). To control its spread, it is crucial to diagnose COVID-19 at an early stage. Accurate molecular diagnostic tests are necessary and valuable to confirm the rapid diagnosis of COVID-19, thus providing valid information for decision-making by the patient, healthcare facilities, and public health organizations (Islam and Iqbal, 2020; Safiabadi Tali et al., 2021; Yuce et al., 2021). Notably, large-scale genome-wide association studies have demonstrated specific disease-related elements in the population (Badua et al., 2021; Robishaw et al., 2021), and there are numerous reports on genes and the prevalence of COVID-19, improving the early diagnosis and clinical management of COVID-19.

Typical pathophysiological mechanisms of COVID-19 include cell death, immune response, oxidative stress, and metabolic activity (Tay et al., 2020; Gattinoni et al., 2021; van Eijk et al., 2021). Cell death and immune infiltration serve a critical role in the formation and progression of COVID-19 (Lee et al., 2020), and immunogenic cell death (ICD) is regarded as a regulative process that could affect cell death and immune infiltration simultaneously (Kroemer et al., 2022). ICD facilitates the activation and recruitment of antigen-presenting cells, thus activating innate and adaptive immune responses (Galluzzi et al., 2020; Minute et al., 2020). Initiating adaptive immunity in ICD is not only enhancing antitumor benefits but is essential for the optimal elimination of infectious etiologies (Galluzzi et al., 2017; Li et al., 2021). Therefore, ICD regulators may have potential diagnostic and therapeutic applications for COVID-19. However, earlier reports focused on only some immune cells and immunological molecules in COVID-19, lacking a perspective on the ICD in COVID-19.

To date, there is growing evidence that multiple cell death modalities, such as ferroptosis, NETosis, and necroptosis, play an important role in the development of COVID-19 (Kang and Wang, 2022; Nishiga et al., 2022; Zhang et al., 2023). Likewise, ICD may also regulate the progression of COVID-19. However, no similar studies have previously explored this specific process, and this study is the first exploration to analyze the relationship between ICD and COVID-19.

In this work, we aimed to determine the vital ICD regulators in the development of COVID-19, thus identifying valuable biomarkers for COVID-19 diagnosis. Subsequently, the correlation between ICD regulators and the infiltrating immune landscape was investigated. Based on the diagnosis biomarkers, the ceRNA network was established, and targeted small molecular agents for COVID-19 were explored. Finally, we validated the identified diagnosis biomarkers with the external dataset. This work offers an in-depth understanding of the developmental mechanisms of COVID-19 at the molecular level and identifies valuable biomarkers.

Materials and methods

Data acquisition

The gene expression profiles for COVID-19 and control samples were downloaded from the GEO databases (Barrett et al., 2013). The GSE157103 dataset comprised 100 COVID-19 samples and 26 control samples (Overmyer et al., 2021), and this dataset was deemed to be the training cohort for the principal analysis of this research. The GSE171110 dataset comprising 44 COVID-19 samples and 10 control samples was applied to validate the expression of the feature genes (Levy et al., 2021). Also, the 34 ICD-related genes included in this research were acquired from the previous report (Wang et al., 2021).

Differential expression analysis

The expression profiles of ICD-related genes in the samples were extracted from the GSE157103 dataset. Next, the limma package was utilized to identify the ICD-related differentially expressed genes (DEGs) between different types of samples (Ritchie et al., 2015). Genes with value of $p < 0.05$ were deemed significant.

Enrichment of functionality

The expression data were analyzed with functional enrichment to evaluate the potential functionality of promising targets. Gene ontology (GO) is a common approach for identifying the functions of genes, including molecular functions, biological pathways, and cellular components (The Gene Ontology C, 2019; Warwick Vesztrocy and Dessimoz, 2020). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment was performed to investigate the genomic information of the DEGs (Kanehisa et al., 2017). Reactome enrichment analysis is also applied to explore gene functions, similar to GO and KEGG analysis (Good et al., 2021). These enrichment approaches were performed with the cluster profiler package (Yu et al., 2012). Additionally, disease ontology was utilized to annotate genes from a disease perspective (Schriml et al., 2022).

Identification of feature genes

The ICD-related DEGs were further utilized to identify significant feature genes, thus diagnosing COVID-19. The feature identification approach is a procedure of limiting the number of factors, specifically vital for establishing a predictive model (Lee et al., 2021). LASSO regression, Random Forest (RF) algorithm, and SVM-RFE were included in this study to explore feature genes. The “glmnet” package was applied to conduct minimum LASSO regression, thus choosing the linear model and keeping the reliable variables (Engelbrechtsen and Bohlin, 2019). Binomial distribution variables were further presented in the LASSO categorization, with a standard error value as the minimum parameter. Next, according to various dependent decision trees from a training pool, the RF algorithm promotes the precision of the model by randomly limiting the overfitting of individual decision trees (Yang et al., 2020). SVM-RFE can identify the optimal parameters by removing the SVM-derived eigenvectors (Sanz et al., 2018). An SVM module based on the “e1071” package was created to further evaluate the diagnostic value of the selected biomarker in COVID-19 (Jiang et al., 2020). The intersected genes, as the most significant feature genes from these three algorithms, were identified for subsequent analysis. Meanwhile, the predictive reliability of feature genes was evaluated by the receiver operating characteristic (ROC) curve, and the area under the curve (AUC) was further obtained (Park et al., 2004). A logistic regression signature with these feature genes was also established to assess diagnostic ability, and ROC curve was utilized to present this result.

Pathway correlation analysis

To further investigate the potential pathways of the feature genes, single-gene gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were performed with the GSEA package and GSVA package, respectively (Subramanian et al., 2005; Hanzelmann et al., 2013). GSEA was applied to evaluate the distribution landscape of the feature genes of a predetermined collection to explore their attribution to the phenotype. For GSVA, we also applied the KEGG pathway set to conduct enrichment analysis for each feature gene. Next, the limma package was utilized to discuss the difference in the GSVA score of the feature gene's up- and down-regulated groups (Ritchie et al., 2015).

TABLE 1 ICD-related DEGs between COVID-19 samples and control samples.

Gene	<i>p</i> -value	Type
ATG5	0.002	Up
BAX	0.003	Down
CASP1	<0.001	Up
CASP8	0.046	Up
CD4	0.011	Down
EIF2AK3	<0.001	Up
ENTPD1	0.028	Up
FOXP3	0.001	Down
HMGB1	<0.001	Up
HSP90AA1	0.004	Up
MYD88	<0.001	Up
PIK3CA	0.023	Up
TLR4	0.038	Up
TNF	0.021	Down

Immune cell infiltration analysis

The CIBERSORT algorithm was applied to analyze the normalized gene expression data in the GSE157103 dataset, and the fraction of immune cells was identified (Chen et al., 2018; Overmyer et al., 2021). Violin plots were displayed to present the expressional difference of the immune infiltrating cells. And Spearman correlation analysis was executed to investigate the association between diverse immune infiltrating cells (Fujita et al., 2009). Meanwhile, the correlation between feature genes and immune cells was also explored with a similar method. These results were visualized with the “ggplot2” package (Ito and Murphy, 2013). *p*-value < 0.05 demonstrated statistical significance.

Drug prediction

Drug Gene Interaction Database (DGIdb¹) integrates existing literature on drug-gene interactions to provide clinical guidance for personalized treatment of disease. In this report, DGIdb was applied to identify gene-targeted drugs, and DrugBank database was further applied to identify drugs' structural elements (Wagner et al., 2016; Wishart et al., 2018).

Establishment of a ceRNA network

Three databases (miRanda, miRDB, TargetScan) were utilized to predict miRNAs (Enright et al., 2003; Agarwal et al., 2015; Chen and Wang, 2020), and the intersections of the predicted miRNAs were identified as the target miRNAs. Afterward, spongeScan was applied to identify the corresponding lncRNAs for miRNAs (Furio-Tari et al.,

2016). A ceRNA network using Cytoscape was established predicated on the cross-talk between mRNAs, miRNAs, and lncRNAs (Shannon et al., 2003; Wu et al., 2020).

Results

Identification of ICD-related DEGs

Based on the GSE157103 dataset, 14 ICD-related differentially expressed genes (DEGs) were identified between COVID-19 and control samples. Detailed information on these DEGs were presented in Table 1. Among them, the expression levels of ENTPD1, HMGB1, HSP90AA1, ATG5, CASP8, EIF2AK3, PIK3CA, CASP1, MYD88, and TLR4 were up-regulated, whereas BAX, TNF, CD4, and FOXP3 were down-regulated in COVID-19 samples than that in non-COVID-19 controls (Figure 1A). Next, the correlation between these genes was displayed in Figures 1B,C, and the significant interaction of most genes was observed. Furthermore, Figure 1D also indicated the pattern of the CNV alterations in ICD-related DEGs on their respective chromosomes.

Enrichment analyses for the ICD-related DEGs

We conducted GO, KEGG, Reactome, and DO enrichment analysis on ICD-related DEGs, thus discovering the potential molecular biological characteristics of COVID-19. The GO enrichment analyses indicated that these genes were largely involved in the regulation of cytokines (Figure 2A). KEGG pathway analyses showed that the positive regulation of cytokine production, response to lipopolysaccharide, response to molecules of bacterial origin, and T cell activation were enriched (Figure 2B). Moreover, Reactome terms were enriched in signaling by interleukins, programmed cell death, and regulated necrosis (Figure 2C). Interestingly, the ICD-related DEGs were also markedly enriched in inflammation and tumor-related signatures (Figure 2D). These findings demonstrated that ICD-related DEGs may serve a critical role in the etiology of COVID-19 by engaging in the regulation of cytokines and diverse cell death processes.

Identification of ICD-related diagnostic genes for COVID-19

For exploring the potential pathogenesis of COVID-19, we further evaluated the diagnostic values of ICD-related DEGs. Machine learning algorithms (LASSO, RF, and SVM-RFE) were selected and executed to identify the significant ICD-related DEGs, thus distinguishing COVID-19 from control samples. LASSO regression analysis was performed to select nine feature genes with statistically significant univariate parameters (Figures 3A,B). RF incorporating feature selection was utilized to investigate the relationship between error rate, number of classification trees, and 14 genes with relative importance (Figure 3C). Meanwhile, the respective importance of these genes was also presented in Figure 3D. Next, SVM-RFE analysis indicated that the SVM model based on five feature genes had an optimum generalization performance (Figures 3E,F). The feature genes from the

¹ <https://dgidb.org/>

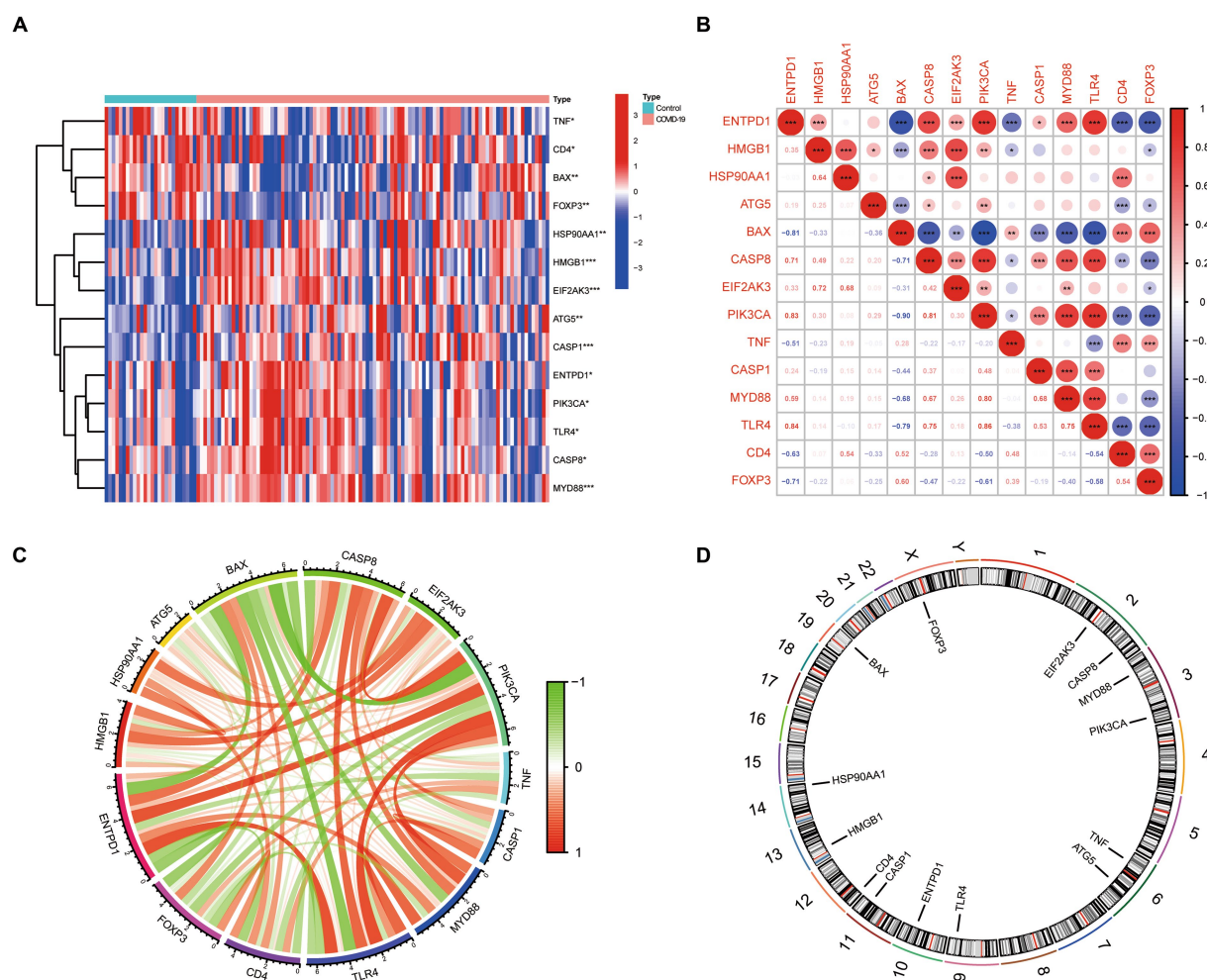


FIGURE 1

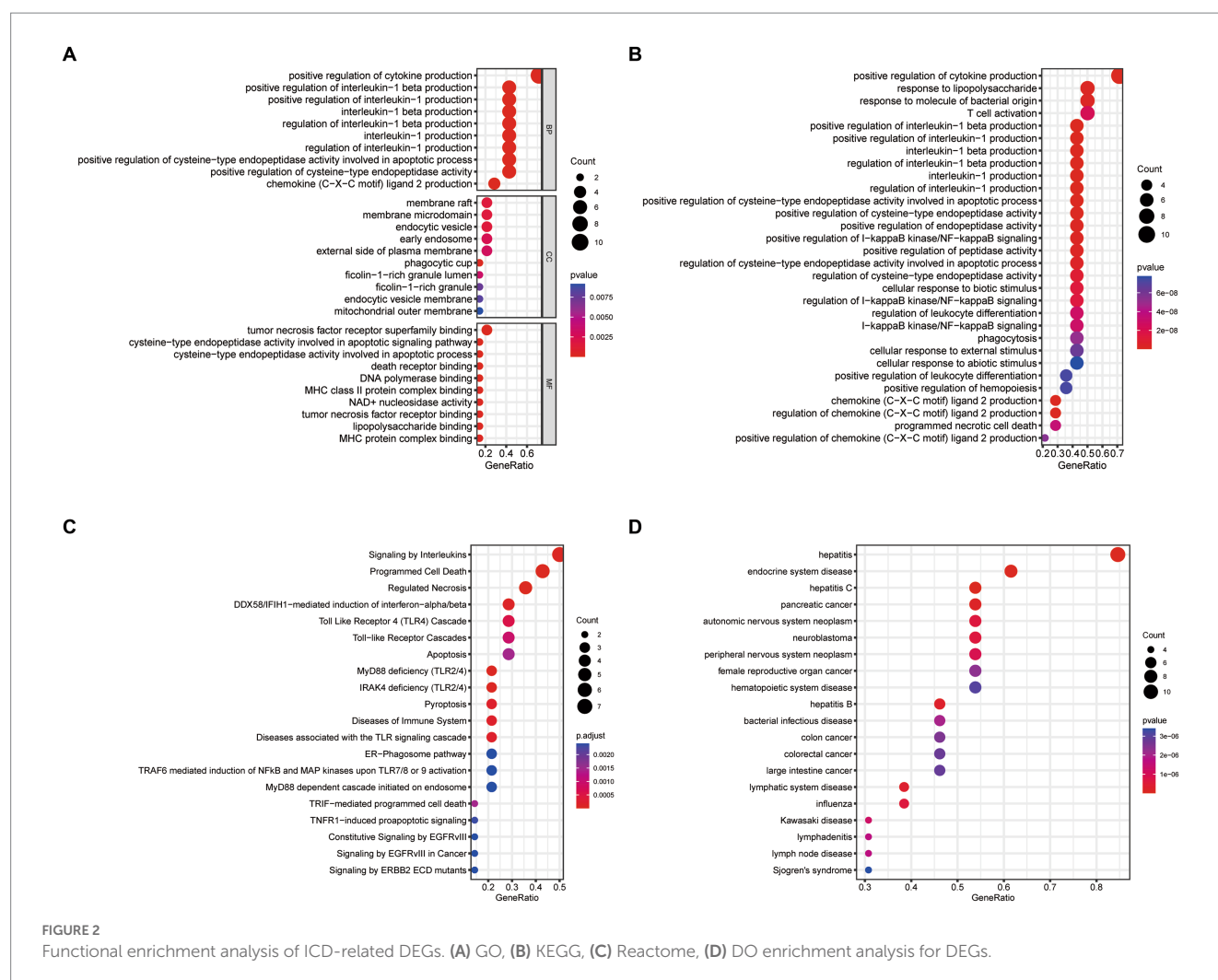
The expression landscape of 14 ICD-related DEGs. (A) Heatmap of DEGs between the COVID-19 samples and control samples. (B,C) The correlation diagram of DEGs. (D) Chromosomal positions of DEGs.

abovementioned algorithms were intersected, and three feature genes (CASP1, CD4, and EIF2AK3) were determined for subsequent analysis (Figure 3G). To reveal the diagnostic accuracy of feature genes in distinguishing COVID-19 from non-COVID-19 controls, ROC curves were plotted for the three feature genes. As depicted in Figure 3H, the AUCs for the three feature genes were higher than 0.6, indicating that these genes have good diagnostic performance. Furthermore, we established a logistic regression model based on the three feature genes, and the ROC curves suggested that the logistic regression model differentiated COVID-19 and control samples with AUC=0.899 (Figure 3I). This evidence demonstrated that the three feature genes can be considered valuable diagnostic genes for COVID-19, and the diagnostic model incorporating the three feature genes may have greater accuracy than individual feature genes.

GSEA and GSVA analysis

To further determine the underlying role of feature genes to identify characteristic differences between COVID-19 samples and control samples, a single-gene GSEA-KEGG pathway analysis was

performed. The top six pathways enriched for individual feature genes were presented in Figures 4A–C. After a systematic analysis, we observed that CASP1 and CD4 were activated in immune response (such as antigen processing and presentation) and RIG-I-like receptor signaling pathway. Meanwhile, EIF2AK3 was closely associated with the metabolism-related activity. Afterward, we found the differentially enriched GSVA pathways between the high- and low-expression groups of each feature gene (Figure 5). The findings indicated that the high expression of CASP1 in COVID-19 may induce this disease by “tyrosine metabolism,” “base excision repair,” and “glyoxylate and dicarboxylate metabolism,” while down-regulation of CASP1 promoted “DORSO_VENTRAL_AXIS_FORMATION,” “O_GLYCAN_BIOSYNTHESIS,” and “VALINE_LEUCINE_AND_ISOLEUCINE_BIOSYNTHESIS.” CD4, whose expression was limited in COVID-19 samples, was significantly correlated with the immune response (“AUTOIMMUNE THYROID DISEASE,” “GRAFT VERSUS HOST DISEASE,” and “ALLOGRAFT REJECTION”). Moreover, the high expression of CD4 activated the pathways such as “GLYCINE SERINE AND THREONINE METABOLISM,” and “PANTOTHENATE AND COA BIOSYNTHESIS.” Similarly, in the EIF2AK3 over-expression group, “GLYCOSAMINOGLYCAN



DEGRADATION,” and metabolism-related pathways (“TAURINE AND HYPOTAURINE METABOLISM,” “PHENYLALANINE METABOLISM,” “TYROSINE METABOLISM”) were enriched. The low-expressed EIF2AK3 served a critical function in the protein export, and biosynthesis process (such as N-GLYCAN biosynthesis and AMINOACYL TRNA biosynthesis).

Immunological infiltration analysis

Based on the CIBERSORT algorithm, the infiltration levels of 22 types of immune cells between COVID-19 and control samples were evaluated. As presented in Figure 6A, compared with the control samples, naïve B cells, follicular helper T cells, Tregs, activated NK cells, monocytes, and activated mast cells were less enriched, while plasma cells, CD4 naïve T cells, activated CD4 memory T cells, $\gamma\delta$ T cells, resting dendritic cells, and activated dendritic cells were more enriched in COVID-19 sample. Subsequently, after performing a correlation analysis of infiltrating immune cells, we identified diverse pairs of interacting immune cells (Figure 6B). Similarly, the proportions of different immune cells between the COVID-19 and control samples had a significant difference (Figure 6C). Furthermore, we also explored the relevance of three diagnostic genes to immune cells. As shown in Figure 6D, Pearson correlation analysis suggested

that activated dendritic cells, neutrophils, activated NK cells, activated CD4 memory T cells and $\gamma\delta$ T cells were all associated with three feature genes (CASP1, CD4, and EIF2AK3). These results suggested that alterations in the immune microenvironment of COVID-19 samples correlated with these three genes.

Prediction of feature gene-targeted drugs

DGIdb was utilized to identify the underlying drug or molecular compounds that could regulate the expression of feature genes in the setting of COVID-19. As demonstrated in the drug-gene interaction network (Figure 7), 33 drugs or molecular compounds targeting feature genes were identified, including 23 for CASP1, 9 for CD4, and 1 for EIF2AK3. The results suggest that CASP1 is more likely to be a valuable target for drug development against COVID-19. Interestingly, the drugs that interacted with different genes were not observed.

A ceRNA network based on diagnostic genes

Recently, the ceRNA network is emerging as a hot topic and serves a pivotal role in controlling gene activity. The ceRNA network was

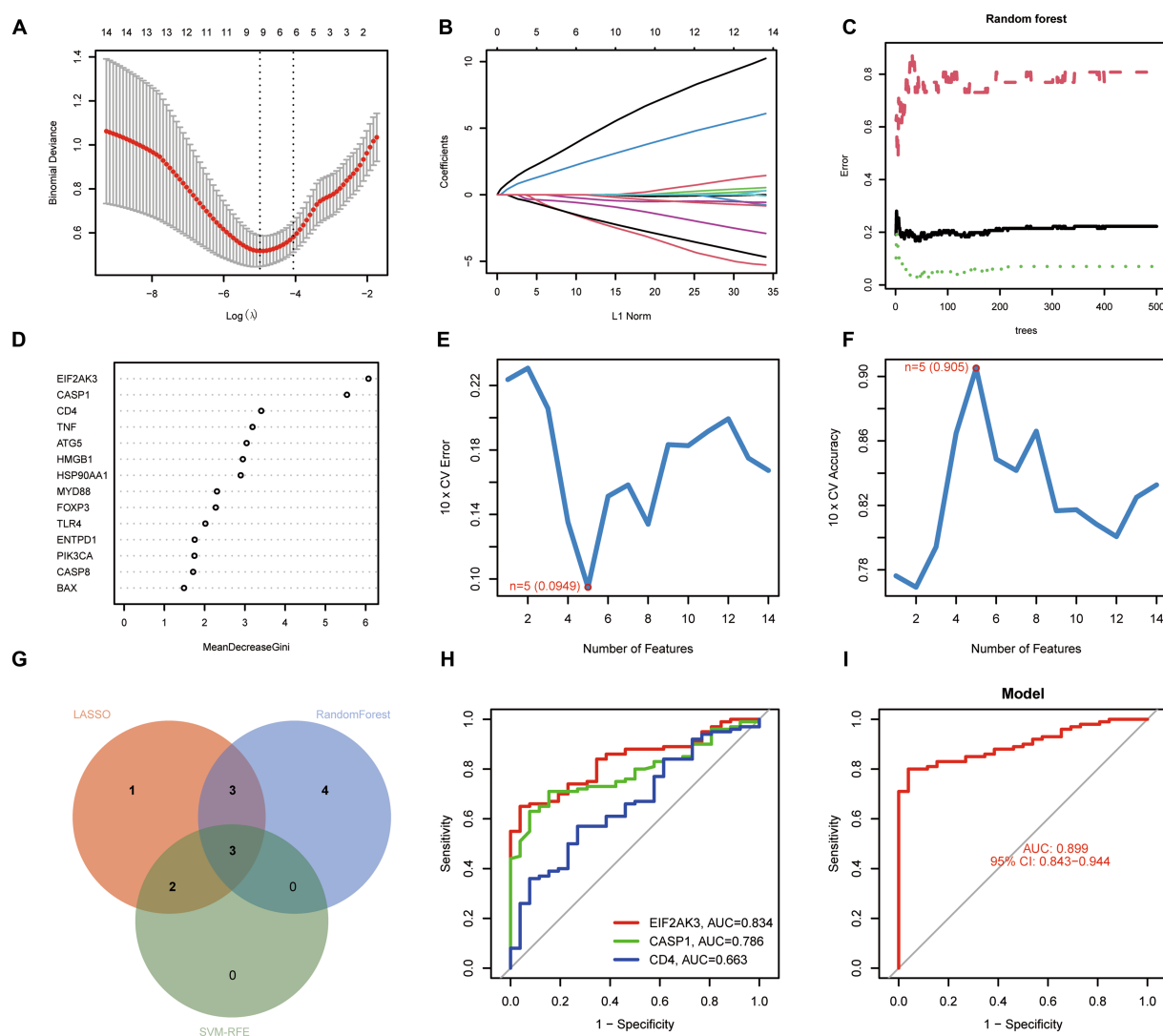


FIGURE 3 Identification of diagnostic genes for COVID-19. **(A,B)** Establishment of the LASSO model. **(C,D)** Screening biomarkers based on random forest (RF) algorithm. **(E,F)** Results of screening biomarkers based on SVM-RFE algorithm. **(G)** Venn diagram showing the intersected genes. **(H)** ROC curves for the feature genes. **(I)** Logistic regression model to identify the AUC of disease samples.

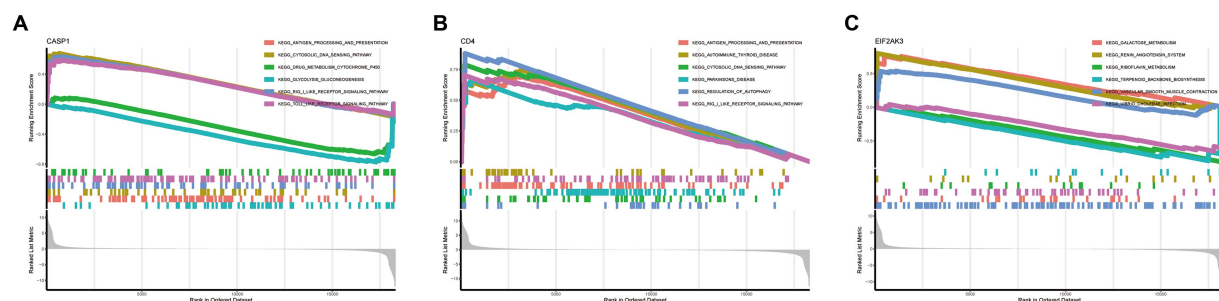
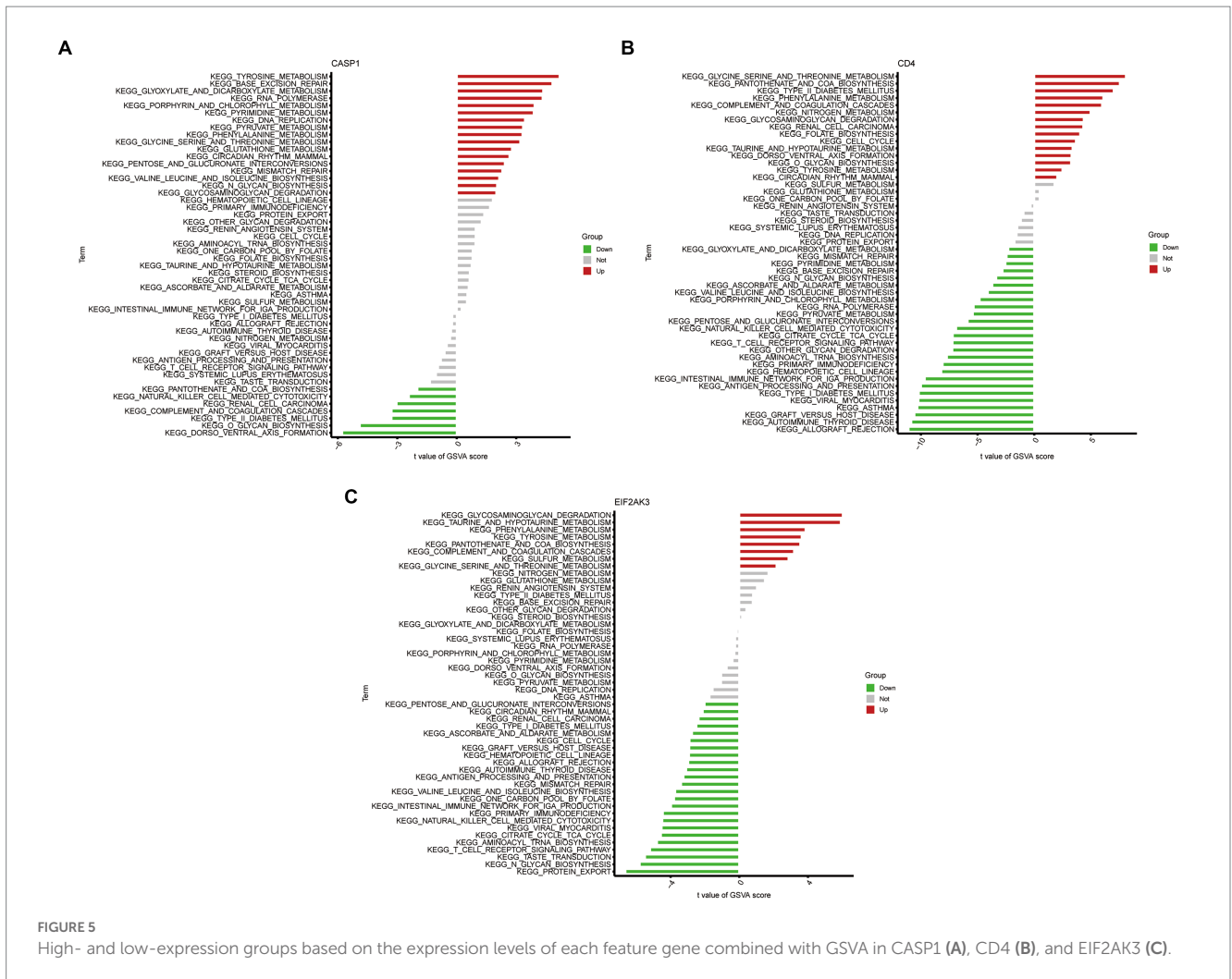


FIGURE 4 Single-gene GSEA-KEGG pathway analysis in CASP1 **(A)**, CD4 **(B)**, and EIF2AK3 **(C)**.

established with miRNA as a bridge to determine the relationship between target gene, mRNA and lncRNA by interacting miRNA response factors. Referring to epigenetic regulators, miRNAs are considered valuable therapeutic options for diverse diseases. The

miRNAs associated with 3 feature genes were explored from three online databases, and the interaction demonstrated by these databases were also identified. A total of 63 miRNAs and 3 mRNAs formed 67 interaction pairs. Additionally, we investigated the lncRNAs that



gene. Hence, the management of COVID-19 patients cannot be implemented by a method, instead diverse methods, including multi-targeted agents and multi-modalities treatment (Fahmi et al., 2021; Rantam et al., 2021). It is essential to explore novel and promising biomarkers for the early diagnosis and management of COVID-19 (Zhong et al., 2020; Gattinoni et al., 2021). Notably, virous cell death forms have been observed to be significantly correlated with the progression of COVID-19, but the association between COVID-19 and ICD has not yet been clarified. Herein, we attempted to identify diagnostic biomarkers pertaining to COVID-19 and delve into the role exerted by ICD within COVID-19.

In this work, we identified 14 ICD-related DEGs between COVID-19 and normal samples, including 10 up-regulated and 4 down-regulated genes. Afterward, functional enrichment analysis of these DEGs was executed, and these genes presented close associations to the regulation of cytokines and cell death signals (such as regulation of cysteine-type endopeptidase activity involved in the apoptotic process, programmed cell death, and regulated necrosis). According to these findings, COVID-19 displayed immunological processes and cell death involvement, thus inducing inflammation and infection.

We applied three different machine learning algorithms (LASSO, RF, and SVM-RFE), each with its advantages. LASSO regression

COVID-19 has been ravaging the world for years, and its specific cause remains unclear. The mutational and variable nature of COVID-19 renders it impossible to explain it with changes in a single

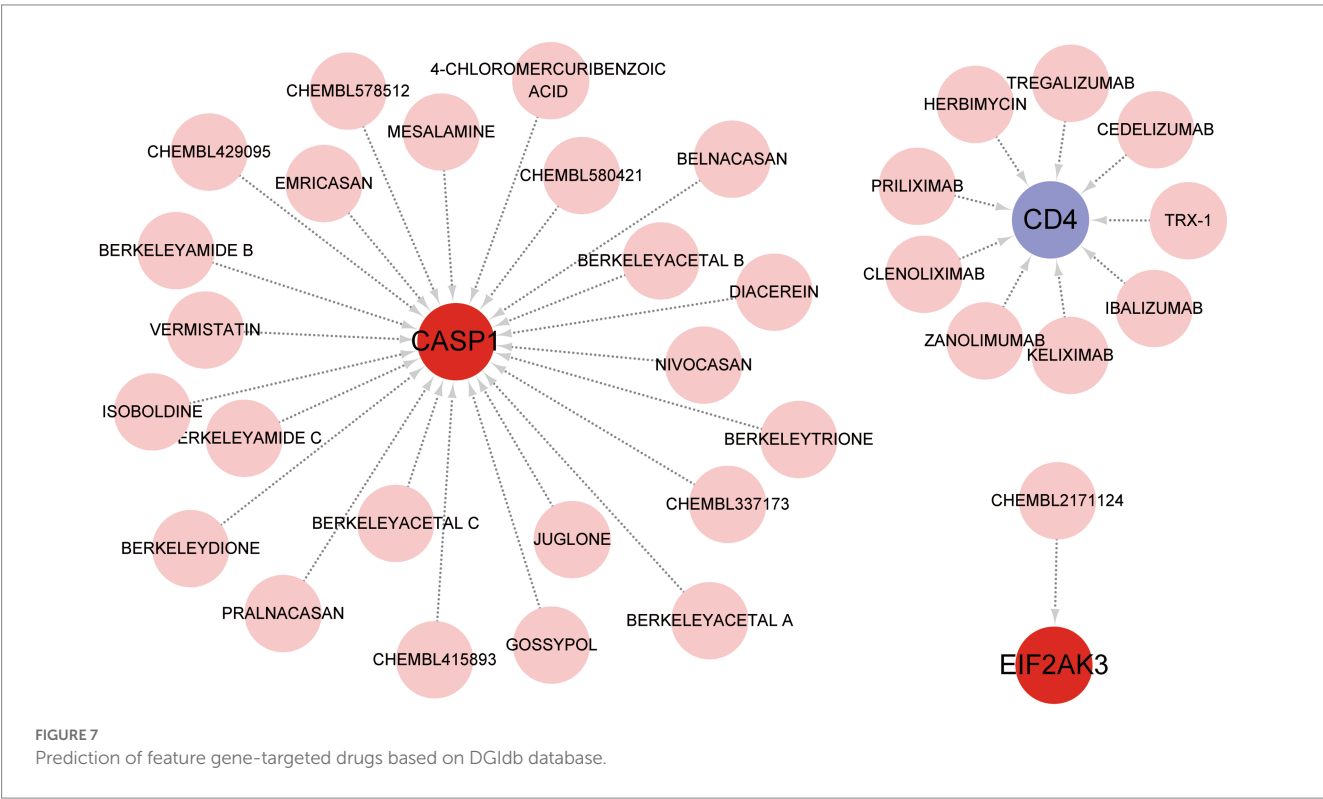
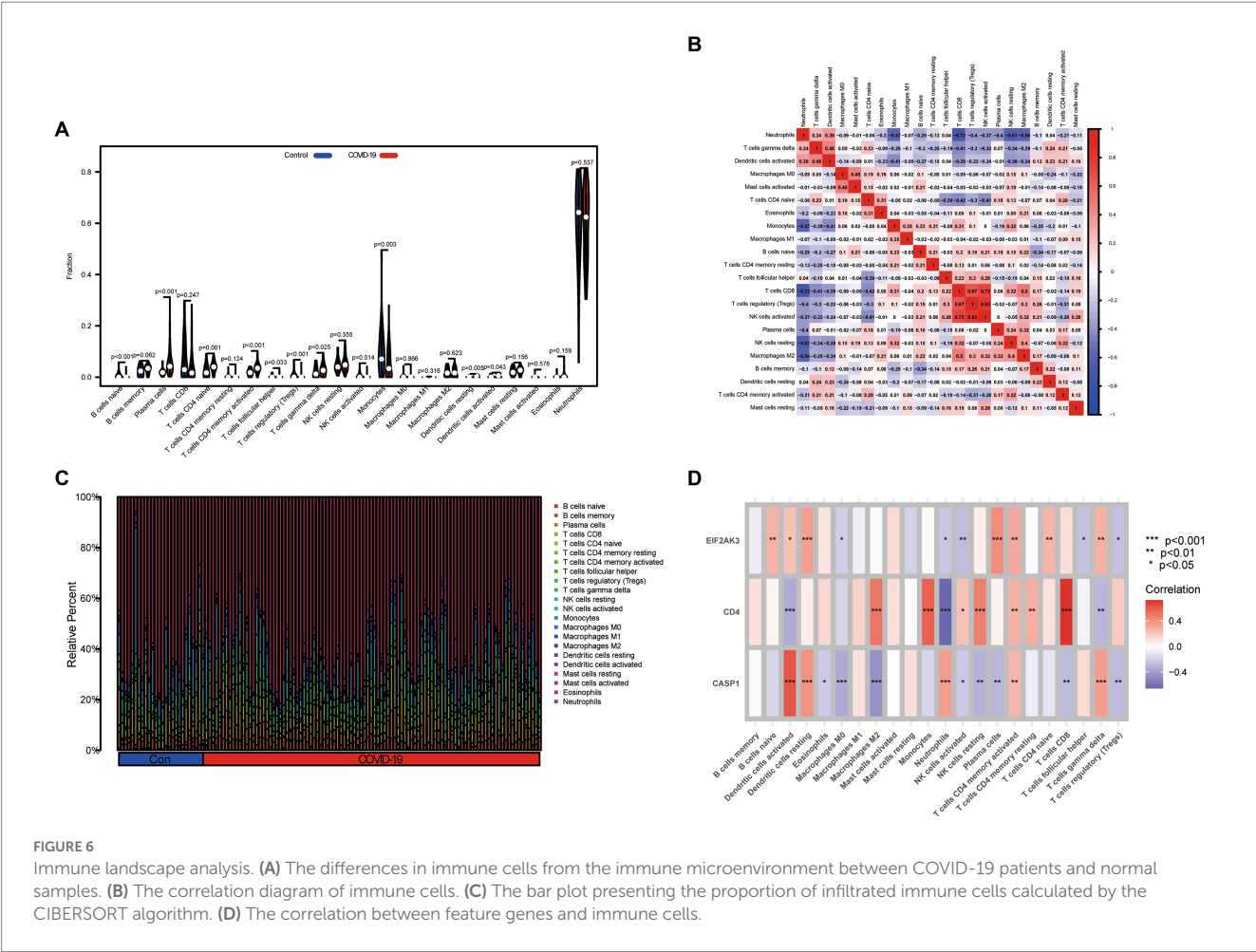




FIGURE 8
A ceRNA network based on three feature genes.

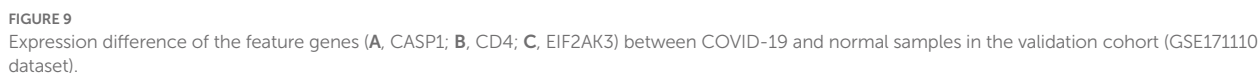


FIGURE 9
Expression difference of the feature genes (**A**, CASP1; **B**, CD4; **C**, EIF2AK3) between COVID-19 and normal samples in the validation cohort (GSE171110 dataset).

were identified and were valuable for further explorations, which demonstrated that our prediction displayed reliability with the integrated approaches.

CASP1, a member of the caspase family, serves a critical role in the execution process of cell apoptosis (Kapplusch et al., 2019). The expression of CASP1 was regulated, and the inflammasome

activated by CASP1-NLRP3 dependent pathway can subsequently secrete IL-1 β and IL-18, thus leading to dysfunctional autophagy (Wang et al., 2017). Upregulation of CASP1 in COVID-19 was reported in a prior report, consistent with our findings (Yang et al., 2022). CD4 encodes the CD4 membrane glycoprotein of T lymphocytes, and CD4+ T cell plays a critical role in anti-viral immunity (Matthias et al., 2002; Sharma et al., 2005). CD4 mediates various biological processes of immune cells. Meanwhile, immune cells, primarily comprising T cells, B cells, and macrophages, crucially affect the pathogenesis of COVID-19 (Tay et al., 2020). EIF2AK3, as a metabolic stress-sensing protein kinase, inhibits protein translation and regulated pro-survival autophagy (You et al., 2021). These genes displayed significantly different expressions between COVID-19 and normal samples, and this result was also observed in the validation cohort. Furthermore, GSEA and GSVA for three feature genes were performed, and we observed that these genes were significantly associated with immune response and metabolism-related activities. Based on the abovementioned findings, CASP1, CD4, and EIF2AK3 have the potential to influence the progression of COVID-19 and serve as diagnostic biomarkers, whereas this finding remains to be confirmed by further clinical trials.

To more comprehensively investigate the effects played by the abundance of immune cells in COVID-19, this research conducted CIBERSORT for evaluating the infiltrating immune status within COVID-19. The abundance of naïve B cells, follicular helper T cells, Tregs, activated NK cells, monocytes, and activated mast cells decreased, while the abundance of plasma cells, CD4 naïve T cells, activated CD4 memory T cells, $\gamma\delta$ T cells, resting dendritic cells, and activated dendritic cells increased, potentially demonstrating the relevance of COVID-19 initiation and development. It is common knowledge that these cells are an essential part of human adaptive immunity, and these cells may be potential factors for COVID-19 pathogenesis (Sette and Crotty, 2021). Meanwhile, three feature genes presented a significant correlation with diverse immune cells, which also confirmed the autoimmunogenicity of these genes. Although this is widely recognized, the molecular mechanisms and functions of immune cell infiltration in COVID-19 need to be further investigated urgently.

Finally, the feature gene for gene-targeted drugs and the ceRNA network were explored. A diagram of 33 drugs with promising therapeutic efficacy against COVID-19 was displayed. The performance of the selected drugs was validated for diverse diseases; however, the mechanisms of many drugs were unclear. Further molecular experiments and clinical investigations are necessary to identify the drugs that are valuable for COVID-19 treatment. The target miRNAs and the target lncRNAs were predicted for CASP1, CD4, and EIF2AK3, and a ceRNA network was established with Cytoscape. This network proposes a potential mechanism for selected genes to be regulated at the transcriptome level.

Inevitably, these are some restrictions of this study. Firstly, this study was performed on public databases, and additional laboratory experiments are necessary. Secondly, the association between these results in this study and clinical variables was worthy of future analysis and validation. However, this study could still reflect convincing evidence to further investigate the significance of ICD-related genes in the treatment and diagnosis of COVID-19.

Conclusion

Taken together, we presented that differentially expressed genes: CASP1, CD4, and EIF2AK3 might be considered as the diagnostic biomarker for COVID-19 and characteristic immunological infiltration could influence the systemic immune microenvironment in the pathogenesis of COVID-19. This work may reveal potential biomarkers related to immune regulation that ultimately led to COVID-19 infection. Additional leading-edge approaches, such as single-cell sequencing method, will further offer promising perspectives into COVID-19 development and drug targets.

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.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

JZ and CY: conceptualization. JZ and KW: methodology. ZS: software and data curation. JZ, KW, and ZS: validation. CY: formal analysis, writing—review and editing, and project administration. JZ: investigation, writing original draft preparation, visualization, and supervision. KW: resources. All authors contributed to the article and approved the submitted version.

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.

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

- Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *elife* 4:e05005. doi: 10.7554/eLife.05005
- Araf, Y., Akter, F., Tang, Y. D., Fatemi, R., Parvez, M. S. A., Zheng, C., et al. (2022). Omicron variant of SARS-CoV-2: genomics, transmissibility, and responses to current COVID-19 vaccines. *J. Med. Virol.* 94, 1825–1832. doi: 10.1002/jmv.27588
- Badua, C., Baldo, K. A. T., and Medina, P. M. B. (2021). Genomic and proteomic mutation landscapes of SARS-CoV-2. *J. Med. Virol.* 93, 1702–1721. doi: 10.1002/jmv.26548
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193
- Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi: 10.1007/978-1-4939-7493-1_12
- Chen, Y., and Wang, X. (2020). miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 48, D127–D131. doi: 10.1093/nar/gkz757
- Engelbrechtsen, S., and Bohlin, J. (2019). Statistical predictions with glmnet. *Clin. Epigenet.* 11:123. doi: 10.1186/s13148-019-0730-1
- Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol.* 5:R1. doi: 10.1186/gb-2003-5-1-r1
- Fahmi, M., Kharisma, V. D., Ansori, A. N. M., and Ito, M. (2021). Retrieval and investigation of data on SARS-CoV-2 and COVID-19 using bioinformatics approach. *Adv. Exp. Med. Biol.* 1318, 839–857. doi: 10.1007/978-3-030-63761-3_47
- Fujita, A., Sato, J. R., Demasi, M. A., Sogayar, M. C., Ferreira, C. E., and Miyano, S. (2009). Comparing Pearson, spearman and Hoeffding's D measure for gene expression association analysis. *J. Bioinform. Comput. Biol.* 7, 663–684. doi: 10.1142/s0219720009004230
- Furio-Tari, P., Tarazona, S., Gabaldon, T., Enright, A. J., and Conesa, A. (2016). spongeScan: a web for detecting microRNA binding elements in lncRNA sequences. *Nucleic Acids Res.* 44, W176–W180. doi: 10.1093/nar/gkw443
- Galluzzi, L., Buque, A., Kepp, O., Zitvogel, L., and Kroemer, G. (2017). Immunogenic cell death in cancer and infectious disease. *Nat. Rev. Immunol.* 17, 97–111. doi: 10.1038/nri.2016.107
- Galluzzi, L., Vitale, I., Warren, S., Adjemian, S., Agostinis, P., Martinez, A. B., et al. (2020). Consensus guidelines for the definition, detection and interpretation of immunogenic cell death. *J. Immunother. Cancer* 8:e000337. doi: 10.1136/jitc-2019-000337
- Gattinoni, L., Gattarello, S., Steinberg, I., Busana, M., Palermo, P., Lazzari, S., et al. (2021). COVID-19 pneumonia: pathophysiology and management. *Eur. Respir. Rev.* 30:210138. doi: 10.1183/16000617.0138-2021
- Good, B. M., Van Aukun, K., Hill, D. P., Mi, H., Carbon, S., Balhoff, J. P., et al. (2021). Reactome and the gene ontology: digital convergence of data resources. *Bioinformatics* 37, 3343–3348. doi: 10.1093/bioinformatics/btab325
- Hadj, H. I. (2022). Covid-19 vaccines and variants of concern: a review. *Rev. Med. Virol.* 32:e2313. doi: 10.1002/rmv.2313
- Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* 14:7. doi: 10.1186/1471-2105-14-7
- Islam, K. U., and Iqbal, J. (2020). An update on molecular diagnostics for COVID-19. *Front. Cell. Infect. Microbiol.* 10:560616. doi: 10.3389/fcimb.2020.560616
- Ito, K., and Murphy, D. (2013). Application of ggplot2 to Pharmacometric graphics. *CPT Pharmacometrics Syst. Pharmacol.* 2:e79. doi: 10.1038/psp.2013.56
- Jiang, H., Gu, J., Du, J., Qi, X., Qian, C., and Fei, B. (2020). A 21-gene support vector machine classifier and a 10-gene risk score system constructed for patients with gastric cancer. *Mol. Med. Rep.* 21, 347–359. doi: 10.3892/mmr.2019.10841
- Kai-Wang To, K., Sridhar, S., Chiu, K. H., Hung, D. L., Li, X., Hung, I. F., et al. (2021). Lessons learned 1 year after SARS-CoV-2 emergence leading to COVID-19 pandemic. *Emerg. Microbes. Infect.* 10, 507–535. doi: 10.1080/22221751.2021.1898291
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092
- Kang, Y., and Wang, Q. (2022). Potential therapeutic value of necroptosis inhibitor for the treatment of COVID-19. *Eur. J. Med. Res.* 27:283. doi: 10.1186/s40001-022-00913-7
- Kapplusch, F., Schulze, F., Rabe-Matschewsky, S., Russ, S., Herbig, M., Heymann, M. C., et al. (2019). CASP1 variants influence subcellular caspase-1 localization, pyroptosome formation, pro-inflammatory cell death and macrophage deformability. *Clin. Immunol.* 208:108232. doi: 10.1016/j.clim.2019.06.008
- Kroemer, G., Galassi, C., Zitvogel, L., and Galluzzi, L. (2022). Immunogenic cell stress and death. *Nat. Immunol.* 23, 487–500. doi: 10.1038/s41590-022-01132-2
- Lee, S., Channappanavar, R., and Kanneganti, T. D. (2020). Coronaviruses: innate immunity, inflammasome activation, inflammatory cell death, and cytokines. *Trends Immunol.* 41, 1083–1099. doi: 10.1016/j.it.2020.10.005
- Lee, Y. W., Choi, J. W., and Shin, E. H. (2021). Machine learning model for predicting malaria using clinical information. *Comput. Biol. Med.* 129:104151. doi: 10.1016/j.combiomed.2020.104151
- Levy, Y., Wiedemann, A., Hejblum, B. P., Durand, M., Lefebvre, C., Surenau, M., et al. (2021). CD177, a specific marker of neutrophil activation, is associated with coronavirus disease 2019 severity and death. *iScience* 24:102711. doi: 10.1016/j.isci.2021.102711
- Li, Y., Liu, X., Zhang, X., Pan, W., Li, N., and Tang, B. (2021). Immunogenic cell death inducers for enhanced cancer immunotherapy. *Chem. Commun. (Camb.)* 57, 12087–12097. doi: 10.1039/d1cc04604g
- Matthias, L. J., Yam, P. T., Jiang, X. M., Vandegraaff, N., Li, P., Pombourios, P., et al. (2002). Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. *Nat. Immunol.* 3, 727–732. doi: 10.1038/ni815
- Minute, L., Teixeira, A., Sanchez-Paulete, A. R., Ochoa, M. C., Alvarez, M., Otano, I., et al. (2020). Cellular cytotoxicity is a form of immunogenic cell death. *J. Immunother. Cancer* 8:e000325. doi: 10.1136/jitc-2019-000325
- Nishiga, M., Jahng, J. W. S., and Wu, J. C. (2022). Ferroptosis of pacemaker cells in COVID-19. *Circ. Res.* 130, 978–980. doi: 10.1161/CIRCRESAHA.122.320951
- Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., Bernstein, M. N., Peters-Clarke, T. M., et al. (2021). Large-scale multi-omic analysis of COVID-19 severity. *Cell Syst.* 12, 23–40.e7. doi: 10.1016/j.cels.2020.10.003
- Park, S. H., Goo, J. M., and Jo, C. H. (2004). Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J. Radiol.* 5, 11–18. doi: 10.3348/kjr.2004.5.1.11
- Rantam, F. A., Kharisma, V. D., Sumartono, C., Nugraha, J., Wijaya, A. Y., Susilowati, H., et al. (2021). Molecular docking and dynamic simulation of conserved B cell epitope of SARS-CoV-2 glycoprotein Indonesian isolates: an immunoinformatic approach. *F1000Res* 10:813. doi: 10.12688/f1000research.54258.1
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Robshaw, J. D., Alter, S. M., Solano, J. J., Shih, R. D., DL, D. M., Maki, D. G., et al. (2021). Genomic surveillance to combat COVID-19: challenges and opportunities. *Lancet Microbe* 2, e481–e484. doi: 10.1016/S2666-5247(21)00121-X
- Safiabadi Tali, S. H., LeBlanc, J. J., Sadiq, Z., Oyewunmi, O. D., Camargo, C., Nikpour, B., et al. (2021). Tools and techniques for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)/COVID-19 detection. *Clin. Microbiol. Rev.* 34:e00228-20. doi: 10.1128/CMR.00228-20
- Sanz, H., Valim, C., Vegas, E., Oller, J. M., and Reverter, F. (2018). SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinform.* 19:432. doi: 10.1186/s12859-018-2451-4
- Schriml, L. M., Munro, J. B., Schor, M., Olley, D., McCracken, C., Felix, V., et al. (2022). The human disease ontology 2022 update. *Nucleic Acids Res.* 50, D1255–D1261. doi: 10.1093/nar/gkab1063
- Sette, A., and Crotty, S. (2021). Adaptive immunity to SARS-CoV-2 and COVID-19. *Cells* 184, 861–880. doi: 10.1016/j.cell.2021.01.007
- 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
- Sharma, D., Balamurali, M. M., Chakraborty, K., Kumaran, S., Jeganathan, S., Rashid, U., et al. (2005). Protein minimization of the gp120 binding region of human CD4. *Biochemistry* 44, 16192–16202. doi: 10.1021/bi051120s
- Simsekler, M. C. E., Alhashmi, N. H., Azar, E., King, N., Luqman, R., and Al, M. A. (2021). Exploring drivers of patient satisfaction using a random forest algorithm. *BMC Med. Inform. Decis. Mak.* 21:157. doi: 10.1186/s12911-021-01519-5
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Tay, M. Z., Poh, C. M., Renia, L., MacAry, P. A., and Ng, L. F. P. (2020). The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* 20, 363–374. doi: 10.1038/s41577-020-0311-8
- The Gene Ontology C (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338. doi: 10.1093/nar/gky1055
- Tibshirani, R. (1997). The lasso method for variable selection in the cox model. *Stat. Med.* 16, 385–395. doi: 10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3
- van Eijk, L. E., Binkhorst, M., Bourgonje, A. R., Offringa, A. K., Mulder, D. J., Bos, E. M., et al. (2021). COVID-19: immunopathology, pathophysiological mechanisms, and treatment options. *J. Pathol.* 254, 307–331. doi: 10.1002/path.5642
- Wagner, A. H., Coffman, A. C., Ainscough, B. J., Spies, N. C., Skidmore, Z. L., Campbell, K. M., et al. (2016). DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res.* 44, D1036–D1044. doi: 10.1093/nar/gkv1165

- Wang, X., Wu, S., Liu, F., Ke, D., Wang, X., Pan, D., et al. (2021). An immunogenic cell death-related classification predicts prognosis and response to immunotherapy in head and neck squamous cell carcinoma. *Front. Immunol.* 12:781466. doi: 10.3389/fimmu.2021.781466
- Wang, D., Zhang, J., Jiang, W., Cao, Z., Zhao, F., Cai, T., et al. (2017). The role of NLRP3-CASP1 in inflammasome-mediated neuroinflammation and autophagy dysfunction in manganese-induced, hippocampal-dependent impairment of learning and memory ability. *Autophagy* 13, 914–927. doi: 10.1080/15548627.2017.1293766
- Warwick Vesztrocy, A., and Dessimoz, C. (2020). Benchmarking gene ontology function predictions using negative annotations. *Bioinformatics* 36, i210–i218. doi: 10.1093/bioinformatics/btaa466
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi: 10.1093/nar/gkx1037
- Wu, X., Sui, Z., Zhang, H., Wang, Y., and Yu, Z. (2020). Integrated analysis of lncRNA-mediated ceRNA network in lung adenocarcinoma. *Front. Oncol.* 10:554759. doi: 10.3389/fonc.2020.554759
- Yang, C. A., Huang, Y. L., and Chiang, B. L. (2022). Innate immune response analysis in COVID-19 and Kawasaki disease reveals MIS-C predictors. *J. Formos. Med. Assoc.* 121, 623–632. doi: 10.1016/j.jfma.2021.06.009
- Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., et al. (2020). Study of cardiovascular disease prediction model based on random forest in eastern China. *Sci. Rep.* 10:5245. doi: 10.1038/s41598-020-62133-5
- You, K., Wang, L., Chou, C. H., Liu, K., Nakata, T., Jaiswal, A., et al. (2021). QRICH1 dictates the outcome of ER stress through transcriptional control of proteostasis. *Science* 371:eabb6896. doi: 10.1126/science.abb6896
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Yuce, M., Filiztekin, E., and Ozkaya, K. G. (2021). COVID-19 diagnosis -a review of current methods. *Biosens. Bioelectron.* 172:112752. doi: 10.1016/j.bios.2020.112752
- Zhang, R., Sun, C., Han, Y., Huang, L., Sheng, H., Wang, J., et al. (2023). Neutrophil autophagy and NETosis in COVID-19: perspectives. *Autophagy* 19, 758–767. doi: 10.1080/15548627.2022.2099206
- Zhong, P., Xu, J., Yang, D., Shen, Y., Wang, L., Feng, Y., et al. (2020). COVID-19-associated gastrointestinal and liver injury: clinical features and potential mechanisms. *Signal Transduct. Target. Ther.* 5:256. doi: 10.1038/s41392-020-00373-7



OPEN ACCESS

EDITED BY

Arli Aditya Parikesit,
Indonesia International Institute for
Life-Sciences (i3L), Indonesia

REVIEWED BY

Surjit Singh,
Sister Nivedita University, India
Anand Anbarasu,
VIT University, India

*CORRESPONDENCE

Tuna Toptan
✉ tuna.toptangrabmair@kgu.de
Rainer Koenig
✉ rainer.koenig@uni-jena.de

[†]These authors have contributed equally to this work and share last authorship

RECEIVED 24 March 2023

ACCEPTED 16 May 2023

PUBLISHED 05 June 2023

CITATION

Kelch MA, Vera-Guapi A, Beder T, Oswald M, Hiemisch A, Beil N, Wajda P, Ciesek S, Erfle H, Toptan T and Koenig R (2023) Machine learning on large scale perturbation screens for SARS-CoV-2 host factors identifies β -catenin/CBP inhibitor PRI-724 as a potent antiviral. *Front. Microbiol.* 14:1193320. doi: 10.3389/fmicb.2023.1193320

COPYRIGHT

© 2023 Kelch, Vera-Guapi, Beder, Oswald, Hiemisch, Beil, Wajda, Ciesek, Erfle, Toptan and Koenig. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Machine learning on large scale perturbation screens for SARS-CoV-2 host factors identifies β -catenin/CBP inhibitor PRI-724 as a potent antiviral

Maximilian A. Kelch¹, Antonella Vera-Guapi², Thomas Beder³, Marcus Oswald⁴, Alicia Hiemisch⁴, Nina Beil⁵, Piotr Wajda⁵, Sandra Ciesek^{1,6,7}, Holger Erfle⁵, Tuna Toptan^{1*†} and Rainer Koenig^{4*†}

¹Institute for Medical Virology, University Hospital Frankfurt, Goethe University Frankfurt, Frankfurt, Germany, ²Institute of Biochemistry II, University Hospital, Frankfurt, Germany, ³Medical Department II, Hematology and Oncology, University Hospital Schleswig-Holstein, Kiel, Germany, ⁴Institute for Infectious Diseases and Infection Control, Jena University Hospital, Jena, Germany, ⁵Advanced Biological Screening Facility (ABSF), High-Content Analysis of the Cell (HiCell), BioQuant, Heidelberg University, Heidelberg, Germany, ⁶German Centre for Infection Research (DZIF), External Partner Site Frankfurt, Frankfurt, Germany, ⁷Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Frankfurt, Germany

Expanding antiviral treatment options against SARS-CoV-2 remains crucial as the virus evolves under selection pressure which already led to the emergence of several drug resistant strains. Broad spectrum host-directed antivirals (HDA) are promising therapeutic options, however the robust identification of relevant host factors by CRISPR/Cas9 or RNA interference screens remains challenging due to low consistency in the resulting hits. To address this issue, we employed machine learning, based on experimental data from several knockout screens and a drug screen. We trained classifiers using genes essential for virus life cycle obtained from the knockout screens. The machines based their predictions on features describing cellular localization, protein domains, annotated gene sets from Gene Ontology, gene and protein sequences, and experimental data from proteomics, phospho-proteomics, protein interaction and transcriptomic profiles of SARS-CoV-2 infected cells. The models reached a remarkable performance suggesting patterns of intrinsic data consistency. The predicted HDF were enriched in sets of genes particularly encoding development, morphogenesis, and neural processes. Focusing on development and morphogenesis-associated gene sets, we found β -catenin to be central and selected PRI-724, a canonical β -catenin/CBP disruptor, as a potential HDA. PRI-724 limited infection with SARS-CoV-2 variants, SARS-CoV-1, MERS-CoV and IAV in different cell line models. We detected a concentration-dependent reduction in cytopathic effects, viral RNA replication, and infectious virus production in SARS-CoV-2 and SARS-CoV-1-infected cells. Independent of virus infection, PRI-724 treatment caused cell cycle deregulation which substantiates its potential as a broad spectrum antiviral. Our proposed machine learning concept supports focusing and accelerating the discovery of host dependency factors and identification of potential host-directed antivirals.

KEYWORDS

coronavirus, SARS-CoV-2, influenza A virus, host dependency factors, CRISPR/Cas9 knockout screen, beta-catenin, machine learning, PRI-724

1. Introduction

Genetic plasticity of SARS-CoV-2 and herd immunity-derived evolutionary pressure have led to the emergence of variants of concern (VOC; Van Egeren et al., 2021; Flores-Vega et al., 2022). Some of the variants and recombinant lineages have evaded vaccine-elicited antibodies and monoclonal antibody-based treatments (Dejnirattisai et al., 2021; Planas et al., 2021; Wang P. et al., 2021; Iketani et al., 2022). Although antiviral drugs such as Remdesivir, Molnupiravir, or Nirmatrelvir/Ritonavir (Paxlovid) have significantly improved disease outcome in patients (Beigel et al., 2020; Hammond et al., 2022; Jayk Bernal et al., 2022; Wong et al., 2022), these treatments might select for more resistant variants over time (Gandhi et al., 2022; Hogan et al., 2023). Consequently, our current virus-directed countermeasures have limitations and there is an urgent need for developing host-directed antivirals (HDA) that can target host-dependency factors (HDF) required for the life cycle of the virus. HDA might provide variant-independent treatment opportunities to reduce disease severity and complement virus-directed antiviral therapies (Edinger TO et al., 2014; Mahajan et al., 2021; Wagoner et al., 2022). In addition, drug repurposing can offer an expedited timeline to bring HDA therapies into clinical settings in a cost-effective and timely manner. Still, a common challenge is the robust identification of relevant host factors among varying *in vitro* conditions, although recent advances in functional genomics have simplified the use of high-throughput technology platforms based on CRISPR/Cas9 and siRNAs (Daniloski et al., 2021; Wang R. et al., 2021; Wei et al., 2021; Zhu et al., 2021). However, when comparing the results of these experimental screens for SARS-CoV-2 host dependency factors, there is only a marginal overlap of hits. This may be due to different viral strains, host cells and/or different multiplicity of infection (MOI) or further different experimental settings such as the selection of knockout library constructs used. Interestingly, when grouping the identified HDF into gene sets of common function or cellular processes, the consistency increases suggesting that these lists of HDF in their specific contexts contain consistent patterns on a more complex level.

Machine learning methods can identify such common patterns in experimental data, as, e.g., the identification of common regulators explaining large scale transcription profiles (Alipanahi et al., 2015; Hörhold et al., 2020) and allow for integration of a broad variety of omics and genomics data. In this study, we followed a machine learning approach to identify HDF of SARS-CoV-2. From the knockout screens (Daniloski et al., 2021; Wang R. et al., 2021; Wei

et al., 2021; Zhu et al., 2021), we assembled a gold standard of genes required for virus replication. With this gold standard, we trained and evaluated the classifier. The classifier based its predictions on features of the genes describing their encoded proteins' cellular localization, protein domains, annotated cellular processes, and gene/protein sequences. Furthermore, we embedded experimental data comprising proteomic, PPI, and transcriptomic profiles of SARS-CoV-2 infected cells. In addition, another classifier was trained based on a gold standard derived from a drug screen (Ellinger et al., 2021), in which more than 6,000 drugs were screened to determine whether they had a protective effect on cells infected with SARS-CoV-2. Both classifiers delivered a list of predicted HDF. For a few, well selected HDF, we interrogated drug databases, and identified β -catenin and its small molecule inhibitor PRI-724, particularly antagonizing the interaction between β -catenin and CBP.

The β -catenin/CBP interaction displays a pivotal part of a switch-like signaling network within the Wnt/ β -catenin pathway that controls the well-balanced interplay between cell proliferation and differentiation (Akiyama, 2000; Liu et al., 2022). Activation of Wnt signaling leads to a sequester of the β -catenin destruction complex enabling cytosolic accumulation of β -catenin and, in turn, its nuclear translocation where it interacts with CBP or p300 to drive either proliferation or differentiation, respectively (Bordonaro and Lazarova, 2016). In this study, we identified β -catenin as a SARS-CoV-2 host factor using machine-learning and sought to analyze the efficacy of one of its inhibitors, PRI-724, against different pandemic-related RNA viruses *in vitro*.

2. Materials and methods

2.1. Gene prioritizations

Two sets of machines were set up to predict HDF, one based on data from CRISPR/Cas9 based knockout screens, and one on data from a drug screen. For defining the gold standard for the first set of machines, the experimental results from four genome-wide CRISPR/Cas9 knockout screens were considered (Daniloski et al., 2021; Wang R. et al., 2021; Wei et al., 2021; Zhu et al., 2021; Table 1). The ranking of the screened genes was taken from the Supplementary material of the respective studies. The rank product was calculated for each gene for which information was available. The $n = 500$ top ranking genes of this new list were selected and used as the positive class for the

TABLE 1 HDF screens serving for a gold standard to train and validate the machines.

	Daniloski et al. (2021)	Wei et al. (2021)	Wang R. et al. (2021)	Zhu et al. (2021)
Number of screened genes	19,050	21,673	20,915	19,114
Cell line	A549-ACE2	Vero-E6	Huh7.5.1 with ACE2 and TMPRSS2 overexpression	A549-ACE2
SARS-CoV-2 strain	USA-WA1/2020 (NR-52281)	US-WA1/2020	USA/WA-1/2020	nCoV-SH01 (patient isolate)
MOI	0.01 (low) 0.3 (high)	0.1	0.01	0.3
CRISPR libraries	GeCKOv2 with 122,411 sgRNAs 6 sgRNAs/gene	<i>C. saeba</i> us with 83,963 sgRNAs 4/sgRNAs gene	GeCKOv2	lentiCRISPRv2 76,441 sgRNAs

classifier, labeled as HDF. The bottom 1,000 genes were used as the negative class (“Non-HDF”). The rest of the genes in the list were used for prediction, i.e., assessing an HDF or non-HDF label. Daniloski et al. provided data of two screens with different MOI (0.01 and 0.3). To get a combined list of genes, we calculated the rank product for each gene listed in their score tables and used this combined ranking for our study. In addition, a second gold standard was established based on the screening of 5,632 compounds and their capacity to inhibit the cytopathic effect (CPE) of SARS-CoV-2 in a Caco-2 cell line (Ellinger et al., 2021). In line with the authors of the original study, compounds that blocked viral CPE by 75% were considered as hits. This led to 273 compounds inhibiting SARS-CoV-2 infection without inducing toxicity. Then, their respective targets were identified using the information of different drug databases [Drugbank (Wishart et al., 2018), ChEMBL (Mendez et al., 2019), TTD (Wang et al., 2020), PharmGKB (Thorn et al., 2013), and BindingDB (Gilson et al., 2016)]. By this, 178 gene targets were found and labeled as “HDF.” For the remaining drugs with no inhibitory effect, their 1,881 targets were labeled as “non-HDF.” Effective drugs with more than 11 targets were considered to be “promiscuous,” and their targets were not included in the gold standard.

2.2. Feature generation

The procedure for feature generation was similar as published earlier for essential gene prediction (Acencio and Lemke, 2009; Aromolaran et al., 2020; Beder et al., 2021). Each gene served as a sample in the machine learning procedure, either labeled as HDF or non-HDF, or was not used for training the classifiers. For each gene, we generated a comprehensive set of 60,593 features comprising the seven categories of (1) gene sequence, (2) protein sequence, (3) protein domains, (4) gene sets from Gene Ontology, (5) conservation, (6) topology in the protein interaction network, and (7) subcellular location of the protein. We distinguished between intrinsic and extrinsic features. Intrinsic features denote features, which can be directly derived from gene and protein sequences, whereas extrinsic features describe network topology, homology, cellular localization of the expressed protein and functional (interaction) domains. Employing BioMart (Smedley et al., 2009), we downloaded the gene and protein sequences from Ensembl (v102). Protein and gene sequence features (categories 1 and 2) were calculated with several software tools [seqinR (Charif et al., 2005), protr (Xiao et al., 2015)], CodonW,¹ and rDNase² spanning a large range of descriptors as explained in the following. Using seqinR, we calculated the number and percentage of each of the 20 amino acids in a protein. In addition, it was used to calculate other amino acid sequence features including the number of residues, the percentage of each of the physico-chemical classes and the theoretical isoelectric point. Using protr, we calculated the autocorrelation, Conjoint Triad Descriptors (CTD), quasi-sequence order and pseudo amino acid composition. CodonW was used calculating general gene features

like gene length and GC content, and frequencies of optimal codons (Hershberg and Petrov, 2009) and the number of codons in the coding sequence. rDNase calculated additional gene features, like auto covariance, pseudo nucleotide composition, and kmer frequencies ($n = 2-7$). It also calculated the sequence attribute distribution. Amino acids were grouped into three classes: (1) polar, (2) neutral, or (3) hydrophobic. The second digit in the feature name accounted for the class. More fine grained, seven attributes comprising (1) hydrophobicity, (2) normalized van-der-Waals volume, (3) polarity, (4) polarizability, (5) charge, (6) secondary structure, and (7) solvent accessibility were represented by the first digit in the feature name. The last three digits described the location of the attribute in the sequence, i.e., either at the beginning of the sequence (000), around the 25% quantile of residues (025), 50% (050), 75% (075), or at the end of the sequence (100). For example, seq.attribute.distribution 42100 is the sequence attribute of amino acids being polarizable (4), neutral (2) and are located at the end of the sequence (100). The domain features (category 3) were calculated using BioMart yielding protein family domains (pfam domains), the number of coiled coils structures, prediction of membrane helices, post-translational modifications, β -turns, cofactor binding, acetylation and glycosylation sites, signal peptides and transmembrane helices, and the number and lengths of UTRs. To calculate the gene set features (category 4), gene sets of all terms from Gene Ontology (Biological Process, Molecular Function, Cellular Localization) were obtained from the gene annotation of Ensembl (v102). First, to each gene, its direct neighbors were added using the STRING database (Szklarczyk et al., 2019), and with this list a gene set enrichment test performed employing Fisher’s exact tests. The negative log10 value of the p -value P was used as a feature. Highly redundant (overlapping) gene sets were removed by the following method. Overlap of genes among each pair of gene sets V and W was quantified by Jaccard similarity coefficients,

$$J(V, W) = \frac{|V \cap W|}{|V \cup W|} \quad (1)$$

Pairs with $J(V, W)$ above the threshold $\alpha = 0.3$ were included into a model and represented as an undirected graph $G(X, E)$, in which X were the n gene sets and pairs above α the edges E . A Mixed Integer Linear Programming problem was set up,

$$\text{Maximize} \quad \sum w_i X_i \quad (2)$$

Subject to

$$X_i + X_j \leq 1 \text{ for every } \{i, j\} \in E \quad (3)$$

$$X_i \in \{0, 1\} \text{ for } 1 \leq i \leq n, \quad (4)$$

in which w_i was the weight of a gene set and calculated from the significance value (p -value P) of the according Fisher’s Exact test,

¹ <http://codonw.sourceforge.net>

² <https://github.com/wind22zhu/rDNase>

$$w_i = \frac{1 - \log_{10}(P)}{100} \quad (5)$$

The problem was solved using Gurobi™ (version 7.5.1³). By this, up to one gene set of each pair with high overlap was selected and the gene set with the higher enrichment was privileged. Gene conservation (category 5) was assessed by sequence alignment of the protein sequence of each gene with all proteins listed in RefSeq (Pruitt et al., 2005) using PSI-Blast (Altschul et al., 1997). The number of obtained proteins with E-value cutoffs from $1e-5$ to $1e-100$ served as feature values of the respective features. For category 5, typical topology features like degree, degree distribution, several centrality descriptors and Page rank was obtained using protein interaction data ($n = 697,802$ interactions) from the BIOGRID database (Oughtred et al., 2019). In addition, we assembled a second network considering also the interactions between SARS-CoV-2 proteins and proteins of the human host cell. For this, in total 22,586 interactions between SARS-CoV-2 and human proteins were obtained from the BioGRID COVID-19 Coronavirus Curation Project (Oughtred et al., 2019). Viral proteins were added to the generic network and an interaction was added to the network between a host protein and a viral protein if the link was listed in the protein interactions between SARS-CoV-2 and human proteins. Using both protein interaction networks, in summary, 24 topological features were generated using the igraph R package for network analysis.⁴ Subcellular localization (category 7) was derived using the prediction software DeepLoc (Almagro Armenteros et al., 2017), which assigns probability scores to 11 eukaryotic cellular compartments (cytoplasm, nucleus, mitochondria, ER, Golgi apparatus, lysosome, vacuole and peroxisome, plasma membrane, extracellular, chloroplast).

Furthermore, 69 features from experimental data derived from gene expression, proteomics and phospho-proteomics profiles of Caco-2 cells post infection with SARS-CoV-2 at different time points were used. Proteomics data was taken from a previous publication (Bojkova et al., 2020) and z-score transformed peptide counts for each time point post infection (0, 2, 6, 10, 24h, MOI = 0.01) vs. time point zero taken. In line, own gene expression data from Caco-2 cells infected with SARS-CoV-2 (FFM1, Wuhan wildtype; 0, 3, 6, 12 and 24hpi, MOI = 0.01). For both datasets, as features, \log_{10} differential expression values between each time point post infection and time point zero (no infection) was taken. The phosphoproteomics data was taken from a previous study (SARS-CoV-2 infected Caco-2 cells, 24hpi, MOI = 1) and features calculated similarly as for the proteomics data (Klann et al., 2020).

2.3. Machine learning pipeline

Each feature was normalized using z-score transformation. We trained and validated two different sets of machines, one based on the gold standard of the knockout screens (knockout screens based classifier) and one based on the gold standard from the drug screen (drugs screen based classifier). The machine learning procedure was

the same for both. For each, a 5-fold cross-validation (CV) was performed in which 4/5 of the data was used to train the model, and the remaining 1/5 of the data (validation set) was used to assess the performance leaving the test set unseen. In addition, we repeated these cross-validations five times and averaged the results over these five independent runs. The first step in the training procedure was feature selection. For this, Least Absolute Shrinkage and Selection Operator (LASSO) regression was performed using the “glmnet” R package (cv.glmnet function with parameters $\alpha = 1$, $\text{type.measure} = \text{“auc”}$). In a second feature selection step, highly correlating features with Pearson’s correlation coefficients $r > 0.70$ were removed. On average, 155 features remained. These features were fed into a Random Forest (RF) classifier using the caret package in R ($\text{tuneLength} = 10$, $\text{metric} = \text{“Kappa”}$, $\text{nthread} = 20$, $\text{ntree} = 500$). All trained models were used to predict the complete set of 20,007 genes to find novel HDF candidates, and to confirm or reject the HDF annotation of the gold standard. We defined the top 10% of the predictions from a voting scheme as predicted HDF (predHDF) of the classifiers, i.e., of the knockout based classifier and of the drug based classifier, leading to 2,182 and 1,989 predHDF, respectively. To evaluate the performance of the learned models, the area under the receiver operating characteristic (ROC-AUC) and area under the precision-recall curve (PR-AUC) of the results from the validation data was calculated using the PRROC package in R. Average and standard variances of the performance estimates were calculated using the results of the individual cross validation runs. Most discriminative features of the model were obtained using the varImp function from the caret package which ranked the features based on their impact on the decision trees accuracy.

2.4. Gene set enrichment analyses

The lists of predHDF were investigated for enriched gene sets using the R package gprofiler2 (Raudvere et al., 2019). The statistical significance threshold for all GO terms was set to a false discovery rate $p < 0.05$. To focus on non-too-general and non-too-specific GO terms, only terms with more than 10 and less than 200 genes were considered. Redundancy of the terms was removed as described above (section 2.2) using $\alpha = 0.4$.

2.5. Drug selection

Aiming for repurposing known drugs including clinically approved drugs for the identified targets, a manually assembled data repository of drugs was set up based on seven public drug databases: Drugbank, ChEMBL, TTD, PharmGKB, BindingDB, IUPHAR/BPS, and DrugCentral (Thorn et al., 2013; Gilson et al., 2016; Ursu et al., 2017; Wishart et al., 2018; Mendez et al., 2019; Armstrong et al., 2020; Wang et al., 2020). The databases provide information about the mechanisms of actions of the drugs, drug-target interactions and therapeutic indications. We considered only drugs approved by either of the following regulatory agencies: Food and Drug Administration (FDA), Health Canada, European Medicines Agency (EMA), and Japan Pharmaceutical and Medical Devices Agency (PMDA), and drugs currently tested in clinical trials.

³ www.gurobi.com

⁴ <https://igraph.org>

2.6. Tissue culture and viruses

Lung adenocarcinoma cell line A549 stably expressing ACE2 and TMPRSS2 (A549-AT; Widera et al., 2021a) was cultured in Minimum Essential Medium (MEM) supplemented with 10% (v/v) fetal calf serum (FCS), 100 U/mL penicillin + 100 µg/mL streptomycin (Thermo Fisher; Waltham, Massachusetts, United States) and 2% (v/v) L-glutamine. Calu-3 was cultured in Dulbecco's Modified Eagles Medium (DMEM)/Ham's F12 (Thermo Fisher; Waltham, Massachusetts, United States) supplemented with 10% (v/v) FCS, 100 U/mL penicillin + 100 µg/mL streptomycin. Vero E6 was cultivated in DMEM (Thermo Fisher; Waltham, Massachusetts, United States) supplemented with 10% (v/v) FCS and 100 U/mL penicillin + 100 µg/mL streptomycin. Antibiotic concentrations were chosen based on the manufacturer's recommendations. Cells were incubated at 37°C, 5% CO₂. If not otherwise indicated, culture reagents were purchased from Sigma (St. Louis, MO, United States).

For all viruses used in this study, the corresponding GenBank accession numbers, and references are listed in Table 2. Briefly, coronavirus isolates were obtained from nasopharyngeal swabs of infected individuals. Swab material was suspended in 1.5 mL phosphate-buffered saline (PBS) and propagated on Caco-2 cells. Cell-free aliquots were stored at −80°C and titers were determined by tissue-culture infectious dose (TCID₅₀) assay in A549-AT cells. The H5N1 influenza strain Hongkong/213/03 was received from the World Health Organization (WHO) Influenza Centre (National Institute for Medical Research, London, United Kingdom). Virus stocks were prepared by infecting Vero cells and aliquots were stored at −80°C. Titers were determined as described above.

According to the committee for Biological Safety (ZKBS), the entirety of infectious work presented in this study was conducted under BSL-3 conditions.

2.7. Inhibitors and chemicals

Compounds used in this study comprise Remdesivir (#HY-104077), ICG-001 (#HY-14428; MedChem Express; Monmouth

Junction, New Jersey, United States), C-82 (#S0990) and PRI-724 (#S8968; Selleck Chemicals; Houston, Texas, United States). Compounds were resuspended in DMSO (Carl Roth; Karlsruhe, Germany) and aliquots were stored at −80°C. Recombinant human Wnt-3a (rhWnt-3a; #5036-WN, R&D Systems; Minneapolis, Minnesota, United States) was resuspended according to manufacturer's instructions and stored at −80°C.

2.8. Compound treatment and infection assay

Inhibitors were diluted serially on confluent A549-AT or Calu-3 cells in 1% (v/v) serum MEM and were subsequently infected at an MOI of 0.1 with viruses for 48 h. Cells were fixed with 3% (v/v) paraformaldehyde (PFA) for 20 min and stored at 4°C in PBS until analysis. Cytopathic effects were quantified by confluency measurement using a Spark® Cyto 400 multimode plate reader (Tecan Group Ltd.; Zürich, Switzerland).

2.9. Scratch assay

Scratch assay was implemented to visualize growth hindrance in A549-AT cells upon treatment with PRI-724. 6·10⁵ cells were seeded in a 12-well format 1 day prior to assay. After applying scratches with a pipette tip, cells were washed with 1x PBS and treated with 10 µM PRI-724 and the respective amount of DMSO in 1% FCS MEM. Cells were subjected to live cell imaging 0, 24, 48 and 72 h post treatment using a Spark® Cyto 400 multimode plate reader (Tecan Group Ltd.; Zürich, Switzerland).

2.10. Live cell imaging

In order to monitor the effectiveness of PRI-724 in terms of blocking syncytia formation and cytopathic effects in general, live cell imaging was carried out in A549-AT cells in a 96-well format. 3.5·10⁴

TABLE 2 Clinical coronavirus isolates used for infection experiments in this study.

Virus	Specification	GenBank ID	References
SARS-CoV-2	B, FFM5/2020	MT358641	Toptan et al. (2020)
SARS-CoV-2	B.1, FFM7/2020	MT358643	Toptan et al. (2020)
SARS-CoV-2	B.1.1.7, FFM-UK7931/2021	MZ427280	Widera et al. (2021b)
SARS-CoV-2	B.1.351, FFM-ZAF1/2021	MW822592	Widera et al. (2021b)
SARS-CoV-2	B.1.617.2, FFM-IND8424/2021	MZ315141	Wilhelm et al. (2021)
SARS-CoV-2	P.2, FFM-BRA1/2021	MW822593	Widera et al. (2021b)
SARS-CoV-2	B.1.429, FFM-CALsprt/2021	MZ317896.2	Wilhelm et al. (2021)
SARS-CoV-2	B.1.617.1, FFM-IND5881/2021	MZ315140	Wilhelm et al. (2021)
SARS-CoV-2	B.1.1.529 BA.1, FFM-ZAF0396/2021	OL800703	Wilhelm et al. (2022)
SARS-CoV-2	B.1.1.529 BA.2, FFM-BA.2-3833/2022	OM617939	Wilhelm et al. (2022)
SARS-CoV-2	B.1.1.529 BA.5, FFM-BA.5-7501/2022	OQ568810	This study
SARS-CoV-1	Frankfurt 1	AY291315	Drosten et al. (2003)
MERS-CoV	EMC/2012	JX869059	Zaki et al. (2012)
Influenza A	H5N1/Hongkong/213/03	NCBI:txid432070	This study

cells were seeded and 1 day after were treated with 10, 3, 1, and 0.3 μM of PRI-724 just prior to infection. DMSO and 3 μM Remdesivir served as negative and positive inhibition controls, respectively. Infections with SARS-CoV-2 B.1.617.2 and SARS-CoV-1 Frankfurt-1 were carried out at an MOI of 0.1. Live cell imaging was performed by using a Spark[®] Cyto 400 multimode plate reader (Tecan Group Ltd.; Zürich, Switzerland) with hourly measurements of confluence and surface roughness for 48 h.

2.11. Immunofluorescence microscopy

A549-AT cells were infected with SARS-CoV-2 B.1.617.2, SARS-CoV-1 Frankfurt-1 and IAV H5N1/Hongkong/213/03 at an MOI of 0.1 for 18 h. Post fixation with 3% PFA, cells were washed and permeabilized using ice-cold MeOH. Epitopes were saturated by applying blocking solution (20% (v/v) goat serum, 2% (w/v) BSA, 0.3 M glycine, 0.1% Tween-20, 0.002% thimerosal) for at least 1 h. Primary antibodies (Supplementary Table S11) were diluted 1:1000 in blocking solution containing only 1% (v/v) goat serum and were then incubated on cells O/N at 4°C. Cells were washed three times with 1x PBS and were subsequently incubated with secondary anti-IgG-A488 (1:1000) together with 0.2 $\mu\text{g}/\text{mL}$ DAPI and for 1 h. Fluorescence imaging was performed with an Operetta CLS[™] High Content Analysis System (Perkin Elmer; Waltham, Massachusetts, United States) using integrated Harmony[®] software v4.9.

2.12. RNA kinetics and RT-qPCR

For assessing intracellular virus RNA replication upon treatment with PRI-724, medium of infected cells was removed 2 hpi and wells were washed three times with 1x PBS and were supplied with fresh 1% serum MEM. Lysates were collected 2, 4, 6, 8, 10, 12, and 24 hpi in A549-AT and 2, 4, 6, 8, 10, 12, 24, 36, and 48 hpi in Calu-3. Samples were prepared and RNA was extracted using RNeasy QIAcube HT (Qiagen; Hilden, Germany) kit according to manufacturer's instructions. RNA was stored at -20°C until analysis. RT-qPCR was carried out using Reliance One-Step Supermix (Bio-Rad; Hercules, CA, United States) to measure total viral N RNA and sg-N RNA (Supplementary Table S10, P1–P12) in a multiplex experiment⁵ (Veleanu et al., 2022). Quantifications of cellular genes *CDC25A*, *BIRC5*, *RNase P* (Supplementary Table S10, P13–P19; Spandidos et al., 2008) were carried out using Luna Universal One-Step RT-qPCR Kit (New England Biolabs; Frankfurt (Main), Germany) according to manufacturer's protocol. Fold changes were calculated using the $2^{-\Delta\Delta\text{Ct}}$ method (Livak and Schmittgen, 2001).

2.13. TCID₅₀ and plaque assay

For the quantification of infectious virus, supernatants were snap-frozen at -80°C . After thawing, supernatants were diluted

on confluent Vero E6 in 96-well plates. Four days post infection, dilutions were analyzed for CPE development. TCID₅₀ was calculated according to the Spearman and Kärber algorithm. For plaque assays, supernatants were diluted on confluent Vero E6 in 24-well plates. After 1 h, supernatants were coated with overlay medium (1.5% methyl cellulose (w/v; Merck KgaA; Darmstadt, Germany), 1x MEM, 1% (v/v) FCS, 100 U/mL penicillin + 100 $\mu\text{g}/\text{mL}$ streptomycin). When plaques reached sufficient size, cells were washed and fixed with 3% PFA for 20 min. Cells were stained with CV staining solution (1% (w/v) crystal violet, 20% (v/v) MeOH).

2.14. Luciferase assay

A549-AT cells were seeded in a 96-well format and were transfected with 0.1 ng of pRL-SV40 (Addgene; #27163) together with either 0.1 μg M50 Super 8x TOPFlash (Addgene; #12456) or M51 Super 8x FOPFlash (TOPFlash mutant; Addgene; #12457) using TransIT[®]-LT1 Transfection Reagent (Mirus Bio LLC; Madison, Wisconsin, United States). After 24 h, cells were treated with either of the following substances: 10 mM LiCl (Merck KgaA; Darmstadt, Germany), 100 $\mu\text{g}/\text{mL}$ rhWnt-3a, 10/3/1/0.3 μM PRI-724. Additionally, cells were treated simultaneously with LiCl and PRI-724. After 24 h, cells were analyzed for luciferase activity using a Dual Luciferase Reporter Assay Kit (Promega; Madison, Wisconsin, United States) according to manufacturer's instructions.

2.15. Western blot

For preparation of protein extracts, cells were lysed using RIPA buffer (150 mM NaCl, 50 mM Tris/HCl pH 8.0, 1% (v/v) Triton X-100, 0.5% (v/v) sodium deoxycholate, 0.3% (v/v) SDS, 2 mM MgCl_2). Buffer was supplemented with 5 mM NaF, 1 mM Na_3VO_4 , 20 mM β -glycerophosphate, 1x cOmplete Mini, EDTA-free protease inhibitor cocktail (Merck KgaA; Darmstadt, Germany) and 1 U/mL Pierce[™] Universal Nuclease (Thermo Fisher; Waltham, Massachusetts, United States). Lysates were stored at -20°C . Protein concentrations were determined using DC Protein Assay Kit (Bio-Rad; Hercules, CA, United States). Proteins were separated by SDS-PAGE and were blotted onto an Amersham[™] Protran[®] nitrocellulose membrane (Merck KgaA; Darmstadt, Germany) for 1.5 h using const. 120 V. Blots were washed and blocked for at least 1 h using 5% (w/v) BSA/TBS-t (150 mM NaCl, 20 mM Tris, 0.1% (v/v) Tween20). Primary antibody (Supplementary Table S11) incubation was performed overnight at 4°C in 5% BSA-TBS-t. After washing, secondary antibody was incubated for 1 h in 5% BSA-TBS-t. Blots were imaged using CLx imaging device (LI-COR; Lincoln, Nebraska, United States). All full-length western blots are presented in Supplementary Figure S5.

2.16. Cytotoxicity measurement by fluorescence microscopy

Confluent A549-AT cells were treated with 10, 3, 1, 0.3 μM PRI-724 and with the highest corresponding amount of DMSO for

⁵ <https://www.who.int/docs/default-source/coronaviruse/whoinhouseassays.pdf>

48 h in 1% MEM. Cells were fixed with 3% PFA and stained with 0.5 µg/mL Hoechst 33342. Imaging was performed with Olympus IX 81 scanning a total of 9 fields/well with a 3 × 3 scatter at a total magnification of 10x (pixel size = 0.645 µm). Screening was performed using ScanR Acquisition v2.4.0.13 software.

The analysis was performed using KNIME and KNIP (Dietz and Berthold, 2016). The images were filtered using a gaussian convolution ($\sigma = 2$ px). A consecutive image thresholding assigned pixels with values greater than the image-mean to the foreground. Cells were detected and labeled by connected component analysis. Cell clumps were afterwards split into single cells using the Wahlby algorithm (Wahlby et al., 2004). The resulting components were filtered based on their size, components too large or too small were rejected as they most likely do not depict cells. Cell count was measured for the different PRI-724 concentrations (Supplementary Figure S4).

2.17. PI staining and flow cytometry

In a 6-well plate, confluent A549-AT cells were treated with 10 µM PRI-724 and the corresponding amount of DMSO for 24 h and were infected with SARS-CoV-2 (B.1.617.2 isolate, MOI 0.01) in 1% MEM. After that, cells were gently detached and fixed in ice-cold 70% ethanol for 1 h. Cells were washed three times with FACS-PBS (2% (v/v) FCS, 0.1% NaN₃) and were then stained overnight at 4°C by applying PI staining buffer (50 µg/mL PI, 2% (v/v) FCS, 200 µg/mL Monarch® RNase A (New England Biolabs; Frankfurt (Main), Germany), 0.1% Igepal CA-630). Cells were analyzed for PI staining using FACSVerse™ Flow Cytometer (BD Biosciences; Mississauga, ON, Canada).

2.18. Statistics

The curve fittings presented for dose-responses to Remdesivir and PRI-724 presented in Figure 1A and Supplementary Figure S1 were performed by applying robust non-linear regression comprising a sigmoidal 4-parameter model for a total of three biological replicates per concentration:

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + \left(\frac{\text{IC}_{50}}{X}\right)^{\text{HillSlope}}} \quad (6)$$

The exact IC₅₀ values calculated from these analyses are listed in Figure 1A and Supplementary Tables S7, S8. For simple group comparisons such as data presented in Figures 1B, 2, 3B, C, 4A, B, 5, one-way ANOVA was generally applied to identify significant differences between treatments. In all tests, equal variances according to the mean were assumed and further analyses were not corrected for multiple testing (Fisher's LSD test). Results were rated significant when $p < 0.05$. For RNA kinetics depicted in Figure 3A, two-way ANOVA was applied to identify significant differences in vRNA levels according to the timepoint of collection and the sort of treatment. Again, equal variances according to the mean were assumed and further analyses were not corrected for multiple testing (Fisher's LSD test). Results were rated significant when $p < 0.05$.

3. Results

3.1. Host dependency factors identified by genome scale screens showed marginal overlap on the gene level, but improved consistency for common cellular processes and functions

First, the overlap of the results from published screening studies were investigated, in the following denoted as Daniloski et al. (2021), Wang R. et al. (2021), Wei et al. (2021), and Zhu et al. (2021). For a balanced comparison, we regarded the top 500 genes with the highest scores of every screen (Figure 6A). Only ACE2 was common among all candidates. The highest overlap ($n = 37$ genes) was observed between the screens of Daniloski et al., and Zhu et al. Besides ACE2, only 11 genes were common among Wang et al., and Wei et al. All these overlaps were not significantly enriched (using Fisher's Exact Test). Similarly, we found very low correlations while analyzing the ranking of all genes of the screens (independent of the cutoff; Figure 6B). Next, we investigated how grouping of genes into gene sets of common cellular processes and gene functions affect the results. For this, gene set enrichment tests were performed with the same lists of the 500 highest scoring genes (Figure 6C). Interestingly, the overlaps were considerably higher but still not significant. Daniloski et al. and Zhu et al. shared 44/51 gene sets, Daniloski et al., Wang et al., and Zhu et al. shared 5/51 gene sets and 2/51 were identical among Daniloski et al., Wei et al., and Zhu et al. The overlapping gene sets included vesicle-mediated transport to the plasma membrane, endosomal and lysosomal transport, phagosome maturation, transforming growth factor-beta production, post-Golgi vesicle-mediated transport, transition metal ion transport, negative regulation of cell growth, and phosphatidylinositol biosynthetic process (Supplementary Table S1). In summary, we observed limited overlap among the investigated screens. The overlap was higher when regarding gene sets of common cellular processes or molecular function. This supported our concept to set up a machine learning procedure to identify more common patterns among these different datasets.

3.2. Classifiers based on data from the knockout screens and a drug based screen performed well in predicting HDF from the gold standards

We assembled a list of 500 genes commonly high ranking among the investigated knockout screens (top 500 genes of the rank products, see Methods). This list served as the input for the machine learning classifier. Similarly, we selected negative controls from the lowest ranking genes. Furthermore, we set up a descriptor for gene predictions by compiling a comprehensive list of more than 60,000 features for each gene describing its nucleic acid and protein sequences, potential associated cellular processes, its associated compartments, functional domains, molecular functions, its conservation, and its network topology within a protein interaction network. Here, also other omics data of SARS-CoV-2 infected cells, like gene transcription profiles, proteomics and interactions between viral and host proteins were integrated. Performing a cross-validation,

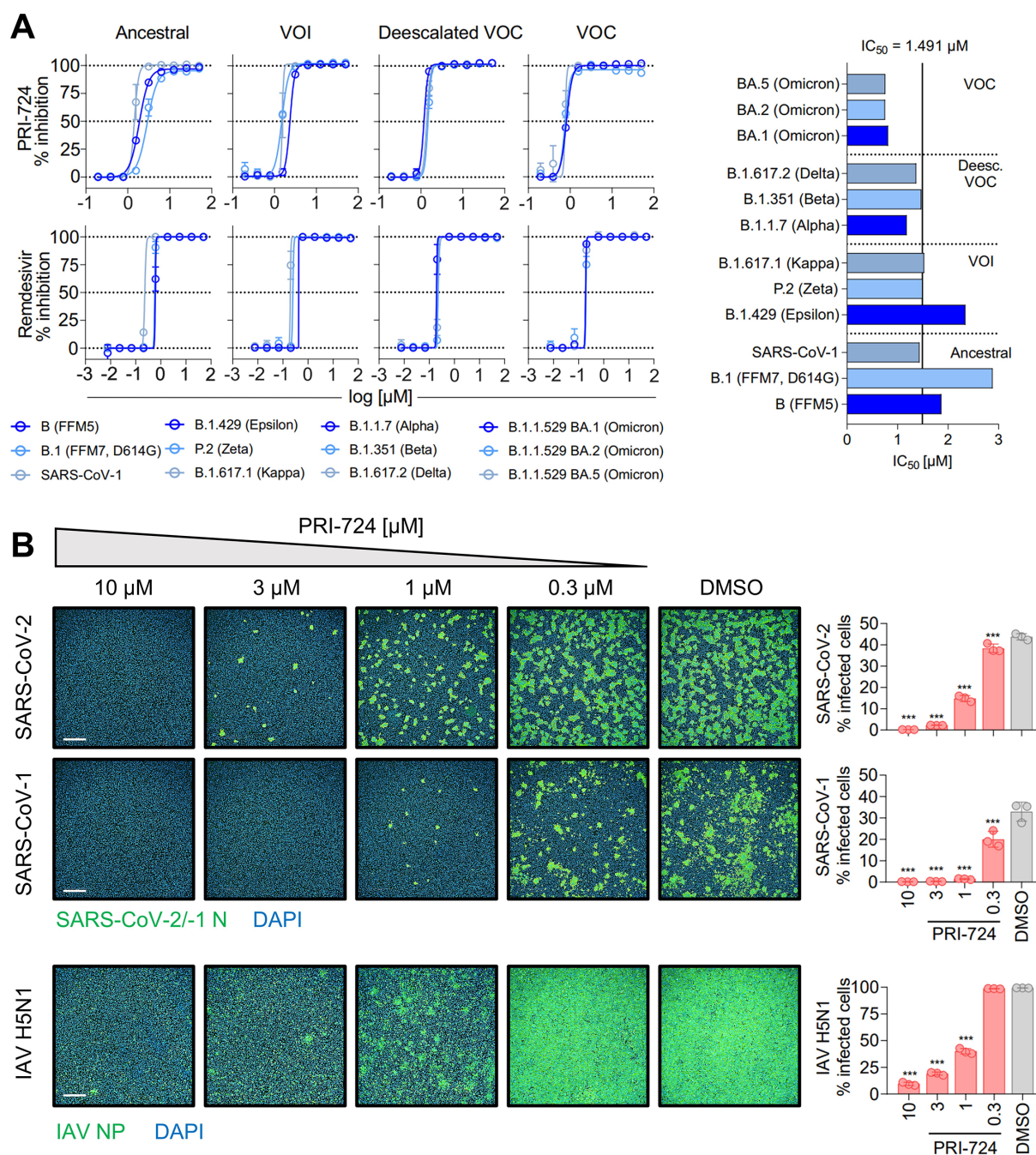


FIGURE 1

Inhibition of SARS-CoV-2, SARS-CoV-1 and IAV H5N1 by PRI-724. (A) Dose–response curves of diverse SARS-CoV-2 mutational variants and SARS-CoV-1 to PRI-724 and Remdesivir. Cell confluence was analyzed 48hpi using a Spark® Cyto Imaging System. Dose–response curves were fit to normalized % inhibition. Data represent mean \pm SD of three biological replicates. The experiment was repeated twice with similar results. IC₅₀ values were calculated by applying robust non-linear regression. Mann–Whitney *U*-test confirmed no significant changes between any dose–responses analyzed. (B) Staining of SARS-CoV-2 (B.1.617.2 strain) and SARS-CoV-1 N protein, as well as IAV H5N1 NP 18hpi. Data points represent mean \pm SD of three biological replicates. The experiment was repeated twice with similar results. Infected cells were determined by Alexa 488/DAPI co-localization using an Operetta® High-Throughput Imaging system. Scale bars=500 μm .

in which training and validation data is strictly separated, we observed good performance results with an area under the curve (AUC) of a Receiver Operator Characteristics of 0.82 ($1\sigma=0.03$; Figure 7A). In addition, the same procedure was performed and a classifier learned based on data of a drug screen (in the following drug screen based

classifier; Ellinger et al., 2021). Here, we obtained slightly worse performance compared to the knockout-based classifier receiving an AUC of the Receiver Operator Characteristics of 0.71 ($1\sigma=0.11$). Additional performance parameters are supplemented (Supplementary Table S2).

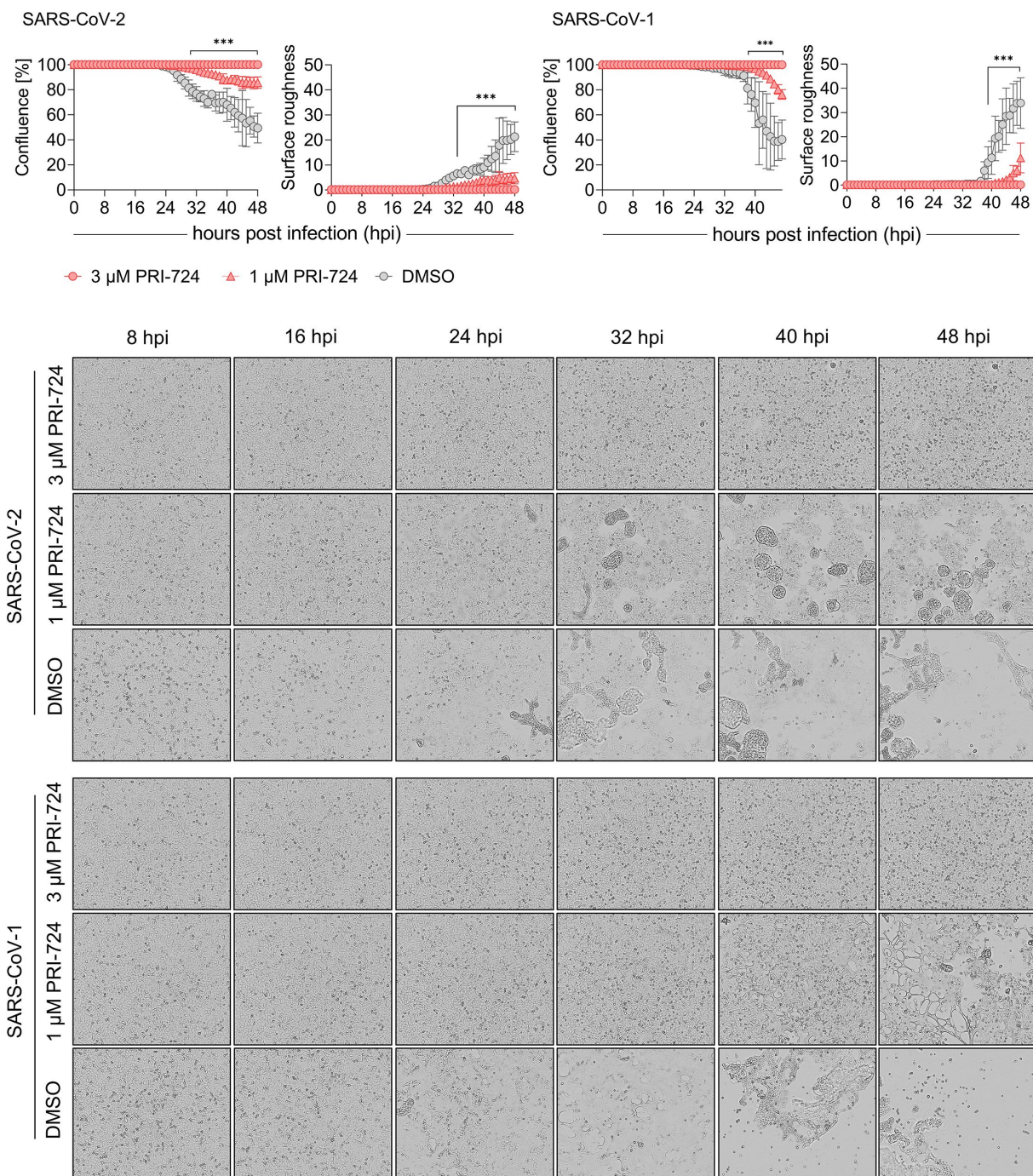


FIGURE 2

Growth kinetics of PRI-724 treated A549-AT cells. A549-AT cells treated with high (3 μM) and low (1 μM) dose of PRI-724 were subsequently infected with SARS-CoV-2 (B.1.617.2 isolate) and SARS-CoV-1 (FFM1 isolate; MOI=0.1). Cells were incubated at 37°C and 5% CO₂ for 48h in a Spark® Cyto Imaging system. **Top:** Graphical presentation of % confluence and surface roughness for both SARS-CoV-2 and SARS-CoV-1 are shown for measurements hourly performed. **Bottom:** Respective brightfield microscopy images represent critical time points of CPE formation over the course of infection during treatment with PRI-724 and without. Scale bar: 200 μm. See also [Supplementary Videos S1–S6](#). Every condition comprises three biological replicates and graphs represent means ± SD; significant differences (***) $p < 0.001$ are indicated by asterisks obtained by performing multiple *t*-tests. The experiment was performed twice yielding similar results.

3.3. Predicted HDF are prominently enriched in gene sets related to morphogenesis and development

We selected 2,182 and 1,989 top-scoring genes identified by knockout- and drug screen-based classifiers, respectively. To identify

the most involved cellular mechanisms in the life cycle of the virus, we performed gene set enrichment analysis for each list illuminating 1,098 and 688 gene sets of the respective classifiers ([Supplementary Tables S3, S4](#)). Interestingly, we observed a high number of common gene sets ($n = 313$) showing consistency among both classifiers ([Figure 7B](#)). Regarding these common gene sets,

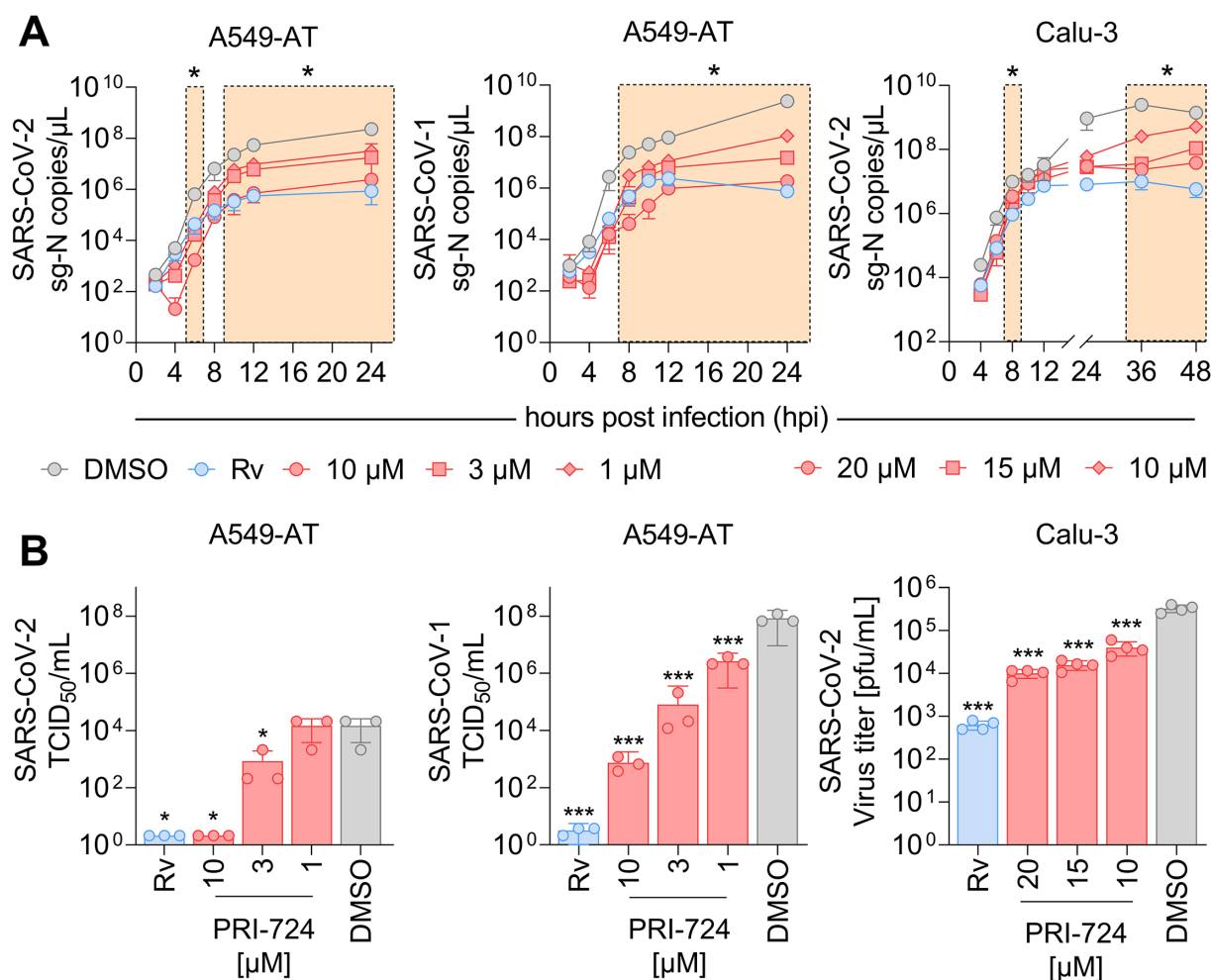


FIGURE 3

RNA kinetics and quantification of infectious virus upon PRI-724 treatment. A549-AT and Calu-3 cells were treated with 3 μM Remdesivir (Rv, blue) and indicated concentrations of PRI-724 [red; A549-AT: 10 μM (circle), 3 μM (square), 1 μM (rhombus); Calu-3: 20 μM (circle), 15 μM (square), 10 μM (rhombus)], as well as DMSO (gray) and were subsequently infected with SARS-CoV-2 (B.1.617.2 isolate) and SARS-CoV-1 (FFM1 isolate) at MOI=0.1. **(A)** Viral sg-N RNA levels in total RNA lysates over time in A549-AT infected with SARS-CoV-2 (left) and SARS-CoV-1 (middle), as well as in Calu-3 infected with SARS-CoV-2 (right). Data points represent mean and SD of three biological replicates. The experiments were repeated twice with similar results. Significant differences between treatments and control determined by two-way ANOVA are indicated through asterisks; * $p < 0.05$ (Supplementary Table S9). **(B)** Infectious SARS-CoV-2 and SARS-CoV-1 virus titers 24hpi in A549-AT determined by TCID₅₀ assay (left, middle) and in Calu-3 determined by plaque assay (right). Data points represent mean and SD of three biological replicates. The experiments were repeated twice with comparable results. Significant differences between treatments and control are indicated through asterisks; * $p < 0.05$, *** $p < 0.001$.

we further found a high fraction of these gene sets (69 out of 313 gene sets) to be related to morphogenesis and development, followed by 60 gene sets being related to neural processes (Supplementary Table S5). Besides these gene sets, further gene sets were related to signaling, gene regulation, and immune and stress response (Supplementary Table S5).

During development, a large variety of distinct different cellular identities need to be established and maintained in the embryo. Particularly, during developmental lineage reprogramming, a somatic cell can be reprogrammed into a distinct cell type by forced expression of lineage-determining factors (Vierbuchen and Wernig, 2012). Similarly, coronaviruses reprogram their host cell for their specific needs, including their replication (Spagnolo and Hogue, 2000; Shi and Lai, 2005; Reggiori et al., 2010; Nagy and Pogany, 2011; Pizzato et al., 2022). We followed this interesting parallelism and selected a gene list

from the gene sets of development and morphogenesis for further prioritization. Aiming to interfere “reprogramming” with a high impact, we prioritized genes coding for proteins with high connectivity (number of nearest neighbors) and centrality measures in a protein–protein interaction (PPI) network to enhance the impact of a treatment by small molecule inhibitors. This led to a short list of genes with highest ranks (employing rank products) based on these measures in this PPI network. Table 3 lists the top 10, Supplementary Table S6 lists the top ranking genes and their connectivity and centrality values when protein interactions to SARS-CoV-2 proteins are also taken into account. Out of these, β-catenin (CTNNB1) was selected for further analysis due to its high gene expression in SARS-CoV-2 infected cells. Interrogating publicly available compound and drug databases led to the selection of PRI-724 for experimental follow up described in the next sections.

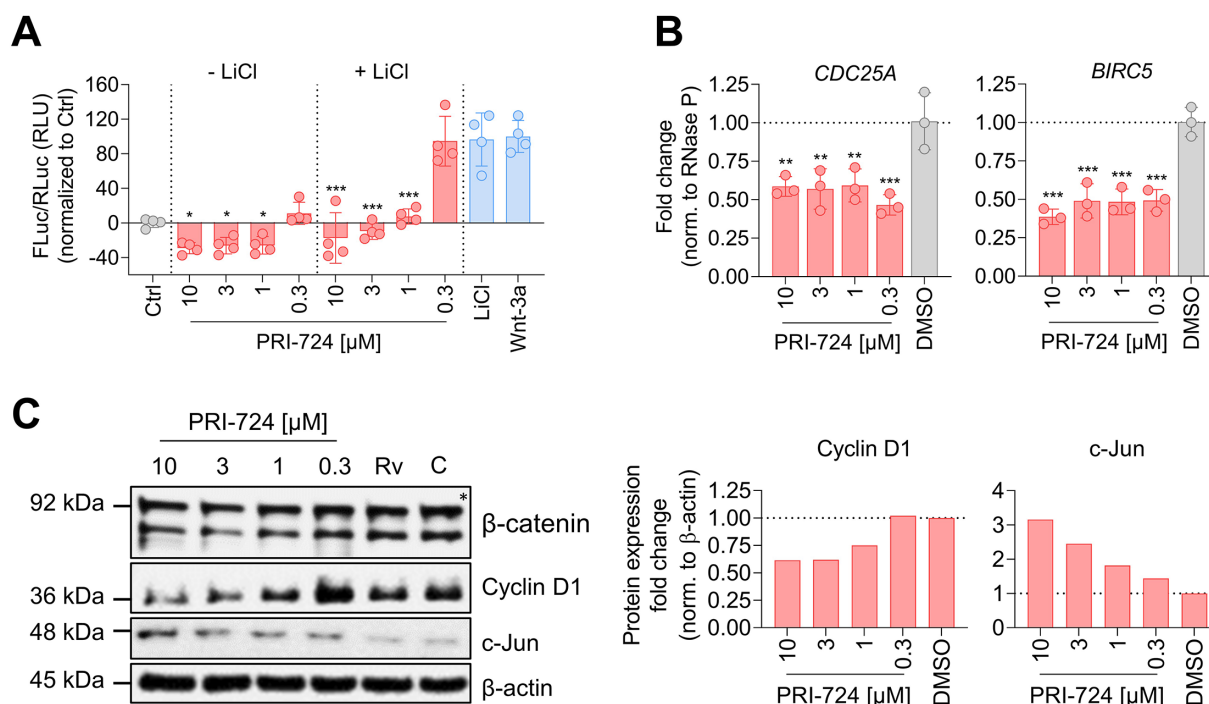


FIGURE 4
PRI-724 shuffles β -catenin/CBP and β -catenin/p300 activity distribution. **(A)** A549-AT cells were transfected with Super M50 8x TOPFlash and pRL-SV40. After 24h, cells were treated with 10mM LiCl, 100 μ g/mL rhWnt-3a (light blue), and 10 μ M, 3 μ M, 1 μ M and 0.3 μ M PRI-724 (red). Luciferase activity was measured after 24h. Firefly luciferase (FLuc) was normalized to Renilla luciferase (RLuc) and the non-treated control (gray) for every sample. Data represent mean and SD of three biological replicates. Samples treated with PRI-724 were compared to the non-treated control (gray) while samples treated with PRI-724+LiCl were compared to the LiCl-treated control (light blue) using ordinary one-way ANOVA. Significant differences are indicated by asterisks; * p <0.05, *** p <0.001. The experiments were repeated twice with comparable results. **(B)** A549-AT cells were treated with indicated concentrations of PRI-724 and DMSO. 24h post treatment, RNA was isolated and RT-qPCR was implemented to analyze *CDC25A* (left), *BIRC5* (right), and *RNase P* expression. Data represent mean and SD of three biological replicates. Significant differences are indicated by asterisks; ** p <0.002, *** p <0.001. Experiments were repeated twice with comparable results. **(C)** Cells were treated as described in **(B)**. Signals were normalized to β -actin expression. The asterisk designates the predicted band size of β -catenin. Rv, Remdesivir; C, Control.

3.4. PRI-724 inhibits SARS-CoV-2 variants, SARS-CoV-1, and influenza a virus in A549-AT cells

In order to validate *in silico* findings regarding PRI-724 in the context of an infection, various SARS-CoV-2 isolates [Ancestral variants: B (FFM5), B.1 (FFM7, D614G); variants of interest (VOI): P.2 (Zeta), B.1.429 (Epsilon), B.1.617.1 (Kappa); deescalated variants of concern (VOC): B.1.1.7 (Alpha), B.1.351 (Beta), B.1.617.2 (Delta); VOC: B.1.1.529 BA.1, BA.2, and BA.5 (Omicron)], and SARS-CoV-1 Frankfurt-1 were tested. A549-AT cells were infected at an MOI of 0.1 and treated with PRI-724 or Remdesivir for 48h. CPE-related confluency changes were measured using automated label-free brightfield microscopy (Figure 1A). For all viruses tested, a mean IC_{50} of 1.491 μ M (95%CI 1.087–1.896) and IC_{50} of 0.2916 μ M (95%CI 0.1948–0.3885) was observed for PRI-724 and Remdesivir, respectively. No significant differences were detected among SARS-CoV-2 variants, even though the B.1 and B.1.429 variants showed ~1.5-fold and ~2-fold higher IC_{50} , respectively, and BA.1, BA.2, and BA.5 variants had ~2-fold lower IC_{50} (Figure 1A; Supplementary Table S7). The overall response pattern among different variants remained similar for Remdesivir. Treatment with the PRI-724 active metabolite C-82 and the analog ICG-001 likewise

inhibited SARS-CoV-2 (B.1.617.2 isolate) in A549-AT with IC_{50} values of 1.005 and 14.85 μ M, respectively (Supplementary Figure S1A). In addition to that, dose-responses to PRI-724 for SARS-CoV-2 variants B.1, B.1.1.7, B.1.351, P.2, and B.1.617.2 were obtained in Calu-3 cells, with overall increased IC_{50} of 8.448 μ M (95%CI 7.431–9.465) for PRI-724 and mean IC_{50} of 0.4962 (95%CI 0.2686–0.7237) for Remdesivir (Supplementary Figure S1B; Supplementary Table S8). In Calu-3, PRI-724 further demonstrated inhibition of MERS-CoV (IC_{50} =22.4 μ M) and showed significantly decreased SARS-CoV-1 nucleocapsid (N) expression upon treatment with 10 μ M, 30 μ M, and 100 μ M PRI-724 (Supplementary Figure S1C). In an attempt to validate our results and to test whether PRI-724 can be used against a broader spectrum of RNA viruses, we infected A549-AT cells with Influenza A virus (IAV), SARS-CoV-2 (B.1.617.2 strain), or SARS-CoV-1 (Frankfurt-1 strain), and performed immunofluorescence staining for IAV nucleoprotein (NP) and the coronavirus N (Figure 1B). Significant concentration-dependent reduction of NP expression was evident for treatment with 10, 3, and 1 μ M PRI-724. Treatment with 10 and 3 μ M PRI-724 led to ~10-fold and ~4-fold reduction of H5N1 NP⁺ cells, whereas for these treatments no cells stained positive for SARS-CoV-1 N. Treatment with 3 μ M reduced SARS-CoV-2 N⁺ cells ~15-fold and staining was negative with 10 μ M PRI-724.

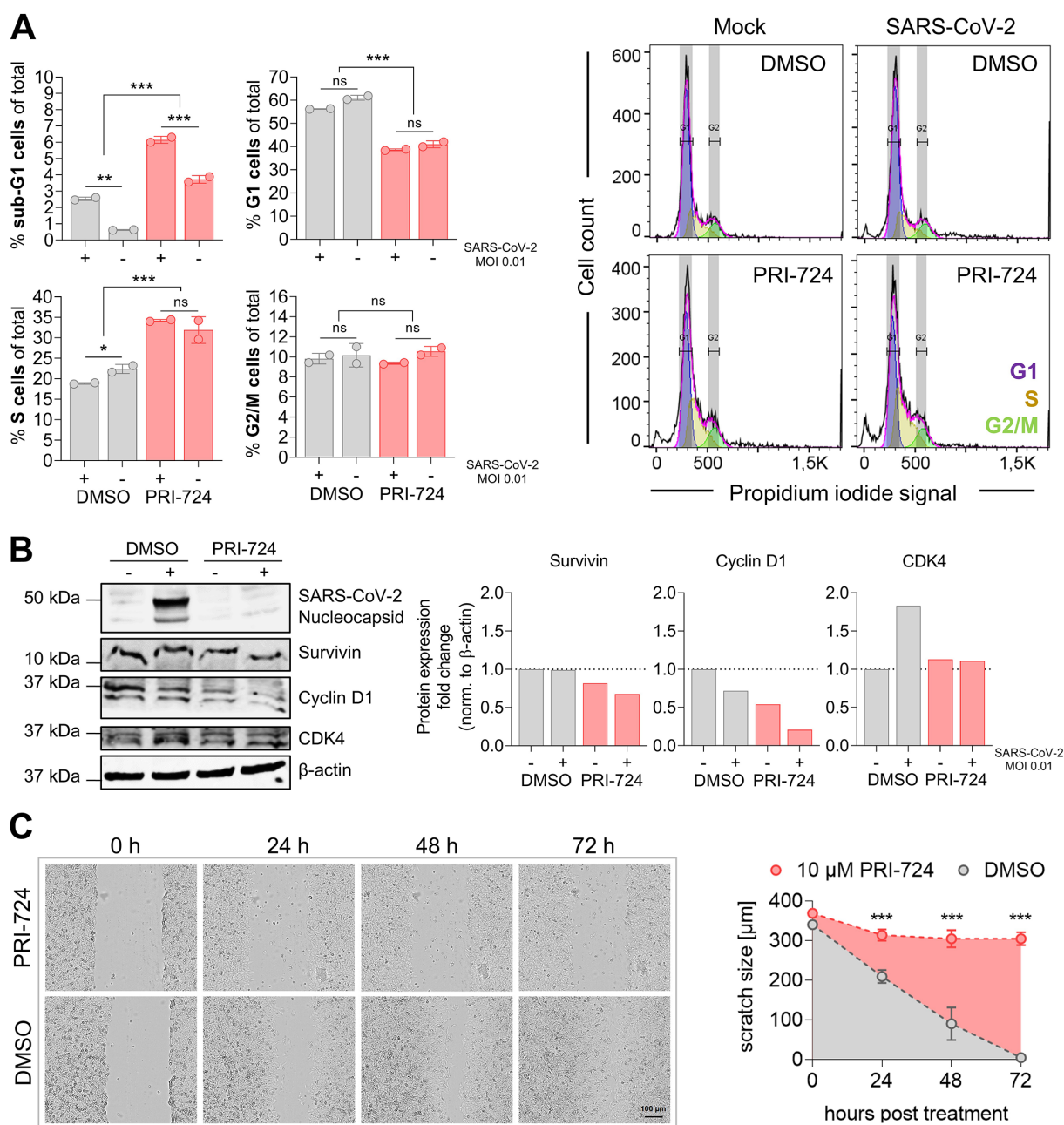


FIGURE 5

Cell cycle analysis in A549-AT upon treatment with PRI-724. **(A)** A549-AT cells were treated with 10 μ M PRI-724 or corresponding amounts of DMSO, and were subsequently infected with SARS-CoV-2 (B.1.617.2 isolate, MOI 0.01). After 24 h, cells were stained with propidium iodide (PI) following flow cytometric analysis. G1, S and G2/M populations were calculated using the Watson model implemented in FlowJo (v10.8.1). Sub-G1 populations were measured manually. Additionally, cell lysates were analyzed by western blot for Nucleocapsid, survivin, cyclin D1 and CDK4 expression. Expression levels were normalized to β -actin expression. **(B)** Confluent A549-AT cells were scratched prior to treatment with 10 μ M PRI-724 or corresponding DMSO amounts. Live cell imaging was performed 0, 24, 48, and 72 h post treatment to assess scratch healing properties using a Spark® Cyto imaging system. All data represent mean and SD of three biological replicates. Experiments were repeated twice. Significant results obtained by one-way ANOVA are indicated by asterisks; * p < 0.05, ** p < 0.002, *** p < 0.001.

3.5. PRI-724 inhibits syncytium formation and CPE onset in A549-AT cells

After confirming PRI-724-mediated inhibition of CPE in A549-AT cells at 48 hpi, we examined the timeline of CPE onset and development. Label-free live-cell brightfield imaging was implemented to address viral growth kinetics upon treatment with PRI-724. Cells

were treated with 3 μ M (high-dose), 1 μ M (low-dose) PRI-724, and the highest corresponding amount of DMSO and were subsequently infected with SARS-CoV-2 (B.1.617.2 isolate) or SARS-CoV-1 (Frankfurt-1 strain; MOI = 0.1). Imaging, confluence measurements and surface roughness measurements were performed over the course of 48 h. A549-AT confluence as well as surface roughness were mostly comparable among all treatment conditions until 24 hpi for both

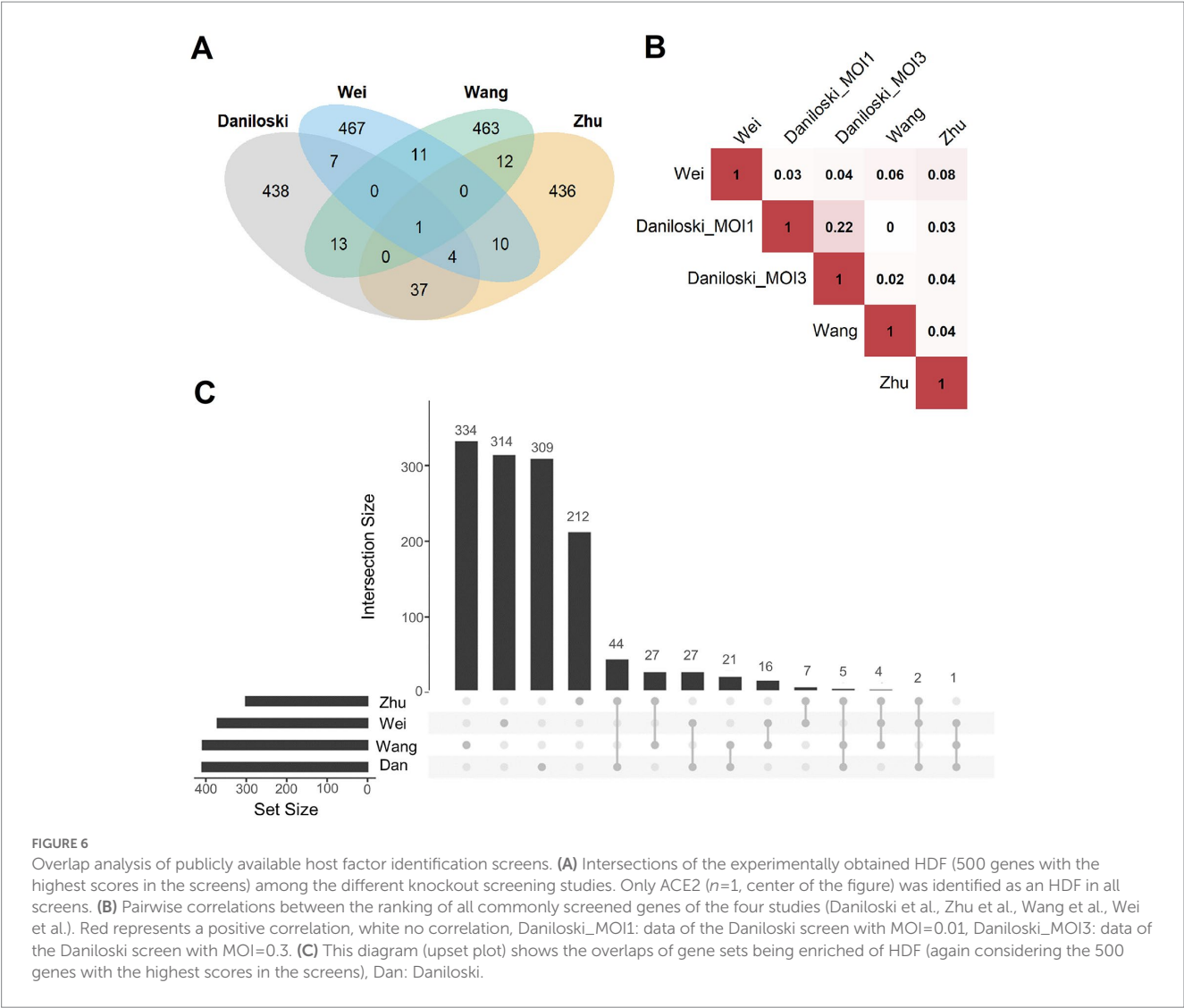


FIGURE 6 Overlap analysis of publicly available host factor identification screens. **(A)** Intersections of the experimentally obtained HDF (500 genes with the highest scores in the screens) among the different knockout screening studies. Only ACE2 ($n=1$, center of the figure) was identified as an HDF in all screens. **(B)** Pairwise correlations between the ranking of all commonly screened genes of the four studies (Daniloski et al., Zhu et al., Wang et al., Wei et al.). Red represents a positive correlation, white no correlation, Daniloski_MOI1: data of the Daniloski screen with MOI=0.01, Daniloski_MOI3: data of the Daniloski screen with MOI=0.3. **(C)** This diagram (upset plot) shows the overlaps of gene sets being enriched of HDF (again considering the 500 genes with the highest scores in the screens), Dan: Daniloski.

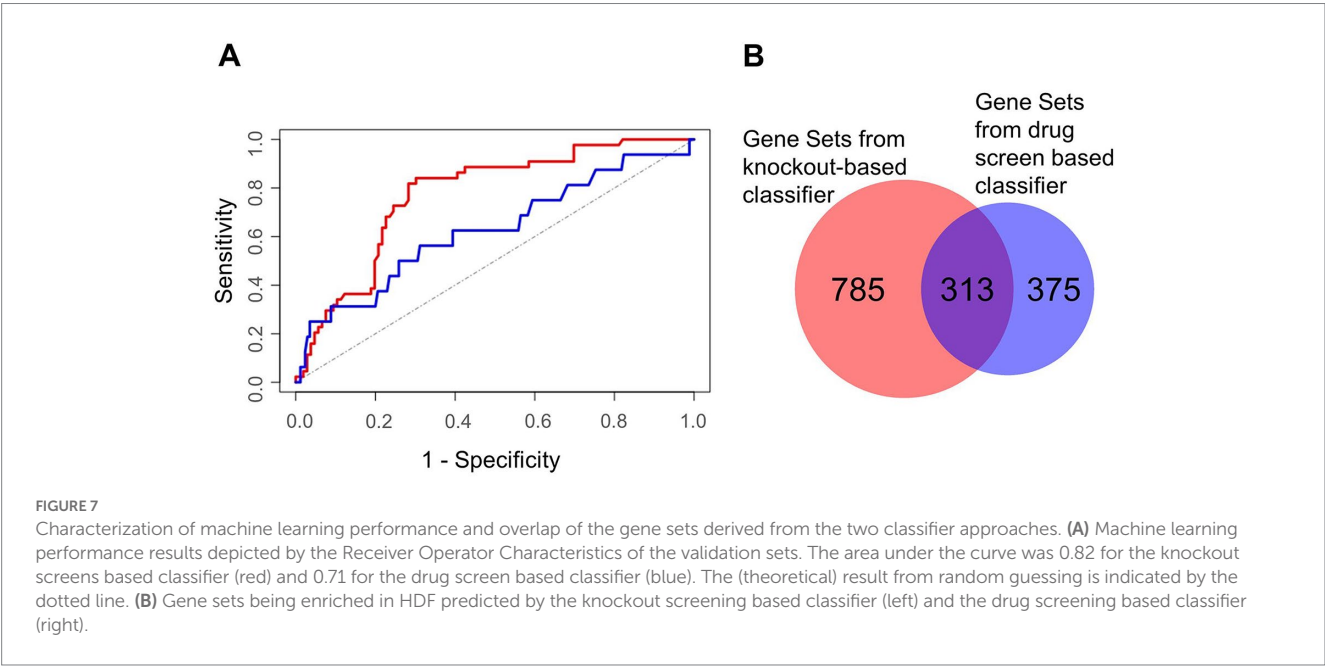


FIGURE 7 Characterization of machine learning performance and overlap of the gene sets derived from the two classifier approaches. **(A)** Machine learning performance results depicted by the Receiver Operator Characteristics of the validation sets. The area under the curve was 0.82 for the knockout screens based classifier (red) and 0.71 for the drug screen based classifier (blue). The (theoretical) result from random guessing is indicated by the dotted line. **(B)** Gene sets being enriched in HDF predicted by the knockout screening based classifier (left) and the drug screening based classifier (right).

TABLE 3 Selected genes of morphogenesis and development with high connectivity, closeness and betweenness centrality.

Gene symbol	Degree ^{1,2}	Closeness centrality ¹⁻³	Betweenness centrality ¹⁻³	Quantile gene expression ⁴
EP300	87	0.421	0.104	74
CUL3	82	0.405	0.148	89
TP53	65	0.390	0.068	82
AR	60	0.383	0.052	93
FASN	53	0.373	0.078	94
CTNNB1	40	0.394	0.046	98
HDAC1	57	0.381	0.047	58
MYCN	53	0.368	0.050	Not expressed
STAT3	39	0.382	0.046	94
AKT1	44	0.369	0.041	93

¹Based on a generic protein interaction network of developmental genes. ²Determined using Python package Networkx. ³Decimal places are rounded. ⁴Based on the expression levels of the listed gene across several time points of SARS-CoV-2 infected A549 cells (details, see Materials and Methods).

viruses. After 24 hpi, virus-induced cell lysis gradually reduced the confluence and increased surface roughness in DMSO-treated cells. SARS-CoV-2 and SARS-CoV-1 infected cells treated with low-dose PRI-724 started to lyse at 29 and 39 hpi, respectively, while high-dose PRI-724 treated cells showed no substantial loss in confluence over the course of the experiment (Figure 2, top; Supplementary Videos S1–S6). These findings were confirmed by brightfield microscopy at indicated time points (Figure 2, bottom). Here, cell–cell-fusion, i.e., syncytium formation, caused by both viruses, was initially detectable at 16 and 24 hpi, respectively, in DMSO-treated cells and was not present in high-dose PRI-724 treated cells. In low-dose PRI-724 treated cells, SARS-CoV-2 syncytia and CPE were detectable, but much smaller compared to the control (Figure 2, bottom, row 2–3). Although, we found likewise results for SARS-CoV-1, cell lysis initiated at 32 hpi compared to the control (Figure 2, bottom, row 5–6).

3.6. PRI-724 reduces virus RNA replication and infectious virus progeny

Since PRI-724 inhibited CPE development, we questioned whether it impaired virus replication and/or progeny virus production. For the replication analysis, the expression of intracellular subgenomic nucleocapsid (sg-N) RNA was monitored over the course of 24 and 48 h in A549-AT and Calu-3 cells, respectively (Figure 3A). In A549-AT cells, intracellular SARS-CoV-2 (B.1.617.2 isolate) sg-N RNA was significantly reduced at 6, 10, 12, and 24 hpi when treated with 10, 3, and 1 μ M PRI-724 just prior to infection. Most remarkable changes were detected 6 and 24 hpi where vRNA was reduced ~370-fold/~96-fold, ~38-fold/~13-fold, and ~20-fold/~7-fold (Supplementary Table S9) compared to the control when treated with 10, 3, and 1 μ M PRI-724, respectively. Similarly, in Calu-3, SARS-CoV-2 vRNA was significantly reduced at 8, 36, and 48 hpi, with most noticeable changes at 36 hpi, namely ~105-, ~69-, and ~10-fold (Supplementary Table S9) after treatment with 20, 15, and 10 μ M PRI-724, respectively. For SARS-CoV-1 in A549-AT, PRI-724 treatment reduced vRNA at 8, 10, 12, and 24 hpi. In comparison to SARS-CoV-2, the impact on SARS-CoV-1 vRNA levels was more pronounced. 24 hpi SARS-CoV-1 vRNA was reduced ~1,288-, ~157-,

and ~22-fold (Supplementary Table S9) upon treatments with 10, 3, and 1 μ M PRI-724, respectively. In A549-AT and Calu-3, Remdesivir treatment appeared more effective in terms of vRNA reduction compared to PRI-724, but timepoints of significant reductions were comparable between both treatments. After 24 h in A549-AT, Remdesivir reduced SARS-CoV-2 and SARS-CoV-1 vRNA 270-fold and ~3,142-fold, respectively (Supplementary Table S9). In Calu-3, SARS-CoV-2 RNA was reduced ~249-fold (Supplementary Table S9) at 24 hpi upon treatment compared to the DMSO control (Figure 3A). In general, reduced sg-N copies correlated with lower total N RNA in A549-AT and Calu-3 cells (Supplementary Figure S2).

To quantify infectious virus in supernatant, TCID₅₀ and plaque assays were implemented (Figure 3B). In the supernatant of SARS-CoV-2 infected A549-AT cells, infectious virus was not detectable following treatment with Remdesivir and 10 μ M PRI-724. Virus titer was ~17.5-fold reduced upon treatment with 3 μ M PRI-724 and no significant reduction was detected with 1 μ M PRI-724, which is in line with the obtained RNA levels (Figure 3A). SARS-CoV-1 titer was reduced ~10⁸-, ~10⁵-, 10³-, and ~31-fold upon treatment with Remdesivir, 10, 3, and 1 μ M PRI-724, respectively. Quantification of SARS-CoV-2 in Calu-3 supernatants at 24 hpi by plaque assay revealed 32.5-, 20.8-, and 8.1-fold reduction in infectious virus titer, when treated with 20, 15, and 10 μ M PRI-724, respectively. Overall, a concentration-dependent reduction of SARS-CoV-2 and SARS-CoV-1 vRNA and infectious virus titers upon PRI-724 treatment in A549-AT and Calu-3 cells was observed.

3.7. PRI-724 downregulates β -catenin/CBP target genes *CDC25A*, *BIRC5* and cyclin D1 protein levels and upregulates β -catenin/p300 target c-Jun

After illuminating the protective *in vitro* properties of PRI-724 during coronavirus infection in A549-AT and Calu-3 cells, we sought to investigate the underlying mechanism. In order to assess β -catenin/CBP activity in A549-AT cells, we implemented a TOPFlash reporter assay (Figure 4A). The TOPFlash plasmid contains TCF/LEF binding sites coupled to a firefly luciferase reporter. Nuclear translocated

β -catenin binds to TCF/LEF transcription factors to recruit the co-factor CBP to the regulatory element (Lee et al., 2010) and thereby activates transcription. Treatment with 10, 3, and 1 μ M PRI-724 inhibited TOPFlash reporter activity significantly in comparison to the uninduced control. Moreover, induction of TOPFlash activity by 10 mM LiCl was effectively blocked when using the same concentrations. The control construct with mutant TCF/LEF binding sites, showed almost no luciferase activity and no differences among treatments (Supplementary Figure S3).

We further verified the effect of PRI-724 on β -catenin/CBP activity by RT-qPCR and western blot analysis. The expression of the direct target genes *CDC25A* and *BIRC5* were reduced ~2-fold in PRI-724-treated A549-AT cells (Figure 4B). Also, t protein levels of cyclin D1 were reduced when treated with 10 and 3 μ M PRI-724. Reciprocally, β -catenin/p300 target c-Jun was increased in a concentration dependent fashion (Figure 4C). β -catenin expression remained unchanged upon treatments.

3.8. PRI-724 causes cell cycle deregulation in A549-AT cells

It has been reported that PRI-724 causes cell cycle deregulation and block proliferation in several cancer cell lines, such as hepatocellular carcinoma (Gabata et al., 2020), pancreatic (Martinez-Font et al., 2020), and colon cancer (Emami et al., 2004; Kleszcz et al., 2020). We therefore determined whether PRI-724 and/or SARS-CoV-2 infection (MOI 0.01) effects host cell cycle progression. Flow cytometry analysis revealed a significant reduction in G1 (55–60% to 40%) and an increase in S cells (20–22% to 32–35%) upon treatment with PRI-724 compared to DMSO. No significant differences were obtained in G2/M populations. Sub-G1 populations were significantly elevated upon PRI-724 treatment and infected samples showed significantly higher sub-G1 populations than non-infected samples. SARS-CoV-2 infection did not seem to have any significant influence on the cell cycle distribution, although we noticed a slight but significant reduction in the S population of DMSO-treated infected cells compared to non-infected cells. Furthermore, western blot confirmed decreased amounts of Survivin and Cyclin D1 upon treatment with PRI-724. Also, infected cells treated with DMSO but not with PRI-724 showed slightly increased CDK4 expression (Figure 5A). PRI-724-treated A549-AT cells also showed reduced proliferation (Figure 5B). To visualize whether PRI-724 induces lateral growth hindrance, a scratch assay was implemented. 10 μ M PRI-724 significantly hampered scratch closure at 24 h post treatment. After 24 h, the scratch size remained at ~310 μ m over the course of the experiment while the scratch was almost entirely closed up in the control after 72 h (Figure 5B).

4. Discussion

In this study we provide a comprehensive and robust machine learning-based host factor identification strategy (Xiao et al., 2015) using HDF screens performed for SARS-CoV-2 (Daniloski et al., 2021; Wang R. et al., 2021; Wei et al., 2021; Zhu et al., 2021) and a drug screen (Ellinger et al., 2021) along with validation of the predicted hit in the wet-lab. HDF screens of human or African green monkey cells

infected with SARS-CoV-2 showed limited overlap in their hits. A better consistency was observed when regarding cellular pathways enriched in observed HDF of the screens. We reason this observation by variances among the different screens due to diverse experimental/technical settings or/and observed biological entities, such as different host cell types or different viral strains. To some extent, such differences may be cleaned out, when regarding a gene set representing a cellular pathway which is essential for the virus. It may contain parts of signaling cascades supporting each other making knocked out genes replaceable in specific conditions. Such interdependencies are highly relevant for discovering drug targets and serve as a promising research venue for future investigations. We took another path here and let machines learn which genes are indispensable, independent of the experimental settings in an automated way. Our machines returned lists of predicted HDF that reconstruct HDF and non-HDF class labels well from unseen data when cross-validated. This proves that the machines can capture the data structure in the gold standard lists well. We followed up on investigating the predictions for their biological content, and particularly their enrichments in known cellular processes and functions. Interestingly, we found a major portion of enriched gene sets being related to development and morphogenesis, followed by gene sets being related to neural processes. This is an intriguing observation in itself, suggesting that cellular processes for development and morphogenesis in particular are hijacked by the virus to facilitate reprogramming of the host cell from cellular proliferation to the proliferation of virus particles.

Focusing on gene sets related to development and morphogenesis, we selected β -catenin due to its top ranking connectivity and centrality scores in a constructed protein–protein interaction network, and it was highly expressed during viral infection in our observed cells. β -catenin has been intensively studied, particularly in oncology. It is known for its highly ambivalent roles. In its classical role, it migrates to the nucleus and acts as a cofactor for p300 or CBP (Teo and Kahn, 2010). Investigating hematopoietic stem cells (Rebel et al., 2002) and an embryonal murine stem cell line (F9; Kawasaki et al., 1998; Ugai et al., 1999), it was shown that the choice of the interaction partner CBP or p300 matters for the critical decision between proliferation/(pluri)potency and initiation of differentiation. When interrogating publicly available small compound databases, we found PRI-724 and ICG-001 as β -catenin inhibitors. More specifically, they inhibit the interaction of β -catenin with CBP. PRI-724 is a second-generation structural derivative of ICG-001. We sought to explore the effect of both of these CBP/ β -catenin inhibitors. PRI-724 performed better than ICG-001 (Supplementary Figure S1) and showed an inhibitory effect on pathogenic viruses such as SARS-CoV-2 variants, SARS-CoV-1, MERS-CoV, and IAV, in two different cell culture models. The inhibition doses were cell-type dependent which might be due to varying proliferation rates of the cells and different basal levels of β -catenin and CBP.

Viruses manipulate the cell cycle to create resources and cellular conditions that are advantageous for viral replication and assembly. Since PRI-724 is also known to affect cell cycle distribution, we hypothesized that the PRI-724-dependent reduction in virus replication and viral progeny production may be due to cell cycle deregulation. Sui et al. observed that SARS-CoV-2 induces dose dependent S and G2/M arrest at the early phase of infection, particularly, when higher MOIs were used, to facilitate virus replication (Sui et al., 2023), and, in turn, at a higher MOI (0.1), a significant decline of cells in S and G2/M phase at the late phase of infection (24 and 48 hpi). In our experiments we did not observe

any significant difference 24 hpi. This is most likely due to lower MOI (0.01) used in our study and is hence in line with the late phase observations from Sui et al. We detected that PRI-724-dependent G1 population decrease and S population increase is independent of the infection and thus it can effectively reduce virus replication and production for different SARS-CoV-2 variants, SARS-CoV-1, MERS-CoV and IAV. Regarding the drug toxicity, the total number of cells was lower after PRI-724 treatment (Supplementary Figure S4) and flow cytometry analysis showed a slight increase in sub-G1 population which was significantly pronounced following SARS-CoV-2 infection (Figure 5A). The cell distribution pattern we found differs from the previously published work showing an increase in G0/G1 cell population (Arensman et al., 2014; Gabata et al., 2020) while using ICG-001 and C-82 (active compound of PRI-724) when applied to non-infected cells. This observation seems to be cell line or compound type dependent, because another study using colon carcinoma HCT116 cells also found a decrease in G1 and an increase in S cell population following PRI-724 treatment (Kleszcz et al., 2020). In summary, SARS-CoV-2 can reprogram the cell cycle in host cells. Our investigation showed that blocking the interaction between β -catenin and CBP with PRI-724 significantly reduced cytopathic effects, viral RNA replication, and infectious virus production. The treatment caused cell cycle deregulation in the host cell, which appeared to be unrelated to the infection. This suggests that our treatment effectively disrupts the virus-induced reprogramming of the cell cycle, which is essential for viruses to complete their life cycles. PRI-724 treatment has been also clinically investigated. PRI-724 was shown to reduce hepatitis C virus-induced liver fibrosis in mice (Bojkova et al., 2020). These investigations were followed up in two clinical trials of the same group (phase I and IIa, respectively; Kimura et al., 2017, 2022). The authors concluded that PRI-724 treatment was well tolerated by patients with HBV and HCV induced liver-cirrhosis and showed improvements of the pathology of concern in several patients. Besides this, blockade of β -catenin/CBP reversed pulmonary fibrosis (Henderson et al., 2010), which is a central COVID-19 complication (Zou et al., 2021). A recent study confirmed overall higher β -catenin levels in SARS-CoV-2 infected patients and targeting reduced virus shedding *in vitro* (Chatterjee et al., 2022).

A limitation of our study is that the machine learning procedure bases on screening data of the four initial screens which were at hand when we developed the computational method. New screens have been, and certainly more will be published as means to better understand the mechanisms of this wide spread and evolving virus. As a future outlook, we would like to repeat the computational analysis and include new screens that have been published since the initiation of this project. We thereby might identify additional HDF that are relevant for the SARS-CoV-2 life cycle. We would like to emphasize that further investigations with 3D models of primary cells, e.g., human bronchial epithelial cells (HBEpC) and human nasal epithelial cells (HNEpC) are crucial to validate our observations. 3D models have the ability to replicate cellular behavior *in vivo* more effectively than immortalized or cancer cell lines. These cell models can also offer a more accurate representation of host-pathogen interactions *in vitro* compared to 2D systems, resulting in more precise data and a superior prediction of drug efficacy and toxicity.

In conclusion, the machine learning approach brought up interesting connections between viral replication and the reprogramming of host cells in the context of SARS-CoV-2 infection. It may also suit for host factor studies of other viruses and a comparison between different virus entities may lead to intriguing new aspects in

virus biology. We observed a strong correlation between the inhibition of CPE, vRNA, and infectious virus production for the tested concentrations in two different cell lines, making PRI-724 a promising HDA against SARS-CoV-2 and SARS-CoV-1 in our models.

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 authors.

Author contributions

TT, HE, and RK conceived the study. AV-G, AH, TB, MO, and RK performed data analysis and machine learning. MK and TT designed and performed the experiments. NB and PW performed the experiments on cytotoxicity measurement by fluorescence microscopy. TT, SC, RK, and HE acquired funding. All authors contributed to the article and approved the submitted version.

Funding

MK and AH were supported by Research Grants of Merck KGaA, Darmstadt, Germany. TT was funded by Willy Robert Pitzer Foundation. MK and TT were supported by the Federal Ministry of Education and Research (BMBF), Germany, FKZ: 01KI2015OB. MO and AH were supported by the Federal Ministry of Education and Research (BMBF), Germany, FKZ: 01EO1502 (CSCC), 13N15711 (LPI), 13N15745 (SARS-CoV2-DX), and 01KI2015OA (SARSirRNA). TB was supported by the Deutsche Forschungsgemeinschaft (KO-3678/5-1). We acknowledge support by the German Research Foundation Project number 512648189 and the Open Access Publication Fund of the Thueringer Universitaets- und Landesbibliothek Jena.

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.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1193320/full#supplementary-material>

References

- Acencio, M. L., and Lemke, N. (2009). Towards the prediction of essential genes by integration of network topology, cellular localization and biological process information. *BMC Bioinformatics*. 10:290. doi: 10.1186/1471-2105-10-290
- Akiyama, T. (2000). Wnt/beta-catenin signaling. *Cytokine Growth Factor Rev.* 11, 273–282. doi: 10.1016/S1359-6101(00)00011-3
- Alipanahi, B., Delong, A., Weirauch, M. T., and Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 33, 831–838. doi: 10.1038/nbt.3300
- Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* 33, 3387–3395. doi: 10.1093/bioinformatics/btx431
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Arensman, M. D., Telesca, D., Lay, A. R., Kershaw, K. M., Wu, N., Donahue, T. R., et al. (2014). The CREB-binding protein inhibitor ICG-001 suppresses pancreatic cancer growth. *Mol. Cancer Ther.* 13, 2303–2314. doi: 10.1158/1535-7163.MCT-13-1005
- Armstrong, J. F., Faccenda, E., Harding, S. D., Pawson, A. J., Southan, C., Sharman, J. L., et al. (2020). The IUPHAR/BPS guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV guide to MALARIA PHARMACOLOGY. *Nucleic Acids Res.* 48, D1006–D1021. doi: 10.1093/nar/gkz951
- Aromolaran, O., Beder, T., Oswald, M., Oyelade, J., Adebisi, E., and Koenig, R. (2020). Essential gene prediction in *Drosophila melanogaster* using machine learning approaches based on sequence and functional features. *Comput. Struct. Biotechnol. J.* 18, 612–621. doi: 10.1016/j.csbj.2020.02.022
- Beder, T., Aromolaran, O., Dönitz, J., Tapanelli, S., Adedeji, E. O., Adebisi, E., et al. (2021). Identifying essential genes across eukaryotes by machine learning. *NAR Genomics Bioinformatics* 3:110. doi: 10.1093/nargab/lqab110
- Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., et al. (2020). Remdesivir for the treatment of Covid-19—final report. *N. Engl. J. Med.* 383, 1813–1826. doi: 10.1056/NEJMoa2007764
- Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., et al. (2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* 583, 469–472. doi: 10.1038/s41586-020-2332-7
- Bordonaro, M., and Lazarova, D. L. (2016). Determination of the role of CBP- and p300-mediated Wnt signaling on colonic cells. *JMIR Res Protoc.* 5:e66. doi: 10.2196/resprot.5495
- Charif, D., Thioulouse, J., Lobry, J. R., and Perrière, G. (2005). Online synonymous codon usage analyses with the ade4 and seqinR packages. *Bioinformatics* 21, 545–547. doi: 10.1093/bioinformatics/bti037
- Chatterjee, S., Keshry, S. S., Ghosh, S., Ray, A., and Chattopadhyay, S. (2022). Versatile β -catenin is crucial for SARS-CoV-2 infection. *Microbiol Spectr.* 10:e0167022. doi: 10.1128/spectrum.01670-22
- Daniloski, Z., Jordan, T. X., Wessels, H. H., Hoagland, D. A., Kasela, S., Legut, M., et al. (2021). Identification of required host factors for SARS-CoV-2 infection in human cells. *Cells* 184, 92–105.e16. doi: 10.1016/j.cell.2020.10.030
- Dejnirattisai, W., Zhou, D., Supasa, P., Liu, C., Mentzer, A. J., Ginn, H. M., et al. (2021). Antibody evasion by the P.1 strain of SARS-CoV-2. *Cells* 184, 2939–2954.e9. doi: 10.1016/j.cell.2021.03.055
- Dietz, C., and Berthold, M. R. (2016). KNIME for open-source bioimage analysis: a tutorial. *Adv. Anat. Embryol. Cell Biol.* 219, 179–197. doi: 10.1007/978-3-319-28549-8_7
- Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1967–1976. doi: 10.1056/NEJMoa030747
- Edinger TOPohl, M. O., and Stertz, S. (2014). Entry of influenza A virus: host factors and antiviral targets. *J. Gen. Virol.* 95, 263–277. doi: 10.1099/vir.0.059477-0
- Ellinger, B., Bojkova, D., Zaliani, A., Cinatl, J., Claussen, C., Westhaus, S., et al. (2021). A SARS-CoV-2 cytopathicity dataset generated by high-content screening of a large drug repurposing collection. *Sci Data* 8:70. doi: 10.1038/s41597-021-00848-4
- Emami, K. H., Nguyen, C., Ma, H., Kim, D. H., Jeong, K. W., Eguchi, M., et al. (2004). A small molecule inhibitor of beta-catenin/CREB-binding protein transcription [corrected]. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12682–12687. doi: 10.1073/pnas.0404875101
- Flores-Vega, V. R., Monroy-Molina, J. V., Jiménez-Hernández, L. E., Torres, A. G., Santos-Preciado, J. I., and Rosales-Reyes, R. (2022). SARS-CoV-2: evolution and emergence of new viral variants. *Viruses* 14:653. doi: 10.3390/v14040653
- Gabata, R., Harada, K., Mizutani, Y., Uchi, H., Yoshimura, K., Sato, Y., et al. (2020). Anti-tumor activity of the small molecule inhibitor PRI-724 against β -catenin-activated hepatocellular carcinoma. *Anticancer Res* 40, 5211–5219. doi: 10.21873/anticancer.14524
- Gandhi, S., Klein, J., Robertson, A. J., Peña-Hernández, M. A., Lin, M. J., Roychoudhury, P., et al. (2022). De novo emergence of a remdesivir resistance mutation during treatment of persistent SARS-CoV-2 infection in an immunocompromised patient: a case report. *Nat. Commun.* 13:1547. doi: 10.1038/s41467-022-29104-y
- Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. (2016). BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* 44, D1045–D1053. doi: 10.1093/nar/gkv1072
- Hammond, J., Leister-Tebbe, H., Gardner, A., Abreu, P., Bao, W., Wisemandle, W., et al. (2022). Oral Nirmatrelvir for high-risk, nonhospitalized adults with Covid-19. *N. Engl. J. Med.* 386, 1397–1408. doi: 10.1056/NEJMoa2118542
- Henderson, W. R., Chi, E. Y., Ye, X., Nguyen, C., Tzang, T. Y., Zhou, B., et al. (2010). Inhibition of Wnt/beta-catenin/CREB binding protein (CBP) signaling reverses pulmonary fibrosis. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14309–14314. doi: 10.1073/pnas.1001520107
- Hershberg, R., and Petrov, D. A. (2009). General rules for optimal codon choice. *PLoS Genet.* 5:e1000556. doi: 10.1371/journal.pgen.1000556
- Hogan, J. I., Duerr, R., Dimartino, D., Marier, C., Hochman, S. E., Mehta, S., et al. (2019). Remdesivir resistance in transplant recipients with persistent coronavirus disease 2019. *Clin Infect Dis Off Publ Infect Dis Soc Am.* 76, 342–345. doi: 10.1093/cid/ciac769
- Hörhold, F., Eisel, D., Oswald, M., Kolte, A., Röhl, D., Osen, W., et al. (2020). Reprogramming of macrophages employing gene regulatory and metabolic network models. *PLoS Comput. Biol.* 16:e1007657. doi: 10.1371/journal.pcbi.1007657
- Iketani, S., Liu, L., Guo, Y., Liu, L., Chan, J. F. W., Huang, Y., et al. (2022). Antibody evasion properties of SARS-CoV-2 omicron sublineages. *Nature* 604, 553–556. doi: 10.1038/s41586-022-04594-4
- Jayk Bernal, A., Gomes da Silva, M. M., Musungaie, D. B., Kovalchuk, E., Gonzalez, A., Delos Reyes, V., et al. (2022). Molnupiravir for Oral treatment of Covid-19 in nonhospitalized patients. *N. Engl. J. Med.* 386, 509–520. doi: 10.1056/NEJMoa2116044
- Kawasaki, H., Eckner, R., Yao, T. P., Taira, K., Chiu, R., Livingston, D. M., et al. (1998). Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation. *Nature* 393, 284–289. doi: 10.1038/30538
- Kimura, K., Ikoma, A., Shibakawa, M., Shimoda, S., Harada, K., Saio, M., et al. (2017). Safety, tolerability, and preliminary efficacy of the anti-fibrotic small molecule PRI-724, a CBP/ β -catenin inhibitor, in patients with hepatitis C virus-related cirrhosis: a single-center, open-label, dose escalation phase 1 trial. *EBioMedicine* 23, 79–87. doi: 10.1016/j.ebiom.2017.08.016
- Kimura, K., Kanto, T., Shimoda, S., Harada, K., Kimura, M., Nishikawa, K., et al. (2022). Safety, tolerability, and anti-fibrotic efficacy of the CBP/ β -catenin inhibitor PRI-724 in patients with hepatitis C and B virus-induced liver cirrhosis: an investigator-initiated, open-label, non-randomised, multicentre, phase 1/2a study. *EBioMedicine* 80:104069. doi: 10.1016/j.ebiom.2022.104069
- Klann, K., Bojkova, D., Tascher, G., Ciesek, S., Münch, C., and Cinatl, J. (2020). Growth factor receptor signaling inhibition prevents SARS-CoV-2 replication. *Mol. Cell* 80, 164–174.e4. doi: 10.1016/j.molcel.2020.08.006
- Kleszcz, R., Krajka-Kuźniak, V., and Paluszczak, J. (2020). Porcupine and CBP/ β -catenin are the most suitable targets for the inhibition of canonical Wnt signaling in colorectal carcinoma cell lines*. *Postępy Hig Med Dośw.* 74, 224–235. doi: 10.5604/01.3001.0014.2497
- Lee, C. H., Hung, H. W., Hung, P. H., and Shieh, Y. S. (2010). Epidermal growth factor receptor regulates beta-catenin location, stability, and transcriptional activity in oral cancer. *Mol. Cancer* 9:64. doi: 10.1186/1476-4598-9-64
- Liu, J., Xiao, Q., Xiao, J., Niu, C., Li, Y., Zhang, X., et al. (2022). Wnt/ β -catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduct. Target. Ther.* 7:3. doi: 10.1038/s41392-021-00762-6
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Mahajan, S., Choudhary, S., Kumar, P., and Tomar, S. (2021). Antiviral strategies targeting host factors and mechanisms obliging +ssRNA viral pathogens. *Bioorg. Med. Chem.* 46:116356. doi: 10.1016/j.bmc.2021.116356
- Martinez-Font, E., Pérez-Capó, M., Ramos, R., Felipe, I., Garcías, C., Luna, P., et al. (2020). Impact of Wnt/ β -catenin inhibition on cell proliferation through CDC25A downregulation in soft tissue sarcomas. *Cancer* 12:2556. doi: 10.3390/cancers12092556
- Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Félix, E., et al. (2019). ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* 47, D930–D940. doi: 10.1093/nar/gky1075
- Nagy, P. D., and Pogany, J. (2011). The dependence of viral RNA replication on co-opted host factors. *Nat. Rev. Microbiol.* 10, 137–149. doi: 10.1038/nrmicro2692
- Oughtred, R., Stark, C., Breitkreutz, B. J., Rust, J., Boucher, L., Chang, C., et al. (2019). The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 47, D529–D541. doi: 10.1093/nar/gky1079
- Pizzato, M., Baraldi, C., Boscato Soppetto, G., Fiozzi, D., Gentile, C., Gentile, M. D., et al. (2022). SARS-CoV-2 and the host cell: a tale of interactions. *Front Virol* 1:388. doi: 10.3389/fviro.2021.815388

- Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M. M., et al. (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* 596, 276–280. doi: 10.1038/s41586-021-03777-9
- Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005). NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 33, D501–D504. doi: 10.1093/nar/gki025
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). G:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. doi: 10.1093/nar/gkz369
- Rebel, V. I., Kung, A. L., Tanner, E. A., Yang, H., Bronson, R. T., and Livingston, D. M. (2002). Distinct roles for CREB-binding protein and p300 in hematopoietic stem cell self-renewal. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14789–14794. doi: 10.1073/pnas.232568499
- Reggiori, F., Monastyrska, I., Verheije, M. H., Cali, T., Ulasli, M., Bianchi, S., et al. (2010). Coronaviruses hijack the LC3-I-positive EDEMosomes, ER-derived vesicles exporting short-lived ERAD regulators, for replication. *Cell Host Microbe* 7, 500–508. doi: 10.1016/j.chom.2010.05.013
- Shi, S. T., and Lai, M. M. C. (2005). Viral and cellular proteins involved in coronavirus replication. *Curr. Top. Microbiol. Immunol.* 287, 95–131. doi: 10.1007/3-540-26765-4_4
- Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., et al. (2009). BioMart-biological queries made easy. *BMC Genomics* 10:22. doi: 10.1186/1471-2164-10-22
- Spagnolo, J. F., and Hogue, B. G. (2000). Host protein interactions with the 3' end of bovine coronavirus RNA and the requirement of the poly(a) tail for coronavirus defective genome replication. *J. Virol.* 74, 5053–5065. doi: 10.1128/JVI.74.11.5053-5065.2000
- Spandidos, A., Wang, X., Wang, H., Dragnev, S., Thurber, T., and Seed, B. (2008). A comprehensive collection of experimentally validated primers for polymerase chain reaction quantitation of murine transcript abundance. *BMC Genomics* 9:633. doi: 10.1186/1471-2164-9-633
- Sui, L., Li, L., Zhao, Y., Zhao, Y., Hao, P., Guo, X., et al. (2023). Host cell cycle checkpoint as antiviral target for SARS-CoV-2 revealed by integrative transcriptome and proteome analyses. *Signal Transduct. Target. Ther.* 8:21. doi: 10.1038/s41392-022-01296-1
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131
- Teo, J. L., and Kahn, M. (2010). The Wnt signaling pathway in cellular proliferation and differentiation: a tale of two coactivators. *Adv. Drug Deliv. Rev.* 62, 1149–1155. doi: 10.1016/j.addr.2010.09.012
- Thorn, C. F., Klein, T. E., and Altman, R. B. (2013). PharmGKB: the pharmacogenomics Knowledge Base. *Methods Mol Biol* 1015, 311–320. doi: 10.1007/978-1-62703-435-7_20
- Toptan, T., Hoehl, S., Westhaus, S., Bojkova, D., Berger, A., Rotter, B., et al. (2020). Optimized qRT-PCR approach for the detection of intra- and extra-cellular SARS-CoV-2 RNAs. *Int. J. Mol. Sci.* 21:4396. doi: 10.3390/ijms21124396
- Ugai, H., Uchida, K., Kawasaki, H., and Yokoyama, K. K. (1999). The coactivators p300 and CBP have different functions during the differentiation of F9 cells. *J. Mol. Med. Berl Ger.* 77, 481–494. doi: 10.1007/s001099900021
- Ursu, O., Holmes, J., Knockel, J., Bologa, C. G., Yang, J. J., Mathias, S. L., et al. (2017). DrugCentral: online drug compendium. *Nucleic Acids Res.* 45, D932–D939. doi: 10.1093/nar/gkw993
- Van Egeren, D., Novokhodko, A., Stoddard, M., Tran, U., Zetter, B., Rogers, M. S., et al. (2021). Controlling long-term SARS-CoV-2 infections can slow viral evolution and reduce the risk of treatment failure. *Sci. Rep.* 11:22630. doi: 10.1038/s41598-021-02148-8
- Veleanu, A., Kelch, M. A., Ye, C., Flohr, M., Wilhelm, A., Widera, M., et al. (2022). Molecular analyses of clinical isolates and recombinant SARS-CoV-2 carrying B.1 and B.1.617.2 spike mutations suggest a potential role of non-spike mutations in infection kinetics. *Viruses* 14:2017. doi: 10.3390/v14092017
- Vierbuchen, T., and Wernig, M. (2012). Molecular roadblocks for cellular reprogramming. *Mol. Cell* 47, 827–838. doi: 10.1016/j.molcel.2012.09.008
- Wagoner, J., Herring, S., Hsiang, T. Y., Ianevski, A., Biering, S. B., Xu, S., et al. (2022). Combinations of host- and virus-targeting antiviral drugs confer synergistic suppression of SARS-CoV-2. *Microbiol. Spectr.* 10:e0333122. doi: 10.1128/spectrum.03331-22
- Wählby, C., Sintorn, I. M., Erlandsson, F., Borgefors, G., and Bengtsson, E. (2004). Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *J. Microsc.* 215, 67–76. doi: 10.1111/j.0022-2720.2004.01338.x
- Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., et al. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 593, 130–135. doi: 10.1038/s41586-021-03398-2
- Wang, R., Simoneau, C. R., Kulsuptrakul, J., Bouhaddou, M., Travisano, K. A., Hayashi, J. M., et al. (2021). Genetic screens identify host factors for SARS-CoV-2 and common cold coronaviruses. *Cells* 184, 106–119.e14. doi: 10.1016/j.cell.2020.12.004
- Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., et al. (2020). Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* 48, D1031–D1041. doi: 10.1093/nar/gkz981
- Wei, J., Alfajaro, M. M., DeWeirdt, P. C., Hanna, R. E., Lu-Culligan, W. J., Cai, W. L., et al. (2021). Genome-wide CRISPR screens reveal host factors critical for SARS-CoV-2 infection. *Cells* 184, 76–91.e13. doi: 10.1016/j.cell.2020.10.028
- Widera, M., Wilhelm, A., Hoehl, S., Pallas, C., Kohmer, N., Wolf, T., et al. (2021b). Limited neutralization of authentic severe acute respiratory syndrome coronavirus 2 variants carrying E484K in vitro. *J. Infect. Dis.* 224, 1109–1114. doi: 10.1093/infdis/jiab355
- Widera, M., Wilhelm, A., Toptan, T., Raffel, J. M., Kowarz, E., Roesmann, F., et al. (2021a). Generation of a sleeping beauty transposon-based cellular system for rapid and sensitive screening for compounds and cellular factors limiting SARS-CoV-2 replication. *Front. Microbiol.* 12:701198. doi: 10.3389/fmicb.2021.701198
- Wilhelm, A., Toptan, T., Pallas, C., Wolf, T., Goetsch, U., Gottschalk, R., et al. (2021). Antibody-mediated neutralization of authentic SARS-CoV-2 B.1.617 variants harboring L452R and T478K/E484Q. *Viruses* 13:1693. doi: 10.3390/v13091693
- Wilhelm, A., Widera, M., Grikscheit, K., Toptan, T., Schenk, B., Pallas, C., et al. (2022). Limited neutralisation of the SARS-CoV-2 omicron subvariants BA.1 and BA.2 by convalescent and vaccine serum and monoclonal antibodies. *EBioMedicine* 82:104158. doi: 10.1016/j.ebiom.2022.104158
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi: 10.1093/nar/gkx1037
- Wong, C. K. H., Au, I. C. H., Lau, K. T. K., Lau, E. H. Y., Cowling, B. J., and Leung, G. M. (2022). Real-world effectiveness of early molnupiravir or nirmatrelvir-ritonavir in hospitalised patients with COVID-19 without supplemental oxygen requirement on admission during Hong Kong's omicron BA.2 wave: a retrospective cohort study. *Lancet Infect. Dis.* 22, 1681–1693. doi: 10.1016/S1473-3099(22)00507-2
- Xiao, N., Cao, D. S., Zhu, M. F., and Xu, Q. S. (2015). Protr/ProtrWeb: R package and web server for generating various numerical representation schemes of protein sequences. *Bioinformatics* 31, 1857–1859. doi: 10.1093/bioinformatics/btv042
- Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E., and Fouchier, R. A. M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367, 1814–1820. doi: 10.1056/NEJMoa1211721
- Zhu, Y., Feng, F., Hu, G., Wang, Y., Yu, Y., Zhu, Y., et al. (2021). A genome-wide CRISPR screen identifies host factors that regulate SARS-CoV-2 entry. *Nat. Commun.* 12:961. doi: 10.1038/s41467-021-21213-4
- Zou, J. N., Sun, L., Wang, B. R., Zou, Y., Xu, S., Ding, Y. J., et al. (2021). The characteristics and evolution of pulmonary fibrosis in COVID-19 patients as assessed by AI-assisted chest HRCT. *PLoS One* 16:e0248957. doi: 10.1371/journal.pone.0248957



OPEN ACCESS

EDITED BY

Sinosh Skariyachan,
St. Pius X College, India

REVIEWED BY

Chandrabose Selvaraj,
Saveetha University, India
Bruno Andrade,
Universidade Estadual do Sudoeste da Bahia,
Brazil

*CORRESPONDENCE

Soumya V. Menon
✉ sweetsou_02@yahoo.com
Biswaranjan Paital
✉ biswaranjanpaital@gmail.com
Dipak Kumar Sahoo
✉ dsahoo@iastate.edu;
✉ dipaksahoo11@gmail.com

RECEIVED 27 March 2023

ACCEPTED 05 June 2023

PUBLISHED 23 June 2023

CITATION

Prabhakaran P, Hebbani AV, Menon SV, Paital B,
Murmu S, Kumar S, Singh MK, Sahoo DK and
Desai PPD (2023) *In silico* generation of novel
ligands for the inhibition of SARS-CoV-2 main
protease (3CL^{pro}) using deep learning.
Front. Microbiol. 14:1194794.
doi: 10.3389/fmicb.2023.1194794

COPYRIGHT

© 2023 Prabhakaran, Hebbani, Menon, Paital,
Murmu, Kumar, Singh, Sahoo and Desai. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

In silico generation of novel ligands for the inhibition of SARS-CoV-2 main protease (3CL^{pro}) using deep learning

Prejwal Prabhakaran^{1,2}, Ananda Vardhan Hebbani³,
Soumya V. Menon^{4*}, Biswaranjan Paital^{5*}, Sneha Murmu⁶,
Sunil Kumar⁶, Mahender Kumar Singh⁷, Dipak Kumar Sahoo^{8*}
and Padma Priya Dharmavaram Desai⁹

¹Department of Biotechnology, New Horizon College of Engineering, Bangalore, India, ²Faculty of Biology, Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany, ³Department of Biochemistry, Indian Academy Degree College (Autonomous), Bangalore, India, ⁴Department of Chemistry and Biochemistry, School of Sciences, Jain (Deemed-to-be) University, Bangalore, India, ⁵Redox Regulation Laboratory, Department of Zoology, College of Basic Science and Humanities, Odisha University of Agriculture and Technology, Bhubaneswar, India, ⁶ICAR-Indian Agricultural Statistics Research Institute, PUSA, New Delhi, India, ⁷DBT-National Brain Research Centre, Gurugram, India, ⁸Department of Veterinary Clinical Sciences, College of Veterinary Medicine, Iowa State University, Ames, IA, United States, ⁹Department of Basic Sciences, New Horizon College of Engineering, Bangalore, India

The recent emergence of novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing the coronavirus disease (COVID-19) has become a global public health crisis, and a crucial need exists for rapid identification and development of novel therapeutic interventions. In this study, a recurrent neural network (RNN) is trained and optimized to produce novel ligands that could serve as potential inhibitors to the SARS-CoV-2 viral protease: 3 chymotrypsin-like protease (3CL^{pro}). Structure-based virtual screening was performed through molecular docking, ADMET profiling, and predictions of various molecular properties were done to evaluate the toxicity and drug-likeness of the generated novel ligands. The properties of the generated ligands were also compared with current drugs under various phases of clinical trials to assess the efficacy of the novel ligands. Twenty novel ligands were selected that exhibited good drug-likeness properties, with most ligands conforming to Lipinski's rule of 5, high binding affinity (highest binding affinity: −9.4 kcal/mol), and promising ADMET profile. Additionally, the generated ligands complexed with 3CL^{pro} were found to be stable based on the results of molecular dynamics simulation studies conducted over a 100 ns period. Overall, the findings offer a promising avenue for the rapid identification and development of effective therapeutic interventions to treat COVID-19.

KEYWORDS

SARS-CoV-2, recurrent neural network, deep learning, 3CL^{pro}, admet

1. Introduction

Coronavirus disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has become a global public health crisis. Vaccines saved many lives despite numerous clinical trials for medicines against SARS-CoV-2 is under process (World Health Organization, 2020; Das et al., 2023). With nearly 765 million cases and 6.9 million deaths worldwide as of 3rd May 2023,¹ there exists a vital need to identify or develop novel therapeutic interventions. Various studies have shown promising results in using repurposed drugs (reusing existing approved drugs for new medical indications) to inhibit the virus at different target sites (Elmezyen et al., 2020; Sarma et al., 2020). Among the target sites being considered, the 3-Chymotrypsin-like protease (3CL^{pro}), is hypothesized to be a crucial target for the development of drugs (Khan et al., 2020; Tahir ul Qamar et al., 2020). 3CL^{pro} is responsible for the cleavage of polyproteins to produce non-structural proteins essential for viral replication (Elmezyen et al., 2020). Therefore, targeting 3CL^{pro} can inhibit the maturation and replication of the virus. 3-Chymotrypsin-like protease (3CL^{pro}) and papain-like protease (PL^{pro}) are essential enzymes in the peptide chain processing reaction. They cleave the C-terminus of the polypeptide chain at 11 sites and the N-terminus of the polypeptide chain at three sites. The cleavage products include structural proteins and some important non-structural proteins, such as RNA-dependent RNA polymerase (RdRp) and helicase. With more cleavage sites, 3CL^{pro} serves as an attractive non-structural protein for the development of drugs targeting SARS-CoV-2 (Li et al., 2020). The structure details are attached as a separate Supplementary file.

This protease contains several highly conserved substrate-binding sites within the active site of the enzyme, making it an attractive target for developing a diverse range of inhibitors. It is also exciting that the structures of 3CL^{pro} in SARS-CoV-2 and SARS-CoV differ by only 12 amino acids with comparable ligand binding efficiency (Macchiagodena et al., 2020). The 3-D structure and other details of the protease are attached as a PDBfile (RCSB, 2022). Jin et al. (2020) utilized the SARS-CoV2-PPC (protease pharmacophore clusters) to identify six principal protease flexible conformations and active sites. The diverse druggable environments of the PPCs were explained by the presence of different sets of PPC consensus anchors in various PPCs, which affirmed the functionality of the PPCs. When a compound is present in a PPC, its protease binding affinities improve with an increasing number of occupied anchors, leading to a greater number of interactions (Pathak et al., 2021). The 3D crystalline structure of 3CL^{pro} was submitted to Protein Data Bank (PDB) in January 2020 under the PDB ID: 6LU7 (Jin et al., 2020), and it was

complexed with an N3 inhibitor. Thus, the active site of the N3 inhibitor could be chosen as the site for designing ligands that can potentially inhibit the activity of the protease (Corbeil et al., 2012; Paital et al., 2022).

During this period of a global pandemic, the drug discovery and development process must be accelerated, but one of the greatest impediments to this is the lead discovery process (Kadurin et al., 2017). To combat this issue, *in-silico* methods such as deep learning have emerged as a promising alternative, offering the potential to not only reduce costs but also significantly compress the timeline (Paital et al., 2015; Jin et al., 2020). These models can learn to generate new data that closely resembles the training data by extracting high-level features from the data (Kadurin et al., 2017). Deep learning has been successfully applied to generate novel molecules (Prykhodko et al., 2019) and has been reported to produce effective lead candidates in very little time (Gupta et al., 2017; Vanhaelen et al., 2017; Zhavoronkov et al., 2019).

In this study, a deep learning model based on a Recurrent Neural Network (RNN) was used to generate new ligands that could potentially act as inhibitors of 3CL^{pro}. RNNs are highly effective in modeling sequential data with a temporal relationship, where each data point depends on the previous one. In this case, the RNN was trained on chemical molecules represented as SMILES strings. The model learns the relationship between each ASCII character and its temporal dependence in the input SMILES strings and predicts the ASCII character in the SMILES string based on the previous characters. A Long Short Term Memory (LSTM) network was specifically selected, as vanilla RNNs suffer from the vanishing gradient problem, where the gradient becomes smaller and smaller for large sequences of data (Menon, 2022).

Molecular docking was performed by virtual screening to identify the best hits against the viral protease. Evaluation of the molecular properties of the ligands and absorption, distribution, metabolism, excretion, and toxicity (ADMET) analysis were performed to study the biological activity and pharmacokinetic properties of the generated ligands. Additionally, molecular dynamic (MD) simulation was employed to investigate the stability and interaction of the ligand-protease complex for a duration of 100 nanoseconds. Finally, the properties of the generated novel ligands were compared to drugs that are currently in clinical trials as a therapeutic intervention for COVID-19. This study evaluates the effectiveness and potential of the newly generated ligands in inhibiting the main viral protease (3CL^{pro}) of SARS-CoV-2.

2. Materials and methods

2.1. Technical implementation

The RNN was implemented using Tensorflow (v2.0²) and Keras (v2.3³) in Python (v3.7⁴) and RDkit⁵ was used for the processing of the molecules.

¹ <https://covid19.who.int/>

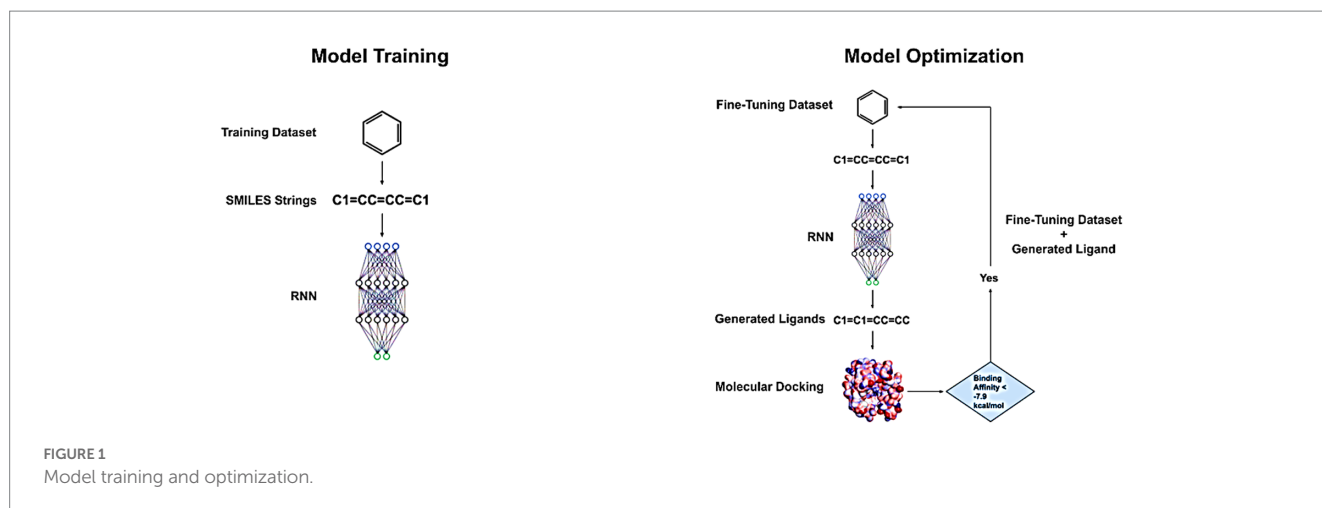
Abbreviations: 3CL^{pro}, 3 Chymotrypsin-like protease; ADMET, Absorption, distribution, metabolism, excretion and toxicity; BSEP, Bile salt export pump; HBA, Hydrogen bond acceptors; HBD, Hydrogen bond donors; MW, Molecular weight; NRB, Number of rotatable bonds; OCT, Organic cation transporters; RNN, Recurrent neural network; SDF, Structure data file format; SMILES, Sequence of the molecule in the simplified molecular-input line-entry system; TPSA, Total polar surface area; hERG, Human ether-a-go-go.

² <https://www.tensorflow.org>

³ <https://keras.io>

⁴ <https://www.python.org>

⁵ <https://www.rdkit.org>



2.2. Recurrent neural network

To accelerate the synthesis of potential inhibitors against 3CL^{pro}, a transfer learning approach was applied. Transfer learning is a machine learning technique where a pre-trained model is used as a starting point for training a new model with a similar task or domain. This allows the model to leverage the knowledge and experience gained from the pre-training to adapt to the new data and tasks more quickly and efficiently. In other words, transfer learning allows for faster and more accurate model development by building on top of previously learned representations.

Here, a publicly available model named LSTM_Chem (License: CC BY-NC-ND 4.0) was used (Gupta et al., 2017). The model consists of two LSTM layers with a 256-sized hidden state vector. It is regularized, having dropout layers. The two layers are followed by a final dense output layer with the softmax activation function. The model input is a bit array sequence of the molecule in the simplified molecular-input line-entry system (SMILES) format. This model was initially trained to produce novel TRPM8 inhibitors.

2.3. Dataset curation

For the LSTM_Chem model to generate potential inhibitors for 3CL^{pro}, it was necessary to retrain and optimize the model on a ligand dataset that exhibits a certain degree of activity against 3CL^{pro}. The training process involved two stages with distinct datasets. The first stage-trained the model on a training dataset to learn the latent space features of chemical molecules. The second stage involved fine-tuning the model using a separate dataset to enable it to generate ligands that possess the chemical features of protease inhibitors for COVID-19.

The training dataset consists of a large volume of diverse ligands from which the RNN learns to produce valid ligands with high accuracy (Figure 1). The dataset was obtained from ChEMBL22⁶ and contained 556,134 SMILES strings, which were processed to remove duplicates, salts, and stereochemical information, resulting in a

collection of unique ligands. Furthermore, only SMILES strings that had lengths between 34 and 74 tokens were retained, leading to a final size of 439,217 SMILES strings. This methodology was chosen following the work done by Gupta et al. (2017). The SMILES string length was constrained as having very long strings would result in the vanishing gradient problem, and the network would not learn anything. Although LSTMs are good at tackling the vanishing gradient, they are not completely immune to it (Moret et al., 2019). Additionally, the LSTM_Chem model accepts a bit array sequence as input, which was obtained by converting the SMILES strings using the Morgan algorithm in RDKit Open-source cheminformatics; (Open-Source Cheminformatics, see Footnote 5). The Morgan algorithm is a graph relaxation algorithm used for molecule canonicalization, which assigns a unique identifier to a molecule regardless of its representation. However, the Morgan algorithm has known issues that can result in noncanonical atom orderings and can be problematic when used with large molecules such as proteins. Therefore, restricting the length of the SMILES strings to fall within a range of 34–74 tokens limits the size of the molecules to small ligands and reduces the likelihood of encountering issues with the Morgan algorithm (Schneider et al., 2015; Moret et al., 2019; Schneider, 2019).

To create a dataset for fine-tuning the model, 845 drugs undergoing clinical trials and drugs that demonstrated activity in different biological assays for COVID-19 were collected from PubChem.⁷ This dataset included compounds that showed activity against not only 3CL^{pro} but other drug targets of SAR-CoV-2 as well. In PubChem BioAssay, “PUBCHEM_ACTIVITY_OUTCOME” is a column that reports the outcome of a specific assay run for a given compound. It describes whether the tested compound showed activity (i.e., produced a measurable effect) against the target of interest or not. Only compounds that were termed “active” were selected. The ligand structures were retrieved in SDF format using compound IDs, via PUG_REST, an application program interface (API) for accessing PubChem (Kim et al., 2018), and SMILES strings were generated using the *MolToSmiles* function from RDKit. This dataset was preprocessed similarly to the training dataset, resulting in 639 ligands

⁶ <https://www.ebi.ac.uk/chembl>

⁷ <https://pubchem.ncbi.nlm.nih.gov/#query=covid-19&tab=bioassay>

that were used for fine-tuning the model. However, unlike the training dataset, no restriction was set on the string length of the SMILES in the fine-tuning dataset.

2.4. Training and optimization

The model underwent training for 50 epochs on the training dataset, followed by 25 epochs of fine-tuning on the fine-tuning dataset. During fine-tuning, 1,000 SMILES strings were generated and evaluated for their validity. The validated ligands were then subjected to docking onto 3CL^{pro}. The ligands that exhibited a binding affinity greater than that of the native ligands were incorporated back into the fine-tuning dataset. The fine-tuning dataset underwent preprocessing and shuffling, with this entire fine-tuning process being repeated for three cycles. This process of adding validated ligands with high-binding affinity back into the dataset and repeating the fine-tuning process is a way to iteratively improve the performance of the model and helps it identify the characteristics of ligands that contribute to their strong binding affinity with 3CL^{pro}. A schematic overview of the model training and optimization process is provided in Figure 1.

2.5. Virtual screening

Virtual screening involves docking ligand libraries to a target macromolecule to discover a lead that would confer a biological function. The virtual screening was done using AutoDock Vina in PyRx (Dallakyan and Olson, 2014).

The generated ligands were converted from SMILES to SDF using openBabel-GUI (O'Boyle et al., 2011). To obtain the lowest free energy of the ligand, the Merck molecular force field (mmff94) parameter was used in PyRx. Finally, the ligands were converted to PDBQT format, preparing the ligand for molecular docking (Morris et al., 1998; Huey et al., 2007).

The 3D crystalline structure of SARS-CoV-2 main protease or 3CL^{pro} (PDB: 6LU7) was obtained from PDB⁸; this served as the target for docking. The target was prepared by removing the native ligand present (N3 Inhibitor) and water molecules using Biovia Discovery Studio (Biovia, 2017).

The native ligand was docked onto the target molecule, and the binding affinity was found to be -7.9 kcal/mol. The amino acid residues involved in binding with the native ligand were obtained using the 2D structure in Biovia Discovery Studio. The amino acid residues are Thr24, Thr26, Phe140, Asn142, Gly143, Cys145, His163, His164, Glu166, His172. The grid box was then positioned over the binding site (center: $x = -10.606$, $y = 17.214$, $z = 64.716$, total size: $x = 24.074$ Å, $y = 24.134$ Å, $z = 19.174$ Å).

Further post-docking analysis and visualization of the ligand-target complex were carried out in Biovia Discovery Studio (Biovia, 2017).

2.6. Evaluation of molecular properties

To evaluate the molecular properties of the ligands, an online tool, Molinspiration⁹ was used by uploading the ligands in SMILES format. Molinspiration also provides bioactivity scores for drug targets such as Ion channel modulators, GPCR (G protein-coupled receptor) ligands, kinase inhibitors, nuclear receptor inhibitors, protease inhibitors, and enzyme inhibitors. The bioactivity score of Molinspiration is calculated by a machine learning-based model that predicts the probability of a molecule being active against a particular target. The model is trained on a large database of known active and inactive compounds and uses various molecular descriptors, such as physicochemical properties and substructure information, to make predictions. These bioactivity scores provide an additional metric for evaluating the drug-like properties of the ligands (Vardhan and Sahoo, 2020). Another online tool Molsoft¹⁰ was used to evaluate the drug-likeness score of the ligands (Prabhavathi et al., 2020).

2.7. Evaluation of ADMET profile

The absorption, distribution, metabolism, elimination, and toxicity (ADMET) are some of the important pharmacokinetic properties that must be evaluated. An online tool called admetSAR¹¹ was used to obtain the ADMET profile of the ligands (Yang et al., 2018). Some of the properties calculated include Ames mutagenesis, blood-brain barrier penetration, BSEP inhibition, Caco-2, Carcinogenicity, cytochrome p450 substrate and inhibitors, glucocorticoid receptor binding, hepatotoxicity, human ether-a-go-go (hERG) inhibition, p-glycoprotein inhibitors and substrate, human intestinal absorption, and human oral bioavailability.

2.8. Curating the reference dataset

To evaluate the capability of the generated ligands as potential anti-COVID drugs, a reference dataset comprising 20 drugs currently in clinical trials for COVID-19 treatment was obtained from PubChem. This included drugs such as Remdesivir, Ritonavir, Galidesivir, etc. (For full list of drugs—Appendix A). The properties of the generated ligands were compared to those of the reference drugs to assess their potential as anti-COVID agents. This comprehensive comparison of the ligands and clinical trial drugs facilitates the assessment of the ligand's properties.

2.9. Molecular dynamics simulations

Molecular Dynamics simulation is a sophisticated computational tool for predicting and analyzing the dynamic behavior of molecules (Verdonk et al., 2003; Radinnurafiqah et al., 2016; Girdhar et al., 2019; Choubey et al., 2022; Mishra et al., 2022). The stabilities of six selected protein-ligand complexes were assessed using GROMACS 2021

⁸ <https://www.rcsb.org/pdb/>

⁹ <http://www.molinspiration.com>

¹⁰ <http://molsoft.com/mprop>

¹¹ <http://lmmd.ecust.edu.cn/admetSar2/>

package through Molecular Dynamics (MD) simulations (Van Der Spoel et al., 2005). The complexes included SARS_COV2_MOL_1, SARS_COV2_MOL_3, SARS_COV2_MOL_9, SARS_COV2_MOL_10, SARS_COV2_MOL_17, and SARS_COV2_MOL_20. The ligand topology parameter for CHARMM forcefield (Vanommeslaeghe et al., 2010) was created using the CGenFF server.¹² A cubic box of TIP3P water models was used to solvate all the complexes. To maintain the periodic boundary conditions, the distance between the protein and the box edge was kept at 1 nm. The systems were neutralized by adding 0.15 M NaCl. Energy minimization was performed using the steepest descent method followed by the conjugate gradient method with maximum number of minimization 50,000 per algorithm. The Particle Mesh Ewald (PME) method was employed to calculate long-range interactions (Abraham and Gready, 2011). The first phase of equilibration was carried out with an NVT ensemble, where the temperature was equilibrated using 50,000 iterations of 2 fs each. In the second phase, the pressure was equilibrated at 300 K with an NPT ensemble using Parrinello-Rahman, a pressure coupling method. The temperature inside the system was regulated using V-rescale, a modified Berendsen thermostat. Finally, a production run of 100 ns was established to gain insights into the dynamic behavior of the complex.

2.10. Trajectory analysis

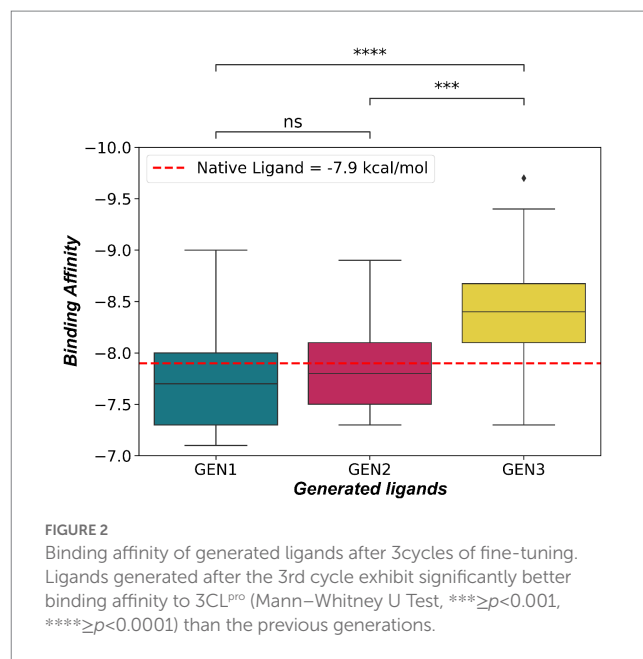
The obtained trajectories after the MD simulations were analyzed for calculations such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), and inter-molecular hydrogen bond using the in-built tools of the GROMACS package. To compute the RMSD in the protein backbone, the *rms* module of GROMACS was employed. RMSD of the ligands were also calculated using the same module, whereas the *rmsf* module was used to determine the RMSF in the atomic positions of the protein C α backbone. In addition, modules like h-bond, gyrate, and SASA were used to calculate the number of hydrogen bonds, Rg, and SASA, respectively.

3. Results

A deep learning model called LSTM_Chem was trained, using transfer learning, to produce novel ligands that could inhibit 3CL^{pro}, the main viral protease of SAR-CoV-2. The ligands' ability to inhibit the protease is evaluated through molecular docking, ADMET analysis, and molecular dynamics simulation.

3.1. Selection of generated ligands

After the first stage of training on the training dataset, a final loss of 0.427 on the training set and 0.567 on the validation set (20% of data from the training dataset) was obtained. Additionally, the model had an accuracy of 82% in generating valid ligands, i.e.,



out of every 100 ligands the model produces, 82 are valid molecules.

The model then underwent three cycles of fine-tuning on the fine-tuning dataset, and after each cycle, the binding affinity of 30 randomly selected ligands was evaluated. Figure 2 depicts the distribution of the binding affinities across the three cycles, indicating that the 3rd generation of molecules had a significantly better binding affinity to 3CL^{pro} than the previous two generations (Mann Whitney U Test, $p < 0.001$). The average binding affinity for the 3rd generation was -8.406 ± 0.087 kcal/mol, and 26 (86.67%) of the ligands had a value higher than the binding affinity of the native Ligand (-7.9 kcal/mol).

After evaluating the binding affinity of 30 ligands generated by the model, the top 20 ligands were chosen for additional investigations. To confirm the novelty of these molecules, a search was conducted in the PubChem database, which did not yield any results for these ligands. Therefore, it can be inferred that the generated molecules are novel. The molecules were named SARS_COV2_MOL_1 - SARS_COV2_MOL_20 as an identifier.

3.2. Comparative analysis

Various molecular and ADMET properties of the generated ligands and drugs in the reference dataset were calculated and contrasted to assess the efficacy of the generated ligand to serve as a potential inhibitor to the 3CL^{pro} protease (Tables 1, 2).

3.2.1. Assessment based on Lipinski's rule

Lipinski's rule of 5 provides a set of criteria to estimate the solubility and permeability of a ligand. This has become a crucial criterion for assessing the oral bioavailability of any drug during the drug development process. The criterion for oral activity is based on the molecular properties of drugs such as molecular weight ($MW \leq 500$), partition coefficient ($\log P \leq 5$), hydrogen bond donors ($HBD \leq 5$), hydrogen bond acceptors ($HBA \leq 10$), and the number of

¹² <https://cgenff.umaryland.edu>

TABLE 1 Comparison of Lipinski's parameters between generated and reference ligands.

Property	% compliance with Ro5 (generated ligands)	Mean \pm SD	% compliance with Ro5 (reference drugs)	Mean \pm SD	% difference
MW \leq 500	65	527.18 \pm 29.55	70	434.44 \pm 48.13	–5
HBD \leq 5	95	3.0 \pm 0.40	90	3.45 \pm 0.42	5
HBA \leq 10	85	8.55 \pm 0.59	80	8.25 \pm 0.81	5
NRB \leq 10	60	10.8 \pm 0.81	70	7.55 \pm 1.13	–10
LogP \leq 5	80	3.85 \pm 0.41	85	2.61 \pm 0.73	–5
TPSA \leq 140	75	825.51 \pm 709.55	60	120.01 \pm 10.78	15

TABLE 2 Comparison of bioactivity scores between reference and generated ligands.

	Mean reference ligands	Reference ligands score \geq 0.5 (%)	Mean generated ligands	Generated ligands score \geq 0.5 (%)
Enzyme inhibitor	–0.15 \pm 0.26	80	–0.35 \pm 0.18	80
Ion channel modulator	–0.52 \pm 0.25	75	–0.62 \pm 0.23	65
Kinase inhibitor	–0.25 \pm 0.21	50	–0.58 \pm 0.20	65
GPCR ligand	–0.22 \pm 0.25	80	–0.12 \pm 0.15	85
Nuclear receptor inhibitor	–0.90 \pm 0.26	45	–0.61 \pm 0.21	65
Protease inhibitor	–0.21 \pm 0.21	80	0.05 \pm 0.11	90

rotatable bonds (NRB \leq 10) (Lipinski et al., 1997). Table 1 and Figure 3 show the various molecular properties of generated and reference ligands plotted to assess their compliance with Lipinski's Ro5.

According to Lipinski's Rule, ligands having less than or equal to 1 violation of Lipinski's criteria can be considered to have oral bioactivity. 16 (80%) of generated ligands exhibit 0 or 1 violation, and all ligands show less than or equal to 2 violations. 19 (95%) generated ligands have less than 5 H Donors, and 17 (85%) have less than 10 H Acceptors. Octanol-water partition coefficient or logP is used as a measure of molecular lipophilicity. Lipophilicity affects drug absorption, bioavailability, hydrophobic drug-receptor interactions, metabolism of molecules, as well as their toxicity. It is one of the key parameters that determine the drug-likeness of compounds (Amézqueta et al., 2020). 16 (80%) compounds among the generated ligands exhibit a LogP $<$ 5.00.

For the oral bioavailability of compounds, the molecular weight of the compound should be \leq 500 Da. 13 (65%) generated ligands and 14 (70%) ligands in the reference dataset were found to have a molecular weight less than 500. The average molecular weight among the generated ligands was found to be 527.176 \pm 29.552 Da. Refer to Table 1 for additional information.

3.2.2. Assessment based on bioactivity score

Molinspiration was used to obtain the bioactivity scores of the generated ligands and the reference drugs. Bioactivity here refers to a quantitative estimate of the compound's potency and efficacy in inhibiting or activating various targets. Compounds with a bioactivity score of more than 0 are considered biologically active, while values between –0.50 and 0.00 are considered moderately active, and less than –0.50 are inactive (Khan et al., 2017). Figure 4 represents the distribution of the bioactivity score

for the generated and reference ligands. It can be referred from Table 2, that the generated ligands and reference drugs have comparable bioactivity scores. 14 (70%) generated ligands show a bioactivity score greater than 0 as a protease inhibitor. This suggests that the ligands share structural characteristics with other protease inhibitors, indicating a high likelihood of their potential as protease inhibitors.

3.2.3. Assessment based on docking

The generated ligands exhibit a strong binding affinity toward the target protease 3CL^{pro} (6LU7), as evidenced by molecular docking results presented in Figure 5 and Appendix A. These results indicate that the binding affinities of the generated ligands are higher than that of the N3 inhibitor in the 6LU7 structure of 3CL^{pro}, which is –7.9 kcal/mol. On average, the generated ligands display a binding affinity of –8.515 \pm 0.091 kcal/mol toward the target protease 3CL^{pro}. The highest binding affinity among the generated ligand was –9.4 kcal/mol, and the lowest was –7.5 kcal/mol.

Among the reference drugs, Nafamostat had the highest binding affinity at –8.6 kcal/mol, while Fingolimod and Favipiravir had the lowest at –5.0 kcal/mol. The average binding affinity of the reference drugs was –6.81 \pm 0.238, which is significantly lower than the binding affinity of the native ligand.

The generated ligands were further evaluated for their binding energy with two additional SARS-CoV-2 3CL^{pro} structures, 7EN8, and 7JKV in the PDB database. The results showed that the binding affinities of the ligands to these structures were significantly better than the reference drugs (Mann Whitney U Test, $p <$ 0.001) (Figure 6A). The average binding affinity of the generated ligands for all three structures was –8.928 \pm 0.091 kcal/mol (Figure 6B).

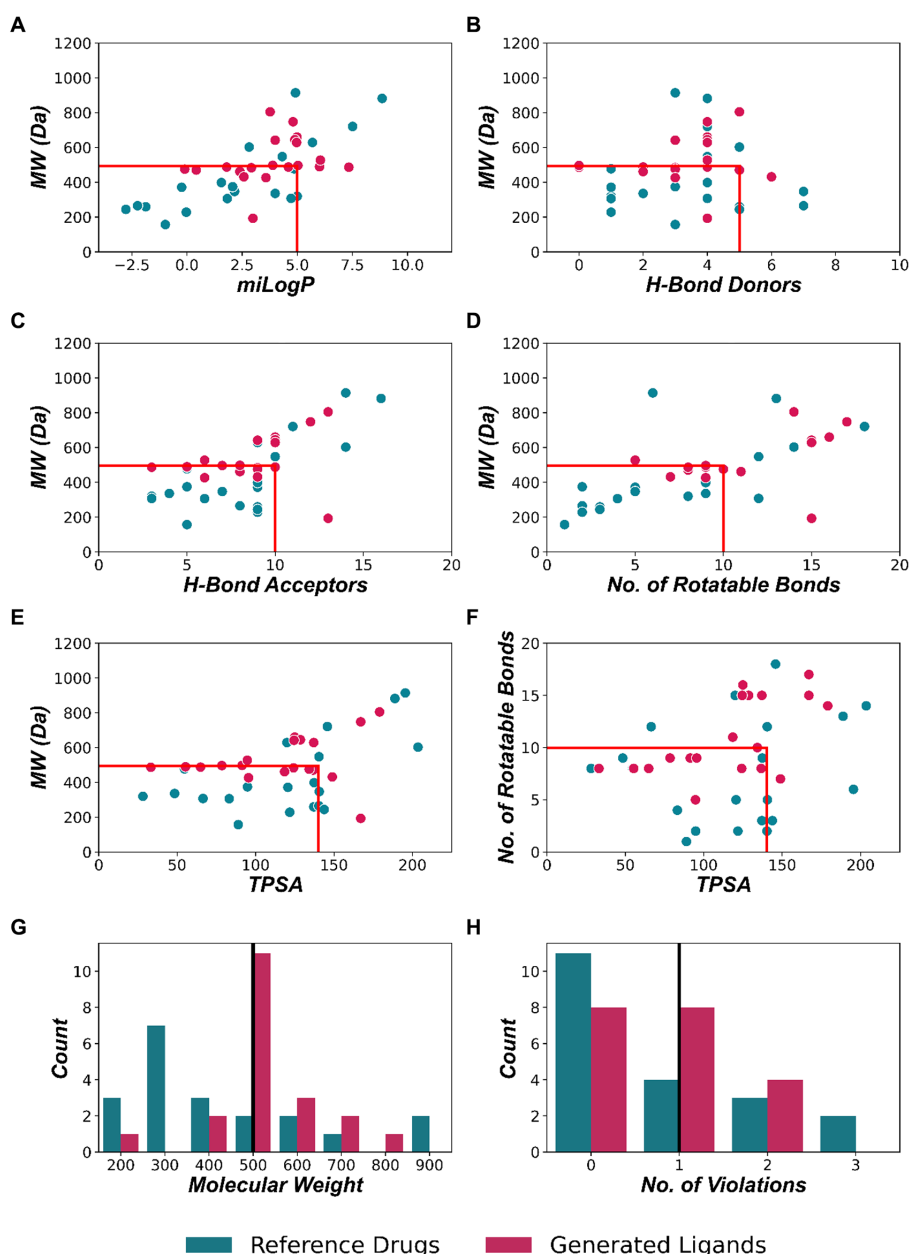


FIGURE 3

Comparative evaluation of structural properties between reference drugs and generated ligands based on Lipinski's rule of 5. (A) Lipophilicity (miLogP), (B) number of H-Bond donors, (C) number of H-bond acceptors, (D) number of rotatable bonds, (E) total polar surface area is compared to the molecular weight (MW). (F) The total polar surface area (TPSA) vs. number of rotatable bonds. (A–F) The red box indicates the ligands that comply with Lipinski's rule. (G) Distribution of molecular weight and the black line depicts the Lipinski's criteria for MW < 500 Da. (H) Distribution of the number of violation to Lipinski's rule. Having 1 or less violation of Lipinski's criteria imply molecules with drug-like properties (depicted by the black line).

3.2.4. Assessment based on ADMET properties

To evaluate the pharmacokinetic properties of the reference and generated ligands admetSAR was used. Figure 7 provides an overview of the various ADMET properties assessed. The generated ligands exhibit a good degree of human intestinal absorption (95%) compared to the reference ligands (66.67%). Although the generated ligands show low oral bioavailability (15%) and none permeate through Caco-2 monolayer, from Lipinski's Ro5 and TPSA predicted earlier (Figure 3), it can

be inferred that the ligands will be absorbed effectively after administration.

P-glycoprotein (P-gp) is an efflux transporter found in various organs and it plays a vital role in the distribution of drugs. 16 (80%) of the generated ligands of the present study were found to be acting as substrates for P-gp. Cytochrome P450 is known to be one of the most important drug-metabolizing families of enzymes. Out of the 57 different CYP genes in the human body, it is established that only about a dozen gene products mediate most of the

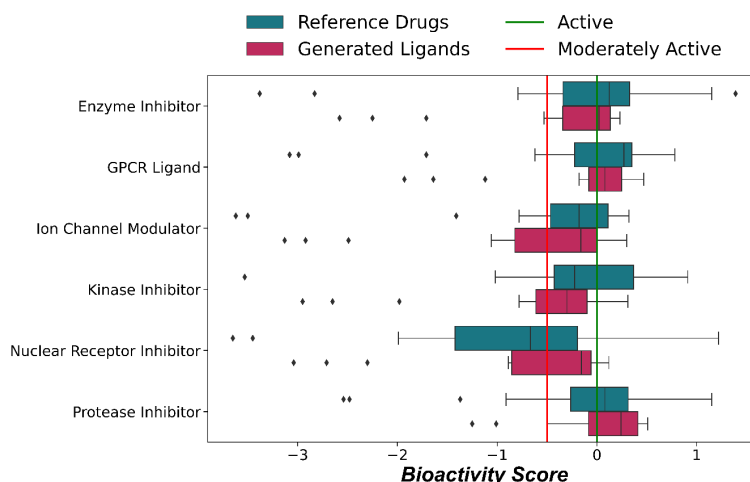


FIGURE 4

Distribution of bioactivity scores from Molinspiration between generated Ligands and reference drugs. The green line represents a bioactivity score of 0; ligands above this line are considered to be active, the red line represents a bioactivity score of -0.5 , below which ligands are considered to be inactive; in the region between the red and green line, ligands are considered to be moderately active.

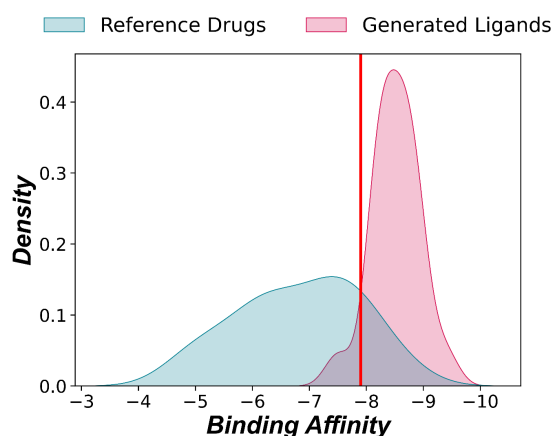


FIGURE 5

Distribution of binding affinities of generated and reference ligands to 3CL^{pro} (Structure: 6LU7). The red line represents the binding affinity of the native ligand (-7.9 kcal/mol).

biotransformation reactions against foreign substances. 100% of the generated ligands were substrates for CYP3A4, which is responsible for the metabolism of nearly 50% of all drugs in clinical use (Zanger and Schwab, 2013), and 35 and 10% of the ligands were also CYP2C9 and CYP2D6 substrates, respectively. One of the major drug excretion routes is the renal organic cation transporters (OCT) (Yin and Wang, 2016), and inhibitors of OCT are known to cause renal toxicity leading to excess drug accumulation. Only 10% of the generated ligands inhibit OCT1, and 15% inhibit OCT2.

Nearly 85% of ligands show some degree of hepatotoxicity; this could be due to the inhibition of the bile salt export pump (BSEP), as all the generated ligands inhibit BSEP (Kenna et al., 2018). However, most (>90%) of the generated ligands are non-carcinogenic and non-mutagenic for Ames mutagenesis.

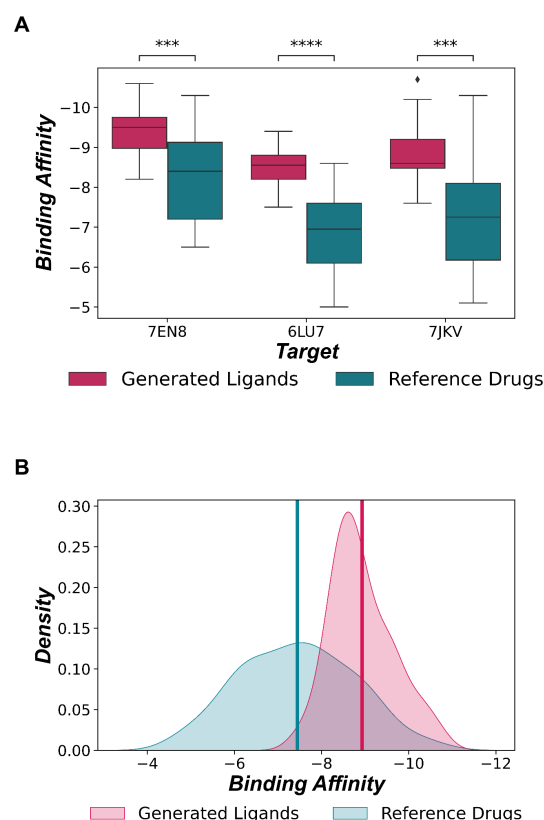


FIGURE 6

(A) Evaluation of binding affinities for three 3CL^{pro} PDB structures. The generated ligands show significantly better binding affinity than the reference drugs (Mann–Whitney U Test, $*** \geq p < 0.001$, $**** \geq p < 0.0001$). (B) Distribution of binding affinity across all three 3CL^{pro}. The mean binding affinity of generated ligands (-8.928 ± 0.091 kcal/mol) and reference drugs (-7.441 ± 0.164 kcal/mol) are depicted by red and green lines, respectively.

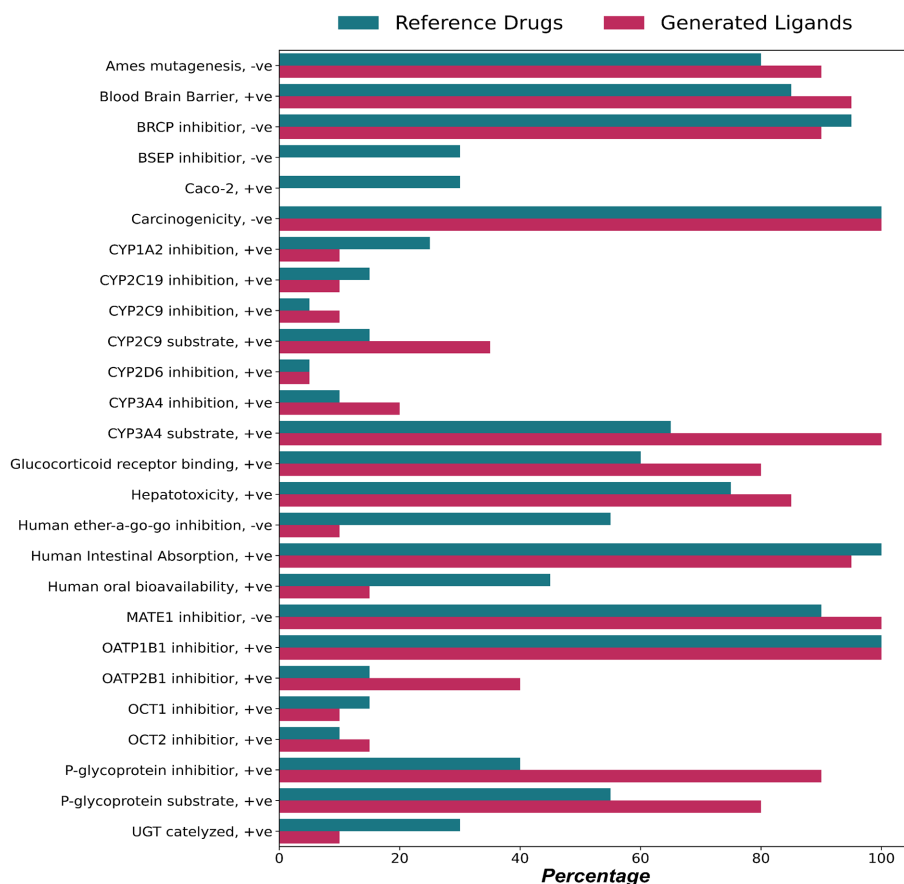


FIGURE 7

Absorption, distribution, metabolism, excretion, toxicity (ADMET) properties of generated and reference ligands.

3.3. Drug likeness score prediction

The drug-likeness score of the generated ligands predicted by MolSoft are shown in Figure 8A. The average drug-likeness score was found to be 0.663 ± 0.118 , and the scores of the generated ligands were found to be in the range with FDA reference drugs used by MolSoft. Drugs that have drug-likeness scores greater than 0 are considered to have drug-like properties (Prabhavathi et al., 2020). In Figure 8B, ligands that fall in the green region have a binding affinity greater than -7.9 kcal/mol and have a drug-likeness score greater than 0.16 (80%) ligands fall under the green region, indicating that these ligands have a good potential to be developed as lead molecules.

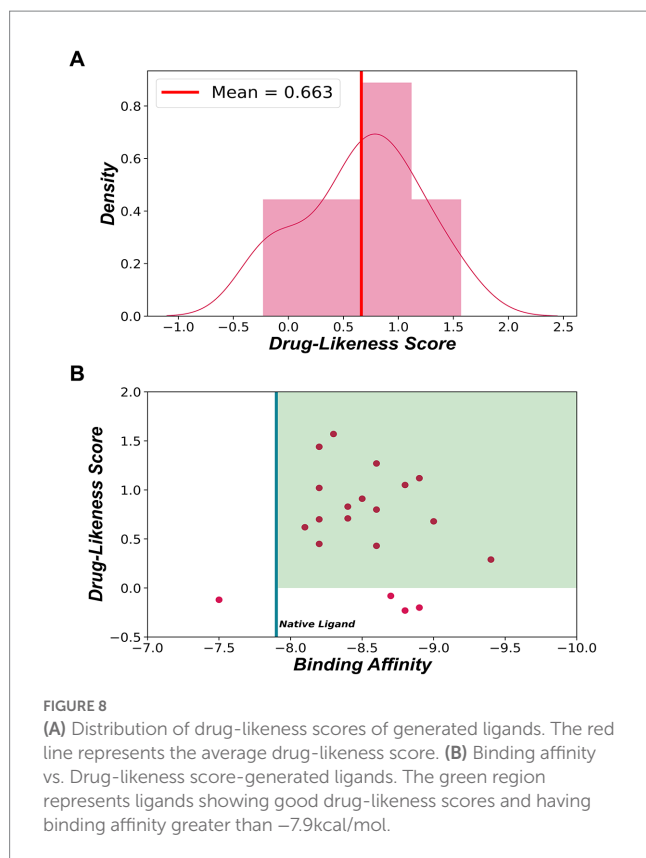
3.4. Post docking analysis

From the 20 generated ligands, 6 were selected to study their interactions with 3CL^{pro} using Discovery Studio. Figures 9A–F, displays the interaction of the six ligands with the protease, and the amino acid residues interacting with the ligand have been annotated. Hydrogen bonds are an essential factor that determines the stability of the docked complex. Most of the ligands can be seen interacting with receptor residues Leu 141, Asn 142, Gly 143, and Ser 144 via a hydrogen bond. SARS_COV2_MOL_17 (Figure 9E) shows the highest

binding affinity of -9.4 kcal/mol and a drug-likeness score of 0.29. The ligand showed hydrogen bond interaction with receptor residues at Leu 141, Gly 143, and Ser 144. However, there is an unfavorable donor-donor interaction at Cys 145. Unfavorable bonds greatly hinder the stability of the protein-ligand complex. SARS_COV_MOL_20 (Figure 9F) has the highest drug-likeness score of 1.57 and a binding affinity of -8.3 kcal/mol. This ligand can be seen interacting with hydrogen bonds at receptor residues Thr 26, Glu 166, and Gln 189. However, among the six ligands, SARS_COV2_MOL_1, 3, 9, and 10 (Figures 9A–D respectively) have shown high binding affinity with a good drug-likeness score and less than 1 violation of Lipinski's Rule. SARS_COV2_MOL_9 has shown a high drug-likeness score of 1.27 and is also a good protease inhibitor (Protease inhibitor bioactivity score - 0.26 (refer Appendix-A)). It also exhibits a high number of hydrogen bond interactions at the receptor (Thr 24, Thr 26, Gly 143, Glu 166, and Gln 189) and interacts with a pi-pi stacked bond at His 41 and pi-alkyl bond at Pro 168, and Met 49. In conclusion, the generated ligands are seen to be forming stable complexes with 3CL^{pro} protease having high binding affinities.

3.5. MD simulations analysis

To further validate the structural stability of the generated ligands, MD simulations was performed on the docked complexes for 100 ns.

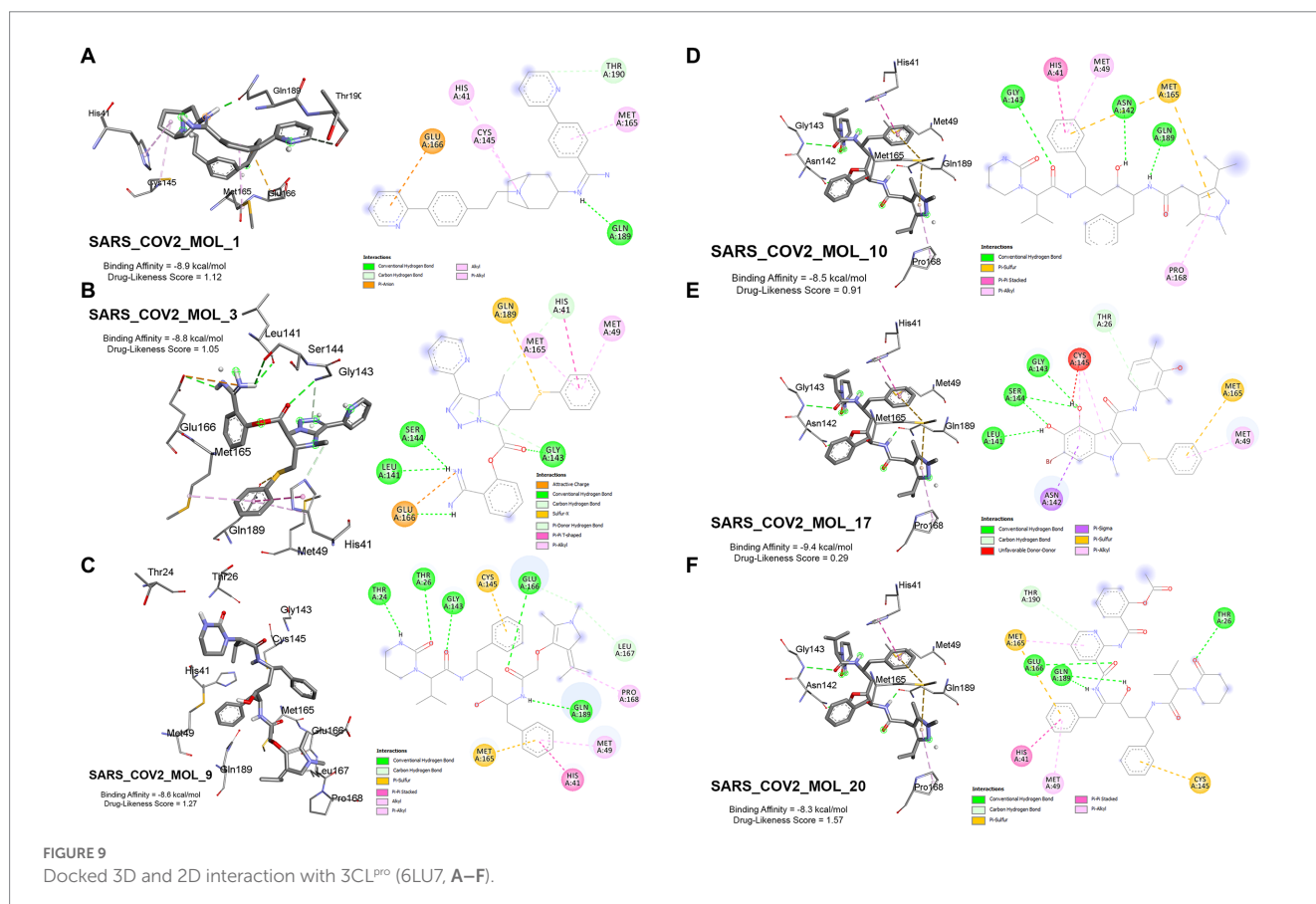


The trajectories obtained after the simulations were analyzed to calculate RMSD, RMSF, hydrogen bonds, Rg, and SASA to assess the stability of the simulated systems.

3.6. Root-mean-square deviation

RMSD is commonly used to evaluate docked complex stability (Martínez, 2015; Sargsyan et al., 2017). It measures the difference between the initial position and the final conformation of the protein backbone. From Figure 10, the SARS_COV2_MOL_3 (red) complex showed the lowest average RMSD value, around 0.18 nm, among all six complexes. The average RMSD values of the other five complexes with SARS_COV2_MOL_1 (black), SARS_COV2_MOL_9 (green), SARS_COV2_MOL_10 (blue), SARS_COV2_MOL_17 (yellow), and SARS_COV2_MOL_20 (brown) was estimated to be ~0.2 nm same as that of the apoprotein. The RMSD values remained nearly constant over the 100 ns period, indicating that the protein-ligand complexes were structurally stable (Figure 9A). The low RMSD values suggest that the protein-ligand interactions were energetically favorable and contributed to the stability of the complexes. The SARS_COV2_MOL_20 and SARS_COV2_MOL_1 complexes showed slight deviations of ~0.4 and ~0.38 nm around 50 and 66 ns, respectively. These fluctuations were further confirmed by observing local changes at the residue level using the RMSF plot.

The ligand RMSD ranged between 0.16–0.39 nm as shown in Figure 9B. SARS_COV2_MOL_17 had the least average RMSD which



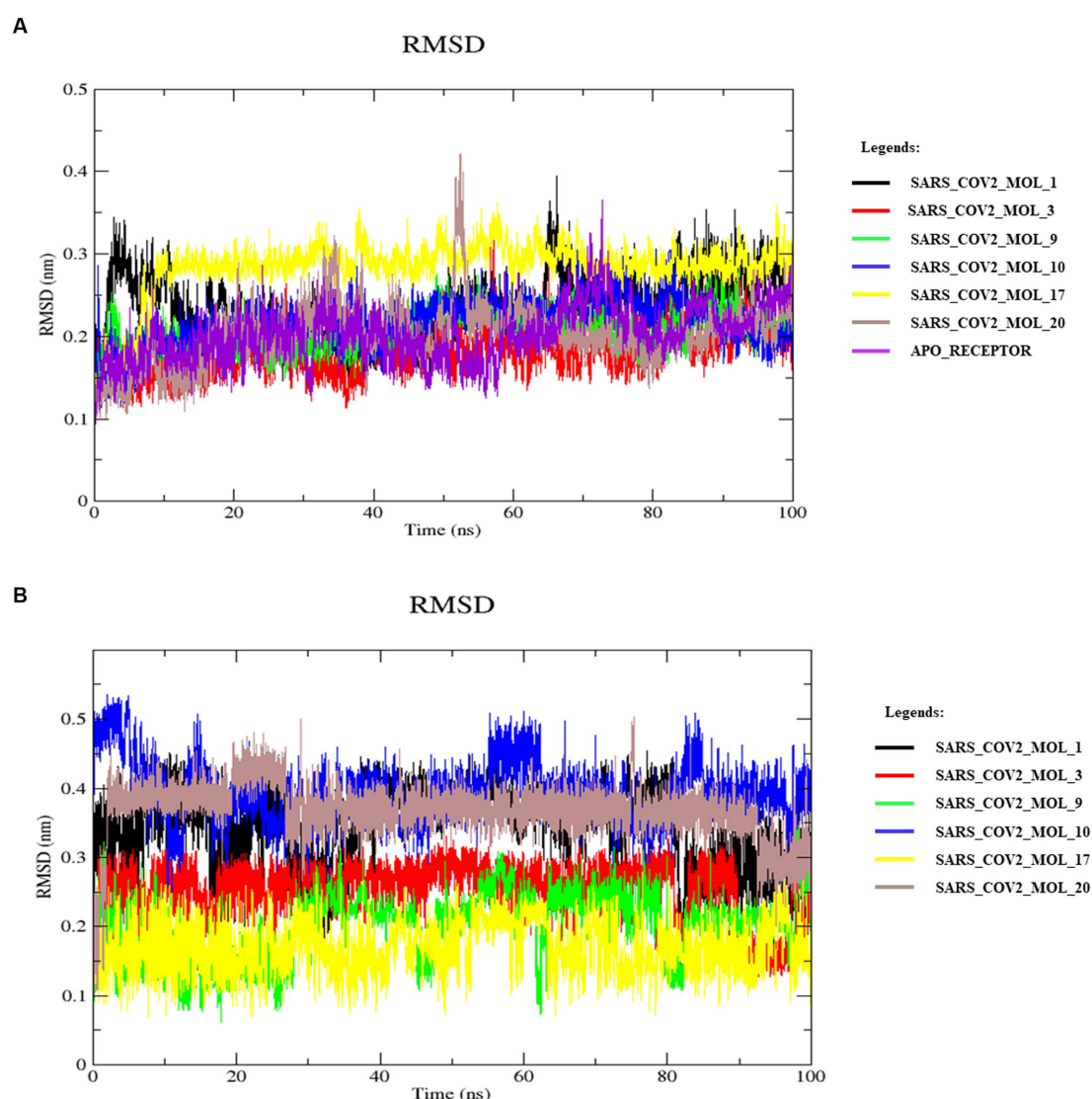


FIGURE 10
RMSDs of the receptor (A) backbone atoms and ligands (B) during MD simulation.

suggest the stability of the protein-ligand system when bound with SARS_COV2_MOL_17. Rests of the ligands also showed no sharp deviation and were stable throughout the period of simulation. It indicates the stability of the complexes. Alignment of the post-MD complexes with their respective initial-docking poses corroborate the low deviations observed, as illustrated in Figure 11.

3.7. Root-mean-square fluctuation

The fluctuations in the protein can be determined by calculating RMSF, which measures the flexibility of each residue over time. The stability of the protein-ligand complexes can be inferred from the RMSF scores, with higher values indicating less stability and more flexibility. The RMSF of C α atoms was calculated for all complexes, and the resulting average values for the six ligands were between 0.103 and 0.150 nm, as shown in Figure 12. The RMSF of the apoprotein was around 0.134 nm with no major fluctuation with respect to the

protein-ligand complexes. However, the residues in the range of 46–50 showed slight fluctuation of average 0.37 nm. These low RMSF values suggest that the protein-ligand complexes are relatively stable and exhibit a moderate degree of flexibility. The close proximity of the complexes with the apoprotein also indicates about the stability of the complexes. This indicates that the ligands can bind to the protein without causing significant changes in its conformation. So, the predicted system appears to be stable (Farmer et al., 2017).

3.8. Radius of gyration and solvent accessible surface area analysis

The compactness and stability of protein structures can be measured using the Rg, which represents the mass-weighted root mean square distance of the atomic distribution from their mutual center of mass. The Rg values depict the inclusive dimensions of the protein and protein-ligand complexes and reflect their appropriate

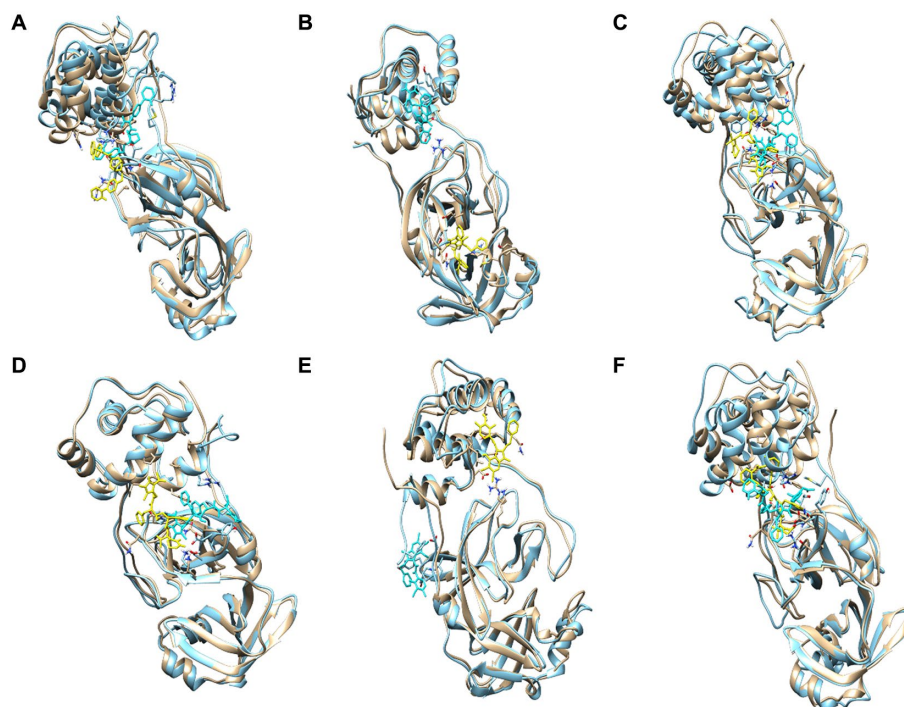


FIGURE 11

Superimposition of the post-MD complexes of (A) SARS_COV2_MOL_1, (B) SARS_COV2_MOL_3, (C) SARS_COV2_MOL_9, (D) SARS_COV2_MOL_10, (E) SARS_COV2_MOL_17, and (F) SARS_COV2_MOL_20, with initial docking pose of the respective complexes.

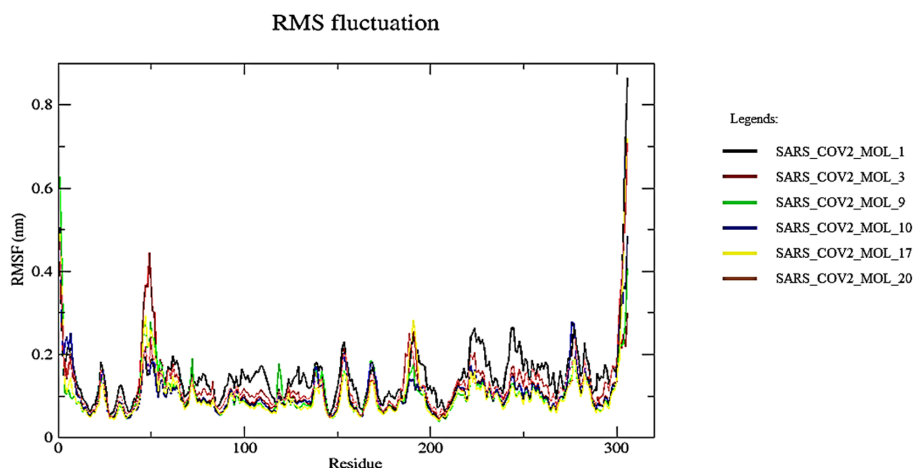


FIGURE 12

RMSF analysis of C α during MD simulation.

interactions. The protein-ligand complexes that displayed the least radius of gyration are considered to be more compact and stable. The Rg values of all the complexes were analyzed, and the solvent-accessible surface area (SASA) was also computed for all the proteins for 100 ns. The SASA is an important measure to determine the area of the receptor exposed to the solvents during the simulation. As shown in Figure 13A, all the systems exhibited similar Rg values, ranging from 2.2 to 2.6 nm throughout the simulation, indicating their stability. The estimated SASA values also displayed a similar pattern, varying between 150.59 and 153.93 nm², as depicted in Figure 13B, with the highest value observed for the SARS_COV2_MOL_3

complex. These observations confirm the stability and compactness of all the protein-ligand complexes, as smaller deviations in average Rg and SASA values (Figure 13) Suggesting a stronger binding between the protein and ligand (Shaji, 2016).

3.9. Hydrogen bond analysis

The stability and molecular recognition process of a protein-ligand complex is affected by the intermolecular hydrogen bonds (H-bonds) between interacting atom pairs. The number of H-bonds

formed between the receptor protein and selected ligands was determined during the 100 ns MD simulations to ascertain the dynamic stability of each complex. The binding strength and specificity of the protein-ligand complex are determined by hydrogen bonds. Figure 14, represents the number of hydrogen bonds formed between the receptor protein and selected ligands throughout the MD simulation. The complex formed with SARS_COV2_MOL_3 (red) showed a higher number of hydrogen bonds, while the rest of the complexes showed a stable number of hydrogen bonds throughout the 100 ns simulation (Figure 12). The results further suggest the stability of the studied 3CL^{pro} inhibitors (Pereira et al., 2019; Zhu et al., 2022).

4. Discussion

Here, the properties of ligands produced by a deep neural network (RNN-LSTM) to inhibit 3CL^{pro} were assessed and compared to drugs currently undergoing clinical trials as a potential treatment for COVID-19. The findings demonstrate the RNN-LSTM's ability to produce new ligands with strong binding affinities to 3CL^{pro}, outperforming the native ligand and reference drugs in most cases. The selection of ligands was primarily based on their binding affinity to the target protease, which is a crucial parameter here. Similar methods for screening and selecting ligands were employed by Meng et al. (2011) and Hernández-Santoyo et al. (2013). *In-silico* docking studies with drugs such as Ritonavir,

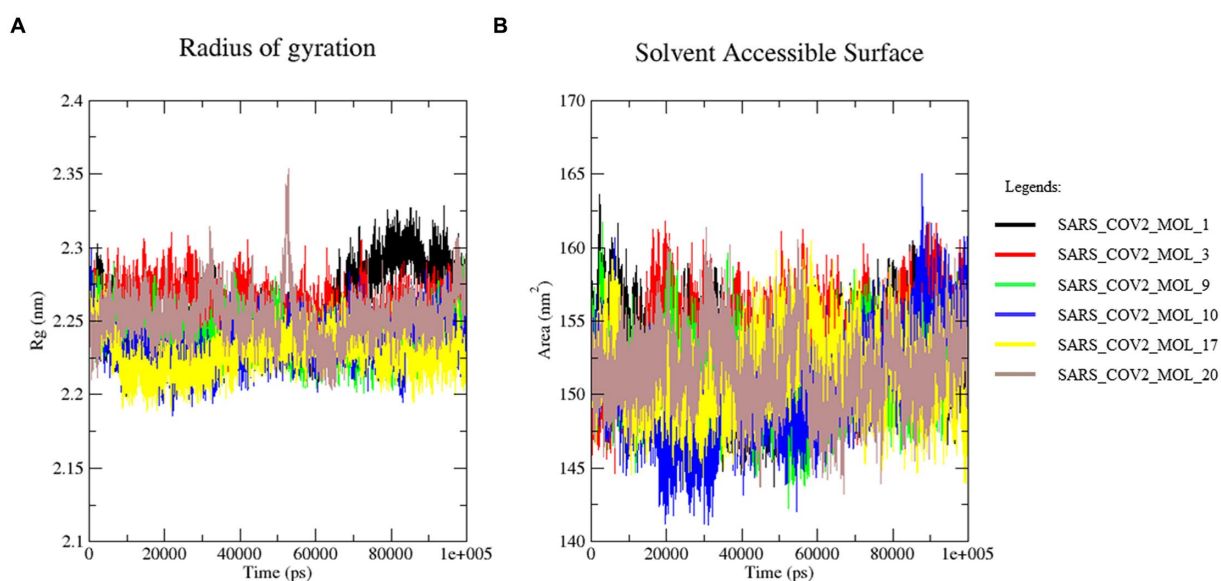


FIGURE 13
(A) Radius of gyration and (B) solvent accessible surface area.

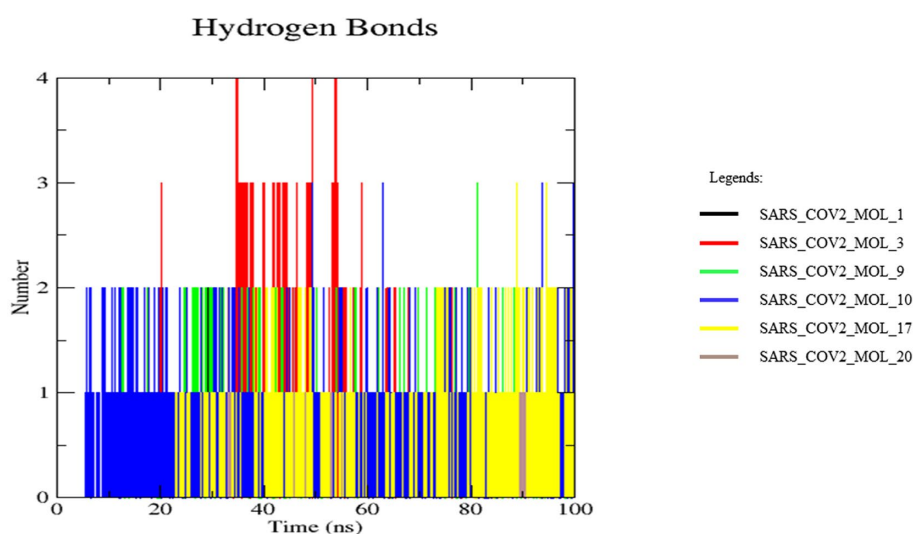


FIGURE 14
Hydrogen bond analysis of the docked complexes.

Lopinavir, Remdesivir, and Benzophenone derivatives have suggested the potential of these drugs to inhibit viral proteases (Bhardwaj et al., 2020; Li et al., 2020; Li and Kang, 2020; Rujuta et al., 2020). Polymerase inhibitors such as Sofosbuvir, Remdesivir, Tenofovir, Ribavirin, and Galidesivir have also been identified as promising inhibitors based on their binding energies (Ju et al., 2020a,b). The binding energy of inhibitors, including nucleotide analogs, is a crucial parameter for assessing their potential to inhibit viral replication (Udofia et al., 2021).

Of the 20 ligands chosen, most exhibit physicochemical properties that satisfy Lipinski's Rule of 5, but some deviate from it in terms of molecular weight. Nonetheless, a molecular weight of 500 Da alone is not a reliable predictor of oral bioavailability and drug-likeness. Studies have shown that compounds having ten or fewer rotatable bonds and a total polar surface area (TPSA) of less than 140 have a higher probability of oral bioavailability and drug-likeness (Veber et al., 2002). And 10 of the generated ligands satisfy this criterion (Figure 3F). The drug-likeness potential was further reinforced by the results from Molinspiration and Molsoft. ADMET analysis revealed that the ligands have an acceptable ADME profile and showed very low toxicity. Leeson et al. (2021) discussed the ability of physicochemical descriptors commonly used to define "drug-likeness" and ligand efficiency measures to differentiate marketed drugs from compounds reported to bind to their efficacious target or targets. The study found that recent drugs approved in 2010–2020 had no overall differences in molecular weight, lipophilicity, hydrogen bonding, or polar surface area from the marketed compounds. However, drugs differed by higher potency, ligand efficiency (LE), lipophilic ligand efficiency (LLE), and lower carboaromaticity (Leeson et al., 2021). The ligands generated by the model can be further improved by repeating the fine-tuning process using a dataset of generated ligands with these desirable properties.

Molecular dynamic simulation studies at 100 ns revealed that the generated ligands formed stable complexes with 3CL^{pro}. The root-mean-square deviation was 0.18 nm and ~0.2 nm, with a root-mean-square fluctuation ranging from 0.103 to 0.150 nm, a solvent-accessible surface area between 150.59 and 153.93 nm², a radius of gyration ranging from 2.2 to 2.6 nm, and a stable number of hydrogen bonds. The results suggest that these ligands form strong and stable complexes with 3CL^{pro}. As docking and simulation studies using *in silico* tools are well accepted for various purposes (Paital et al., 2015; Shaji, 2016; Farmer et al., 2017; Mishra et al., 2019; Pereira et al., 2019; Bulut et al., 2020; Sahoo et al., 2022; Zhu et al., 2022; Hou et al., 2023), the present study may be useful for targeting 3CL^{pro}.

Overall, the generated ligands demonstrated comparable or even superior drug-like properties when compared to the drugs currently undergoing clinical trials, making them a promising candidate for further development in the treatment of COVID-19.

5. Conclusion

In the current study, a deep RNN was trained to produce novel ligands that could potentially inhibit the main viral protease of SARS-CoV-2, 3CL^{pro} (PDB: 6LU7), and 20 novel ligands were identified. The study's results unequivocally indicate that the novel ligands produced by the deep generative model have the potential to serve as effective anti-COVID drugs. Furthermore, the study adds to the growing body of evidence supporting the use of deep learning as a means of expediting drug discovery. However, further testing by *in vitro* and *in vivo* studies is necessary before considering them for human use.

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 authors.

Author contributions

PP, BP, SVM, and DS: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, roles/writing—original draft, and writing—review and editing. AH, SM, MS, SK, and PD: concept, writing—original draft, and writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Schemes (number ECR/2016/001984 from Science Engineering Research Board, DST, Govt. of India and 1188/ST, Bhubaneswar, dated 01.03.17, ST- (Bio)-02/2017 from Department of Biotechnology, DST, Govt. of Odisha, India) to BP are acknowledged.

Acknowledgments

Encouragements and laboratory facilities provided by the honorable Director, College of Basic Science and Humanities, Bhubaneswar, and the honorable Vice Chancellor, Odisha University of Agriculture and Technology Bhubaneswar, and the Central Instrumentation facility are duly acknowledged.

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.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1194794/full#supplementary-material>

References

- Abraham, M. J., and Gready, J. E. (2011). Optimization of parameters for molecular dynamics simulation using smooth particle-mesh Ewald in GROMACS 4.5. *J. Comput. Chem.* 32, 2031–2040. doi: 10.1002/jcc.21773
- Amézqueta, S., Subirats, X., Fuguet, E., Rosés, M., and Ràfols, C. (2020). In *Liquid-Phase Extraction*; Handbooks in Separation Science. eds Poole, C. F and Elsevier, 183–208. doi: 10.1016/B978-0-12-816911-7.00006-2
- Bhardwaj, V. K., Singh, R., Sharma, J., Rajendran, V., Purohit, R., and Kumar, S. (2020). Identification of bioactive molecules from tea plant as SARS-CoV-2 main protease inhibitors. *J. Biomol. Struct. Dyn.* 39, 3449–3458. doi: 10.1080/07391102.2020.2020
- Biovia, D. S. (2017). BIOVIA Discovery studio visualizer. Available at: <https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/visualization/>.
- Bulut, H., Hattori, S. I., Das, D., Murayama, K., and Amp Mitsuya, H. (2020). Crystal structure of SARS-CoV-2 Main protease in complex with an inhibitor GRL-2420. Deposited protein in RCSB database. doi: 10.2210/pdb7jkv/pdb
- Choubey, A., Dehury, B., Kumar, S., Medhi, B., and Mondal, P. (2022). Naltrexone a potential therapeutic candidate for COVID-19. *J. Biomol. Struct. Dyn.* 40, 963–970. doi: 10.1080/07391102.2020.1820379
- Corbeil, C. R., Williams, C. I., and Labute, P. (2012). Variability in docking success rates due to dataset preparation. *J. Comput. Aided Mol. Des.* 26, 775–786. doi: 10.1007/s10822-012-9570-1
- Dallakyan, S., and Olson, A. (2014). Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* 1263, 243–250. doi: 10.1007/978-1-4939-2269-7_19
- Das, K., Pattanaik, M., and Paital, B. (2023). The significance of super intelligence of artificial intelligence agencies in the social savageries of COVID-19: An appraisal. Integrated Science of Global Epidemics, Integrated Science 14, Springer Nature Switzerland AG 2023. eds Rezaei N. doi: 10.1007/978-3-031-17778-1_16
- Elmezayen, A., Al-Obaidi, A., Şahin, A., and Yelekcı, K. (2020). Drug repurposing for coronavirus (COVID-19): *in silico* screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *J. Biomol. Struct. Dyn.* 39, 2980–2992. doi: 10.1080/07391102.2020.1758791
- Farmer, J., Kanwal, F., Nikulsin, N., Tsilimigras, M. C. B., and Jacobs, D. J. (2017). Statistical measures to quantify similarity between molecular dynamics simulation trajectories. *Entropy (Basel)*. 19:646. doi: 10.3390/e19120646
- Girdhar, K., Dehury, B., Singh, M. K., Daniel, V. P., Choubey, A., Dogra, S., et al. (2019). Novel insights into the dynamics behavior of glucagon-like peptide-1 receptor with its small molecule agonists. *J. Biomol. Struct. Dyn.* 37, 3976–3986. doi: 10.1080/07391102.2018.1532818
- Gupta, A., Müller, A., Huisman, B., Fuchs, J., Schneider, P., and Schneider, G. (2017). Generative recurrent networks for De novo drug design. *Mol. Informat.* 37:1700111. doi: 10.1002/minf.201700111
- Hernández-Santoyo, A., Tenorio-Barajas, A. Y., Altuzar, V., Vivanco-Cid, V., and Mendoza-Barrera, C. (2013). “Ogawa, Protein-protein and protein-ligand docking” in *Protein engineering—technology and applications*, (United Kingdom) vol. 196. doi: 10.5772/56376
- Hou, N., Shuai, L., Zhang, L., Xie, X., Tang, K., Zhu, Y., et al. (2023). Development of Highly Potent Noncovalent Inhibitors of SARS-CoV-2 3CLpro. *ACS Cent Sci.* 9:217–227. doi: 10.1021/acscentsci.2c01359
- Huey, R., Morris, G. M., Olson, A. J., and Goodsell, D. S. (2007). A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* 28, 1145–1152. doi: 10.1002/jcc.20634
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582, 289–293. doi: 10.1038/s41586-020-2223-y
- Ju, J., Kumar, S., Li, X., Jockusch, S., and Russo, J. J. (2020a). Nucleotide analogues as inhibitors of viral polymerases. *BioRxiv*. doi: 10.1101/2020.01.30.927574
- Ju, J., Li, X., Kumar, S., Jockusch, S., Chien, M., Tao, C., et al. (2020b). Nucleotide analogues as inhibitors of SARS-CoV polymerase. *BioRxiv*. doi: 10.1101/2020.03.12.989186
- Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., and Zhavoronkov, A. (2017). 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
- Kenna, J., Taskar, K., Battista, C., Bourdet, D., Brouwer, K., Brouwer, K., et al. (2018). Can bile salt export pump inhibition testing in drug discovery and development reduce liver injury risk? An international transporter consortium perspective. *Clin. Pharmacol. Therapeut.* 104, 916–932. doi: 10.1002/cpt.1222
- Khan, T., Dixit, S., Ahmad, R., Raza, S., Azad, I., Joshi, S., et al. (2017). Molecular docking, PASS analysis, bioactivity score prediction, synthesis, characterization and biological activity evaluation of a functionalized 2-butanone thiosemicarbazone ligand and its complexes. *J. Chem. Biol.* 10, 91–104. doi: 10.1007/s12154-017-0167-y
- Khan, S., Zia, K., Ashraf, S., Uddin, R., and Ul-Haq, Z. (2020). Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. *J. Biomol. Struct. Dyn.* 39, 2607–2616. doi: 10.1080/07391102.2020.1751298
- Kim, S., Thiessen, P., Cheng, T., Yu, B., and Bolton, E. (2018). An update on PUG-REST: RESTful interface for programmatic access to PubChem. *Nucleic Acids Res.* 46, W563–W570. doi: 10.1093/nar/gky294
- Leeson, P. D., Bento, A. P., Gaulton, A., Hersey, A., Manners, E. J., Radoux, C. J., et al. (2021). Target-based evaluation of ‘drug-like’ properties and ligand efficiencies. *J. Med. Chem.* 64, 7210–7230. doi: 10.1021/acs.jmedchem.1c00416
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* 382, 1199–1207. doi: 10.1056/NEJMoa2001316
- Li, Q., and Kang, C. (2020). Progress in developing inhibitors of SARS-CoV-2 3C-like protease. *Microorganisms* 8:1250. doi: 10.3390/microorganisms8081250
- Li, Z., Li, X., Huang, Y. Y., Wu, Y., Liu, R., Zhou, L., et al. (2020). Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proc. Natl. Acad. Sci. U. S. A.* 117, 27381–27387. doi: 10.1073/pnas.2010470117
- Lipinski, C., Lombardo, F., Dominy, B., and Feeney, P. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25. doi: 10.1016/S0169-409X(96)00423-1
- Macchiagodena, M., Pagliai, M., and Procacci, P. (2020). Inhibition of the main protease 3CLPro of the coronavirus disease 19 via structure-based ligand design and molecular modelling. *Chem. Phys. Lett.* 750:137489. doi: 10.1016/j.cplett.2020.137489
- Martinez, L. (2015). Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS One* 10:e0119264. doi: 10.1371/journal.pone.0119264
- Meng, X. Y., Zhang, H. X., and Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* 1; 7(2), 146–157. doi: 10.2174/157340911795677602
- Menon, S. V. (2022). COVID-19: review on the biochemical perspective of the structural and non-structural proteins involved in SARS CoV-2 Corona virus. *Biochem. Anal. Biochem.* 11:42. doi: 10.35248/2161-1009.22.11.420
- Mishra, P., Paital, B., Jena, S., Samanta, L., Kumar, S., and Swain, S. (2019). Possible activation of NRF2 by vitamin E/curcumin against altered thyroid hormone induced oxidative stress via NFκB/AKT/mTOR/KEAP1 signaling in rat heart. *Sci. Rep.* 9:7408. doi: 10.1038/s41598-019-43320-5
- Mishra, P., Tandon, G., Kumar, M., Paital, B., Swain, S., Kumar, S., et al. (2022). Promoter sequence interaction and structure based multi-targeted (redox regulatory genes) molecular docking analysis of vitamin E and curcumin in T4 induced oxidative stress model using H9C2 cardiac cell line. *J. Biomol. Struct. Dyn.* 40, 12316–12335. doi: 10.1080/07391102.2021.1970624
- Moret, M., Friedrich, L., Grisoni, F., and Merk, D. (2019). Schneider, G. generating customized compound libraries for drug discovery with machine intelligence. *ChemRxiv*:10119299. doi: 10.26434/chemrxiv.10119299.v1
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., et al. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662. doi: 10.1002/(SICI)1096-987X(19981115)19:143.O.CO;2-B
- O’Boyle, N., Banck, M., James, C., Morley, C., Vandermeersch, T., and Hutchison, G. (2011). Open babel: an open chemical toolbox. *J. Chem.* 3:33. doi: 10.1186/1758-2946-3-33
- Paital, B., Pati, S. G., Panda, F., Jally, S. K., and Agrawal, P. K. (2022). Changes in physicochemical, heavy metals and air quality linked to spot *Aplocheilus panchax* along Mahanadi industrial belt of India under COVID-19-induced lockdowns. *Environ. Geochem. Health* 45, 751–770. doi: 10.1007/s10653-022-01247-3, ISSN 1573-2983
- Paital, B., Sablok, G., Kumar, S., Singh, S. K., and Chainy, G. B. N. (2015). Investigating the conformational structure and potential site interactions of SOD inhibitors on Ec-SOD in marine mud crab *Scylla serrata*: a molecular modeling approach. *Interdiscip. Sci. Comput. Life Sci.* 8, 312–318. doi: 10.1007/s12539-015-0110-2
- Pathak, N., Chen, Y. T., Hsu, Y. C., Hsu, N. Y., Kuo, C. J., Tsai, H. P., et al. (2021). Uncovering flexible active site conformations of SARS-CoV-2 3CL proteases through protease pharmacophore and clusters and covid –19 fdrug repurposing. *ACS Nano* 15, 857–872. doi: 10.1021/acsnano.0c07383
- Pereira, J. E. M., Eckert, J., Rudic, S., Yu, D., Mole, R., Tsapatsaris, N., et al. (2019). Hydrogen bond dynamics and conformational flexibility in antipsychotics. *Phys. Chem. Chem. Phys.* 21, 15463–15470. doi: 10.1039/C9CP02456E
- Prabhavathi, H., Dasegowda, K., Renukananda, K., Lingaraju, K., and Naika, H. (2020). Exploration and evaluation of bioactive phytochemicals against BRCA proteins by *in silico* approach. *J. Biomol. Struct. Dyn.* 39, 5471–5485. doi: 10.1080/07391102.2020.1790424
- Prykhodko, O., Johansson, S., Kotsias, P., Arús-Pous, J., Bjerrum, E., Engkvist, O., et al. (2019). A de novo molecular generation method using latent vector based generative adversarial network. *Journal of Cheminformatics* 11:74. doi: 10.1186/s13321-019-0397-9
- Radinnurafiqah, M., Paital, B., Kumar, S., Abubaker, S., and Tripathy, S. (2016). AgNO₃ dependant modulation of glucose mediated respiration kinetics in *Escherichia coli* at different pH and temperature. *J. Mol. Recognit.* 29, 544–554. doi: 10.1002/jmr.2554

- RCSB protein data bank. Available at: <https://www.rcsb.org/>. (2022) (Accessed Dec 12, 2022).
- Rujuta, R. D., Arpita, P. T., Narendra, N., and Manisha, M. (2020). *In silico* molecular docking analysis for repurposing therapeutics against multiple proteins from SARS-CoV-2. *Eur. J. Pharmacol.* 886:173. doi: 10.1016/j.ejphar.2020.173430
- Sahoo, A., Swain, S. S., Paital, B., and Panda, M. (2022). Combinatorial approach of vitamin C derivative and anti-HIV drug-darunavir against SARS-CoV-2. *Front. Biosci. (Landmark Ed.)* 27:10. doi: 10.52586/f.bll2701010
- Sargsyan, K., Grauffel, C., and Lim, C. (2017). How molecular size impacts RMSD applications in molecular dynamics simulations. *J. Chem. Theory Comput.* 13, 1518–1524. doi: 10.1021/acs.jctc.7b00028
- Sarma, P., Shekhar, N., Prajapat, M., Avti, P., Kaur, H., Kumar, S., et al. (2020). *In-silico* homology assisted identification of inhibitor of RNA binding against 2019-nCoV N-protein (N terminal domain). *J. Biomol. Struct. Dyn.* 39, 2724–2732. doi: 10.1080/07391102.2020.1753580
- Schneider, G. (2019). Mind and Machine in Drug Design. *Nat. Mach. Intell.* 1, 128–130. doi: 10.1038/s42256-019-0030-7
- Schneider, N., Sayle, R. A., and Landrum, G. A. (2015). Get your atoms in order- an open-source implementation of a novel and robust molecular canonicalization algorithm. *J. Chem. Inf. Model.* 55, 2111–2120. doi: 10.1021/acs.jcim.5b00543
- Shaji, D. (2016). The relationship between relative solvent accessible surface area (rASA) and irregular structures in protean segments (ProSs). *Bioinformation* 12, 381–387. doi: 10.6026/97320630012381
- Tahir ul Qamar, M., Alqahtani, S. M., Alamri, M. A., and Chen, L.-L. (2020). Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *J. Pharmaceut. Anal.* 10, 313–319. doi: 10.1016/j.jpha.2020.03.009
- Udofia, I. A., Gbayo, K. O., Oloba-Whenu, O. A., Ogunbayo, T. B., and Isanbor, C. (2021). *In silico* studies of selected multi-drug targeting against 3CL^{pro} and nsp12 RNA-dependent RNA-polymerase proteins of SARS-CoV-2 and SARS-CoV. *Netw. Model. Anal. Health Informat. Bioinformat.* 10:22. doi: 10.1007/s13721-021-00299-2
- 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
- Vanhaelen, Q., Mamoshina, P., Aliper, A., Artemov, A., Lezhnina, K., Ozerov, I., et al. (2017). Design of efficient computational workflows for *in silico* drug repurposing. *Drug Discov. Today* 22, 210–222. doi: 10.1016/j.drudis.2016.09.019
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2010). CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* 31, 671–690. doi: 10.1002/jcc.21367
- Vardhan, S., and Sahoo, S. (2020). *In silico* ADMET and molecular docking study on searching potential inhibitors from limonoids and triterpenoids for COVID-19. *Comput. Biol. Med.* 124:103936. doi: 10.1016/j.combiomed.2020.103936
- Veber, D., Johnson, S., Cheng, H., Smith, B., Ward, K., and Kopple, K. (2002). Molecular properties that influence the Oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615–2623. doi: 10.1021/jm020017n
- Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins* 52, 609–623. doi: 10.1002/prot.10465
- World Health Organization. *Coronavirus disease 2019 (COVID-19) situation report-36*. Available at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>. (2020) (Accessed August 24, 2020).
- Yang, H., Lou, C., Sun, L., Li, J., Cai, Y., Wang, Z., et al. (2018). admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics* 35, 1067–1069. doi: 10.1093/bioinformatics/bty707
- Yin, J., and Wang, J. (2016). Renal drug transporters and their significance in drug-drug interactions. *Acta Pharm. Sin. B* 6, 363–373. doi: 10.1016/j.apsb.2016.07.013
- Zanger, U., and Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* 138, 103–141. doi: 10.1016/j.pharmthera.2012.12.007
- Zhavoronkov, A., Ivanenkov, Y., Aliper, A., Veselov, M., Aladinskiy, V., Aladinskaya, A., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 37, 1038–1040. doi: 10.1038/s41587-019-0224-x
- Zhu, J., Zhang, H., Lin, Q., Lyu, J., Lu, L., Chen, H., et al. (2022). Progress on SARS-CoV-2 3CLpro inhibitors: inspiration from SARS-CoV 3CLpro peptidomimetics and small-molecule anti-inflammatory compounds. *Drug Des. Devel. Ther.* 16, 1067–1082. doi: 10.2147/DDDT.S359009



OPEN ACCESS

EDITED BY

Arli Aditya Parikesit,
Indonesia International Institute for Life-
Sciences (i3L), Indonesia

REVIEWED BY

Yong Guo,
Zhengzhou University, China
Arif Nur Muhammad Ansori,
Airlangga University, Indonesia

*CORRESPONDENCE

Joshuan J. Barboza
✉ jbarbozame@ucvvirtual.edu.pe
Magdi E. A. Zaki
✉ mezaaki@imamu.edu.sa

[†]These authors have contributed equally to this work

RECEIVED 03 April 2023

ACCEPTED 30 May 2023

PUBLISHED 28 June 2023

CITATION

Akash S, Bayil I, Rahman MA, Mukerjee N, Maitra S, Islam MR, Rajkhowa S, Ghosh A, Al-Hussain SA, Zaki MEA, Jaiswal V, Sah S, Barboza JJ and Sah R (2023) Target specific inhibition of West Nile virus envelope glycoprotein and methyltransferase using phytocompounds: an *in silico* strategy leveraging molecular docking and dynamics simulation.

Front. Microbiol. 14:1189786.

doi: 10.3389/fmicb.2023.1189786

COPYRIGHT

© 2023 Akash, Bayil, Rahman, Mukerjee, Maitra, Islam, Rajkhowa, Ghosh, Al-Hussain, Zaki, Jaiswal, Sah, Barboza and Sah. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Target specific inhibition of West Nile virus envelope glycoprotein and methyltransferase using phytocompounds: an *in silico* strategy leveraging molecular docking and dynamics simulation

Shopnil Akash¹, Imren Bayil², Md. Anisur Rahman³,
Nobendu Mukerjee^{4,5†}, Swastika Maitra^{6†}, Md. Rezaul Islam¹,
Sanchaita Rajkhowa⁷, Arabinda Ghosh⁸, Sami A. Al-Hussain⁹,
Magdi E. A. Zaki^{9*}, Vikash Jaiswal¹⁰, Sanjit Sah^{11,12},
Joshuan J. Barboza^{13*} and Ranjit Sah^{14,15,16}

¹Department of Pharmacy, Faculty of Allied Health Sciences, Daffodil International University, Dhaka, Bangladesh, ²Department of Bioinformatics and Computational Biology, Gaziantep University, Gaziantep, Türkiye, ³Department of Pharmacy, Islamic University, Kushtia, Bangladesh, ⁴Department of Microbiology, West Bengal State University, Kolkata, West Bengal, India, ⁵Department of Health Sciences, Novel Global Community Educational Foundation, Hebersham, NSW, Australia, ⁶Department of Microbiology, Adamas University, Kolkata, West Bengal, India, ⁷Centre for Biotechnology and Bioinformatics, Dibrugarh University, Dibrugarh, Assam, India, ⁸Microbiology Division, Department of Botany, Gauhati University, Gwahati, Assam, India, ⁹Department of Chemistry, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, ¹⁰Department of Cardiovascular Research, Larkin Community Hospital, South Miami, FL, United States, ¹¹Global Consortium for Public Health and Research, Datta Meghe Institute of Higher Education and Research, Jawaharlal Nehru Medical College, Wardha, India, ¹²SR Sanjeevani Hospital, Kayanpur, Siraha, Nepal, ¹³Escuela de Medicina, Universidad Cesar Vallejo, Trujillo, Peru, ¹⁴Tribhuvan University Teaching Hospital, Institute of Medicine, Kathmandu, Nepal, ¹⁵Dr. D. Y. Patil Medical College, Hospital and Research Centre, Dr. D. Y. Patil Vidyapeeth, Pune, Maharashtra, India, ¹⁶Department of Public Health Dentistry, Dr. D.Y. Patil Dental College and Hospital, Dr. D.Y. Patil Vidyapeeth, Pune, Maharashtra, India

Mosquitoes are the primary vector for West Nile virus, a flavivirus. The virus's ability to infiltrate and establish itself in increasing numbers of nations has made it a persistent threat to public health worldwide. Despite the widespread occurrence of this potentially fatal disease, no effective treatment options are currently on the market. As a result, there is an immediate need for the research and development of novel pharmaceuticals. To begin, molecular docking was performed on two possible West Nile virus target proteins using a panel of twelve natural chemicals, including Apigenin, Resveratrol, Hesperetin, Fungisterol, Lucidone, Ganoderic acid, Curcumin, Kaempferol, Cholic acid, Chlorogenic acid, Pinocembrin, and Sanguinarine. West Nile virus methyltransferase (PDB ID: 2OY0) binding affinities varied from -7.4 to -8.3 kcal/mol, whereas West Nile virus envelope glycoprotein affinities ranged from -6.2 to -8.1 kcal/mol (PDB ID: 2I69). Second, substances with larger molecular weights are less likely to be unhappy with the Lipinski rule. Hence, additional research was carried out without regard to molecular weight. In addition, compounds 01, 02, 03, 05, 06, 07, 08, 09, 10 and 11 are more soluble in water than compound 04 is. Besides, based on maximum binding affinity, best three compounds (Apigenin, Curcumin, and Ganoderic Acid) has been carried out molecular dynamic simulation (MDs) at 100ns to determine

their stability. The MDs data is also reported that these mentioned molecules are highly stable. Finally, advanced principal component analysis (PCA), dynamics cross-correlation matrices (DCCM) analysis, binding free energy and dynamic cross correlation matrix (DCCM) theoretical study is also included to established mentioned phytochemical as a potential drug candidate. Research has indicated that the aforementioned natural substances may be an effective tool in the battle against the dangerous West Nile virus. This study aims to locate a bioactive natural component that might be used as a pharmaceutical.

KEYWORDS

West Nile virus, molecular docking, molecular dynamic simulation, PCA, drug-likeness, ADMET

1. Introduction

The infectious disease has the potential to cause a global pandemic and has been transmitted regularly throughout human history, such as plague, cholera, the flu, SARS-CoV, and MERS-CoV (Fahmi et al., 2021; Piret and Boivin, 2021). Besides, many infectious diseases may lead to pandemics, including Monkeypox, and Marburg virus (Farahat and Memish, 2022). These infections were introduced to humans due to increased interaction with animals due to activities such as breeding, hunting, and global commerce. Recently, another new viral disease has been affecting people all over the globe, whose name is the West Nile virus. In 1937, researchers discovered the West Nile virus for the first time on the African continent, specifically in the West Nile area of Uganda (Sule et al., 2018). Before 1997, the West Nile virus was not thought to be harmful to birds. However, in Israel around that time, a more dangerous strain of the virus was responsible for the deaths of many bird species with symptoms of encephalitis and paralysis. Since the beginning of the 1950s, several nations worldwide have reported cases of human infections caused by the West Nile virus (Hayes, 2001). Today, the West Nile virus is the most common infectious disease spread by mosquitoes on the American continent. It is transmitted from person to person, most often by the bite of an infected mosquito. The West Nile virus infection was reported in two persons in New York City on August 16, 2022, one in Brooklyn and another in Queens. According to an announcement made by the New York City Health Department, the virus was also identified in a “record number” of infected mosquitoes around the city. According to the declaration made by the health authority, there have been a minimum of 54 instances detected throughout the continent this year, along with four fatalities. The risk of contracting the West Nile virus is highest during the mosquito season, which begins in the summer and lasts into the autumn (New York Times, 2022).

The majority of people who have been diagnosed with the West Nile virus do not experience any symptoms of illness. Fever and other symptoms like headaches, body aches, joint pains, diarrhea, vomiting, or a rash arise in around one of every five infected patients. About one in every 150 infected persons may get a severe disease, which can often be deadly. Over the age of 60, people have the greatest chance of developing this life-threatening disease. Positive cases in New York City had a median age of 62. Those who already have preexisting

health problems, such as cancer, diabetes, or hypertension, may be at a greater risk of developing severe further health issues. In extreme cases of West Nile virus, recovering might require several weeks or months, but impairment to the central nervous system can be incurable (New York Times, 2022).

The West Nile virus infection should be included in the differential diagnosis, and particular laboratory test findings should raise a high index of clinical suspicion. The IgM-capture enzyme-linked immunosorbent test is the most often applied diagnostic technique. This test has a sensitivity range of 95 to 100% in serum and spinal fluid. The IgM is usually detected in the serum and cerebrospinal fluid by the time the sickness manifests itself. When a person has encephalitis or meningitis, a high IgM of West Nile virus antibodies is undoubtedly indicative of infection; however, the IgM may last for many months to more than a year because humoral IgM antibodies do not penetrate the blood–brain barrier, an intrathecal West Nile virus-specific IgM strongly supports a central nervous system infection (Kemmerly, 2003).

The clinical care of West Nile virus diseases is considered supportive since neither vaccination nor any drug can be used to treat or prevent West Nile virus in humans. Patients experiencing severe meningeal symptoms typically require pain relief for their headaches, antiemetic treatment, and rehydration for the nausea and vomiting associated with their condition (CDC, n.d.).

Although, a huge number of people are affecting in recent time, but no authorized medication or potential vaccine is available still to fight the West Nile virus (Bergmann et al., 2022). Therefore, it is of the utmost importance issues to rapidly advance the research and development of potent anti-viral medications to combat this lethal pathogenic viral infection.

To development of a novel drug candidate, required more than 10 years of times, huge amount of money and resources (Rahman et al., 2012). So, the primary goal of *in silico* research is to develop a drug design strategy to predict the inhibitory activity of a series of natural compounds against West Nile virus with the help of advanced *in silico* method such as molecular docking, molecular dynamics simulations, ADMET, principal component analysis (PCA), dynamics cross-correlation matrices (DCCM) analysis, binding free energy and dynamic cross correlation matrix (DCCM). The molecular docking studies was conducted to determine the binding affinity of the selected phytocompounds with the target

proteins of the West Nile virus and then based on maximum affinity, three most potential compounds were carried out to the molecular dynamics simulations and determine the stability and binding energies of the protein-ligand complexes. Ultimately the results of this investigation have the potential to contribute to the development of novel drugs for the treatment of West Nile virus infection.

2. Literature based evidence

2.1. Pharmacological activity of reported compounds in earlier studies and based on which compound are chosen

2.1.1. Apigenin pharmacological activity

Apigenin (APG) is a flavonoid found in high concentrations in many fruits, vegetables, and Chinese medicinal herbs. It has a wide range of physiological effects, including anti-inflammatory, antioxidant, antibacterial, anti-viral, and blood pressure-lowering properties (Yan et al., 2017). Apigenin has been shown to have anti-cancer, anti-diabetic, and antioxidant effects. APG inhibits tumor cell growth and survival by triggering apoptosis, cell cycle arrest, and the creation of ROS. It also prevents metastasis and angiogenesis by changing various cellular signaling pathways (Zhou et al., 2022). APG suppresses Enterovirus-71 infection by interfering with RNA-transacting factor interactions (Zhang et al., 2014).

2.1.2. Resveratrol antiviral activity

Resveratrol is a natural chemical produced by certain plants (Campagna and Rivas, 2010). Resveratrol has anti-viral activity against human and animal viruses (Abba et al., 2015). Resveratrol has also inhibitory effects on various viruses (Zhao et al., 2017).

2.1.3. Hesperetin pharmacological activity

Hesperetin is a flavanone, and in its water-soluble glycoside form of hesperidin, it can be found in various citrus fruits. Hesperetin has shown anti-viral properties *in vitro* against a few RNA viruses (Zandi et al., 2011). Hesperidin also demonstrated anti-viral activity against the growth of the 17D strain of the yellow fever virus (Agrawal et al., 2021). The bioflavonoid Hesperetin inhibits the intracellular reproduction of viruses (Oo et al., 2016).

2.1.4. Fungisterol anti-viral activity

Fungisterol, a compound found in mushrooms, has anti-viral properties (Zhu et al., 2021). Potential anti-viral effects of Fungisterol have been demonstrated against various pathogens, including the human immunodeficiency virus (HIV), influenza, herpes simplex virus (HSV), hepatitis B and C viruses, and others (Raut and Adhikari, 2021).

2.1.5. Lucidone pharmacological activity

Lucidone is a natural compound found in the fruiting body of the edible mushroom *Ganoderma lucidum*. It belongs to the class of organic compounds known as coumarins, which are fragrant and sweet-smelling organic compounds that are commonly found in plants (Boh et al., 2007). Recent research has been reported that the

Lucidone might be composed to have various pharmacological properties, including antioxidant, anti-inflammatory, antitumor, antiviral and neuroprotective activities. It also reported that the Lucidone suppresses dengue viral replication through the induction of heme oxygenase-1 and suppresses Hepatitis C Virus (Chen et al., 2013, 2018).

2.1.6. Ganoderic acid anti-viral activity

Ganoderic acid (GA) has been shown to have anti-viral, antihypertensive, anti-cancer, and immune-modulatory effects (Zhu et al., 2021). Anti-viral effects of Ganoderic acid against Enterovirus 71 (EV71) infection are observed without cytotoxicity (Zhang et al., 2014). There is evidence that Ganoderic acid can inhibit viral replication and reduce the extent of liver damage (Li and Wang, 2006).

2.1.7. Curcumin pharmacological activity

It has also been demonstrated that Curcumin prevents the development of several viruses and cancerous cells. Curcumin can block the production of HSV-1's immediate early genes and the action of the HIV-1 integrase, both of which are required for the virus's replication (Zandi et al., 2010). Through mechanisms present in the infected cells, Curcumin can stop the production of viral genes and, consequently, viral replication (Kutluay et al., 2008).

2.1.8. Kaempferol pharmacological activity

Previous reports indicate flavanol Kaempferol and its glycosides have strong anti-viral properties (Khazdair et al., 2021). Kaempferol could prevent the spread of viruses in the brain, lungs, kidneys, heart, and spleen. According to research, the hepatitis B virus, the H1N1 and H9N2 influenza viruses, and the Herpes simplex virus are all susceptible to the effects of Kaempferol (Zhang et al., 2012; Rahman et al., 2022).

2.1.9. Cholic acid pharmacological activity

Since both Cholic acid and metal ions have pharmacological effects, such as antibacterial, anti-viral, antifungal, antimalarial, antitubercular, anti-cancer, spermicidal, and antiallergic, their organometallic complexes were created to have a synergistic effect (Kishu and Siva, 2011). It has been found that several cholic acid derivatives, including tauroolithocholic acid, lithocholic acid 3-sulfate, and tauroolithocholic acid 3-sulfate, preferentially suppress type 1 human immunodeficiency virus replication (HIV-1) (Baba et al., 1989). Cholic acid can be prevented Cocksackievirus B3 recurrence and produce anti-viral activity (Han et al., 2018).

2.1.10. Chlorogenic acid pharmacological activity

Chlorogenic acid is a widely distributed natural chemical with numerous significant pharmacological actions, including antioxidant, anti-inflammatory, antibacterial, anti-viral, hypoglycemic, lipid-lowering, anti-cardiovascular, antimutagenic, anti-cancer, and immunomodulatory (Naveed et al., 2018). Chlorogenic acid (CHA) treats viral upper respiratory tract infections brought on by influenza, parainfluenza, and respiratory syncytial viruses (Ding et al., 2017). Besides, another *in silico* investigation has reported that Chlorogenic acid might be effective against SARS-CoV-2 (Aini et al., 2022).

2.1.11. Pinocembrin pharmacological activity

Pinocembrin demonstrated potent antifungal action in a dose-dependent manner against *Penicillium italicum* (Peng et al., 2012). Pinocembrin exhibits anti-viral activity against the reproduction of the Zika virus (Ghildiyal et al., 2020; Huang et al., 2022). Pinocembrin inhibits the Zika virus replication cycle's post-entry process, and also, it also prevents the formation of viral envelope protein and RNA (Guang and Du, 2006).

2.1.12. Sanguinarine pharmacological activity

Sanguinarine is a plant alkaloid with considerable antiangiogenic, anti-cancer, and anti-viral potential but comparatively moderate toxicity (Feldman et al., 2018). Sanguinarine promotes anti-viral activity by upregulating TLR expression and its downstream mediator in MDM (Tilgner and Shi, 2004). It demonstrated selectivity for MKP-1 over MKP-3, inhibiting cellular MKP-1 with an IC_{50} of 10 nM (Guang and Du, 2006).

Below Figure 1 has shown the chemical structures of studied compounds and Table 1 displayed the pharmacological evidence more precisely.

2.2. Genome and physiological characteristics

The genus *Flavivirus*, of the family *Flaviviridae*, is host to the West Nile virus. *Flavivirus* virions are round and have a diameter of around 50 nm. The viral envelope and membrane proteins are encased in a lipid bilayer surrounding the nucleocapsid (roughly 30 nm in diameter and composed of capsid protein and genomic RNA). The virion has a single, plus-sense RNA genome about 11 kilobases in length and is divided into a 5' untranslated region (UTR). The conserved dinucleotide AG follows the 5' caps of the flavivirus genome, and the conserved dinucleotide CUOH finishes off the 3' ends of the genome. The 5' UTR and 3' UTR of genomic RNA each have a length of about 100 and 500 to 700 nucleotides, respectively. Viral and cellular proteases co-translate and post-translate the encoded polyprotein into seven nonstructural proteins (NS1, NS2a, NS2b, NS3, NS4a, and NS5) and three structural proteins (capsid, pre-membrane or transmembrane, and envelope). The nonstructural proteins function primarily as replication complex components to ensure the successful replication of viral RNA. For RNA replication

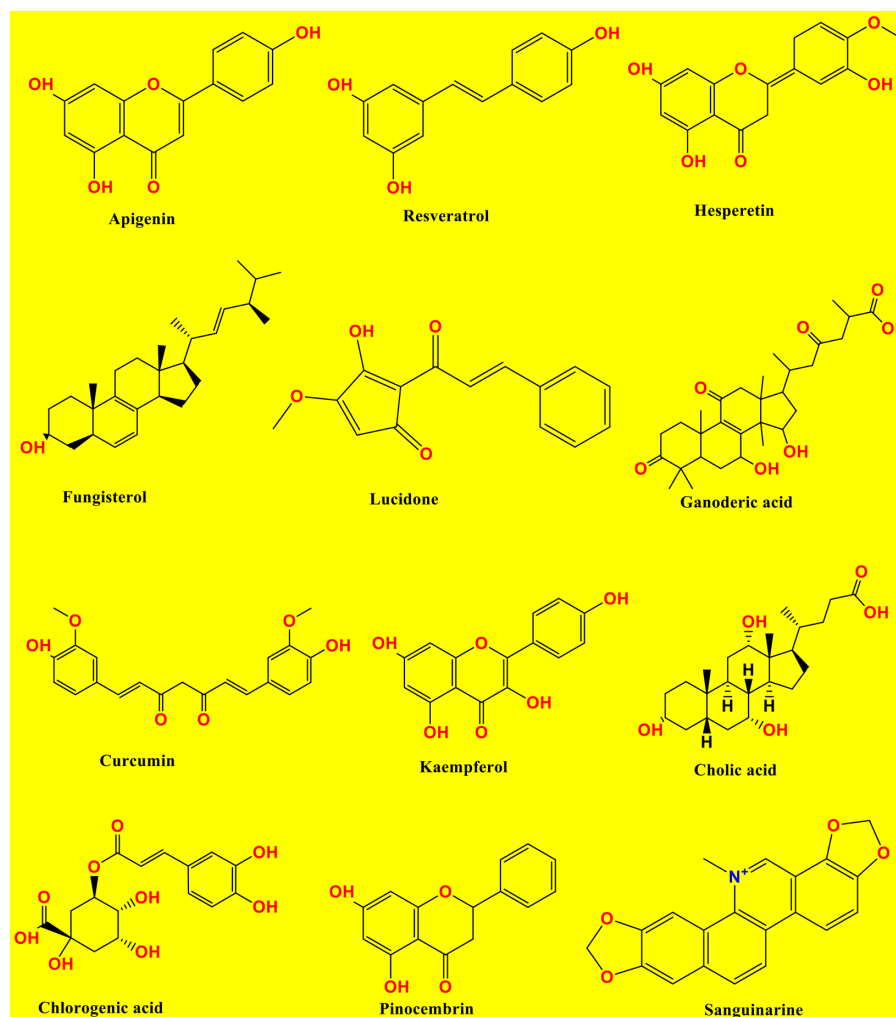


FIGURE 1
Chemical structures of studied compounds.

TABLE 1 Experimental and clinical data on the anti-viral activities of natural compounds against viral strains have been reported in different research investigations.

Drug name	Anti-viral activity	References
Apigenin	Apigenin suppresses enterovirus-71 infection by interfering with RNA-transacting factor interactions	Zhang et al. (2014)
Resveratrol	Resveratrol effectively reduced the cytopathogenic action of AD virus type 7	Guan et al. (2008)
Hesperetin	Hesperetin inhibits the intracellular reproduction of viruses	Oo et al. (2016)
Fungisterol	Fungisterol's anti-viral properties related to the polysaccharides present in the mycelium and fruiting bodies	Raut and Adhikari (2021)
Lucidone	Lucidone inhibits DENV protein production	Senthil Kumar and Wang (2016)
Ganoderic acid	Ganoderic acid are effective against human immunodeficiency virus type 1 (HIV-1)	Lindequist et al. (2005)
Curcumin	Curcumin prevented colon cancer cell lines (HT-29 and HCT-15) from proliferating by gathering cells in the G2-M phase	Syng-Ai et al. (2004)
Kaempferol	The 3a channel can be blocked by kaempferol and kaempferol glycosides, providing anti-viral action.	Schwarz et al. (2014)
Cholic acid	Cholic acid suppresses type 1 human immunodeficiency virus replication (HIV-1)	Baba et al. (1989)
Chlorogenic acid	Chlorogenic acid in reducing the HBV DNA relapse	Zuo et al. (2015)
Pinocembrin	Pinocembrin prevents the formation of viral envelope protein and RNA	Guang and Du (2006)
Sanguinarine	Sanguinarine promotes an anti-viral activity by upregulating TLR expression and its downstream mediator in MDM	Sunthamala et al. (2020)

to take place, NS1 must interact with NS4a. Hydrophobic NS2a was shown to have a role in flavivirus particle formation and release. With NS3, NS2b generated a complex essential for NS3's serine protease activity. A serine protease, 5'-RNA triphosphatase, nucleoside triphosphatase, and RNA helicase are all functions of NS3, a multifunctional protein. Yet, the roles of NS4a and NS4b, two membrane-associated proteins, remain unknown. The NS5 protein is an RNA-dependent RNA polymerase and a methyltransferase that modifies the 5' RNA cap structure. When a flavivirus replicates, a replication complex is assembled on the 3' UTR of the genomic RNA during the process of co-translation. When the positive-sense RNA is copied, the resulting negative-sense RNA is still base-paired with the original replicative form (RF). After that, the RF is reused as a starting point for the asymmetric production of more positive-sense RNA (Figure 2) (Tilgner and Shi, 2004).

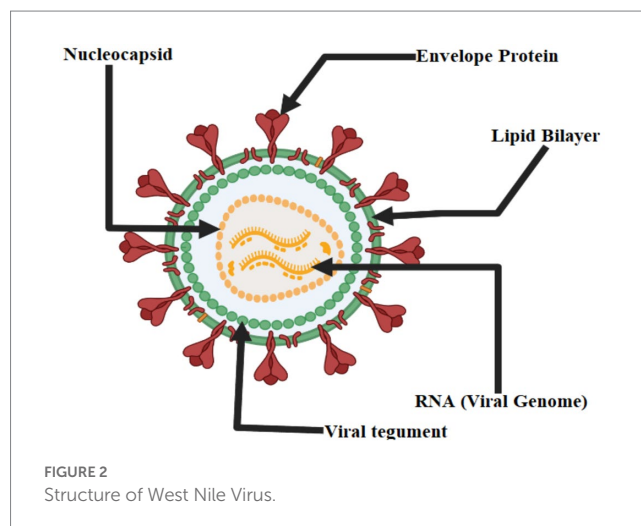
2.3. Transmission and clinical manifestation

The West Nile virus infection is a mosquito-borne disease that may infect people, birds, and animals. According to the CDC, the primary host for the virus is birds, and mosquitoes acquire the virus by biting affected birds. Health experts have announced that it cannot be transmitted from direct contact from person to person (West Nile Virus Found in Another NY County, Bringing Case Total to 5, n.d.). Figure 3 displays the transmission routes of the West Nile virus.

3. Computational method and material

3.1. Preparation of ligand and geometry optimization

By utilizing vibrational frequency from the DMol3 code of Material Studio 08, a method known as DFT functional was used to



accomplish molecular optimization (Yuan et al., 2016; Kumer et al., 2021). The molecular optimization was performed to produce exceptionally accurate results during molecular docking. After optimizing, chemical compounds were generated as a pdb file for molecular docking, drug-likeness, and ADMET (Figure 4).

3.2. Determination of ADMET, Lipinski rule, and drug-likeness

To assess their drug likeliness, we examine the ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) pharmacological characteristics of chosen compounds (Wijaya et al., 2021). With the aid of ADMET, we may predict if a molecule will be successfully implemented while designing new drugs based on their pharmacokinetics qualities. The pkCSM web server and Swiss-ADME were employed for this (AI Azzam, 2023). The Lipinski rule violation, GI (gastrointestinal) absorption, BBB penetration, and solubility are some of the pharmacological issues

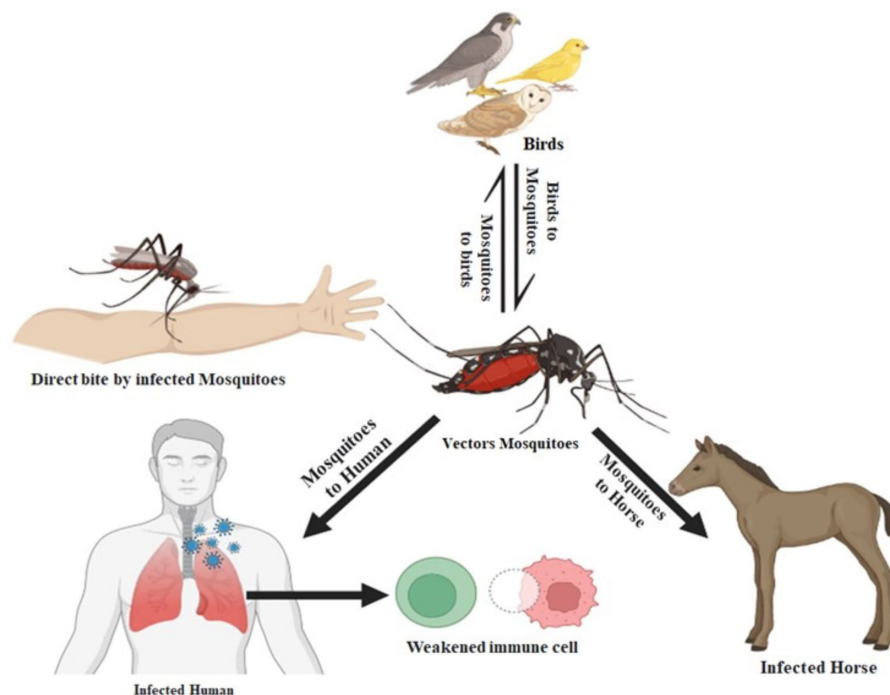


FIGURE 3

The West Nile virus transmission from animals to humans. When an infected mosquito bites an animal or a person (the “host”), the virus is introduced into the host’s circulation and has the potential to severely sick the host. The virus enters the circulation of the mosquito when an infected bird bites it and ultimately travels to its salivary glands.

discussed by the Swiss-ADME online-based web server.¹ We research our compound’s excretion and toxicity using the pkCSM web server (Daina et al., 2017; Wijaya et al., 2021).² Drug likeness is one of a drug compound’s sensory attributes that is frequently used in the drug discovery process. When Lipinski and colleagues released the rule of 5 (Ro5) in 1997, it was based on research of 2,245 drug qualities from the World Drug Index (WDI) databank that had been accepted for phase 2 clinical trials (Sadowski and Kubinyi, 1998; Akash, 2022). This was the first time that pharmacokinetic properties had been approached academically. Molecular weight, the amount of hydrogen bond donors and acceptors, topological polar surface area, consensus log Po/w, and drug similarity features such as lipinski rule violation and synthesis accessibility is all examined in this section.

3.3. Protein preparation, molecular docking analysis

The crystal structure of several west Nile virus strains was retrieved from the Protein Data Bank (PDB) of the RCSB Protein online portal (Burley et al., 2017). Protein purification was carried out carefully by removing ligands and water using Pymol v2.4.1 software³ (Padmi et al., 2022) (illustrated in Figure 5). The crystal

structure of the cleaned protein file is fed into AutoDock tools after the extra water and ligand have been removed, and the instructions are followed to create the pdbqt file format (Yuan et al., 2017; Kumer et al., 2022).

The development of the binding affinity of molecular complex may be predicted *via* the use of a computer approach known as “*in silico* molecular docking” (Agarwal and Mehrotra, 2016). In current studies, the PyRx AutoDock vina tool is applied for molecular docking. Before, performing docking, previously prepared drug, and protein were uploaded in PyRx application, and converted them as autodocking protein, and autodocking protein. Finally, specified grid parameter were selected, and perform the molecular docking studies (Yuliana et al., 2013).

3.4. Molecular dynamic simulations (MDs) methods

MD modeling is used to study the movement of molecules and atoms in dynamic systems, like protein-ligand complexes, in order to understand important physicochemical phenomena. MD simulations were run using receptor-ligand complexes acquired *via* molecular docking research to verify the binding mechanism and stability in a dynamic system. For this purpose, three best-docked compounds were subjected to 100 ns MD simulations: apigenin and curcumin against the West Nile virus methyltransferase with PDB ID 2OY0 as the target receptor, and Ganoderic acid against the West Nile virus envelope glycoprotein (PDB ID 2I69) target receptor. The simulation system was set up using the CHARMM-GUI web-based graphical

¹ <http://www.swissadme.ch/index.php>

² <https://biosig.lab.uq.edu.au/pkcsml/prediction>

³ <https://pymol.org/2/>

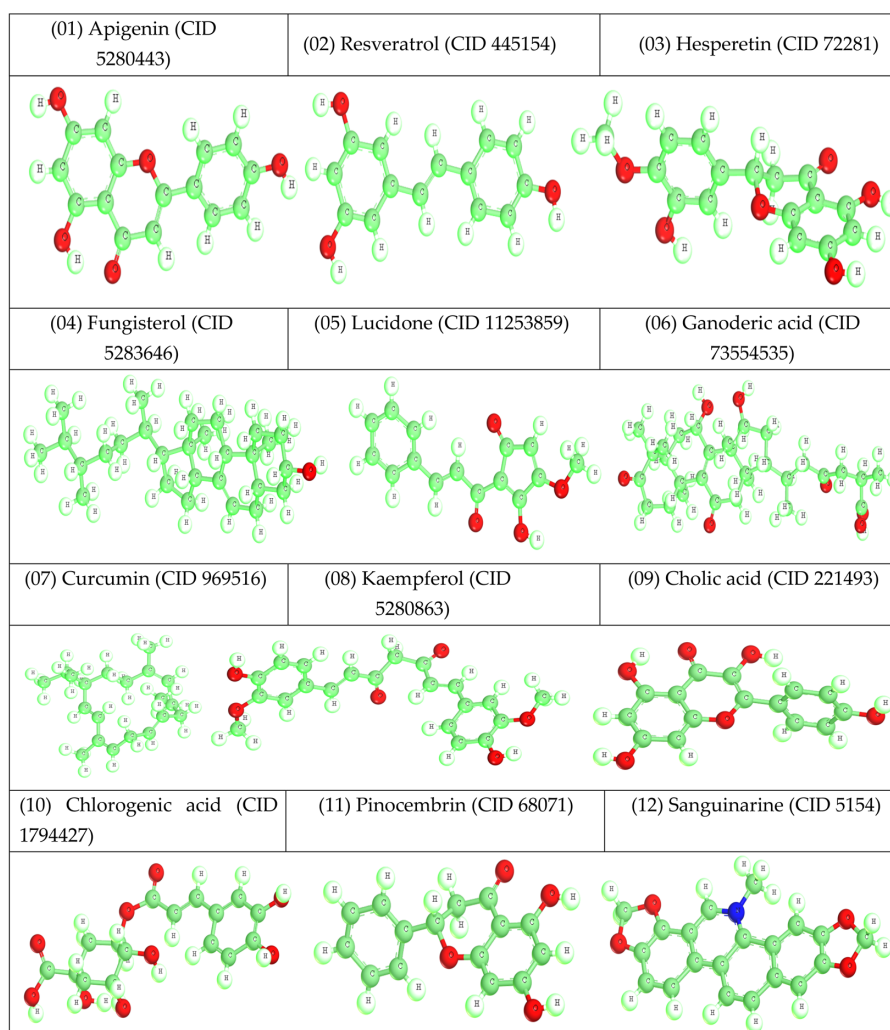


FIGURE 4
Optimized structure of reported natural compounds.

interface and produced the force field for both the ligands and the proteins (Wang et al., 2019). In the simulations that were carried out for 100 ns inside a periodic water box, the CHARMM36 force field and the Gromacs version 2020 software package were both utilized (Akash et al., 2022; Terefe and Ghosh, 2022). The complexes were positioned within a rectangular box that had a buffer distance of 10 in each of the cardinal directions. The box was subsequently dissolved in water molecules containing TIP3P. To keep the systems neutral, sodium and chloride ions were introduced, and then the energy was minimized using the steepest descent approach. Equilibration was performed on each of the complete systems at a temperature of 310 K for a total of 5,000 steps (i.e., 10 PS). The NPT ensemble's performance lasted 100 s. The Lincs technique was used to place constraints on hydrogen, and as a result, the timestep was set at 2 fs. All van der Waals forces were calculated using a switching method between 12 and 14, and 14 was found to be the cutoff value. The particle mesh

Ewald (PME) method with a maximum grid spacing of 1.2 was used to figure out the long-range electrostatic interactions. No multiple-time-stepping strategy was used; instead, PME computations were performed at each step. At a temperature of 310 kelvin, the initial velocities were chosen at random from a Maxwellian distribution. The temperature was maintained at 310 K. The barostat was programmed to have a target of 1 bar for any system size changes that occurred. The integration time step was 2 fs. After then, the simulation's output was re-centered, and the trajectories were subsequently assessed using the VMD (University of Illinois at Urbana-Champaign, Urbana, IL, USA) program, Bio3D, and QTGRACE, respectively. In the investigation of the system's stability, the root means square deviation (RMSD), the root means square fluctuation (RMSF), the radius of gyration (Rg), the number of hydrogen bonds, and the principal component analysis were taken into consideration (Walters, 2012).

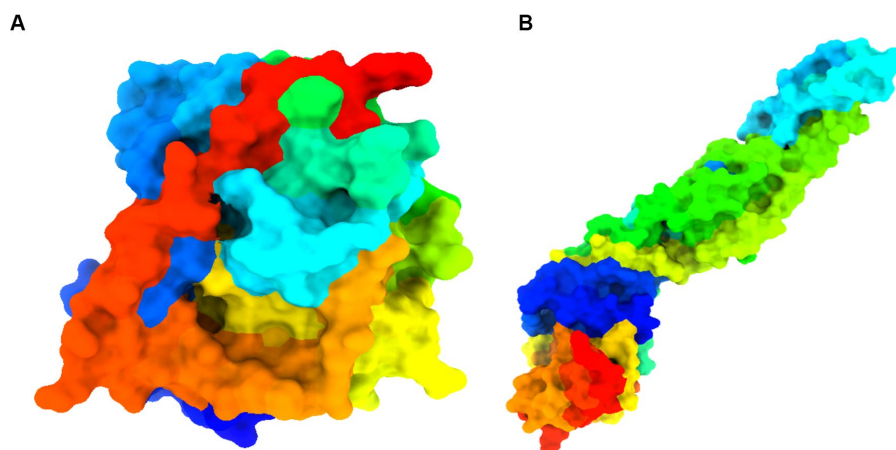


FIGURE 5

Details information about studied proteins. **(A)** West Nile virus methyltransferase (PDB ID 2OY0). Organism: West Nile virus. Method: X-ray diffraction. Resolution: 2.80 (Zhou et al., 2007). **(B)** West Nile virus envelope glycoprotein (PDB ID 2I69). Organism: West Nile virus. Method: X-ray diffraction. Resolution: 3.11 (Kanai et al., 2006).

3.5. Binding free energy calculation using MM-PBSA

In the molecular dynamics simulation, free energy calculation takes a major role in determining the binding energy of ligands inside proteins (Zikri et al., 2020). In this investigation, the MM-PBSA method was used to calculate the free interaction energy between ligands and West Nile virus methyltransferase (PDB ID 2OY0) and West Nile virus envelope glycoprotein (PDB ID 2I69). Calculations using MM-PBSA were performed to determine the binding affinity of protein-ligand complexes. Binding free energy (ΔG) estimation was done by eq. (3) using the script MMPBSA.py of the AMBER package (Wang et al., 2019).

$$[\Delta G_{\text{bind}} G = -\text{complex } G - \text{protein } G - \text{ligand}] \quad (1)$$

G-complex is the free energy of the complex; G-receptor is the free energy of the receptor; G-ligand is the free energy of the ligand (Terefe and Ghosh, 2022).

4. Result and analysis

4.1. Lipinski rule, pharmacokinetics

The Lipinski rule and drug resemblance give important details about the initial stages of medication development and raise the probability of success. According to the Lipinski rule of five, the drug's permeability and bioavailability depend on the molecular weight, the number of hydrogen bond donors and acceptors, and Topological polar surface area (\AA^2). The six-ligand compound can successfully pass the Lipinski rule of five without any violations (Walters, 2012; Akash et al., 2022). However, the other six ligand compound cannot satisfy by Lipinski rule.

Besides, low-molecular-weight ligand compounds are more readily absorbed, dispersed, and transported than those with greater molecular weights (Santos et al., 2016). The polarity of a ligand compound can be determined by a molecule's topological polar surface area (TPSA), which is a significant pharmacokinetics feature. These values help describe the drug transport characteristics. All polar atoms, primarily oxygen and nitrogen with connected hydrogen, make up the polar surface area of molecules. All our ligands showed very promising pharmacokinetics properties shown in the Table 2.

4.2. Molecular docking analysis against West Nile virus

One of the core technologies in computer-aided drug design is molecular docking, which allows for the analysis of protein-ligand complexes' non-bond interactions and energy bindings. Several natural compound derivatives were studied to attach molecularly with the target receptor protein of the West Nile virus. Molecular docking has been used to obtain specific information on protein-ligand complexes' interaction and binding affinity. The compounds were evaluated by utilizing the interaction with the maximum binding affinity scores. The use of computer-based docking studies in drug design is essential because it enables examination of the ligand configuration with the receptor protein's active site and investigation of non-bond interaction. Typically, it was believed that a pharmaceutical drug's effective binding affinity was -6.00 kcal/mol is expected (Kawsar et al., 2022; Rahman et al., 2022). The most incredible binding energy for the West Nile virus methyltransferase with (PDB ID: 2OY0) target receptor is -8.3 kcal/mole , whereas the maximum binding energy for the West Nile virus envelope glycoprotein with (PDB ID: 2I69) target receptor is -8.1 kcal/mol , as observed in molecules 04 and 20 of our medication (Table 3). That suggests our pharmacological compound is working more effectively against these two targets. As a result, our therapeutic

TABLE 2 Predicted data of Lipinski rule, pharmacokinetics.

No.	Weight (g/mol)	Hydrogen bond acceptor	Hydrogen bond donor	Topological polar surface area (Å ²)	Lipinski rule	
					Result	Violations
01	270.24	05	03	90.90	Yes	00
02	228.24	03	03	60.69	No	01
03	302.28	06	03	96.22	Yes	00
04	400.68	01	01	20.23	No	02
05	70.18	04	01	63.60	Yes	00
06	368.38	06	02	93.06	No	02
07	368.38	06	02	93.06	No	02
08	286.24	06	04	111.13	Yes	00
09	408.57	05	04	97.99	No	01
10	354.31	09	06	164.75	Yes	01
11	256.25	04	02	66.76	Yes	00
12	332.33	04	00	40.80	No	01

TABLE 3 Summary of binding affinities.

No.	Name	West Nile virus methyltransferase (PDB ID 2OY0)	West Nile virus envelope glycoprotein (PDB ID 2I69)
		Binding Affinity (kcal/mol)	Binding Affinity (kcal/mol)
01	Apigenin	−8.5	−7.5
02	Resveratrol	−7.7	−7.2
03	Hesperetin	−8.0	−7.1
04	Fungisterol	−7.4	−7.0
05	Lucidone	−7.8	−7.4
06	Ganoderic acid	−7.6	−8.0
07	Curcumin	−8.3	−7.3
08	Kaempferol	−8.1	−6.8
09	Cholic acid	−7.9	−7.6
10	Chlorogenic acid	−8.0	−7.1
11	Pinocembrin	−7.5	−6.5
12	Sanguinarine	−7.2	−6.7

molecule may be particularly successful in treating West Nile virus infection.

4.3. Protein-ligand interaction valuation

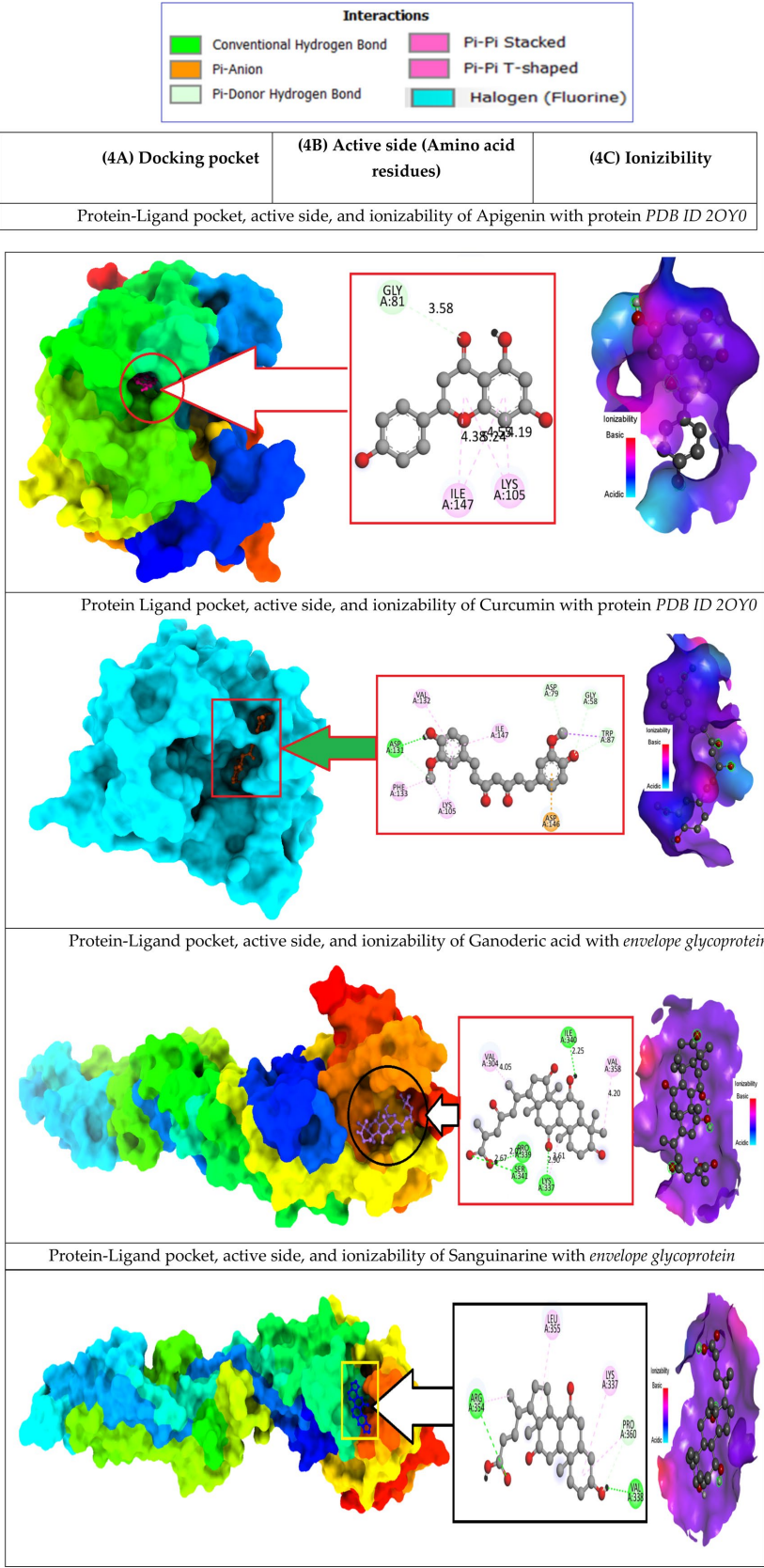
Docking studies of West Nile virus methyltransferase (PDB ID 2OY0) have revealed that drugs like Apigenin, and Curcumin, had the greatest binding score. On the other hand, compounds such as Ganoderic acid, and Cholic acid have been found to dock with the

West Nile virus envelope glycoprotein (PDB ID: 2I69) with the highest binding score. So, the protein-ligand interaction valuation of these mentioned compounds was carried out by BIOVIA Discovery Studio 2016 v16.1.0 and Pymol v2.4.1 software (Wahyuni et al., 2022). Curcumin's binding sites are located at LYS-A:105 and ILE-A:147 on West Nile virus methyltransferase (PDB ID 2OY0), Apigenin binding site are found at VAL-A:132, ASP-A:131, PHE-A:133, LYS-A:105, ASP-A:146, TRP-A:87, GLY-A:58, ASP-A:79. On the other hand, Ganoderic acid binding sites are located at PRO-A:339, VAL-A:304, LYS-A:337, ARG-A:338, VAL-A:385, VAL-A:343, ASA-A:347. Cholic acid binding sites are located at LE-A:340, VAL-A:358, VAL-A:304, LYS-A:337, SER-A:341, and PRO-A:339. Sanguinarine binding sites are located at ALA-A:161, PRO-A:146, SER-A:363, VAL-A:364, ALA-A:369, LYS-A:370, and THR-A:148. This active site was predicted by discovery studio 2021 (Kumer et al., 2022).

Besides, Un-ionized medications are administration into the body more effectively than their ionized counterparts, which are also known as charged drugs. A drug that is a weak acid will be absorbed predominantly in the acidic medium if it is administered orally; on the other hand, a drug that is a weak base will be captured in the alkaline system of the small intestines if it is administered orally (Song et al., 2004; Kumar et al., 2013). In (6C, as depicted in Figure 6) ionizability is displayed. The red color represented basicity, and sky-blue represented acidic condition while the deep blue color means neutral condition. According to the study, it is seen that most of the drugs are slightly basic in nature, which means they may better absorption rate in basic medium.

4.4. MD simulation results

The molecular dynamics simulation was launched to comprehend the stability of the complex in depth. This approach is essential for understanding the structure–function relationship of macromolecules. In this study, Apigenin, Curcumin, and Ganoderic Acid were determined to be the best leaders among all the natural



compounds evaluated due to their consistent performance in terms of docking score and protein-ligand interactions. To validate this further, the docked complexes of selected compounds, i.e., apigenin and curcumin with West Nile virus methyltransferase (PDB ID 2OY0) and Ganoderic acid with West Nile virus envelope glycoprotein (PDB ID 2I69), were studied for the complexes' stability and intermolecular interaction with respect to time by 100 ns molecular dynamics (MD) simulation. When simulations were completed, the root mean square deviation (RMSD), root mean square fluctuation (RMSF), the radius of gyration (Rg), H-bond analysis, and principal component analysis (PCA) were determined for each frame of the trajectory.

4.4.1. Study of stability using root mean square deviation

RMSD refers to the occurrence of deviations that were noticed during the evolution of the simulation (Peng et al., 2012; Ghildiyal et al., 2020; Huang et al., 2022). Furthermore, the RMSD defines the stability of the structural as the smaller the RMSD, the greater the stability (Huang et al., 2022). RMSD calculations for protein backbones, ligands, and complexes during a 100 ns MD simulation of each protein-ligand complex were used to get an insight into conformational changes during protein-ligand interactions, as shown in Figure 7.

Derived from a 100 ns MD simulation, the RMSD of the ligand, protein backbone, and complex RMSD values for apigenin, curcumin, and Ganoderic acid are displayed. The RMSD of the protein backbone is displayed in red. The complex RMSD (root-mean-square deviation) is depicted in black. The RMSD (root mean square deviation) of the ligand is depicted in green.

Figures 7A,B illustrate that the West Nile virus methyltransferase protein with apigenin ligand has a lower RMSD value compared to the West Nile virus methyltransferase protein with curcumin ligand. In Figures 7A,B, Ligand's average RMSD is 0.6785 Å and 2.066 Å, respectively. Although the apigenin ligand bound the West Nile virus methyltransferase protein, the low RMSD value showed large fluctuations throughout the simulation. In the case of the ligand curcumin, the fluctuation was just observed during the first 10 ns of trajectory. Furthermore, the ligand RMSD stabilized again after 10 ns and remained consistent till the end of the simulation time of 100 ns. This indicates that the curcumin ligand contacts remained intact during the simulation. According to Figure 7C, the Ganoderic acid ligand-bound West Nile virus envelope glycoprotein showed a stable RMSD value during the simulation time. The trajectories were analyzed with the VMD program, and the results showed that the ligands curcumin and Ganoderic acid did not jump out of the domain of the protein, indicating that they were located within the binding site. This was discovered when we found that neither of the ligands jumped out of the domain. According to the results of the root-mean-square-difference (RMSD) analysis of the ligands, neither the binding orientation of the curcumin nor the Ganoderic acid ligands changed during the simulation. Apigenin and Ganoderic acid have maintained the same conformational orientation throughout their whole structures and have not deviated from them. This suggests that, in the case of the West Nile virus methyltransferase protein-bound Apigenin-ligand, numerous binding orientations were observed and that this ligand shifted positions throughout MD simulations and departed the binding

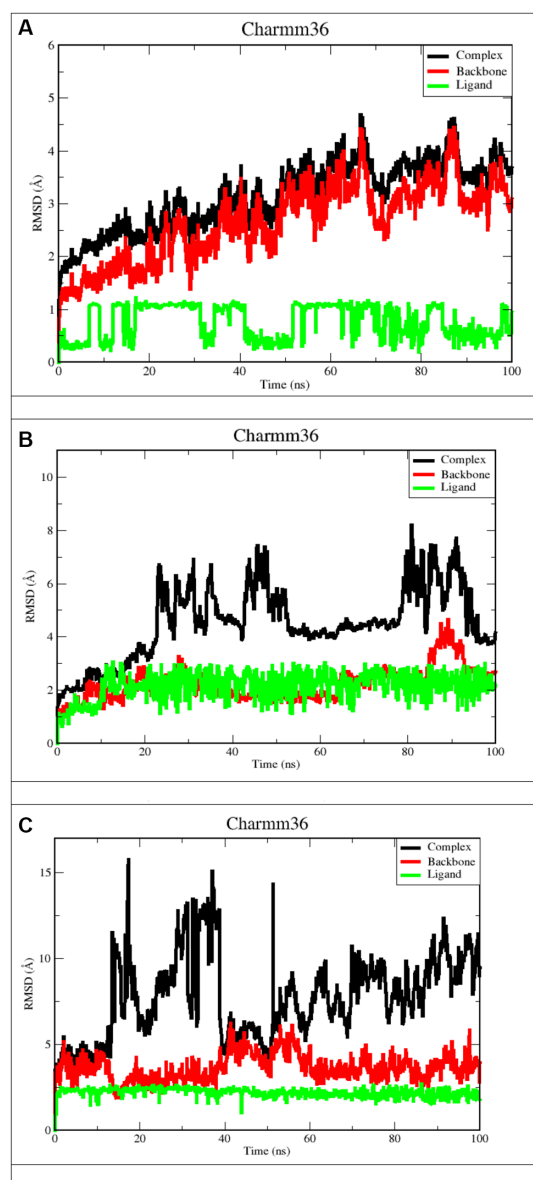


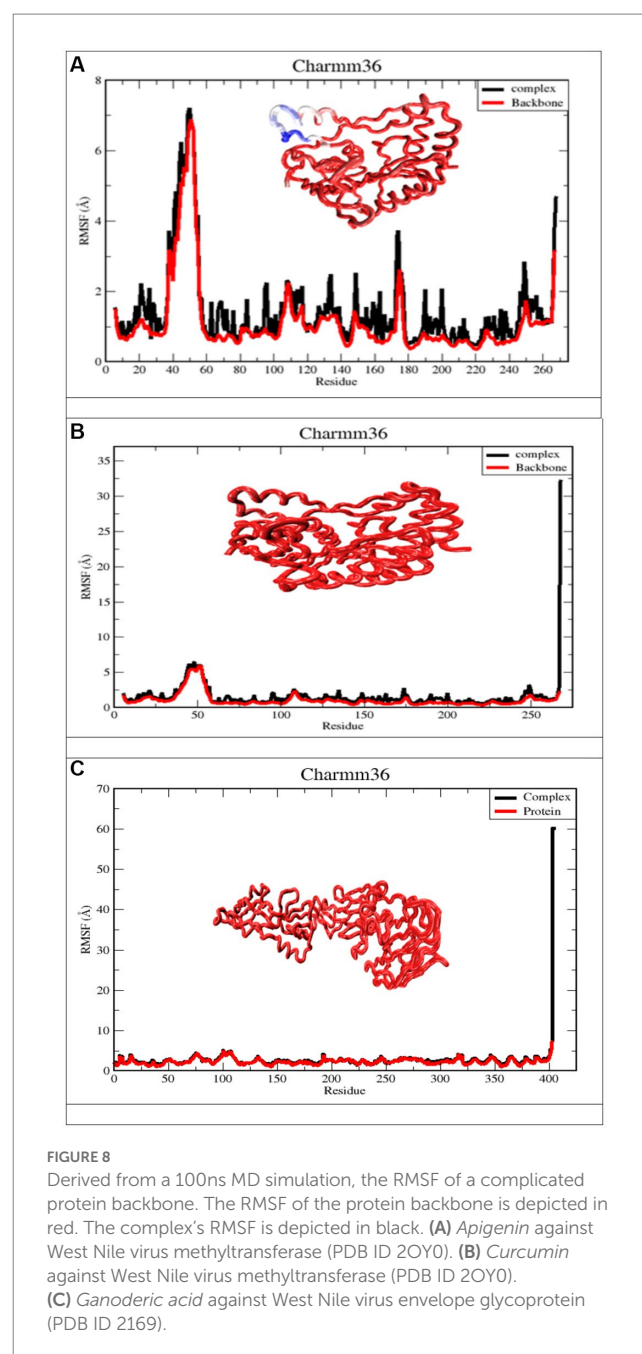
FIGURE 7
RMSD plots after 100 ns run. (A) Apigenin against West Nile virus methyltransferase (PDB ID 2OY0). (B) Curcumin against West Nile virus methyltransferase (PDB ID 2OY0). (C) Ganoderic acid against West Nile virus envelope glycoprotein (PDB ID 2I69).

pocket. The docking position was deemed unacceptable for the candidate with a complex structure. During the 100 ns simulation, the RMSD analysis of backbone atoms for the two proteins, West Nile virus methyltransferase and West Nile virus envelope glycoprotein, revealed different conformational states in terms of the protein's backbone. Indeed, curcumin and Ganoderic acid tend to reach a constant equilibrium, but the RMSD of the apigenin backbone was considerably high. The backbone of Apigenin remained distinct throughout the simulation, resulting in a maximum RMSD of 4.5 Å. This difference in the deviation range explains the change in stability of West Nile virus methyltransferase, which represents the effect of the substituted amino acid on the structure of the protein. The RMSD plots of Apigenin and

Curcumin's backbones have average values of 2.91 and 1.76, respectively. The first 30 ns of the simulation of curcumin showed instability of the protein, but from 30 to 60 ns of simulation, the protein was stable. The first stable conformation of the Ganoderic acid backbone occurs between 20 and 40 nanoseconds, while the second conformation occurs between 60 and 100 nanoseconds. The RMSD remains constant at 1.7 Å, and between 20 ns and 40 ns, RMSD >5 Å is observed to fluctuate significantly. Among the three trajectories, the apigenin backbone protein exhibited the greatest variation and the highest RMSD value. As opposed to Apigenin and Curcumin, the Ganoderic acid protein backbone exhibited less divergent patterns. In addition, according to Figure 7, the average RMSD values for the three complexes were 3.50 Å, 3.97 Å, and 5.50 Å, respectively. As determined by RMSDs, the protein-ligand complex in Figure 7A exhibited a rising fluctuation, whereas in Figure 7B, the complex is equilibrated between 50 and 75 ns. Also, as can be seen in the complicated Figure 7C, the trajectory is initially in equilibrium for 5 ns, then has a big increase to 16 at 20 ns, and finally experiences a substantial reduction to 5.3 at roughly 23 ns. From then on, the complex exhibits a wide range of fluctuations, from nearly stable to extremely unstable, until it reaches 70 ns. In the last 30 ns, the complex has shown persistent, minute fluctuations. As a result, the West Nile virus methyltransferase complex protein-bound apigenin exhibited the lowest RMSD value compared to the curcumin compound and Ganoderic acid compound complexes; however, this did not prove its greater stability and fewer conformational changes compared to other complexes. At the conclusion of the simulation, the ligand appeared highly unstable when complexed with apigenin. The apigenin ligand exhibited substantial variations, indicating the general instability of the complex.

4.4.2. Root mean square fluctuations

RMSF measures the amount by which atomic locations have deviated from their initial positions and illustrates the dynamic nature of the protein-ligand interaction. In other words, RMSF exemplifies how dynamic the interaction between the protein and the ligand is. As a result, the RMSF data of the protein backbone and complexes were plotted to visualize the average fluctuation of all the amino acid residues, as depicted in Figure 8 for a 100 ns MD trajectory. Furthermore, the RMSF value can be used to assess the significance of individual protein residues in preserving the native shape of a protein-ligand complex. A high RMSF number indicates greater flexibility, while a low RMSF value indicates a more stable zone. Hence, a higher number of RMSF residues or groups suggests a higher degree of flexibility, which in turn suggests a higher probability of interaction with ligand molecules. Moreover, reduced RMSF fluctuations are associated with lesser flexibility, resulting in diminished interaction potential. Figure 8A depicts the RMSF analysis for the ligand apigenin. None of the nine residues with a high RMSF value (RMSF value >3) were outside of protein binding sites. In addition, fluctuations of substantial peaks in the RMSF graph were found at LYS45, GLU46, GLY47, ASN48, VAL49, THR50, GLY51, GLY52, and HSD53, which are shown in blue on the protein. Intriguingly, these residues are not part of the binding pocket, and it is hypothesized that they have no significant effect on the ligand-binding process. The curcumin ligand has four different residues that have high



RMSF, but none of them are in the protein binding sites. The Ganoderic acid ligand has not exhibited significant fluctuations with low RMSF values, none of which are in protein-binding sites. As demonstrated in Figure 8A, the RMSF values of the apigenin ligand complex are greater than those of the curcumin structure. The high RMSF values of apigenin indicated a greater degree of flexibility and instability in the protein, whereas the low RMSF values of the curcumin and Ganoderic acid complexes indicated restricted movement of the residues and a rigid structure in the presence of the ligands for West Nile virus methyltransferase. According to the information presented above, the findings of curcumin and Ganoderic acid have demonstrated that, in general, the protein was stable and did not undergo any significant changes. In this investigation, apigenin and protein interactions

with RMSF values greater than 2.5 are regarded as having fewer stable bonds. Figure 8, on the other hand, reveals a negligible distinction between the fluctuation scores of the residues in the protein-curcumin combination and the protein-Ganoderic acid complex. In Figures 8B,C, the RMSF plot of ligand-protein for each ligand reveals that the RMSF value at the C-terminal residues (end) is quite high because these residues represent the tails or ends of the protein structure, which are highly reactive and free to move. Each of the two complex system residues, VAL267 (C-terminal), exhibits significantly higher fluctuations than other residues. This is likely because it is placed at the end of the protein chain, allowing it to be more flexible than other residues. In Figure 8C, the C-terminal residue changed significantly, with RMSF values of 60, whereas for curcumin, it reached 30. Figure 8C depicts the creation of a highly mobile free-end loop in Ganoderic acid, which may account for the significant variation observed.

4.4.3. Radius of gyration analysis

To evaluate structural compression changes, the gyration radius diagram of each structure was recorded over the course of the simulation. To determine the compactness of the system over time, Rg was calculated, with higher Rg values indicating less compactness (more unfolded) with conformational entropy and lower Rg values indicating high compactness with greater structural stability (more folded). Figure 9 demonstrates that the simulation Rg values for the three complexes are 1 Å, 9 Å, 1 Å, 975 Å and 3.5 Å, respectively. From the Apigenin plot, it can be seen that the system exhibited little change until 100 ns, with the exception of a few unfolding events between 40 and 70 ns, after which it stabilized until the end of the simulation. The stability of the protein in the complex was demonstrated by the Rg value, which showed less variation as a result of the fewer changes it experienced. Curcumin, on the other hand, exhibits only very modest fluctuations at the beginning but afterwards demonstrates a significant increase that reaches a Rg value of 2.5 Å between 20 and 40 ns. In the case of Ganoderic acid, the structure showed a sharp increase of Rg up to 3.75 Å at 10 ns, then an immediate decrease of the Rg value towards 3.50 Å, then an increase towards 3.70 Å between 20 ns and 40 ns, then a gradual

fluctuation decreases of Rg towards 3.60 Å until the 80 ns. At 80 ns, there is a sharp decrease in the Rg value towards 3.50 Å. After that, an average Rg value of 3.60 Å was equilibrated to small fluctuations until the simulation ended. The RG findings demonstrated that the binding of these three molecules induces structural modifications. In general, the patterns of change in RG values across all complexes were distinct. According to the results, the Apigenin complex had the smallest Rg value, which may indicate that it is more compact than the other complexes. A loss of compactness may occur as a consequence of a change in the interaction pattern between the protein and the ligand. The dissolution of hydrogen bonds between molecules might possibly be explained by the curcumin complex's elevated Rg value. It is also possible that conformational changes in the protein structure were induced by the interaction of curcumin with proteins, which dramatically altered the curcumin complex's microenvironment. In addition, the Ganoderic acid complex has been reported to have experienced compositional variations, which indicates a less tightly bound structure.

4.4.4. Hydrogen bond analysis

It is worthwhile to investigate hydrogen bonding interactions because they contribute to the binding process and, in particular, persistent hydrogen bonds. Using the VMD hydrogen bond analysis tool, all conceivable hydrogen bonding interactions between the two specified areas, in this case the protein and the ligand, have been explored throughout time. Hydrogen bonds are recognized, and the output contains the overall number of hydrogen bonds as well as their occupancy over time. The 'Percentage occupancy of the Hbond' output of the hydrogen bond analysis tool provides access to these values (Figure 10). In docked complexes, significantly more stable hydrogen bonds are reported to be established. During the Molecular Dynamics (MD) simulations, the H-bonds that were present in the docking structures were not only preserved, but additional H-bonds were also found. Apigenin, curcumin, and Ganoderic acid each have their own individual occupancies of identified H-bonds, which are listed in Table 4.

In the case of the apigenin complex, during MD, the protein-apigenin complex stability was maintained by interactions with

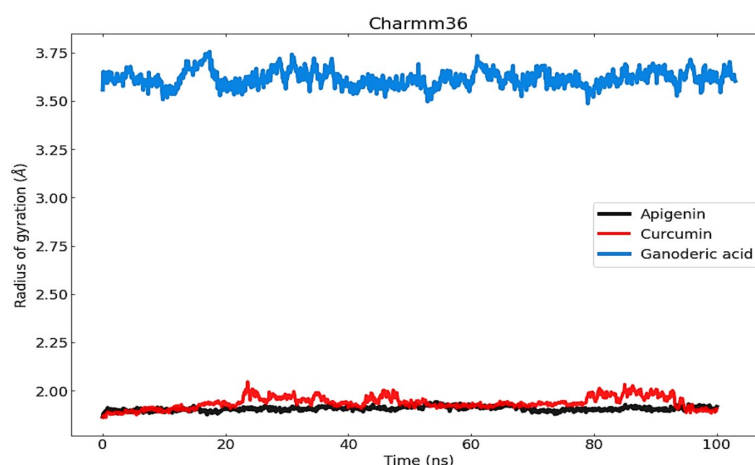


FIGURE 9

The radius of gyration (Rg) was used to measure and comprehend the compactness of protein complexes and their structure.

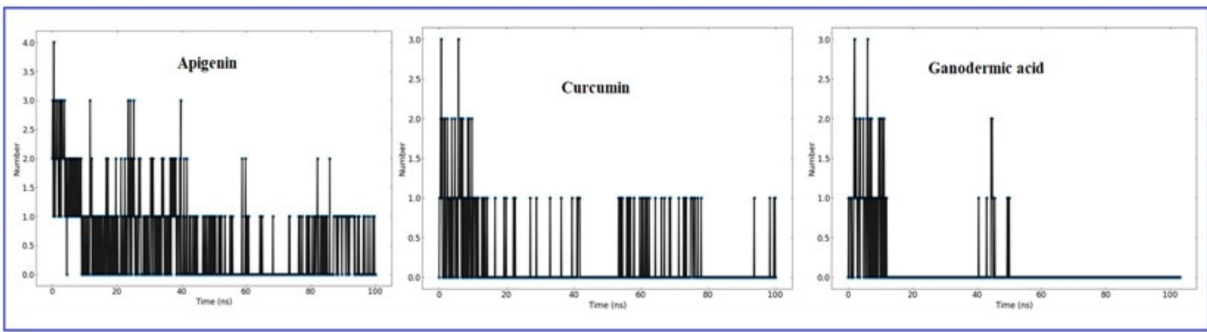


FIGURE 10
The number of H-bonds formed by the ligand molecule, with the proteins (Apigenin, Curcumin and Ganoderic acid) was obtained from 100ns MD simulation.

TABLE 4 Analysis of H-bond occupancies for each ligand during MD simulation.

	Donor acceptor	Occupancy
Apigenin	Apigenin-Side-O3 GLY81-Main-O	0.90%
	Apigenin-Side-O3 ILE147- Main-O	0.10%
	LYS105-Main-N Apigenin-Side-O2	11.28%
	LYS105-Main-N Apigenin-Side-O5	3.19%
	LYS105-Side-N2 Apigenin-Side-O2	10%
Curcumin	Curcumin-Side-O5 ASP131-Side-OD1	40%
	Curcumin-Side-O3 ASP131-Side-OD2	40%
	Curcumin-Side-O3 ASP131-Side-OD1	1.20%
	VAL132-Main-N Curcumin-Side-O3	80%
Ganoderic acid	Ganoderic-Side-O7 ILE340-Main-O	10%
	Ganoderic-Side-O6 ILE340-Main-O	19%
	Ganoderic-Side-O5 PRO339-Main-O	19%
	SER341-Side-OG Ganoderic-Side-O4	19%
	SER341-Side-OG Ganoderic-Side-O5	29%

GLY 81, LYS 105, and ILE 147 residues. It has hydrogen bonds with the highest occupancy of 11.28%, where apigenin forms the hydrogen bond with LYS 105. Then 0.90% occupancy, where apigenin plays a role as an acceptor for GLY 81, and 0.10% with

ILE 147. Although the Apigenin complex has 50 hydrogen bonds more than the other complexes, the compound (Apigenin) formed only two hydrogen bonds, which were found in docked simulation. Curcumin has 28 total hydrogen bonds, with the highest occupancy of 80%, where curcumin and VAL 132 form hydrogen bonds. Then 40% occupancy, where curcumin plays a role as an acceptor of ASP 131, and 1.20% with ASP 131. Only three of the 28 hydrogen bonds were formed in the docked simulation. We conclude that, during MD, the curcumin complex has more stability than the apigenin complex. Ganoderic acid has 32 total hydrogen bonds with the highest occupancy of 29%, where Ganoderic acid has the role of a donor and SER 341 has the role of an acceptor of hydrogen bond formation. Then 19% occupancy where Fisetin plays a role as an acceptor to ILE 340 and 19% with PRO 339. Only two of the 32 hydrogen bonds were formed in the docked simulation. In the system, hydrogen bonds were stable, and most of them appeared between 0 and 15 ns and 40 and 50 ns. Hydrogen bond occupancy with a score of more than 100% indicates that more than one atom pair interacts to form hydrogen bonds. Meanwhile, curcumin has a lower total hydrogen bond but a higher hydrogen bond occupancy than Ganoderic acid. This total hydrogen bond and hydrogen bond occupancy determine the stability of each system. The number and occupancy of hydrogen bonds are the keys to the interaction stabilization of the protein-ligand complex.

4.4.5. Principal component analysis (PCA) and dynamics cross-correlation matrices (DCCM) analysis

Principal component analysis was used to identify and understand significant concerted motions in different regions of the protein. PCA extracts the most variable dynamic motions in simulations, which are required for biological function. In addition, it may be used to examine the effect of various parameters on the collective motion and to reduce the motion's complexity, which is related to the system's stability and protein functions. Also, it can be utilized to characterize the many conformational variations that are associated with the process of protein folding as well as the open-close mechanism of ion channels. The conformational alterations of West Nile virus methyltransferase and West Nile virus envelope

glycoprotein were further analyzed using dynamic cross correlation matrix (DCCM) analysis. Principal component analysis (PCA) and a DCCM were used to identify protein backbone conformation shifts. PCA and DCCM were performed using RStudio and Bio3d; axes, each dot represents a different protein configuration. The distribution of blue and red dots represented the extent of

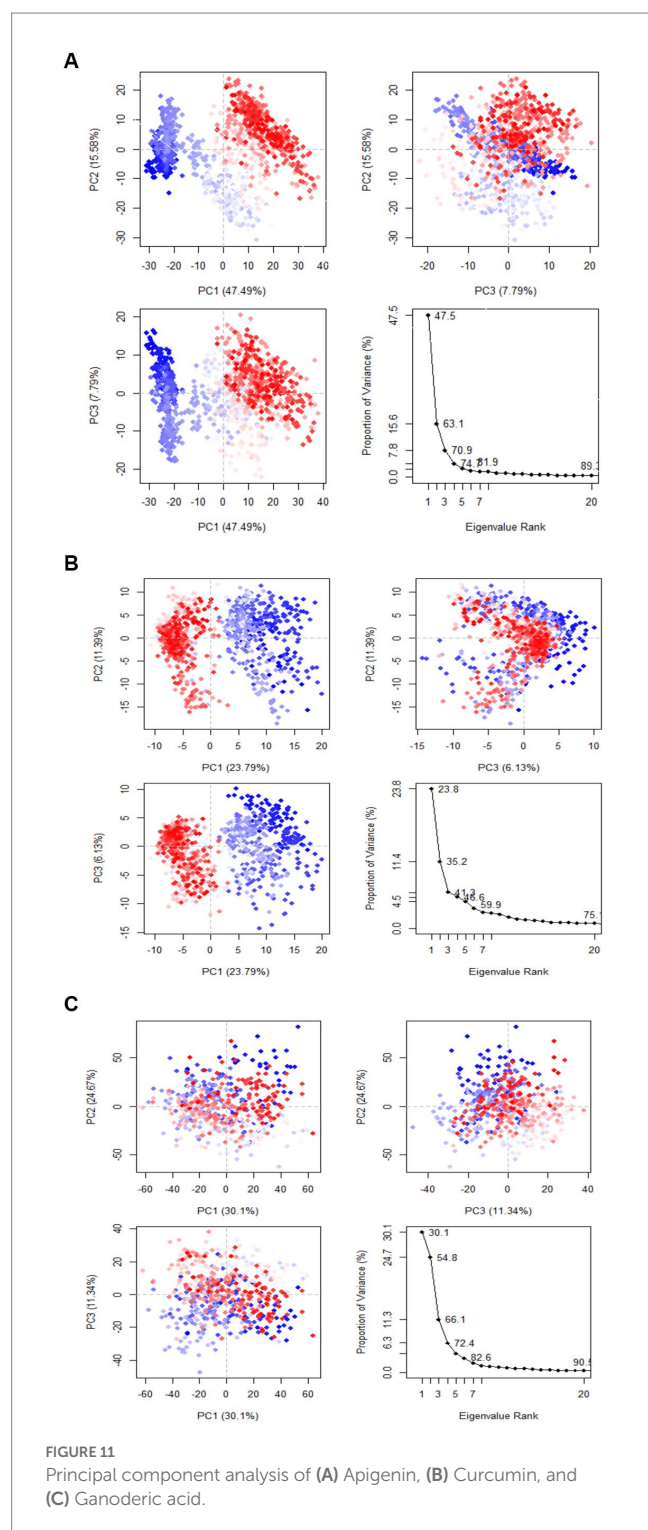
conformational changes in the simulation, where the color ranges from blue to white to red corresponds to simulation duration. The color blue represents the initial timestep, the color white represents the intermediate, and the color red represents the final timestep (illustrated in Figure 11).

The first three PCs predicted the majority of the motion of the protein backbone from the MD trajectories in Figure y. In the apigenin protein, PCA analysis shows that the first three eigenvectors account for 47.49, 15.59, and 7.79% (Figure 12A). In the case of curcumin protein, PCA analysis display that the first three eigenvectors account for 23.79, 11.39, and 6.13% (Figure 12B). For Ganoderic Acid Protein, PCA analysis display that the first three eigenvectors account for 30.1, 24.67, and 11.34% (Figure 12C). The highest PC1 (47.49%) was noticed for the apigenin protein, which indicates that the protein had undergone higher conformational changes. The lowest PC1 (6.13%) was observed for the curcumin protein, indicating that the protein had undergone very few conformational changes compared to apigenin. Comparatively, the Ganoderic acid compound exhibited less conformational change than the apigenin complex but more than the curcumin complex.

Additionally, the resulting dynamical cross-correlation graphs exhibit both positive and negative amino acid correlation effects. Overall correlation was displayed by DCCM, and it was in the range of -1.0 to 1.0 . (From dark purple to dark blue). Various colors were used to denote varying degrees of association between residues, with darker colors indicating stronger correlations. Correlations closer to 1 indicated that the residues were moving in the same direction, while correlations closer to -1 indicated that the residues were moving in the opposite direction. In order to see the relationship between the I and J residue indices, we constructed pairwise correlated graphs. Colors such as dark cyan, white, and pink were used to analyze the predicted map results. Fully correlated pairs are denoted by the color cyan, while anti-correlated pairs are denoted by the color pink. Comparative results reveal that the atomic motions in the Apigenin complex resemble more closely those of the Curcumin structure. But, Comparing the DCCM diagrams of the three systems, it could be found that the correlated motions of the apigenin and curcumin systems were noticeably distinct from the Ganoderic complex. We postulate that the simultaneous appearance of positively and negatively correlated movements destabilizes the domain. For this reason, we observed high anti-correlation in the apigenin complex compared to the curcumin complex, indicating a more compact structure of the apigenin. For the Ganoderic acid complex, which showed less anticorrelated and correlated motions and more noncorrelated motions than the Apigenin and Curcumin complexes in all regions of the protein (illustrated in Figure 12).

4.4.6. Binding free energy analysis

To analyze the molecular binding interaction of protein-ligand complexes, the binding free energy (G) was calculated using MM-PBSA, which considers both bonded and non-bonded (van der Waals and electrostatic) interactions. Using MMPBSA and the final 20 ns of the trajectory, the binding free energies of the protein-ligand complexes were calculated. Using the MMGBSA approach, the binding free energy (ΔG_{bind}) of 12 natural compounds, of which



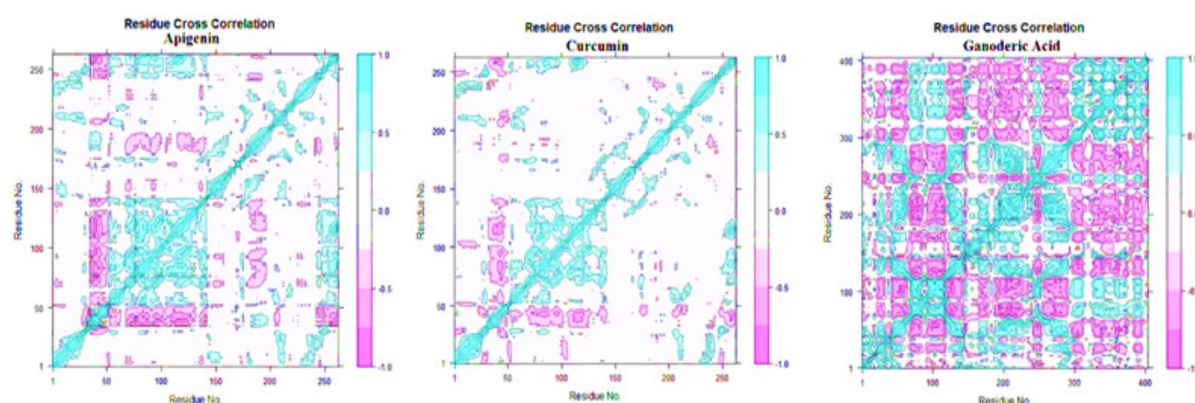


FIGURE 12
Dynamic cross correlation matrix (DCCM) plots for (Apigenin, Curcumin, and Ganoderic Acid).

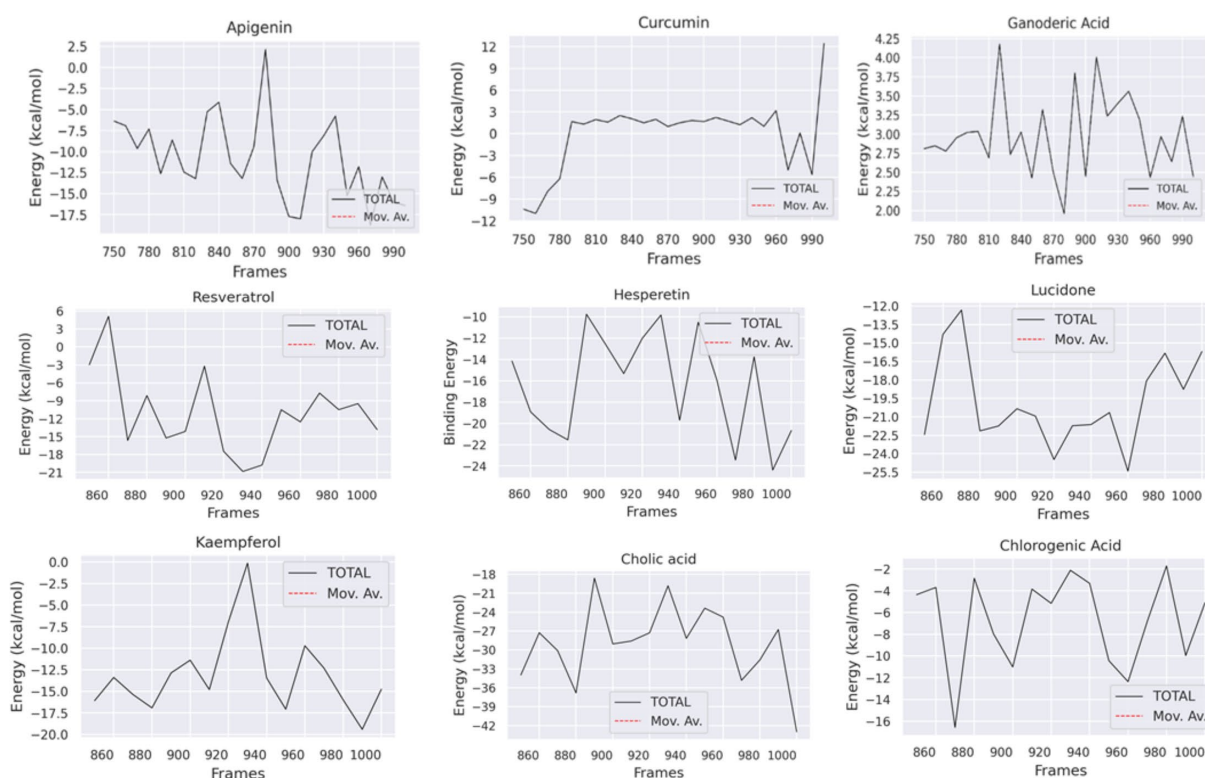


FIGURE 13
Binding free energy plot.

top nine were chosen based on binding affinity score, was determined. The greater the negative values, the more favorable the binding free energy between proteins and ligands. As shown in Table 4 and Figure 13, the free binding energies of the natural compounds agree with the molecular docking results. The Hesperetin, Lucidone, and Cholic acid demonstrated the greatest amount of binding energy (-16.44 kcal/mol, -19.78 kcal/mol and -29.00) in comparison to the other natural chemicals. This energy is mostly different from the curcumin compound, and it means that the interaction in the compound Hesperetin, Lucidone, and Cholic acid complex is the

most stable interaction during simulation, followed by the protein-compound complex. On the other hand, the compound Ganoderic acid complex has the most positive binding energy, indicating that the interaction between compound Ganoderic acid and the target protein active site is the weakest.

Table 5 and Figure 13 highlights the primary components of binding free energy for protein-ligand complexes, including van der Waals (ΔE_{VDW}), electrostatics (ΔE_{EEL}), a polar portion of solvation (ΔG_{PB}), non-polar part of solvation (ΔG_{NP}), dispersion (ΔG_{DISP}), and binding energy (ΔG).

TABLE 5 Binding energy results.

		Δ EVDW (kJ/mol)	Δ EEEL (kJ/mol)	Δ GPB (kJ/mol)	Δ GNP (kJ/mol)	Δ GDISP (kJ/mol)	Δ G Binding (kJ/mol)
01	Apigenin	−21.76	−8.06	21.93	−2.98	0.0	−10.88
02	Resveratrol	−21.27	−19.65	32.80	−2.93	0.0	−11.05
03	Hesperetin	−25.87	−8.51	21.13	−3.19	0.0	−16.44
05	Lucidone	−25.46	−16.61	25.38	−3.09	0.0	−19.78
06	Ganoderic acid	−0.00	0.00	3.05	−0.06	0.0	2.99
07	Curcumin	−0.00	−1.43	6.26	−0.80	0.0	−0.05
08	Kaempferol	−22.78	−16.56	29.34	−3.14	0.0	−13.15
09	Cholic acid	−15.82	−277.93	267.47	−2.72	0.0	−29.00
10	Chlorogenic acid	−33.86	−38.33	69.52	−4.03	0.0	−6.71

4.5. ADME, aquatic and non-aquatic toxicity

By utilizing the pkCSM web server, ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties were performed. The Table 6 displays the reported compound's ADMET values. Drug development fails in between 40 and 60% of cases as a result of inadequate ADMET characteristics. The virtual screening process should take this pharmacokinetics characteristic into consideration as a major factor. The range of the water solubility log S has been reported −2.449 to −6.618. All compounds had significant levels of intestinal absorption in humans excluding 10. The VDss level varies from −2.172 to 0.822, with a maximum Total Clearance rate of 0.653 mL/min/kg. 4,5,6,11 and 12 number compounds have the permeability to the Blood–brain barrier 1,2,3,7,8,9 and 10 compounds cannot pass the Blood–brain barrier. The CYP450 A2 substrate can be inhibited by most of the compounds. The CYP450 2C9 substrate can only be inhibited by 12 specific drugs; other compounds are inactive. Both of these enzymes are crucial for the metabolism of drugs and are frequently present in the human liver. In the disposition and renal clearance of mostly cationic drug molecules, renal OCT2 substrate in the kidney plays a very significant function. Finally, they all are free from AMES toxicity, Skin Sensitization and Hepatotoxicity (Table 6).

5. Conclusion

Our *in-silico* research has been reported that all the natural molecules have better binding affinity, high solubility in aqueous system, free from hepatic toxicity, and skin sensitization, most of them are accepted by lipinski rule. Besides, different types of active amino acid are seen during the formation of drug protein complexes such as VAL-A:132, ASP-A:131, PHE-A:133, LYS-A:105, ASP-A:146, TRP-A:87, GLY-A:58, ASP. The most active compounds were reported Apigenin against *West Nile virus methyltransferase* (PDB ID 2OY0), and Fungisterol, & Sanguinarine against *West Nile virus envelope glycoprotein* (PDB ID 2I69) with maximum binding affinity −8.1.3 kcal/mol, and −8.1 kcal/mol. The drug-likeness properties, and the theoretical ADMET data is accepted by our reported phytochemicals. After that, their stability is confirmed by molecular dynamic

simulation at 100 ns which is also confirmed that the molecules are highly stable when form protein ligands complex. As, there is currently neither an antiviral medicine nor a vaccination that can treat WNV infection in people. In the course of drug discovery and repurposed research, various potential natural molecules have been studied in this investigation that capable to inhibit WNV *in silico* model; however, none of these candidates have performed it to the stage of clinical assessment. Now, further experimental studies should be conducted and determine their practical value to establish them as potential drug candidate for further use.

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/s.

Author contributions

SA, IB, MR, NM, SM, MI, SR, AG, SA-H, MZ, VJ, SS, JB, and RS: data collection, conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Deanship of Scientific Research, Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia.

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.

TABLE 6 Summary of *in silico* ADMET prediction.

S/N	Absorption			Distribution		Metabolism		Excretion		Toxicity		
	Water solubility Log S	Caco-2 Permeability $\times 10^{-6}$	Human Intestinal Absorption (%)	VDss (human)	BBB Permeability	CYP450 1A2 Inhibitor	CYP450 2C9 Substrate	Total Clearance (ml/min/kg)	Renal OCT2 substrate	AMES toxicity	Skin Sensitization	Hepatotoxicity
01	−3.329	1.007	93.25	0.822	No	Yes	No	0.566	No	No	No	No
02	−3.178	1.17	90.935	0.296	No	Yes	No	0.076	No	Yes	No	No
03	−3.407	0.294	70.277	0.746	No	No	No	0.444	No	No	No	No
04	−6.818	1.205	94.757	0.29	Yes	No	No	0.565	No	No	No	No
05	−2.701	1.144	95.474	−0.125	Yes	Yes	No	0.11	No	Yes	No	No
06	−3.058	2.625	66.348	−2.172	Yes	Yes	No	−0.369	No	Yes	No	No
07	−2.892	1.643	78.45	0.011	No	Yes	No	−57.293	No	Yes	No	No
08	−3.04	0.032	74.29	1.274	No	Yes	No	0.477	No	No	No	No
09	−3.763	0.597	61.546	−0.804	No	No	No	0.653	No	No	No	No
10	−2.449	−0.84	36.377	0.581	No	No	No	0.307	No	No	No	No
11	−3.538	1.152	92.417	−0.386	Yes	Yes	Yes	0.122	No	No	No	No
12	−2.892	0.239	82.294	0.011	Yes	Yes	No	−0.416	No	Yes	No	No

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

- Abba, Y., Hassim, H., Hamzah, H., and Noordin, M. M. (2015). Antiviral activity of resveratrol against human and animal viruses. *Adv. Virol.* 2015, 1–7. doi: 10.1155/2015/184241
- Agarwal, S., and Mehrotra, R. (2016). An overview of molecular docking. *JSM Chem.* 4, 1024–1028.
- Agrawal, P. K., Agrawal, C., and Blunden, G. (2021). Pharmacological significance of hesperidin and hesperetin, two citrus flavonoids, as promising antiviral compounds for prophylaxis against and combating COVID-19. *Nat. Prod. Commun.* 16. doi: 10.1177/1934578X211042540
- AI Azzam, K. (2023). SwissADME and pkCSM webserver predictors: an integrated online platform for accurate and comprehensive predictions for in silico ADME/T properties of artemisinin and its derivatives. *Eng. Technol.* 325, 14–21. doi: 10.31643/2023/6445.13
- Aini, N. S., Kharisma, V. D., Widyananda, M. H., Murtadlo, A. A. A., Probojati, R. T., Turista, D. R., et al. (2022). In silico screening of bioactive compounds from *Syzygium cumini* L. and *moringa oleifera* L. against SARS-CoV-2 via tetra inhibitors. *Pharm. J.* 14, 267–272. doi: 10.5530/pj.2022.14.95
- Akash, S. (2022). Computational screening of novel therapeutic and potent molecules from bioactive trehalose and its eight derivatives by different insilico studies for the treatment of diabetes mellitus. *Organic Commun.* 15, 288–296. doi: 10.25135/acg.oc.134.2204.2446
- Akash, S., Kumer, A., Chandro, A., Chakma, U., and Matin, M. M. (2022). Quantum calculation, docking, ADMET and molecular dynamics of ketal and non-ketal forms of D-glucofuranose against bacteria, black & white fungus, and triple-negative breast cancer. *Biointerface Res. Appl. Chem.* 13:374.
- Baba, M., Schols, D., Nakashima, H., Pauwels, R., Parmentier, G., Meijer, D. K., et al. (1989). Selective activity of several cholic acid derivatives against human immunodeficiency virus replication in vitro. 2, 264–271.
- Bergmann, F., Trachsel, D. S., Stoeckle, S. D., Bernis Sierra, J., Lübke, S., Groschup, M. H., et al. (2022). Seroepidemiological survey of West Nile virus infections in horses from Berlin/Brandenburg and North Rhine-Westphalia, Germany. *Viruses* 14:243. doi: 10.3390/v14020243
- Boh, B., Berovic, M., Zhang, J., and Zhi-Bin, L. (2007). Ganoderma lucidum and its pharmaceutically active compounds. *Biotechnol. Annu. Rev.* 13, 265–301. doi: 10.1016/S1387-2656(07)13010-6
- Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., and Velankar, S. J. P. C. (2017). Protein Data Bank (PDB): the single global macromolecular structure archive. *Methods Mol. Biol.*, 627–641. doi: 10.1007/978-1-4939-7000-1_26
- Campagna, M., and Rivas, C. (2010). Antiviral activity of resveratrol. *Biochem. Soc. Trans.* 38, 50–53. doi: 10.1042/BST0380050
- CDC. (n.d.). Treatment & Prevention. Available at: <https://www.cdc.gov/westnile/healthcareproviders/healthCareProviders-TreatmentPrevention.html> (Accessed November 9, 2021).
- Chen, W.-C., Tseng, C.-K., Lin, C.-K., Wang, S.-N., Wang, W.-H., Hsu, S.-H., et al. (2018). Lucidone suppresses dengue viral replication through the induction of heme oxygenase-1. *Virulence* 9, 588–603. doi: 10.1080/21505594.2017.1421893
- Chen, W.-C., Wang, S.-Y., Chiu, C.-C., Tseng, C.-K., Lin, C.-K., Wang, H.-C., et al. (2013). Lucidone suppresses hepatitis C virus replication by Nrf2-mediated heme oxygenase-1 induction. *Antimicrob. Agents Chemother.* 57, 1180–1191. doi: 10.1128/AAC.02053-12
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7:42717. doi: 10.1038/srep42717
- Ding, Y., Cao, Z., Cao, L., Ding, G., Wang, Z., and Xiao, W. (2017). Antiviral activity of chlorogenic acid against influenza A (H1N1/H3N2) virus and its inhibition of neuraminidase. *Sci. Rep.* 7:45723. doi: 10.1038/srep45723
- Fahmi, M., Kharisma, V. D., Ansori, A. N. M., and Ito, M. (2021). “Retrieval and investigation of data on SARS-CoV-2 and COVID-19 using bioinformatics approach” in *Coronavirus disease-COVID-19* (Cham: Springer), 839–857.
- Farahat, R. A., and Memish, Z. A. (2022). Re-emergence of Marburg virus and monkeypox in the shadow of COVID-19 pandemic: current situation and implications—correspondence. *Int. J. Surg.* 106:106923. doi: 10.1016/j.ijsu.2022.106923
- Feldman, N., Kuryakov, V., Sedyakina, N., Gromovych, T., and Lutsenko, S. V. (2018). Preparation of liposomes containing benzophenanthridine alkaloid sanguinarine and evaluation of its cytotoxic activity. *Int. J. Nanotechnol.* 15, 280–287.
- Ghildiyal, R., Prakash, V., Chaudhary, V., Gupta, V., and Gabrani, R. (2020). “Phytochemicals as antiviral agents: recent updates” in *Plant-derived bioactives* (New York, NY: Springer), 279–295.
- Guan, W.-D., Yang, Z.-F., Liu, N., Qin, S., Zhang, F.-X., and Zhu, Y.-T. (2008). In vitro experimental study on the effect of resveratrol against several kinds of respiroviruses. *Zhong Yao Cai* 31, 1388–1390.
- Guang, H.-M., and Du, G.-H. (2006). Protections of pinocembrin on brain mitochondria contribute to cognitive improvement in chronic cerebral hypoperfused rats. *Eur. J. Pharmacol.* 542, 77–83. doi: 10.1016/j.ejphar.2006.04.054
- Han, J.-Y., Jeong, H. I., Park, C.-W., Yoon, J., Ko, J., Nam, S.-J., et al. (2018). Cholic acid attenuates ER stress-induced cell death in Cocksackievirus-B3 infection. *J. Microbiol. Biotechnol.* 28, 109–114. doi: 10.4014/jmb.1708.08009
- Hayes, C. G. (2001). West Nile virus: Uganda, 1937, to New York City, 1999. *Ann. N. Y. Acad. Sci.* 951, 25–37. doi: 10.1111/j.1749-6632.2001.tb02682.x
- West Nile Virus Found in Another NY County, Bringing Case Total to 5. (n.d.). Available at: <https://www.nbcnewyork.com/news/local/west-nile-virus-found-in-another-ny-county-bringing-state-case-total-to-5/3858809/>
- Huang, K.-K., Lin, M.-N., Hsu, H.-C., Hsu, Y.-L., Huang, T.-N., Lu, I., et al. (2022). Pinocembrin reduces keratinocyte activation and ameliorates imiquimod-induced psoriasis-like dermatitis in BALB/c mice through the Heme Oxygenase-1/signal transducer and activator of transcription 3 pathway. *Evid. Based Complement. Alternat. Med.* 2022, 2022:7729836. doi: 10.1155/2022/7729836
- Kanai, R., Kar, K., Anthony, K., Gould, L. H., Ledizet, M., Fikrig, E., et al. (2006). Crystal structure of West Nile virus envelope glycoprotein reveals viral surface epitopes. *J. Virol.* 80, 11000–11008. doi: 10.1128/JVI.01735-06
- Kawsar, S., Kumer, A., Munia, N. S., Hosen, M. A., Chakma, U., and Akash, S. J. O. C. (2022). Chemical descriptors, PASS, molecular docking, molecular dynamics and ADMET predictions of glucopyranoside derivatives as inhibitors to bacteria and fungi growth. *Organ. Communicat.* 15, 184–203. doi: 10.25135/acg.oc.122.2203.2397
- Kemmerly, S. A. (2003). Diagnosis and treatment of West Nile infections. *Ochsner J.* 5, 16–17.
- Khazdair, M. R., Anaegoudari, A., and Agbor, G. A. (2021). Anti-viral and anti-inflammatory effects of kaempferol and quercetin and COVID-19: a scoping review. *Asian Pac. J. Trop. Biomed.* 11:327. doi: 10.4103/2221-1691.319567
- Kishu, T., and Siva, K. (2011). Cholic acid as a Lead molecule: a review. *Asian J. Res. Chem.* 4, 683–684. doi: 10.5958/0974-4150
- Kumar, S., Prasad, A., Iyer, S., and Vaidya, S. (2013). Systematic pharmacognostical, phytochemical and pharmacological review on an ethno medicinal plant, *Basella alba* L. *J. Pharmacogn. Phytother.* 5, 53–58. doi: 10.5897/JPP12.0256
- Kumer, A., Chakma, U., Chandro, A., Howlader, D., Akash, S., Kobir, M., et al. (2022). Modified D-glucofuranose computationally screening for inhibitor of breast cancer and triple breast cancer: chemical descriptor, molecular docking, molecular dynamics and QSAR. *J. Chil. Chem. Soc.* 67, 5623–5635. doi: 10.4067/S0717-97072022000305623
- Kumer, A., Chakma, U., Matin, M. M., Akash, S., Chando, A., and Howlader, D. (2021). The computational screening of inhibitor for black fungus and white fungus by D-glucofuranose derivatives using in silico and SAR study. *Organ. Commun.* 14, 305–322. doi: 10.25135/acg.oc.116.2108.2188
- Kumer, A., Chakma, U., Rana, M. M., Chandro, A., Akash, S., Elseehy, M. M., et al. (2022). Investigation of the new inhibitors by sulfadiazine and modified derivatives of α -D-glucopyranoside for white spot syndrome virus disease of shrimp by in silico: quantum calculations, molecular docking, ADMET and molecular dynamics study. *Molecules* 27:3694. doi: 10.3390/molecules27123694
- Kutluay, S. B., Doroghazi, J., Roemer, M. E., and Triezenberg, S. J. J. V. (2008). Curcumin inhibits herpes simplex virus immediate-early gene expression by a mechanism independent of p300/CBP histone acetyltransferase activity. *Virology* 373, 239–247. doi: 10.1016/j.virol.2007.11.028
- Li, Y.-Q., and Wang, S. F. (2006). Anti-hepatitis B activities of ganoderic acid from *Ganoderma lucidum*. *Biotechnol. Lett.* 28, 837–841. doi: 10.1007/s10529-006-9007-9
- Lindequist, U., Niedermeyer, T. H., and Jülich, W.-D. (2005). The pharmacological potential of mushrooms. *Evid. Based Complement. Alternat. Med.* 2, 285–299. doi: 10.1093/ecam/neh107
- Naveed, M., Hejazi, V., Abbas, M., Kamboh, A. A., Khan, G. J., Shumzaid, M., et al. (2018). Chlorogenic acid (CGA): a pharmacological review and call for further research. *Biomed. Pharmacother.* 97, 67–74. doi: 10.1016/j.biopha.2017.10.064
- New York Times. (2022). *Explained: What to know about the West Nile virus in New York City*. New York Times, New York City.

- Oo, A., Hassandarvish, P., Chin, S. P., Lee, V. S., Bakar, S. A., and Zandi, K. J. P. (2016). In silico study on anti-chikungunya virus activity of hesperetin. *PeerJ* 4:e2602.
- Padmi, H., Kharisma, D. V., Ansori, A. N. M., Sibero, M. T., Widyananda, M. H., Ullah, M., et al. (2022). Jurnal-macroalgae bioactive compounds for the potential antiviral of SARS-CoV-2: an in silico study. *J. Pure Appl. Microbiol.* 16, 1018–1027. doi: 10.22207/JPAM.16.2.26
- Peng, L., Yang, S., Cheng, Y. J., Chen, F., Pan, S., Fan, G. J. F. S., et al. (2012). Antifungal activity and action mode of pinocembrin from propolis against *Penicillium italicum*. 21, 1533–1539.
- Piret, J., and Boivin, G. (2021). Pandemics throughout history. *Front. Microbiol.* 11:631736. doi: 10.3389/fmicb.2020.631736
- Rahman, M., Islam, M., Akash, S., Mim, S., Rahaman, M., Bin Emran, T., et al. (2022). In silico investigation and potential therapeutic approaches of natural products for COVID-19: computer-aided drug design perspective. *Front. Cell. Infect. Microbiol.* 12:929430. doi: 10.3389/fcimb.2022.929430
- Rahman, M. M., Karim, M. R., Ahsan, M. Q., Khalipha, A. B. R., Chowdhury, M. R., and Saifuzzaman, M. (2012). Use of computer in drug design and drug discovery: a review. *Int. J. Pharmaceut. Life Sci.* 1. doi: 10.3329/ijpls.v1i2.12955
- Rahman, M. M., Wang, X., Islam, M. R., Akash, S., Supti, F. A., Mitu, M. I., et al. (2022). Multifunctional role of natural products for the treatment of Parkinson's disease: at a glance. *Front. Pharmacol.* 13:976385. doi: 10.3389/fphar.2022.976385
- Raut, J. K., and Adhikari, M. K., *Mushroom: a true super food*. Khumaltar: Nepal Academy of Science & Technology, (2021).
- Sadowski, J., and Kubinyi, H. (1998). A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* 41, 3325–3329.
- Santos, G. B., Ganesan, A., and Emery, F. S. (2016). Oral administration of peptide-based drugs: beyond Lipinski's rule. *ChemMedChem* 11, 2245–2251. doi: 10.1002/cmde.201600288
- Schwarz, S., Sauter, D., Wang, K., Zhang, R., Sun, B., Karioti, A., et al. (2014). Kaempferol derivatives as antiviral drugs against the 3a channel protein of coronavirus. *Planta Med.* 80, 177–182. doi: 10.1055/s-0033-1360277
- Senthil Kumar, K., and Wang, S.-Y., "Pharmacological applications of lucidone: a naturally occurring cyclopentenone," in *Medicinal plants-recent advances in research and development*, Cham Springer, (2016), 273–295.
- Song, N.-N., Zhang, S.-Y., and Liu, C.-X. (2004). Overview of factors affecting oral drug absorption. *Asian J. Drug Metab. Pharmacokinet.* 4, 167–176.
- Sule, W. F., Oluwayelu, D. O., Hernández-Triana, L. M., Fooks, A. R., Venter, M., and Johnson, N. (2018). Epidemiology and ecology of West Nile virus in sub-Saharan Africa. *Parasit. Vectors* 11, 1–10. doi: 10.1186/s13071-018-2998-y
- Sunthamala, N., Suebsamran, C., Khruaphet, N., Sankla, N., Janpirom, J., Khankhum, S., et al. (2020). Sanguinarine and Chelidonine synergistically induce endosomal toll-like receptor and M1-associated mediators expression. *J. Pure Appl. Microbiol.* 14, 2351–2361. doi: 10.22207/JPAM.14.4.13
- Syng-Ai, C., Kumari, A. L., and Khar, A. (2004). Effect of curcumin on normal and tumor cells: role of glutathione and bcl-2. *Mol. Cancer Ther.* 3, 1101–1108.
- Terefe, E. M., and Ghosh, A. (2022). Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of phytochemicals isolated from *Croton dichogamus* against the HIV-1 reverse transcriptase. *Bioinform. Biol. Insights* 16. doi: 10.1177/11779322221125605
- Tilgner, M., and Shi, P.-Y. (2004). Structure and function of the 3' terminal six nucleotides of the West Nile virus genome in viral replication. *J. Virol.* 78, 8159–8171. doi: 10.1128/JVI.78.15.8159-8171.2004
- Wahyuni, D. K., Wacharasindhu, S., Bankeeree, W., Punnapayak, H., Parikesit, A., and Kharisma, V. (2022). Molecular simulation of compounds from n-hexane fraction of *Sonchus oleraceus* L. leaves as SARS-CoV-2 antiviral through inhibitor activity targeting strategic viral protein. *J. Pharm. Pharmacogn. Res.* 10, 1126–1138. doi: 10.56499/jppres22.1489_10.6.1126
- Walters, W. P. (2012). Going further than Lipinski's rule in drug design. *Expert Opin. Drug Discovery* 7, 99–107. doi: 10.1517/17460441.2012.648612
- Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z., 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
- Wijaya, R. M., Hafidzhah, M. A., Kharisma, V. D., Ansori, A. N. M., and Parikesit, A. A. (2021). COVID-19 in silico drug with *Zingiber officinale* natural product compound library targeting the Mpro protein. *Makara J. Sci.* 25:5. doi: 10.7454/mss.v25i3.1244
- Yan, X., Qi, M., Li, P., Zhan, Y., and Shao, H. J. C. (2017). Apigenin in cancer therapy: anti-cancer effects and mechanisms of action. *Cell Biosci.* 7, 1–16.
- Yuan, S., Chan, H. S., and Hu, Z. (2017). Using PyMOL as a platform for computational drug design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 7:e1298. doi: 10.1002/wcms.1298
- Yuan, X., Luo, K., Zhang, K., He, J., Zhao, Y., and Yu, D. (2016). Combinatorial vibration-mode assignment for the FTIR spectrum of crystalline melamine: a strategic approach toward theoretical IR vibrational calculations of triazine-based compounds. *J. Phys. Chem. A* 120, 7427–7433. doi: 10.1021/acs.jpca.6b06015
- Yuliana, D., Bahtiar, F. I., and Najib, A. (2013). In silico screening of chemical compounds from roselle (*Hibiscus Sabdariffa*) as angiotensin-I converting enzyme inhibitor used PyRx program. *ARPN J. Sci. Technol.* 3, 1158–1160.
- Zandi, K., Ramedani, E., Mohammadi, K., Tajbakhsh, S., Deilami, I., Rastian, Z., et al. (2010). Evaluation of antiviral activities of curcumin derivatives against HSV-1 in Vero cell line. *Nat. Prod. Commun.* 5, 1935–1938. doi: 10.1177/1934578X1000501220
- Zandi, K., Teoh, B.-T., Sam, S.-S., Wong, P.-F., Mustafa, M. R., and AbuBakar, S. (2011). Antiviral activity of four types of bioflavonoid against dengue virus type-2. *Virol. J.* 8, 1–11. doi: 10.1186/1743-422X-8-560
- Zhang, W., Qiao, H., Lv, Y., Wang, J., Chen, X., Hou, Y., et al. (2014). Apigenin inhibits enterovirus-71 infection by disrupting viral RNA association with trans-acting factors. *PLoS One* 9:e110429. doi: 10.1371/journal.pone.0110429
- Zhang, W., Tao, J., Yang, X., Yang, Z., Zhang, L., Liu, H., et al. (2014). Antiviral effects of two *Ganoderma lucidum* triterpenoids against enterovirus 71 infection. *Biochem. Biophys. Res. Commun.* 449, 307–312. doi: 10.1016/j.bbrc.2014.05.019
- Zhang, T., Wu, Z., Du, J., Hu, Y., Liu, L., Yang, F., et al. (2012). Anti-Japanese-encephalitis-viral effects of kaempferol and daidzin and their RNA-binding characteristics. *PLoS One* 7:e30259. doi: 10.1371/journal.pone.0030259
- Zhao, X., Cui, Q., Fu, Q., Song, X., Jia, R., Yang, Y., et al. (2017). Antiviral properties of resveratrol against pseudorabies virus are associated with the inhibition of IκB kinase activation. *Sci. Rep.* 7, 1–11.
- Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., et al. (2007). Structure and function of flavivirus NS5 methyltransferase. *J. Virol.* 81, 3891–3903. doi: 10.1128/JVI.02704-06
- Zhou, Y., Yu, Y., Lv, H., Zhang, H., Liang, T., Zhou, G., et al. (2022). Apigenin in cancer therapy: from mechanism of action to nano-therapeutic agent. *Food Chem. Toxicol.* 168:113385. doi: 10.1016/j.fct.2022.113385
- Zhu, J., Song, S., Sun, Z., Lian, L., Shi, L., Ren, A., et al. (2021). Regulation of glutamine synthetase activity by transcriptional and posttranslational modifications negatively influences ganoderic acid biosynthesis in *Ganoderma lucidum*. *Environ. Microbiol.* 23, 1286–1297. doi: 10.1111/1462-2920.15400
- Zikri, A. T., Pranowo, H. D., and Haryadi, W. (2020). Stability, hydrogen bond occupancy analysis and binding free energy calculation from flavonol docked in DAPK1 active site using molecular dynamic simulation approaches. *Indon. J. Chem.* 21, 383–390. doi: 10.22146/ijc.56087
- Zuo, J., Tang, W., and Xu, Y. (2015). "Anti-hepatitis B virus activity of chlorogenic acid and its related compounds" in *Coffee in health and disease prevention* (Amsterdam: Elsevier), 607–613.



OPEN ACCESS

EDITED BY

Arli Aditya Parikesit,
Indonesia International Institute for
Life-Sciences (i3L), Indonesia

REVIEWED BY

Arif Nur Muhammad Ansori,
Airlangga University, Indonesia
Mohammad Rizki Fadhil Pratama,
Universitas Muhammadiyah
Palangkaraya, Indonesia

*CORRESPONDENCE

Nobendu Mukerjee

✉ nabendu21@rkmvcchhara.org

Vetriselvan Subramaniyan

✉ subramaniyan.vetriselvan@monash.edu

Gobinath Ramachawolran

✉ r.gobinath@rcsiucd.edu.my

[†]These authors have contributed equally to this work

RECEIVED 16 April 2023

ACCEPTED 15 June 2023

PUBLISHED 11 July 2023

CITATION

Akash S, Baeza J, Mahmood S, Mukerjee N, Subramaniyan V, Islam MR, Gupta G, Rajakumari V, Chinni SV, Ramachawolran G, Saleh FM, Albadrani GM, Sayed AA and Abdel-Daim MM (2023) Development of a new drug candidate for the inhibition of Lassa virus glycoprotein and nucleoprotein by modification of evodiamine as promising therapeutic agents. *Front. Microbiol.* 14:1206872. doi: 10.3389/fmicb.2023.1206872

COPYRIGHT

© 2023 Akash, Baeza, Mahmood, Mukerjee, Subramaniyan, Islam, Gupta, Rajakumari, Chinni, Ramachawolran, Saleh, Albadrani, Sayed and Abdel-Daim. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development of a new drug candidate for the inhibition of Lassa virus glycoprotein and nucleoprotein by modification of evodiamine as promising therapeutic agents

Shopnil Akash¹, Javiera Baeza^{2,3}, Sajjat Mahmood⁴, Nobendu Mukerjee^{5,6*†}, Vetriselvan Subramaniyan^{7,8*†}, Md. Rezaul Islam¹, Gaurav Gupta^{9,10}, Vinibha Rajakumari¹¹, Suresh V. Chinni^{12,13}, Gobinath Ramachawolran^{14*}, Fayez M. Saleh¹⁵, Ghadeer M. Albadrani¹⁶, Amany A. Sayed¹⁷ and Mohamed M. Abdel-Daim^{18,19}

¹Department of Pharmacy, Faculty of Allied Health Sciences, Daffodil International, University, Dhaka, Bangladesh, ²Center for Bioinformatics and Molecular Simulation, Universidad de Talca, Talca, Chile, ³Millennium Nucleus of Ion Channels-Associated Diseases (MiNICAD), Universidad de Chile, Santiago, Chile, ⁴Department of Microbiology, Jagannath University, Dhaka, Bangladesh, ⁵Department of Microbiology, West Bengal State University, West Bengal, Kolkata, India, ⁶Department of Health Sciences, Novel Global Community Educational Foundation, Hebersham, NSW, Australia, ⁷Pharmacology Unit, Jeffrey Cheah School of Medicine and Health Sciences, MONASH University, Jalan Lagoan Selatan, Bandar Sunway, Selangor, Malaysia, ⁸Center for Transdisciplinary Research, Department of Pharmacology, Saveetha Institute of Medical and Technical Sciences, Saveetha Dental College and Hospital, Saveetha University, Chennai, Tamil Nadu, India, ⁹School of Pharmacy, Suresh Gyan Vihar University, Jagatpura, Jaipur, India, ¹⁰Uttaranchal Institute of Pharmaceutical Sciences, Uttaranchal University, Dehradun, India, ¹¹Faculty of Foundation, MAHSA University, Selangor, Malaysia, ¹²Department of Biochemistry, Faculty of Medicine, Bioscience, and Nursing, MAHSA University, Selangor, Malaysia, ¹³Department of Periodontics, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India, ¹⁴Department of Foundation, RCSI & UCD Malaysia Campus, Pulau Pinang, Malaysia, ¹⁵Department of Medical Microbiology, Faculty of Medicine, University of Tabuk, Tabuk, Saudi Arabia, ¹⁶Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, ¹⁷Department of Zoology, Faculty of Science, Cairo University, Giza, Egypt, ¹⁸Department of Pharmaceutical Sciences, Pharmacy Program, Batterjee Medical College, Jeddah, Saudi Arabia, ¹⁹Department of Pharmacology, Faculty of Veterinary Medicine, Suez Canal University, Ismailia, Egypt

The Lassa virus (LASV), an RNA virus prevalent in West and Central Africa, causes severe hemorrhagic fever with a high fatality rate. However, no FDA-approved treatments or vaccines exist. Two crucial proteins, LASV glycoprotein and nucleoprotein, play vital roles in pathogenesis and are potential therapeutic targets. As effective treatments for many emerging infections remain elusive, cutting-edge drug development approaches are essential, such as identifying molecular targets, screening lead molecules, and repurposing existing drugs. Bioinformatics and computational biology expedite drug discovery pipelines, using data science to identify targets, predict structures, and model interactions. These techniques also facilitate screening leads with optimal drug-like properties, reducing time, cost, and complexities associated with traditional drug development. Researchers have employed advanced computational drug design methods such as molecular docking, pharmacokinetics, drug-likeness, and molecular dynamics simulation to investigate evodiamine derivatives as potential LASV inhibitors. The results revealed remarkable binding affinities, with many outperforming standard compounds. Additionally, molecular active simulation data suggest stability when bound to target receptors. These promising

findings indicate that evodiamine derivatives may offer superior pharmacokinetics and drug-likeness properties, serving as a valuable resource for professionals developing synthetic drugs to combat the Lassa virus.

KEYWORDS

Lassa fever virus, emerging viral infections, drug discovery, ADMET, molecular docking, molecular dynamics simulation, evodiamine

1. Introduction

The Lassa virus (LASV), an Arenavirus family member, is responsible for causing hemorrhagic fever and multiple organ failure. Transmission to humans occurs through rodents and human-to-human contact and is considered endemic in several West African countries, including Sierra Leone, Liberia, Guinea, and Nigeria (Agbonlahor et al., 2021). The first documented case of the Lassa virus dates back to 1969 in Nigeria, and annual reports suggest an incidence of 100,000–300,000 cases, resulting in approximately 5,000 fatalities (CFDCA Prevention, 2022). However, these numbers are rough estimates due to regional variations in Lassa fever surveillance. In Sierra Leone and Liberia, Lassa fever accounts for 10–16% of hospitalized patients with the infection (CFDCA Prevention, 2022). From week 1 to week 52 of 2022, at least 1,067 LASV cases were diagnosed across 112 Local Government Areas (LGAs) and 27 states of the Nigerian Federation, with 189 fatalities resulting from the disease. According to the latest report from the Nigeria Center for Disease Control (NCDC), 8,202 cases were reported from 26 December 2022 to 1 January 2023, affecting more than 63 healthcare personnel (Vanguard, 2023).

LASV's genome is a single-stranded, bipartite ribonucleic acid (RNA), lacking an arenavirus's typical negative-strand coding configuration (Andersen et al., 2015). The spherical Lassa virus ranges from 70 to 150 nm in size and features a glycoprotein envelope with T-shaped spikes measuring 7–10 nm on its surface (Ogbu et al., 2007).

Although the virus's pathophysiology is not fully understood, research has shown that it primarily targets endothelial and antigen-presenting cells, particularly dendritic cells, and upon entering the human body, the Lassa virus infects most tissues, initially affecting the mucosa, intestine, lungs, and urinary system, followed by the vascular system (Mahanty et al., 2003; Rojek et al., 2008). LASV RNA genome is responsible for encoding a few translational products such as highly glycosylated membrane glycoprotein (MGP), RNA polymerase, a matrix protein, and a nucleoprotein (NP). Among these genomic products membrane glycoprotein plays a crucial role in viral attachment and fusion through endothelial cell surface (Meyer et al., 2002). Lassa fever membrane glycoprotein consists of two subunits (GP1 and GP2), where GP1 is responsible for receptor binding and GP2 plays a significant role in cell membrane fusion (Lenz et al., 2001; Igonet et al., 2011; Borenstein-Katz et al., 2019). MGP incorporates with the α -dystroglycan receptor of the extracellular matrix and initiates LASV entry into the host cell (Bowen et al.,

2000). The pathogen replicates intracellularly by utilizing L-polymerase and nucleocapsid proteins. After that, nucleocapsid proteins synthesize both mRNAs and antigenomic RNAs that are responsible for evading the host immune system (Yun and Walker, 2012). The entire pathophysiology of the Lassa fever virus depends on the successful attachment of the virus with endothelial cells. However, the Lassa virus glycoprotein spike (PDB ID 5FT2) was considered a putative drug target in our study to inhibit viral attachment with host endothelial cells (Li et al., 2016).

However, the Lassa virus NP is critical for both transcription and RNA replication since it encloses viral genomic RNAs into ribonucleoprotein (RNP) complexes (Hass et al., 2004). Though the exact mechanism of NP involvement in the pathogen's pathophysiology is poorly understood, it has been shown in prior research that the NP of the Lassa virus plays an essential role in viral RNA synthesis and host immune system suppression by actively suppressing type I interferon (IFN) (Martinez-Sobrido et al., 2006; Martinez-Sobrido et al., 2007). Generalized immune suppression in the infected host is correlated with severe arenavirus infections, including fatal Lassa cases, which indicates that the NP of the Lassa virus is a key element in Lassa fever pathogenesis (Baize et al., 2009). Therefore, Lassa virus nucleoprotein (PDB ID: 3MX5) was also considered as a potential drug target for our following study (Qi et al., 2010).

The “multimammate rat,” a rodent species, serves as the primary reservoir or host for the Lassa virus. The two predominant modes of Lassa virus transmission to humans are ingestion and inhalation (Tewogbola and Aung, 2020; CFDCA Prevention, 2022). Mastomys rats excrete the virus in their urine and droppings, and direct exposure to these materials—such as handling contaminated objects, consuming tainted food, or contacting open wounds or sores—can result in infection. Lassa virus typically infects humans upon contact with the urine or feces of infected Mastomys rats or through direct contact with the blood, urine, feces, or other bodily secretions of an individual suffering from Lassa fever (McCormick and Fisher-Hoch, 2002; Atkin et al., 2009). Infection can occur when humans come into contact with contaminated rat excrement or when they capture and consume the rodents as food. Lassa fever can be contracted from an infected individual, although this is relatively rare (Asogun et al., 2019). Blood, saliva, urine, and semen are some of the bodily fluids that can spread infection from person to person, occurring in both household and healthcare settings. Those in close contact with the infected individuals are generally only exposed to the patient's symptoms; however, a patient can excrete the virus through their semen for up to 3 months and in their urine for 3–9 weeks, following the onset of their illness (Azeez-Akande, 2016) (Figure 1).

In a world where the menacing Lassa virus poses a deadly threat, we find ourselves with limited therapeutic options and no FDA-approved drugs or vaccines for treating Lassa fever (Baral et al., 2020). The pressing need for effective antiviral treatments has driven researchers to explore new frontiers, turning to computational studies in search of potential drug candidates. Meanwhile, numerous biological actions of evodiamine, including anti-inflammatory, anti-viral anti-tumor, blood pressure reduction, and immunological modulation, have been reported (Dai et al., 2012; Gavaraskar et al., 2015; Yang F. et al., 2017; Zhang et al., 2022). Evodiamine, a naturally occurring indole alkaloid derived from the traditional medicinal plant *Evodia rutaecarpa*, has been increasingly recognized in recent years for its various pharmacological effects, including anti-inflammatory, anti-cancer, and anti-obesity properties, to name a few. Intriguingly, some studies have also shown evodiamine's potential as an antiviral agent, a property that aligns well with the focus of our study. Moreover, evodiamine has been considered a potential medication option for combating liver diseases (Li et al., 2020). However, to the best of our knowledge, there has been no investigation performed to assess the potentiality of evodiamine derivatives to function as a potential medication option for Lassa fever. Harnessing the power of technology, this study aimed to expedite the discovery process while conserving valuable time, resources, and funding required for developing novel therapeutic options for combating the Lassa virus (Rahman et al., 2012). Several evodiamine derivatives were subjected to computational investigation through systematic approaches, including molecular docking, drug-likeness assessment, molecular dynamics simulation, and ADMET analysis. This theoretical study will add a new dimension in considering these evodiamine derivatives as potential treatment options for treating Lassa fever.

2. Computational method

2.1. Determination of the data of ADMET

Many drug-like molecules are eliminated from trial phases for not having proper absorption, distribution, metabolism, excretion, and toxicity (ADMET) profile (Alqahtani, 2017). Therefore, we have calculated ADMET profiles for selected evodiamine derivatives, employing the pkCSM (<https://biosig.lab.uq.edu.au/pkcsm/prediction>) web server (Pires et al., 2015). The pkCSM web server uses a cutting-edge method based on graph-based signatures to predict various pharmacokinetic features. Predictive models may be effectively trained using these signatures for several different ADMET features. The method, known as pkCSM, also offers a platform for the analysis and optimization of pharmacokinetic and toxicity properties implemented in a friendly, open-source web interface, a useful tool to assist medicinal chemists in striking a balance between potency, safety, and pharmacokinetic properties. Assessing the ADMET profiles for selected compounds using this server will help us to select the most suitable compounds that have good absorption, distribution through blood, good metabolic profile, better excretion rate, and lowest toxicity. The pkCSM server accepts SMILES as input; hence, canonical SMILES for compounds containing no chiral carbons and isomeric SMILES

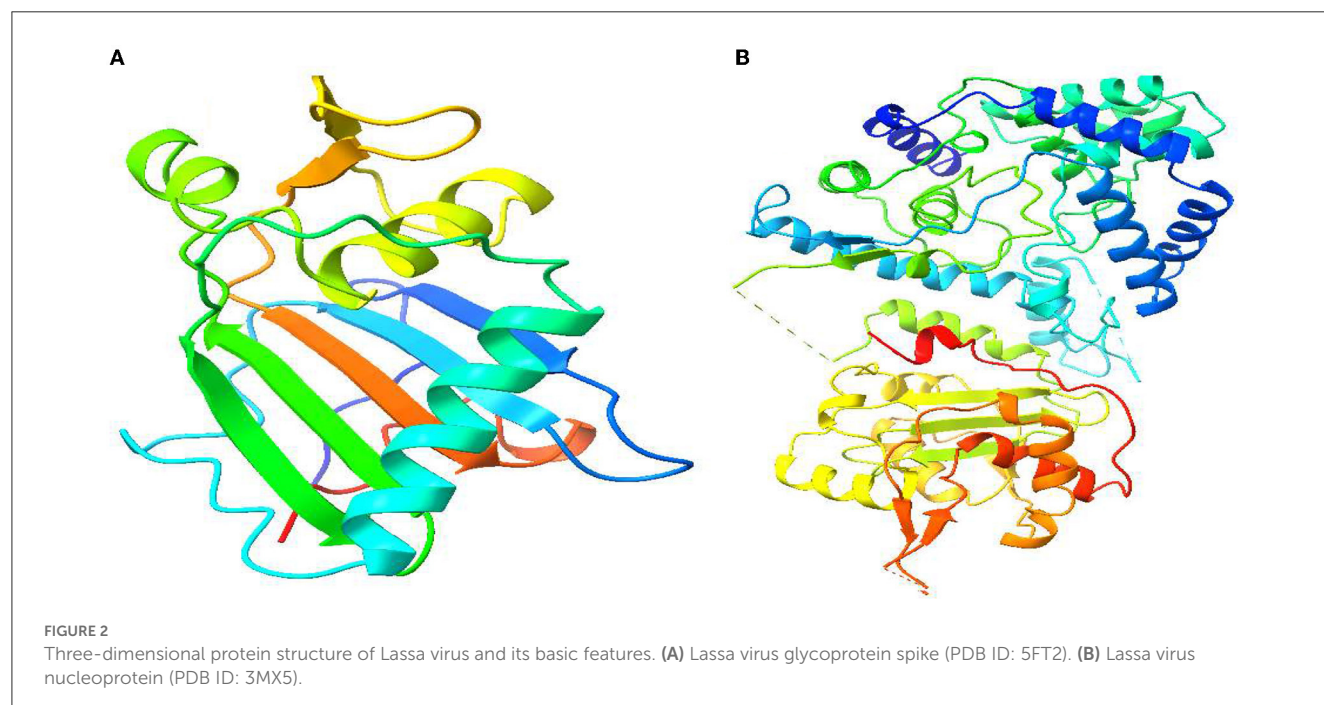
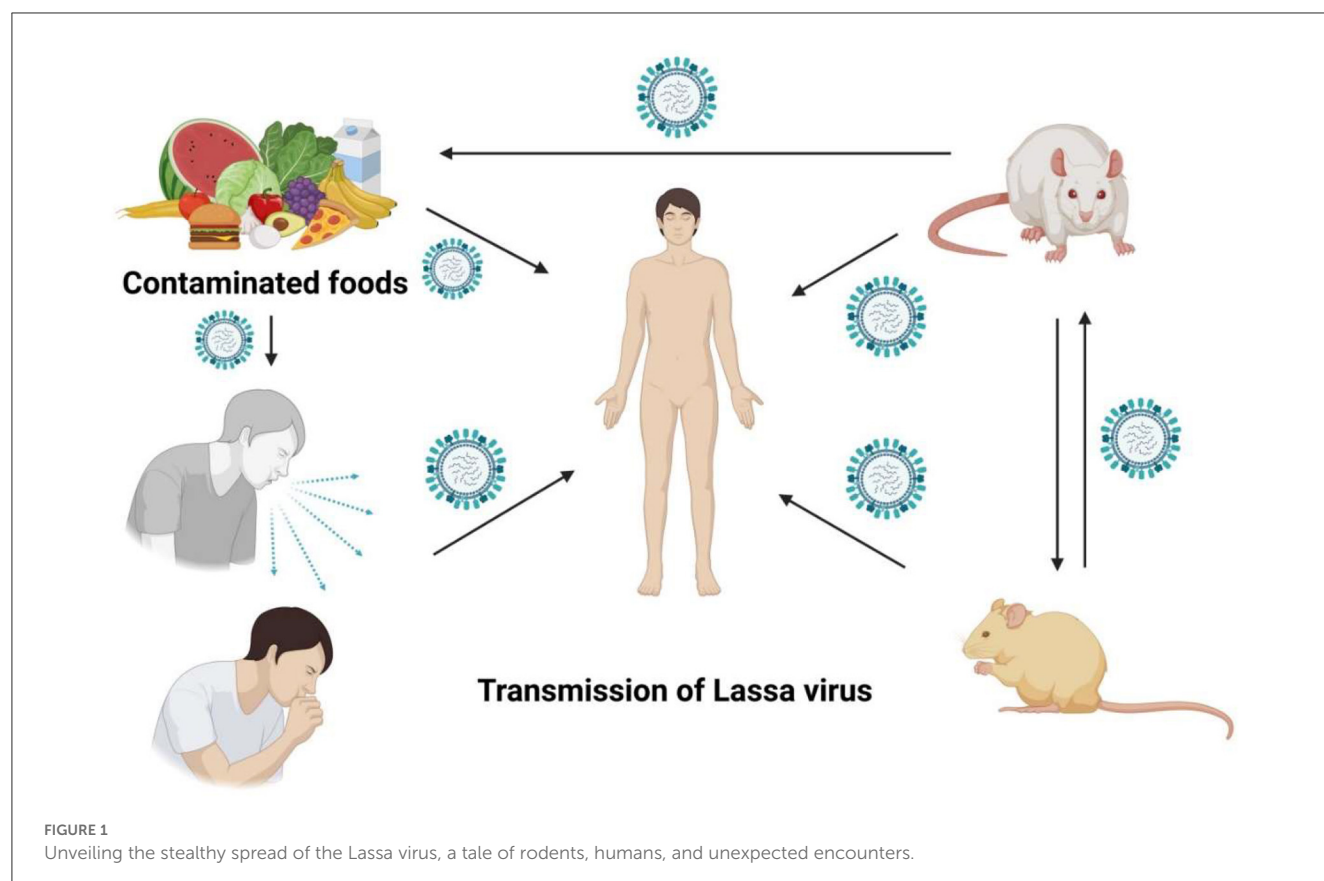
for compounds containing chiral carbons were used to generate desired ADMET values.

2.2. Preparation of ligand and molecular optimization

Before starting molecular docking, we optimized the three-dimensional structures of the selected compounds using the Materials Studio 8.0 software package (Sharma et al., 2019). This program effectively optimizes the overall geometry and chemical structure of ligands to achieve minimum ground-state energy so that these structures could be docked with receptor proteins without any interruption. After importing the ligand structures into the Material Studio 8.0 software, density-functional theory (DFT) was incorporated by applying the DND basis (diffused basis set) semi-core pseudo-potentials (Papajak and Truhlar, 2010; Obot et al., 2015; Ribeiro et al., 2015). In material research, DFT is commonly used to investigate electronic structure organization using a quantum mechanical modeling strategy. Finally, the structures are saved in a PDB format for further computational analysis.

2.3. Preparation of protein and molecular docking studies

Molecular docking analysis is a significant aspect of computational drug design (Jakhar et al., 2020). Therefore, in our study, we have incorporated molecular docking analysis to understand the molecular binding dynamics between selected compounds and selected proteins. First, the three-dimensional structures of the Lassa virus glycoprotein spike (PDB ID 5FT2) and Lassa virus nucleoprotein (PDB ID: 3MX5) were acquired from the RCSB protein data bank in the PDB format (Rose et al., 2016). Users of this platform have access to approximately 200,000 experimentally established PDB structures of biological macromolecules and almost a million computed structure models. The PDB structures (Figure 2) were subjected to energy minimization using Swiss PDB Viewer v4.1.0 (Kaplan and Littlejohn, 2001). After energy minimization, both protein structures were opened in BIOVIA Discovery Studio Visualizer to delete excessive water molecules, surrounding the protein which could interrupt ligand-protein docking (Design, 2014). It will be any unwanted heteroatom attached with protein structure were also deleted, concerning these ligands that could occupy/interfere the ligand protein binding. Water molecules surrounding the macromolecules were also eliminated as they do not play a role in ligand-protein molecular interaction. After completing the protein preparation part, we acquired ligand structure files from the PubChem database in SDF format (Kim et al., 2019). The ligand structures were also energy minimized before molecular docking is started. For molecular docking, we have used AutoDock Vina in PyRx software where the ligand structures used were converted into the PDBQT format, and the grid center points were set to $X = -21.6734$, $Y = -17.1276$, and $Z = 28.1838$ and the box dimensions (Å) $X = 47.70441$, $Y = 64.6983$, and $Z = 51.6696$ for



(PDB ID: 3MX5), and the grid center for (PDB ID: 5FT2) were set to $X = -8.7259$, $Y = -31.6119$, and $Z = -27.5813$ and the box dimension $X = 81.3565$, $Y = 55.5874$, and $Z = 91.9504$ (Trott and Olson, 2010). PyRx software presents the 9 most suitable docking

poses of the ligand–protein complex after the docking is completed (Dallakyan and Olson, 2015). We have selected the most suitable docking poses where the ligands are strongly interacting with the protein's catalytic cavity and visualized them using BIOVIA

TABLE 1 Data of Lipinski rule, pharmacokinetics, and drug-likeness.

Sl. No.	CID	Molecular weight (g/mol)	Consensus Log Po/w	Hydrogen bond acceptor	Hydrogen bond donor	Molar refractivity	Topological polar surface area Å ²	Lipinski rule		Bioavailability Score
								Result	Violation	
01	Evodiamine	303.36	2.7	1	1	97.67	39.34	Yes	0	0.55
02	49806754	342.39	2.52	2	0	107.13	52.27	Yes	0	0.55
03	49806624	427.93	4.54	1	0	132.07	28.48	Yes	1	0.55
04	49806625	427.93	4.48	1	0	132.07	28.48	Yes	1	0.55
05	49806500	437.49	3.76	3	0	133.97	54.78	Yes	0	0.55
06	129710532	321.37	2.44	2	2	100.72	48.57	Yes	0	0.55
07	151289	303.36	2.7	1	1	97.67	39.34	Yes	0	0.55
08	56967508	289.37	3	1	1	97.25	22.27	Yes	0	0.55
09	49804912	425.45	4.03	3	0	127.44	45.55	Yes	1	0.55
Standard: Sofosbuvir		529.45	1.44	11	3	125.53	167.99	No	2	0.17

Discovery Studio Visualizer to have a great insight into ligand binding position in the protein cavity.

2.4. Lipinski rule, pharmacokinetics, and drug-likeness

Determination of pharmacokinetic properties is an effective approach to distinguish between drug-like and non-drug-like small molecules. We used the SwissADME server to calculate important drug-like features such as molecular weight, hydrogen bond acceptor, hydrogen bond donor, molar refractivity, topological surface area, and bioavailability (Daina et al., 2017). All these features were well calculated considering Lipinski’s rule of five proposed by Lipinski (2004). Distinguishing drug-like molecules by considering Lipinski’s rule of five is a globally acceptable approach proposed by Chris Lipinski suggested that any drug-like molecules should follow at least three of the following four rules: (1) A drug-like molecule must have a maximum molecular weight of 500 g/mol or less. (2) The lipophilicity of any drug-like molecules should not cross 5 logP. (3) The maximum number of hydrogen bond acceptors in the drug-like molecule should not cross 10. (4) The maximum number of hydrogen bond donors present in the chemical structure of a drug-like molecule should not cross five. In our investigation, the canonical SMILES of elected evodiamine derivatives and the standard drug sofosbuvir were collected from the PubChem database. The canonical SMILES were inputted into the SwissADME server to calculate selected parameters for the proposed small molecules. The Lipinski rule of five following and violating decisions of selected derivatives was also obtained from SwissADME.

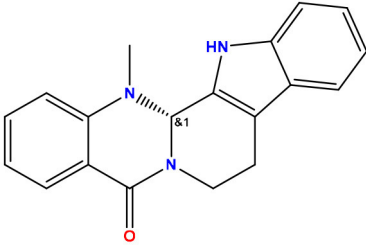
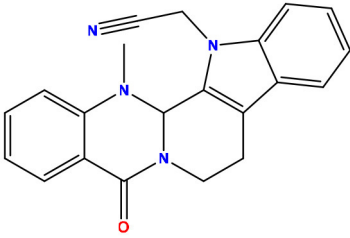
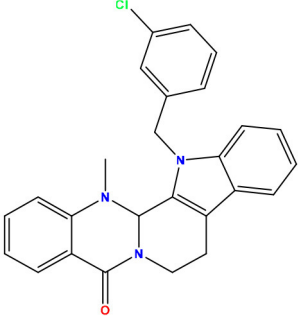
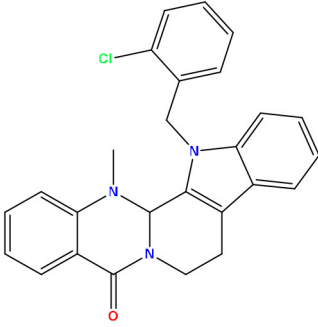
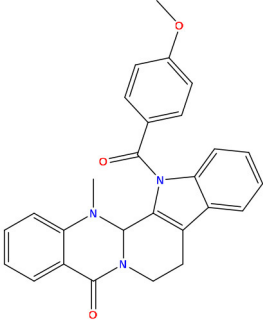
2.5. System preparation

The compounds were parameterized by a general AMBER force field (GAFF) for organic molecules with the ANTECHAMBER module implemented by AMBER. Protein was parametrized by AMBER ff14SB force field. The ligands were bound to Lassa fever virus nucleoprotein (PDB ID: 3MX5) in aqueous solutions with an explicit solvent TIP3P water box. NaCl ions were modeled to neutralize the system. The tutorial for the LEaP program was used for the formation of the protein–ligand complexes and the preparation of the system (Case et al., 2005; Shukla and Tripathi, 2020).

2.6. Molecular dynamics simulation

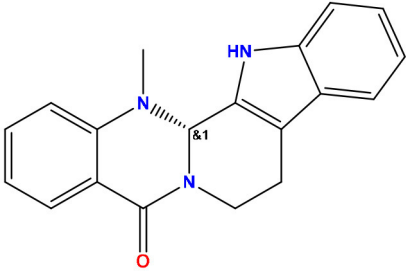
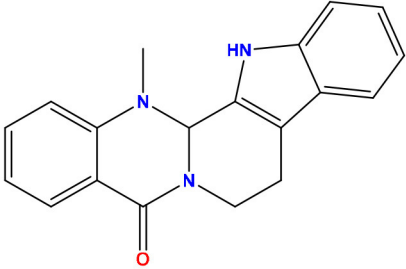
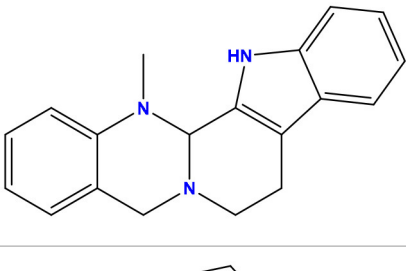
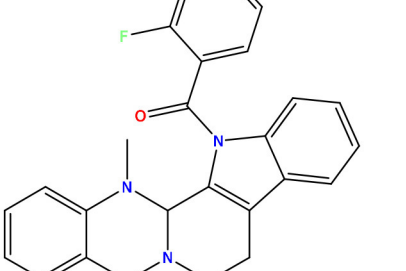
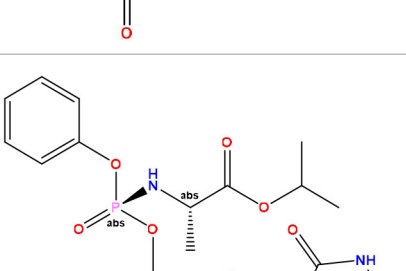
The simulations were well carried out by AMBER 16 using the Particle Mesh Ewald (PME) method in each system (Essmann et al., 1995). The systems were prepared as described above. The SHAKE algorithm was used to constrain hydrogen bonds, allowing the use of an integration time of 2 fs. The following molecular dynamics protocol was used: (I) solvent minimization (30,000 steps), (II) equilibrium to heat the system from 100K

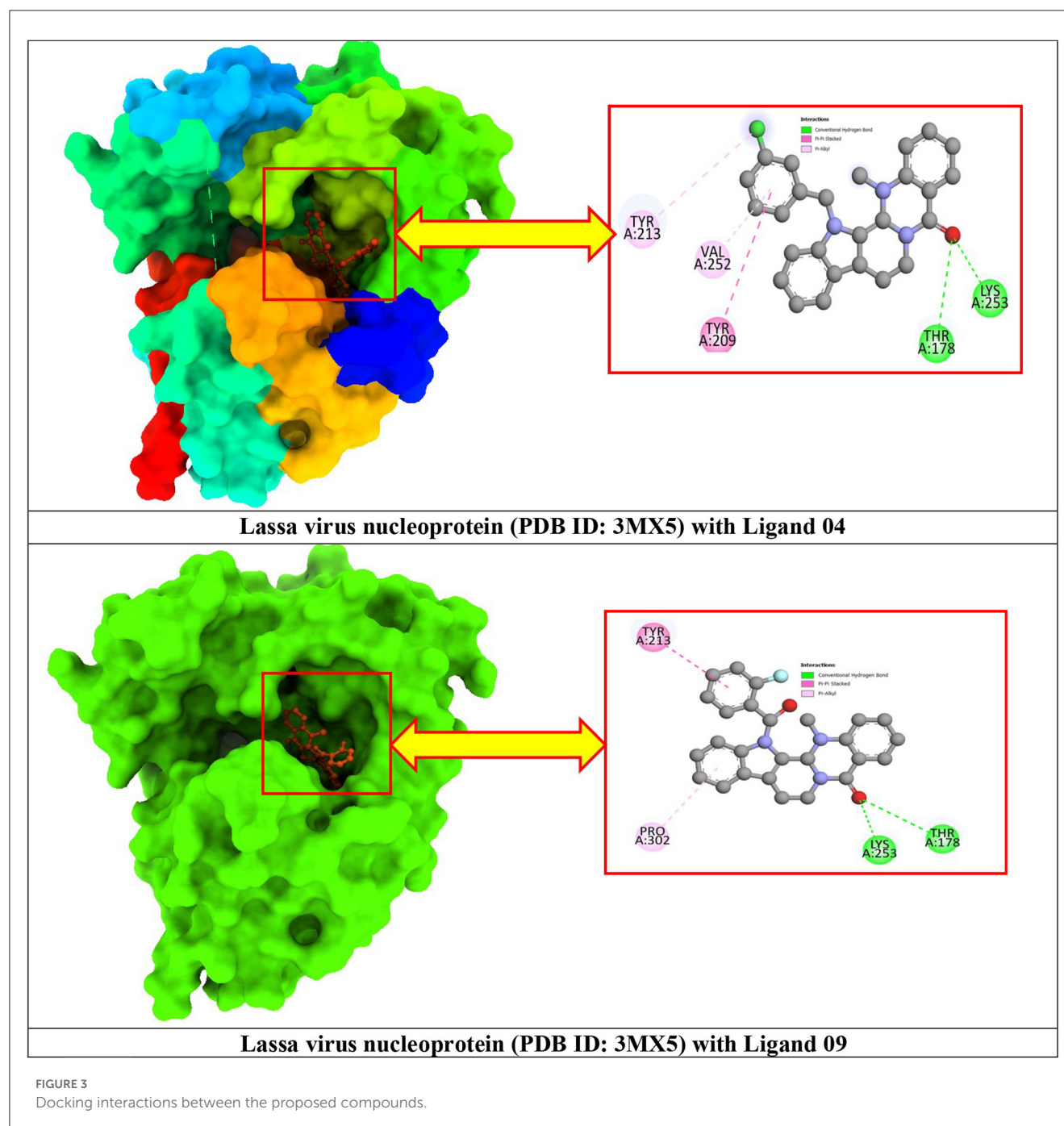
TABLE 2 Molecular docking/binding energy score for all derivatives represented.

Sl. No.	PubChem CID	Chemical structure	Lassa virus glycoprotein spike (PDB ID 5FT2)	Lassa virus nucleoprotein (PDB ID 3MX5)
			Binding affinity (kcal/mol)	Binding affinity (kcal/mol)
01	Evodiamine		−7.2	−10.7
02	49806754		−7.4	−9.1
03	49806624		−8.3	−9.7
04	49806625		−8.5	−11.0
05	49806500		−8.1	−9.8

(Continued)

TABLE 2 (Continued)

Sl. No.	PubChem CID	Chemical structure	Lassa virus glycoprotein spike (PDB ID 5FT2)	Lassa virus nucleoprotein (PDB ID 3MX5)
			Binding affinity (kcal/mol)	Binding affinity (kcal/mol)
06	129710532		−1.7	−2.0
07	151289		−7.8	−10.6
08	56967508		−7.2	−9.4
09	49804912		−8.3	−10.3
Standard Sofosbuvir			−5.9	−7.1



to 298K at a constant volume with restricted proteins (1 ns), (III) equilibrium to relax the system with restricted proteins (1 ns), (IV) relaxation of the system for 1 ns at constant pressure and 298 K with restriction of less than 10 kcal/mol-Å² of the protein, (V) minimization of the system with restrictions only on the protein backbone, (VI) relaxation of the system for 1 ns at constant pressure and 298 K with a restriction of <10 kcal/mol-Å² of the backbone, and (VII) three equilibration steps where the restraint on the backbone was decreased until it was free. Finally, 140 ns of molecular dynamics production was launched.

3. Result and discussions

3.1. Lipinski rule, pharmacokinetics, and drug-likeness

Evodiamine derivatives are well known for their effective application in different disease treatments, such as pulmonary hypertension, gastric cancer, and hepatocellular carcinoma (Zhang et al., 2020; Fan et al., 2021; Liang et al., 2022). Recently, computational drug designing application of evodiamine derivatives was noticed in the potential treatment of viral diseases

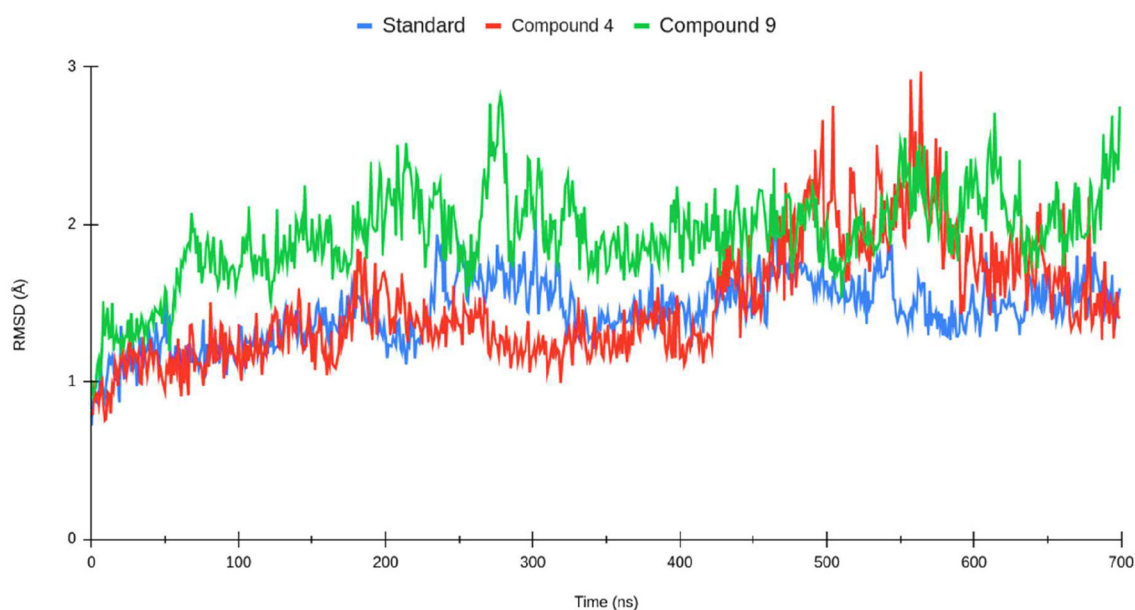


FIGURE 4

Root mean square deviation (RMSD) of the Lassa virus nucleoprotein (PDB ID: 3MX5) backbone for standard sofosbuvir complex (blue), compound 1 (red), and compound 2 (green).

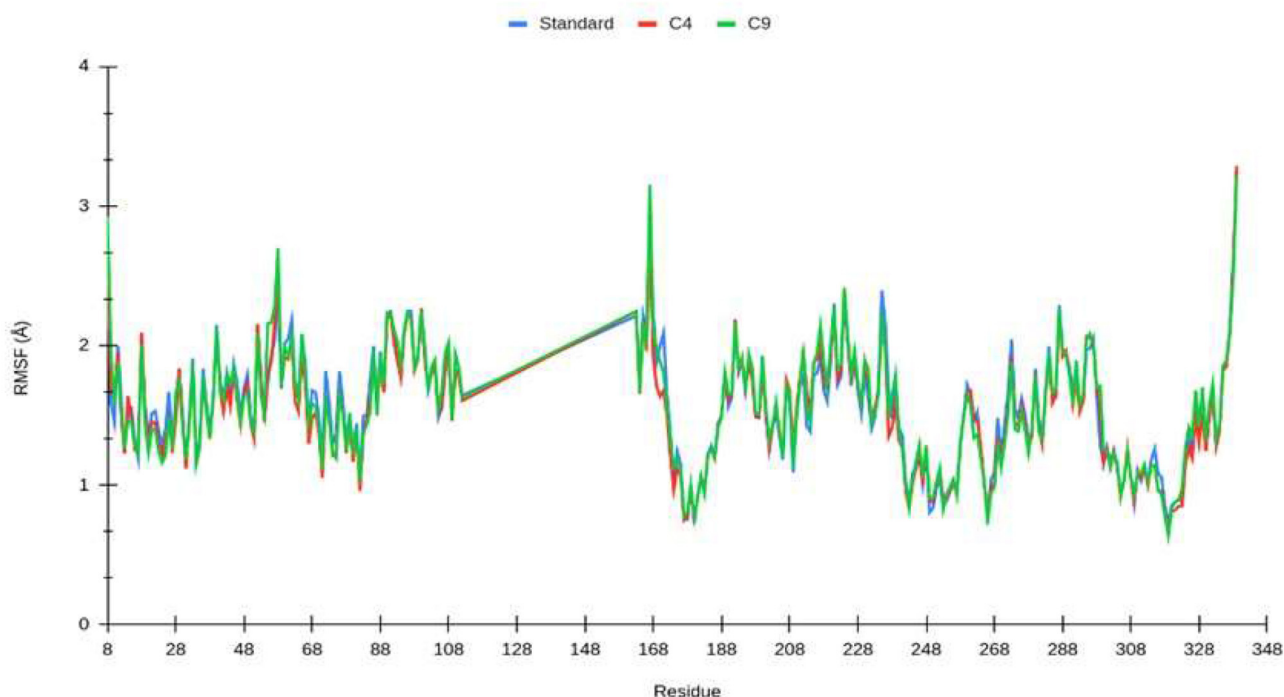


FIGURE 5

Root mean square fluctuation (RMSF) of Lassa fever nucleoprotein backbone over time (140 ns) for standard (blue), compound 4 (red), and compound 9 (green). Residues 113 to 162 are not present in the nucleoprotein structure.

such as COVID-19 caused by the SARS-CoV-2 virus (Belal et al., 2022). On that account, we have investigated the pharmacokinetics and drug-likeness properties (Table 1) of our selected evodiamine derivatives to select the most potent small molecule for Lassa

fever treatment caused by the Lassa virus. The pharmacokinetic profile of the standard drug sofosbuvir was also calculated for the following comparative analysis. According to SwissADME, all nine of our selected compounds followed the Lipinski rule of

five, whereas the compounds named 49806624, 49806625, and 49804912 expressed only one violation of the Lipinski rules. In contrast, the compounds named evodiamine, 49806754, 49806500, 129710532, 151289, and 56967508 all followed Lipinski rules, which prove their credibility to be selected as potential drug candidates for having appropriate pharmacokinetics and drug profile. The calculation of topological polar surface area (TPSA) is a crucial indicator to understand the ability of drug molecule transportation (Ertl et al., 2000). Ideally, a TPSA score of $<130 \text{ \AA}^2$ indicates excellent drug-transporting ability in the host system where the minimum TPSA score of a drug-like molecule should not be $<20 \text{ \AA}^2$. Interestingly, the standard drug sofosbuvir has a TPSA score of 167.99 \AA^2 which crosses the ideal range of TPSA. All nine selected evodiamine compounds had a TPSA score of $<130 \text{ \AA}^2$, indicating their excellent transporting ability as a drug in the host system. All nine selected evodiamine compounds expressed a bioavailability score of 0.55, which was significantly higher than sofosbuvir. The lipophilicity value was also calculated to predict non-aqueous solubility. By assessing the data presented, we predict that compound 129710532 showed the lowest score (2.44 Log Po/w) and compound 49806624 showed the highest lipophilicity score (4.54 Log Po/w). For an ideal drug-like molecule, the molar refractivity score should be between 40 and 130 units. Only three compounds (49806624, 49806625, and 49806500) had slightly higher scores than the ideal range. However, the molar refractivity scores for the other six compounds were satisfying.

3.2. Molecular docking and interaction analysis

Lassa virus glycoprotein spike (PDB ID: 5FT2) and Lassa virus nucleoprotein (PDB ID 3MX5) were docked with evodiamine's nine selected derivatives. These docked complexes were compared with the standard drug sofosbuvir for justifying the significance of conducting this study. According to the molecular docking rules, a stable protein–ligand complex should express minimum binding energy with a strong binding affinity of the ligand with the receptor protein. Sofosbuvir showed binding energy of -5.9 kcal/mol for the Lassa virus glycoprotein spike. However, eight (ligand nos: 01, 02, 03, 04, 05, 07, 08, and 09) out of nine selected evodiamine derivatives expressed higher binding affinity than the standard drug. These docking scores indicate that ligand nos. 03, 04, and 09 have bound much more strongly than sofosbuvir with the Lassa virus glycoprotein spike. As stronger binding has a positive correlation with forming a more stable receptor–ligand complex, we can suppose that our selected compounds will have a better role in stabilizing the target protein than the stronger drug.

In addition, Lassa virus nucleoprotein was also docked with the same ligands. Sofosbuvir showed a binding energy score of -7.1 kcal/mol with this receptor. Except for ligand no. 06, all eight ligands showed better docking scores than the standard drug (Table 2). Ligand no. 04 had a binding energy score of only -11 kcal/mol , indicating excellent binding with the receptor by forming a stable protein–ligand complex. Moreover, ligand nos. 01, 03, 05, 07, and 09 also showed excellent binding affinity, suggesting that they could be also considered for potential future drug

TABLE 3 Binding free energy calculations for the protein–ligand systems based on MMGBSA.

MMGBSA	Delta G	Standard. dev.
Standard	$-29,4833$	$\pm 6,1179$
Compound 01	$-39,4656$	$\pm 7,8665$
Compound 09 09	$-36,3851$	$\pm 3,7805$

Values expressed in kcal/mol.

development. After analyzing all the docked complexes, it could be said that ligand nos. 04 and 09 expressed excellent docking scores with both target receptors, which was much higher than sofosbuvir. These two compounds can be considered very strong candidates for developing future drug development against Lassa Virus.

3.3. Protein–ligand interaction and molecular docking poses

Protein–ligand docked complexes were visualized using the PyMOL program to better understand different types of interactions. For the Lassa virus nucleoprotein (PDB ID: 3MX5), compound nos. 04 (-8.5 kcal/mol) and 09 (-8.3 kcal/mol) had the maximum binding energy. Compound no. 04 formed a Py-Alkyl bond with TYR A:213 and VAL A:252; a conventional hydrogen bond with THR A:178 and LYS A:253; and pi-pi stacked with TYR A:209 (Figure 3). However, compound no. 09 formed a Pi-Alkyl bond with PRO A:302; a conventional hydrogen bond with LYS A:253 and THR A:178; and pi-pi stacked with TYR A:213 (Figure 3). Selected protein ligands with strong molecular interactions were further analyzed by incorporating molecular dynamics simulation to explore the significant roles of selected compounds in stabilizing virulent proteins.

3.4. Molecular dynamics simulation analysis

The production of the molecular trajectory (140 ns) was used to perform the analyses. The root mean square deviation (RMSD) allows one to calculate the divergence between the two overlapping structures, so the lower the value, the higher the similarity between them. The RMSD and RMSF plots were performed using Visual Molecular Dynamics (VMD) (Humphrey et al., 1996a,b). VMD and TCL scripts to establish residues that have contact with the ligand at a distance of fewer than 5 \AA for at least 50% of the molecular trajectory. The binding free energy calculations for the protein–ligand systems were estimated using the Python script MMPBSA.py provided by AMBER.

The root mean square deviation (RMSD) was calculated for the systems described above (shown in Figure 4). The protein–ligand complexes were equilibrated at 40 ns from the molecular dynamics trajectory. In the standard complex, the RMSD had values of $1\sim 2 \text{ \AA}$. The complexes formed by compounds 1 and 2 reached higher RMSD values. Compound 4 had fluctuations of $1\sim 3 \text{ \AA}$; at the 100th nanosecond of molecular dynamics, the protein

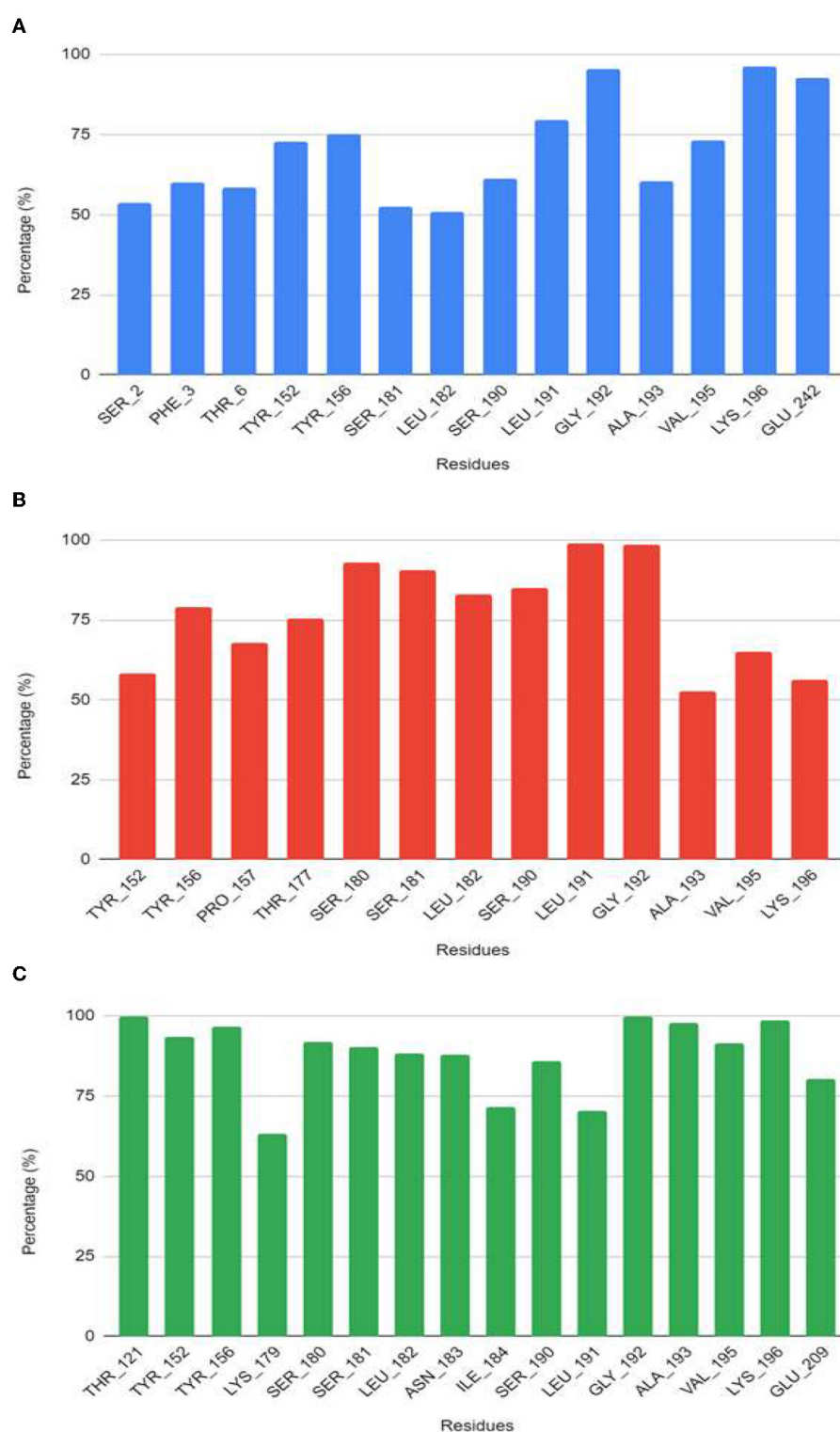


FIGURE 6

Residues that have contacts at <5 Å with the ligand for at least 50% of the molecular dynamics trajectory for the systems (A) sofosbuvir, (B) compound 4, and (C) compound 9.

rotates sharply, which twists the N-terminus and C-terminus of the proteins, justifying the increase of the RMSD value. Compound 2 achieved similar RMSD values as compound 1; however, compound 2 is more stable along the molecular dynamic trajectory.

To characterize the local changes in the interaction motifs close to the ligand, the root mean square fluctuation (RMSF) between the nucleoprotein and the ligand was calculated (Figure 5). The RMSF values are similar in all systems. The

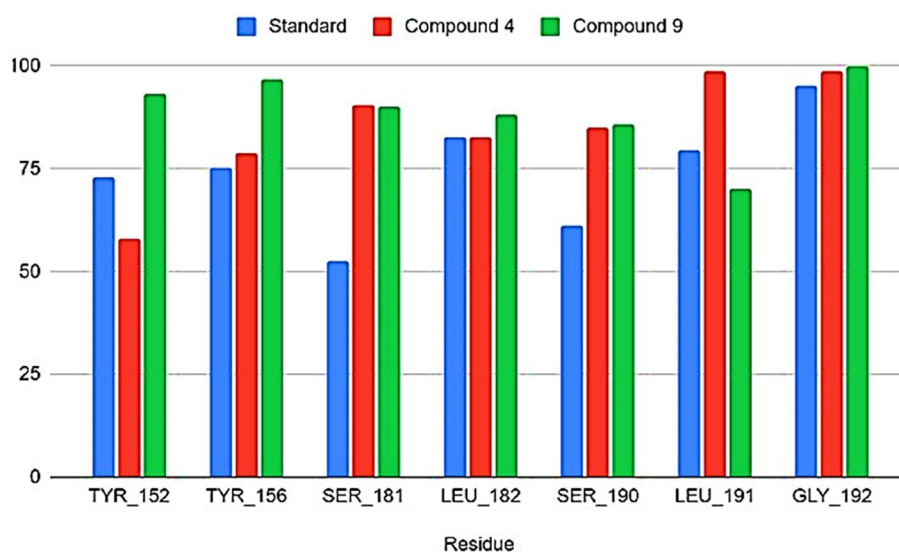


FIGURE 7
Common residues among contacts found.

most stable interaction correspond to the residues that are found to keep interacting with the ligand. The most stable interaction motifs are located near the N-terminus (8–50 aa.) and in the region between residues 180 and 200, corresponding to the amino acids found to have contact with the ligand.

From the molecular dynamics simulation, we estimated the residues that possessed ligand–ligand contacts at 5 Å for at least 50% of the molecular dynamics. We found 14, 13, and 16 protein–ligand contacts for the standard, compound 4, and compound 9 systems, respectively (Figure 5). Common residues TYR: A 209, TYR: A 206, SER: A 238, LEU: A 239, SER: A 247, LEU: A 248, and GLY: A 249 were found forming contacts in the systems. Among these, the GLY: A 249 residue has a prominent role, forming contacts in all systems in at least 95% of the molecular dynamics trajectory.

The free energy of the protein–ligand binding was calculated by the MMGBSA method to estimate which compound binds better to the nucleoprotein. Table 3 shows the union energy residues for each system. A comparison of the binding free energy values shows that Ribavirin is less stable than compounds 4 and 9. From the above analysis, compound 9 was chosen because it has a higher number of contacts with the nucleoprotein and has a favorable binding energy compared with the other ligands.

The free energy of the protein–ligand binding was calculated by the MMGBSA method to estimate which compound binds better to the nucleoprotein. Figures 6, 7 show the union energy residues for each system. A comparison of the binding free energy values shows that sofosbuvir is less stable than compounds 4 and 9. From the above analysis, compound 9 was chosen because it has a higher number of contacts with the nucleoprotein and has a favorable binding energy compared with the other ligands.

3.5. ADMET data investigation

The consumption of drugs of poor ADMET properties can lead to side effects such as allergic reactions, rashes, and organ damage. To avoid these complications, we have calculated the most important ADMET properties for our selected evodiamine derivatives (Table 4). To understand the absorption aspect, we have selected the following three parameters: water solubility, Caco-2 permeability, and human intestinal absorption. According to the data collected from the pkCSM server, the standard drug sofosbuvir has a low human intestine absorption rate of only 60.168%. In contrast to that, all our proposed compounds have higher human interest in absorption rate where compound 05 showed the highest score of 96.562%. According to the water solubility test (calculated in Log S), the range from highly soluble compounds to insoluble compounds is <−10 poorly <−6, moderately <−4 soluble <−2 very <0 <highly. As presented in Table 4, compounds 01, 06, 07, and 08 were declared as soluble. Furthermore, compounds 02, 03, 04, 05, and 09 were declared as moderately soluble. The Caco-2 cell line, which is normally utilized as a reliable model to evaluate the absorption property of any orally delivered medications, is made up of human colorectal adenocarcinoma epithelial cells (Sambuy et al., 2005). According to pkCSM guidelines, a Caco-2 permeability score over 0.9 is considered high, and the standard drug sofosbuvir expressed a low Caco-2 permeability score of 0.454. All nine proposed compounds showed a higher Caco-2 permeability score than the sofosbuvir, where compound 08 showed the highest score of 1.519.

For predicting the distribution, we have selected the volume of distribution of the steady-state method (VDss) and blood–brain barrier (BBB) as our keynote parameters. The VDss values indicate how evenly drugs are distributed between the blood and the tissue. A higher VD score (>0.45) suggests that the therapeutic molecule is dispersed more evenly throughout the

body, whereas a lower (<-0.15) result denotes uneven drug distribution. Compound number 08 showed excellent VDss distribution, and other derivatives also expressed satisfying VDss scores, except for compound 05. The BBB shields our brain from any interactions with outside substances. This implies that BBB permeability assessment is a crucial characteristic in choosing ideal drug-like compounds (Passeleu-Le Bourdonnec et al., 2013). When the BBB permeability score is <-1 , the distribution is poor. In contrast, a score of >0.3 indicates excellent BBB permeability. Compounds 01, 03, 04, 06, 07, and 08 showed positive BBB permeability, whereas compounds 02, 05, and 09 showed poor blood–brain permeability. In metabolic profile analysis, it can be stated that all selected compounds showed positive CYP450 1A2 inhibition, and none of the compounds inhibited the CYP450 2D6 substrate.

Estimating overall clearance and organic cation transporter 2 (OCT2) allowed for the incorporation of excretion analysis (Filipski et al., 2009). Total clearance produces a total clearance score using the combined information from hepatic clearance and renal clearance, which provides a clear excretion profile of any particular drug. All selected derivatives showed better total clearance scores than sofosbuvir. None of the compounds were predicted to be potential renal OCT2 substrates. Finally, in toxicity prediction, we have predicted that none of the substances induce skin sensitization, and special precautions should be taken before recommending these compounds to patients suffering from liver diseases, as all of the selected compounds can induce hepatotoxicity including sofosbuvir.

4. Harnessing evodiamine derivatives for lassa virus intervention—a thought-provoking discussion

In the subsequent study, Lassa virus nucleoprotein and Lassa virus glycoprotein spike were considered potential drug targets to inhibit the Lassa virus. *Evodia rutaecarpa* is a rich source of evodiamine, an alkaloid that has garnered scientific interest for its potent therapeutic effects against various diseases such as anti-obesity, anti-allergenic, analgesic, and anti-ulcerogenic properties (Wang et al., 2008; Tan and Zhang, 2016). Furthermore, in the hippocampus, evodiamine significantly reduces neuroinflammation (TNF- α , IL-1 β , and IL-6) and glial cell activation, rendering it a potential treatment for neurodegenerative diseases such as Alzheimer's disease (Wang et al., 2018). Despite concerns about evodiamine-induced hepatotoxicity and cardiotoxicity, its effectiveness against various cancer cells (lung cancer, gastric cancer, oral cancer, colorectal cancer, and pancreatic cancer) cannot be dismissed (Wei et al., 2012; Sachita et al., 2015; Wen et al., 2015; Zhao et al., 2015; Zou et al., 2015; Yang W. et al., 2017). Additionally, evodiamine exhibits anti-inflammatory and antioxidative stress potency, and one study suggests its potential as a therapeutic lead compound in liver diseases (Zhang et al., 2018, 2022; Li et al., 2020).

Given the broad-spectrum antiviral effects of alkaloids against various DNA and RNA viruses, evodiamine has also

TABLE 4 Theoretical ADMET data analysis.

Sl. No.	Absorption		Distribution		Metabolism		Excretion		Toxicity			
	Water solubility Log S	Caco-2 permeability × 10 ^{−6}	Human intestinal absorption (%)	VDs (human)	BBB permeability	CYP450 1A2 inhibitor	CYP450 2D6 substrate	Total clearance (ml/min/kg)	Renal OCT2 substrate	Max. tolerated the dose (log mg/kg/day)	Skin sensitization	Hepatotoxicity
01	−3.71	1.226	93.70	0.327	0.470	Yes	No	0.355	No	0.051	No	Yes
02	−4.193	1.08	95.43	0.402	0.274	Yes	No	0.468	No	−0.436	No	Yes
03	−4.881	1.054	94.24	0.371	0.437	Yes	No	0.156	No	0.473	No	Yes
04	−4.807	1.053	94.53	0.378	0.424	Yes	No	0.217	No	0.467	No	Yes
05	−5.064	1.3	96.56	−0.243	−0.013	Yes	No	0.213	No	0.193	No	Yes
06	−3.609	0.882	94.96	0.188	0.398	Yes	No	0.360	No	−0.046	No	Yes
07	−3.71	1.226	93.70	0.327	0.470	Yes	No	0.355	No	0.051	No	Yes
08	−3.607	1.519	93.27	0.915	0.373	Yes	Yes	1.123	No	0.276	No	Yes
09	−4.421	1.147	96.34	−0.134	0.173	Yes	No	0.132	No	0.514	No	Yes
10	−3.836	0.454	60.16	−0.744	−1.876	No	No	−0.117	No	0.568	No	Yes

demonstrated promising therapeutic effects against viruses such as influenza A virus (Dai et al., 2012; Abookleesh et al., 2022). Consequently, natural evodiamine derivatives were considered for identifying potential therapeutic agents for Lassa fever treatment in the present study. In light of the comprehensive findings, the ingeniously designed evodiamine demonstrates promising potential in combating both the Lassa virus glycoprotein spike and the Lassa virus nucleoprotein, with favorable ADMET and drug-likeness properties and molecular dynamic simulations validating their stability.

5. Conclusion

Lassa fever, a neglected tropical disease, lacks FDA-approved vaccines and has limited treatment options. This study aimed to identify promising therapeutic candidates, focusing on evodiamine derivatives from the PubChem database. Using the PyRx application, computational docking was performed, followed by theoretical bioavailability and toxicological predictions via pkCSM and Swiss ADME tools. The results revealed that most selected inhibitors demonstrated favorable binding energies with the Lassa virus glycoprotein spike and nucleoprotein. Molecular dynamic simulations supported the stability of the protein–ligand complexes. Meeting Lipinski's criteria, ADME analysis indicated that the investigated inhibitors were generally safe. However, patients with impaired liver function should exercise caution due to potential hepatotoxicity. Overall, evodiamine derivatives showed potential as inhibitors against Lassa virus glycoprotein spike and nucleoprotein, warranting further wet lab validation to confirm these *in silico* findings.

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

SA and NM: conceptualization, writing original draft, and analysis. JB, SM, NM, VS, MI, and GG: initial draft and methodology. VR, SC, GR, and FS: visualization. GA, AS, and MA-D: revised and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R30), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This research was funded by the Ministry of Higher Learning Malaysia under the Fundamental Research Grant Scheme, FRGS/1/2018/STG03/AIMST/02/1. All authors would like to acknowledge their respective departments to conduct the study.

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.

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

- Abookleesh, F. L., Al-Anzi, B. S., and Ullah, A. (2022). Potential antiviral action of alkaloids. *Molecules* 27, 903. doi: 10.3390/molecules27030903
- Agbonlahor, D. E., Akpede, G. O., Happi, C. T., and Tomori, O. (2021). Years of lassa fever outbreaks in Nigeria, 1969–2020: an epidemiologic analysis of the temporal and spatial trends. *Am. J. Trop. Med. Hyg* 105, 974–985. doi: 10.4269/ajtmh.20-1160
- Alqahtani, S. (2017). In silico ADME-Tox modeling: progress and prospects. *Expert. Opin. Drug. Metab. Toxicol.* 13, 1147–1158. doi: 10.1080/17425255.2017.1389897
- Andersen, K. G., Shapiro, B. J., Matranga, C. B., Sealfon, R., Lin, A. E., Moses, L. M., et al. (2015). Clinical sequencing uncovers origins and evolution of Lassa virus. *Cell* 162, 738–750. doi: 10.1016/j.cell.2015.07.020
- Asogun, D. A., Günther, S., Akpede, G. O., Ihekweazu, C., and Zumla, A. (2019). Lassa fever: epidemiology, clinical features, diagnosis, management and prevention. *Inf. Dis. Clin* 33, 933–951. doi: 10.1016/j.idc.2019.08.002
- Atkin, S., Anaraki, S., Gothard, P., Walsh, A., Brown, D., Gopal, R., et al. (2009). The first case of Lassa fever imported from Mali to the United Kingdom, February 2009. *Eurosurveillance* 14, 10. doi: 10.2807/es.14.10.19145-en
- Azeez-Akande, O. (2016). Review of Lassa fever, an emerging old world haemorrhagic viral disease in sub-Saharan Africa. *Afr. J. Clin. Exp. Microbiol* 17, 282–289. doi: 10.4314/ajcem.v17i4.9
- Baize, S., Marianneau, P., Loth, P., Reynard, S., Journeaux, A., Chevallier, M., et al. (2009). Early and strong immune responses are associated with control of viral replication and recovery in lassa virus-infected cynomolgus monkeys. *J. Virol* 83, 5890–5903. doi: 10.1128/JVI.01948-08

- Baral, P., Pavada, E., Gerstman, B. S., and Chapagain, P. P. (2020). In-silico identification of the vaccine candidate epitopes against the Lassa virus hemorrhagic fever. *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-63640-1
- Belal, A., Elsayed, A., Gharib, A. F., Ali Alqarni, M. A., Soliman, A. M., Mehany, A. B., et al. (2022). Toward the discovery of SARS-CoV-2 main protease inhibitors: exploring therapeutic potentials of evodiamine and its derivatives, virtual screening, molecular docking, and molecular dynamic studies. *Nat. Prod. Commun.* 17, 1934578X.221143082. doi: 10.1177/1934578X221143082
- Borenstein-Katz, A., Shulman, A., Hamawi, H., Leitner, O., and Diskin, R. (2019). Differential antibody-based immune response against isolated GP1 receptor-binding domains from Lassa and Junin viruses. *J. Virol.* 93, e00090–e00019. doi: 10.1128/JVI.00090-19
- Bowen, M. D., Rollin, P. E., Ksiazek, T. G., Hustad, H. L., Bausch, D. G., Demby, A. H., et al. (2000). Genetic diversity among Lassa virus strains. *J. Virol.* 74, 6992–7004. doi: 10.1128/JVI.74.15.6992-7004.2000
- 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, 1668–1688. doi: 10.1002/jcc.20290
- CFDCA Prevention (2022). *Lassa. Fever*. Available online at: <https://www.cdc.gov/vhf/lassa/index.html#:~:text=Lassa%20fever%20is%20an%20a> (accessed April 26, 2022).
- Dai, J. P., Li, W. Z., Zhao, X. F., Wang, G. F., Yang, J. C., Zhang, L., et al. (2012). A drug screening method based on the autophagy pathway and studies of the mechanism of evodiamine against influenza A virus. doi: 10.1371/journal.pone.0042706
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 42717. doi: 10.1038/srep42717
- Dallakyan, S., and Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. *Chem. Biol. Methods. Prot.* 2, 243–250. doi: 10.1007/978-1-4939-2269-7_19
- Design, L. J. B. C. (2014). Pharmacophore and ligand-based design with Biovia Discovery Studio,®.
- Ertl, P., Rohde, B., and Selzer, P. (2000). Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* 43, 3714–3717. doi: 10.1021/jm000942e
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., Pedersen, L. G., et al. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593. doi: 10.1063/1.470117
- Fan, X., Deng, J., Shi, T., Wen, H., Li, J., Liang, Z., et al. (2021). Design, synthesis and bioactivity study of evodiamine derivatives as multifunctional agents for the treatment of hepatocellular carcinoma. *Bioorg. Chem.* 114, 105154. doi: 10.1016/j.bioorg.2021.105154
- Filipski, K. K., Mathijssen, R. H., Mikkelsen, T. S., Schinkel, A. H., and Sparreboom, A. (2009). Contribution of organic cation transporter 2 (OCT2) to cisplatin-induced nephrotoxicity. *Clin. Pharm. Therap.* 86, 396–402. doi: 10.1038/clpt.2009.139
- Gavaraskar, K., Dhulap, S., and Hirwani, R. R. (2015). Therapeutic and cosmetic applications of Evodiamine and its derivatives—A patent review. *Fitoterapia* 106, 22–35. doi: 10.1016/j.fitote.2015.07.019
- Hass, M., Glnitz, U., Mller, S., Becker-Zajia, B., and Gnther, S. (2004). Replicon system for Lassa virus. *J. Virol.* 78, 13793–13803. doi: 10.1128/JVI.78.24.13793-13803.2004
- Humphrey, W., Dalke, A., and Schulten, K. (1996a). J. Mol. Graphics 14, 33–38.
- Humphrey, W., Dalke, A., and Schulten, K. (1996b). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5
- Igonet, S., Vaney, M. C., Vonnrhein, C., Bricogne, G., Stura, E. A., Hengartner, H., et al. (2011). X-ray structure of the arenavirus glycoprotein GP2 in its postfusion hairpin conformation. *Proc. Nat. Acad. Sci.* 108, 19967–19972. doi: 10.1073/pnas.1108910108
- Jakhar, R., Dangi, M., Khichi, A., and Chhillar, A. K. (2020). Relevance of molecular docking studies in drug designing. *Curr. Bioinf.* 15, 270–278. doi: 10.2174/1574893615666191219094216
- Kaplan, W., and Littlejohn, T. G. (2001). Swiss-PDB viewer (deep view). *Briefings. Bioinf.* 2, 195–197. doi: 10.1093/bib/2.2.195
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2019). PubChem 2019 update: improved access to chemical. *Data.* 47, D1102–D1109. doi: 10.1093/nar/gky1033
- Lenz, O., ter Meulen, J., Klenk, H. D., Seidah, N. G., and Garten, W. (2001). The Lassa virus glycoprotein precursor GP-C is proteolytically processed by subtilase SKI-1/S1P. *Proc. Nat. Acad. Sci.* 98, 12701–12705. doi: 10.1073/pnas.221447598
- Li, S., Sun, Z., Pryce, R., Parsy, M. L., Fehling, S. K., Schlie, K., et al. (2016). Acidic pH-induced conformations and LAMP1 binding of the Lassa virus glycoprotein spike. *PLoS Pathog.* 12, e1005418. doi: 10.1371/journal.ppat.1005418
- Li, X., Ge, J., Zheng, Q., Zhang, J., Sun, R., Liu, R., et al. (2020). Evodiamine and rutaecarpine from *Tetradium rutilcarpum* in the treatment of liver diseases. *Phytomedicine* 68, 153180. doi: 10.1016/j.phymed.2020.153180
- Liang, Z., Lei, F., Deng, J., Zhang, H., Wang, Y., Li, J., et al. (2022). Design, synthesis and bioactivity evaluation of novel evodiamine derivatives with excellent potency against gastric cancer. *Eur. J. Med. Chem.* 228, 113960. doi: 10.1016/j.ejmech.2022.114530
- Lipinski, C. A. (2004). Lead-and drug-like compounds: the rule-of-five revolution. *Drug. Discovery. Today. Technol.* 1, 337–341. doi: 10.1016/j.ddtec.2004.11.007
- Mahanty, S., Hutchinson, K., Agarwal, S., Mcrae, M., Rollin, P. E., Pulendran, B., et al. (2003). Cutting edge: impairment of dendritic cells and adaptive immunity by Ebola and Lassa viruses. *J. Immunol.* 170, 2797–2801. doi: 10.4049/jimmunol.170.6.2797
- Mart-nez-Sobrido, L., Giannakas, P., Cubitt, B., and Garc-a-Sastre, A., and de la Torre, J. C. (2007). Differential inhibition of type I interferon induction by arenavirus nucleoproteins. *J. Virol.* 81, 12696–12703. doi: 10.1128/JVI.00882-07
- Mart-nez-Sobrido, L., Ziga, E. I., Rosario, D., and Garc-a-Sastre, A., and de la Torre, J. C. (2006). Inhibition of the type I interferon response by the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus. *J. Virol.* 80, 9192–9199. doi: 10.1128/JVI.00555-06
- McCormick, J. B., and Fisher-Hoch, S. P. (2002). Lassa fever. *Arenaviruses I: the Epidemiology. Mol. Cell. Biol. Arenaviruses* 4, 75–109. doi: 10.1007/978-3-642-56029-3_4
- Meyer, B. J., Torre, D., and Southern, J. C. (2002). Arenaviruses: genomic RNAs, transcription, and replication. *Mol. Cell. Biol. Arenaviruses* 2, 139–157. doi: 10.1007/978-3-642-56029-3_6
- Obot, I. B., Macdonald, D. D., and Gasem, Z. M. (2015). Density functional theory (DFT) as a powerful tool for designing new organic corrosion inhibitors. Part 1: an overview. *Corrosion. Sci.* 99, 1–30. doi: 10.1016/j.corsci.2015.01.037
- Ogbu, O., Ajuluchukwu, E., and Uneke, C. J. (2007). Lassa fever in West African sub-region: an overview. *J. Vector. Borne. Dis.* 44, 1.
- Papajak, E., and Truhlar, D. G. (2010). Efficient diffuse basis sets for density functional theory. *J. Chem. Theor. Comput.* 6, 597–601. doi: 10.1021/ct900566x
- Passeleu-Le Bourdonnec, C., Carrupt, P. A., Scherrmann, J. M., and Martel, S. (2013). Methodologies to assess drug permeation through the blood–brain barrier for pharmaceutical research. *Pharm. Res.* 30, 2729–2756. doi: 10.1007/s11095-013-1119-z
- Pires, D. E., Blundell, T. L., and Ascher, D. B. (2015). pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* 58, 4066–4072. doi: 10.1021/acs.jmedchem.5b00104
- Qi, X., Lan, S., Wang, W., Schelde, L. M., Dong, H., Wallat, G. D., et al. (2010). Cap binding and immune evasion revealed by Lassa nucleoprotein structure. *Nature* 468, 779–783. doi: 10.1038/nature09605
- Rahman, M. M., Karim, M. R., Ahsan, M. Q., Khalifa, A. B. R., Chowdhury, M. R., Saifuzzaman, M., et al. (2012). Use of computer in drug design and drug discovery: a review. *Int. J. Pharm. Life. Sci.* 1, 1–15. doi: 10.3329/ijpls.v1i1.12955
- Ribeiro, R. A. P., Lazaro, D., and Pianaro, S. R. (2015). Density functional theory applied to magnetic materials: Mn₃O₄ at different hybrid functionals. *J. Mag. Magnet. Mat.* 391, 166–171. doi: 10.1016/j.jmmm.2015.04.091
- Rojek, J. M., Sanchez, A. B., and Nguyen, N. T., de la Torre, J. C., and Kunz, S. (2008). Different mechanisms of cell entry by human-pathogenic old world and new world arenaviruses. *J. Virol.* 82, 7677–7687. doi: 10.1128/JVI.00560-08
- Rose, P. W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., et al. (2016). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic. Acids. Res.* 5, gkw1000.
- Sachita, K., Kim, Y., Yu, H. J., Cho, S. D., and Lee, J. S. (2015). In vitro assessment of the anticancer potential of evodiamine in human oral cancer cell lines. *Phytotherapy. Res.* 29, 1145–1151. doi: 10.1002/ptr.5359
- Sambuy, Y., Angelis, D., e., Ranaldi, I., Scarino, G., and Stammati, M. L. A., and Zucco, F. (2005). The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell. Biol. Toxicol.* 21, 1–26. doi: 10.1007/s10565-005-0085-6
- Sharma, S., Kumar, P., and Chandra, R. (2019). Applications of BIOVIA materials studio, LAMMPS, and GROMACS in various fields of science and engineering. *Mol. Dyn. Simulation. Nanocomp. BIOVIA. Mat. Studio. Lammops. Gromacs* 5, 329–341. doi: 10.1016/B978-0-12-816954-4.00007-3
- Shukla, R., and Tripathi, T. (2020). Molecular dynamics simulation of protein and protein–ligand complexes. *Comput. Aided. Drug. Design* 5, 133–161. doi: 10.1007/978-981-15-6815-2_7
- Tan, Q., and Zhang, J. (2016). Evodiamine and its role in chronic diseases. *Drug. Discovery. Mother. Nature* 5, 315–328. doi: 10.1007/978-3-319-41342-6_14
- Tewogbola, P., and Aung, N. (2020). Lassa fever: history, causes, effects, and reduction strategies. *Virus* 2, 16. doi: 10.14202/IJOH.2020.95-98
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization,

- and multithreading. *J. Comput. Chem* 31, 455–461. doi: 10.1002/jcc.21334
- Vanguard (2023). *Nigeria. records. 189. Lassa. Fever. Deaths.* —. NCDC. Available online at: <https://www.vanguardngr.com/2023/01/nigeria-records-189-lassa-fever-deaths-ncdc/> (accessed January 9, 2023).
- Wang, D., Wang, C., Liu, L., and Li, S. (2018). Protective effects of evodiamine in experimental paradigm of Alzheimer's disease. *Cognitive. Neurodyn* 12, 303–313. doi: 10.1007/s11571-017-9471-z
- Wang, T., Wang, Y., Kontani, Y., Kobayashi, Y., Sato, Y., Mori, N., et al. (2008). Evodiamine improves diet-induced obesity in a uncoupling protein-1-independent manner: involvement of antiadipogenic mechanism and extracellularly regulated kinase/mitogen-activated protein kinase signaling. *Endocrinology* 149, 358–366. doi: 10.1210/en.2007-0467
- Wei, W. T., Chen, H., Wang, Z. H., Ni, Z. L., Liu, H. B., Tong, H. F., et al. (2012). Enhanced antitumor efficacy of gemcitabine by evodiamine on pancreatic cancer via regulating PI3K/Akt pathway. *Int. J. Biol. Sci* 8, 1. doi: 10.7150/ijbs.8.1
- Wen, Z., Feng, S., Wei, L., Wang, Z., Hong, D., Wang, Q., et al. (2015). Evodiamine, a novel inhibitor of the Wnt pathway, inhibits the self-renewal of gastric cancer stem cells. *Int. J. Mol. Med* 36, 1657–1663. doi: 10.3892/ijmm.2015.2383
- Yang, F., Shi, L., Liang, T., Ji, L., Zhang, G., Shen, Y., et al. (2017). Anti-tumor effect of evodiamine by inducing Akt-mediated apoptosis in hepatocellular carcinoma. *Biochem. Biophys. Res. Commun* 485, 54–61. doi: 10.1016/j.bbrc.2017.02.017
- Yang, W., Ma, L., Li, S., Cui, K., Lei, L., Ye, Z., et al. (2017). Evaluation of the cardiotoxicity of evodiamine in vitro and in vivo. *Molecules* 22, 943. doi: 10.3390/molecules22060943
- Yun, N. E., and Walker, D. H. (2012). Pathogenesis of Lassa fever. *Viruses* 4, 2031–2048. doi: 10.3390/v4102031
- Zhang, T., Lai, Z., Yuan, S., Huang, Y. Y., Dong, G., Sheng, C., et al. (2020). Discovery of evodiamine derivatives as highly selective PDE5 inhibitors targeting a unique allosteric pocket. *J. Med. Chem* 63, 9828–9837. doi: 10.1021/acs.jmedchem.0c00983
- Zhang, Y., Wang, J., Wang, C., Li, Z., Liu, X., Zhang, J., et al. (2018). Pharmacological basis for the use of evodiamine in Alzheimer's disease: antioxidation and antiapoptosis. *Int. J. Mol. Sci* 19, 1527. doi: 10.3390/ijms19051527
- Zhang, Y., Zhang, Y., Zhao, Y., Wu, W., Meng, W., Zhou, Y., et al. (2022). Protection against ulcerative colitis and colorectal cancer by evodiamine via anti-inflammatory effects. *Mol. Med. Rep* 25, 1–14. doi: 10.3892/mmr.2022.12704
- Zhao, L. C., Li, J., Liao, K., Luo, N., Shi, Q. Q., Feng, Z. Q., et al. (2015). Evodiamine induces apoptosis and inhibits migration of HCT-116 human colorectal cancer cells. *Int. J. Mol. Sci* 16, 27411–27421. doi: 10.3390/ijms161126031
- Zou, Y., Qin, X., Xiong, H., Zhu, F., Chen, T., Wu, H., et al. (2015). Apoptosis of human non-small-cell lung cancer A549 cells triggered by evodiamine through MTDH-dependent signaling pathway. *Tumor. Biol* 36, 5187–5193. doi: 10.1007/s13277-015-3174-z



OPEN ACCESS

EDITED BY

K. M. Kumar,
Pondicherry University, India

REVIEWED BY

Joseph Atia Ayariga,
Alabama State University, United States
Mayank Gangwar,
Banaras Hindu University, India

*CORRESPONDENCE

Shopnil Akash
✉ shopnil.ph@gmail.com
Binata Nayak
✉ binatanayak@suniv.ac.in
Amare Mekonnen
✉ amarebitew8@gmail.com

RECEIVED 16 April 2023

ACCEPTED 20 June 2023

PUBLISHED 19 July 2023

CITATION

Akash S, Mir SA, Mahmood S, Hossain S,
Islam MR, Mukerjee N, Nayak B, Nafidi H-A, Bin
Jardan YA, Mekonnen A and Bourhia M (2023)
Novel computational and drug design
strategies for inhibition of monkeypox virus and
Babesia microti: molecular docking, molecular
dynamic simulation and drug design approach
by natural compounds.
Front. Microbiol. 14:1206816.
doi: 10.3389/fmicb.2023.1206816

COPYRIGHT

© 2023 Akash, Mir, Mahmood, Hossain, Islam,
Mukerjee, Nayak, Nafidi, Bin Jordan, Mekonnen
and Bourhia. This is an open-access article
distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Novel computational and drug design strategies for inhibition of monkeypox virus and *Babesia microti*: molecular docking, molecular dynamic simulation and drug design approach by natural compounds

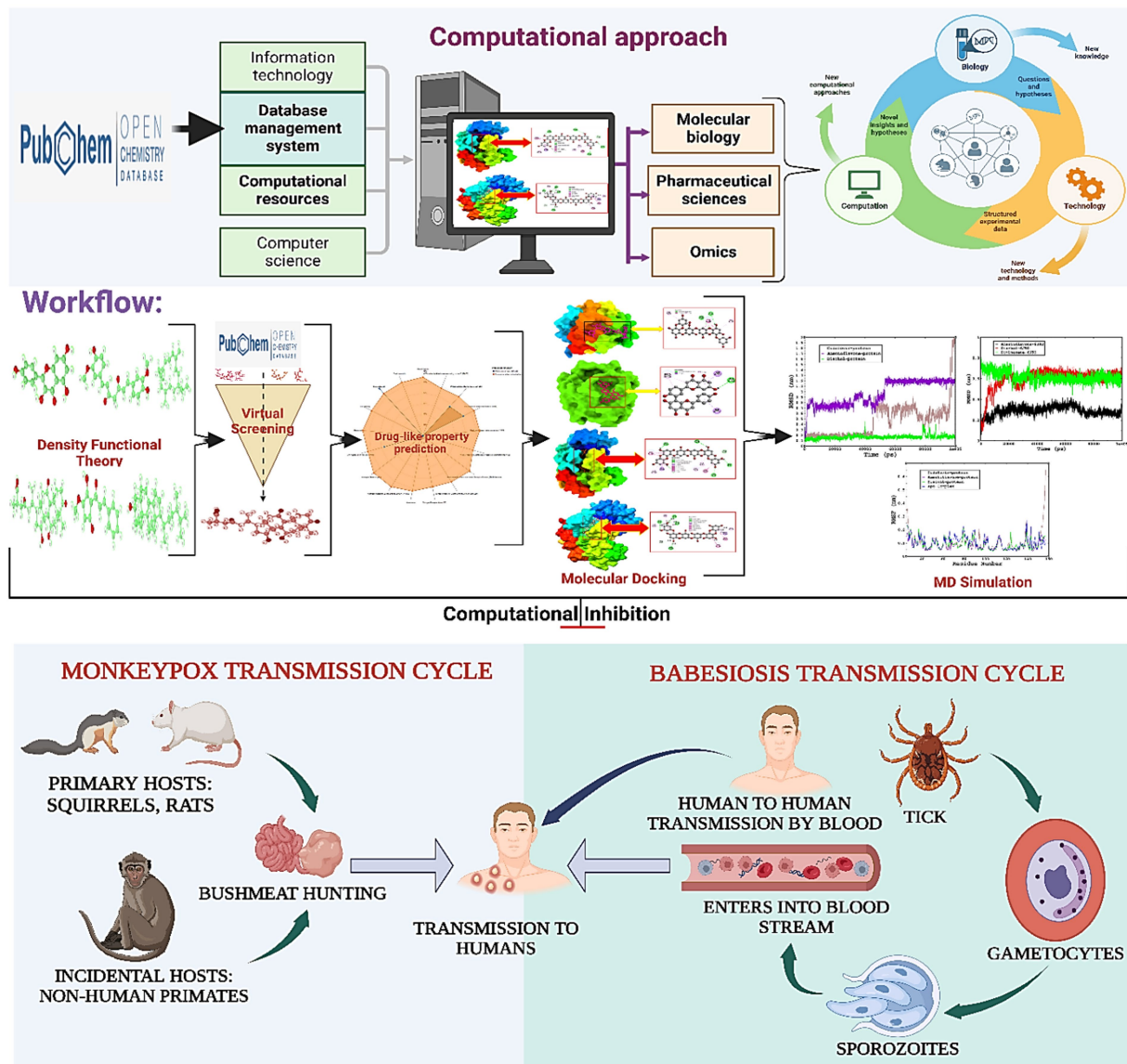
Shopnil Akash^{1*}, Showkat Ahmad Mir², Sajjat Mahmood³,
Saddam Hossain⁴, Md. Rezaul Islam¹, Nobendu Mukerjee⁵,
Binata Nayak^{2*}, Hiba-Allah Nafidi⁶, Yousef A. Bin Jordan⁷,
Amare Mekonnen^{8*} and Mohammed Bourhia⁹

¹Department of Pharmacy, Faculty of Allied Health Sciences, Daffodil International University, Dhaka, Bangladesh, ²School of Life Sciences, Sambalpur University, Sambalpur, Odisha, India, ³Department of Microbiology, Jagannath University, Dhaka, Bangladesh, ⁴Department of Biomedical Engineering, Faculty of Engineering and Technology, Islamic University, Kushtia, Bangladesh, ⁵Department of Microbiology, West Bengal State University, Kolkata, West Bengal, India, ⁶Department of Food Science, Faculty of Agricultural and Food Sciences, Laval University, Quebec City, QC, Canada, ⁷Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia, ⁸Department of Biology, Bahir Dar University, Bahir Dar, Ethiopia, ⁹Department of Chemistry and Biochemistry, Faculty of Medicine and Pharmacy, Ibn Zohr University, Laayoune, Morocco

Background: The alarming increase in tick-borne pathogens such as human *Babesia microti* is an existential threat to global public health. It is a protozoan parasitic infection transmitted by numerous species of the genus *Babesia*. Second, monkeypox has recently emerged as a public health crisis, and the virus has spread around the world in the post-COVID-19 period with a very rapid transmission rate. These two novel pathogens are a new concern for human health globally and have become a significant obstacle to the development of modern medicine and the economy of the whole world. Currently, there are no approved drugs for the treatment of this disease. So, this research gap encourages us to find a potential inhibitor from a natural source.

Methods and materials: In this study, a series of natural plant-based biomolecules were subjected to in-depth computational investigation to find the most potent inhibitors targeting major pathogenic proteins responsible for the diseases caused by these two pathogens.

Results: Among them, most of the selected natural compounds are predicted to bind tightly to the targeted proteins that are crucial for the replication of these novel pathogens. Moreover, all the molecules have outstanding ADMET properties such as high aqueous solubility, a higher human gastrointestinal absorption rate, and a lack of any carcinogenic or hepatotoxic effects; most of them followed Lipinski's rule. Finally, the stability of the compounds was determined by molecular dynamics simulations (MDs) for 100ns. During MDs, we observed that the mentioned compounds have exceptional stability against selected pathogens.



GRAPHICAL ABSTRACT

Conclusion: These advanced computational strategies reported that 11 lead compounds, including dieckol and amentoflavone, exhibited high potency, excellent drug-like properties, and no toxicity. These compounds demonstrated strong binding affinities to the target enzymes, especially dieckol, which displayed superior stability during molecular dynamics simulations. The MM/PBSA method confirmed the favorable binding energies of amentoflavone and dieckol. However, further *in vitro* and *in vivo* studies are necessary to validate their efficacy. Our research highlights the role of Dieckol and Amentoflavone as promising candidates for inhibiting both monkeypox and *Babesia microti*, demonstrating their multifaceted roles in the control of these pathogens.

KEYWORDS

molecular docking, molecular dynamics simulations, drug-likeness, admet, DFT, *Babesia microti*, monkeypox virus

1. Introduction

Human babesiosis is a newly recognized tick-borne infection that is caused by an intraerythrocytic protozoan species belonging to the genus *Babesia* (Chand et al., 2022). In recent decades, the epidemiology of human babesiosis has shifted from a few isolated cases to the emergence of outbreaks in the northeastern and midwestern United States, and research has shown that human babesiosis develops within the red blood cells of humans and small rodents (Tanowitz and Weiss, 2009). There have been more than 100 human occurrences documented in the United States, with the most severe infections occurring in individuals who already have compromised immune systems (Djokic et al., 2018; Kumar et al., 2021). Another investigation suggested that *Babesia microti* is the most prominent causative agent of babesiosis in humans in the United States, especially in the Northeast and upper Midwest, where the disease is naturally occurring. *Babesia* parasites were first identified in 1888 in Romanian cattle (Vannier and Gelfand, 2020). The first human case of babesiosis was documented in the territory of the former Yugoslavia in 1957, and the second human case was reported in California in 1968 (Rosner et al., 1984). A year later, a third patient with babesiosis was discovered, and the species responsible for the disease was found to be *B. microti*. The patient was a native of Nantucket, which is located in the state of Massachusetts, and babesiosis was quickly identified as an epidemic ailment on the island (Kumar et al., 2021).

Human babesiosis can cause acute respiratory distress syndrome, hemolytic anemia, multiple organ failure, and death. While the parasite is transmitted to humans mainly by the bite of an infected tick, a growing number of instances of human-to-human transmission through blood transfusion have been documented (Mohr et al., 2000; Chiu et al., 2021).

More than one hundred species of *Babesia* have been reported, and these parasites can infect a wide variety of wild and domestic animals. Babesiosis is of major concern and pathogenicity, particularly in cattle, and has had a significant economic impact in several cattle-producing nations (Spielman, 1994; Kumar et al., 2021). *Babesia veratrum* is the principal species identified as a human pathogen. Several different genetically similar pathogen substrains have been documented to infect humans. These include the *Babesia divergens*-like and the *Babesia microti*-like viruses (Hunfeldt et al., 2008; Kumar et al., 2021). According to the most recent statistics from the U.S. Centers for Disease Control and Prevention (CDC), more than 16,000 cases of babesiosis have been documented in the United States between 2011 and 2019, with the majority of confirmed cases in the Northeast. Most instances were recorded in New York, Massachusetts, and Connecticut during this period (Yannielli and Alcamo, 2009; U.S.NEWS, 2023). CDC researchers have described babesiosis as not endemic in Maine, New Hampshire, or Vermont, but these states have experienced increases comparable to or greater than those observed in areas where the infection is endemic (U.S.NEWS, 2023).

Furthermore, the viral zoonotic disease known as monkeypox is caused by the MPOX virus, which is related to the variola virus (which causes smallpox). Skin lesions or rashes that are typically limited to the face, hands, and feet are the hallmarks of the monkeypox infection. In 1970, a person in the Democratic Republic of the Congo was identified as the first human case of monkeypox. The subject was nine months old, and the incident occurred in a region of the country where smallpox had been eradicated as recently as 1968. Since then, the vast majority of reports have come from remote, tropical areas of the Congo Basin, primarily in the Democratic Republic of the Congo, with increasing evidence that the

disease is spreading throughout Central and West Africa (Parker et al., 2007; Farasani, 2022). In 1970, clinical isolates of monkeypox were found in 11 African countries, namely Benin, Cameroon, the Central African Republic, the Democratic Republic of the Congo, Gabon, Liberia, Nigeria, Sierra Leone, and South Sudan (Beer and Rao, 2019). The potential severity of monkeypox in humans is still unknown. For example, in 1996 and 1997, the Democratic Republic of the Congo experienced an outbreak with unusually high incidence rates but a lower-than-usual case fatality rate. Since 2017, more than 500 new cases, 200 of which were confirmed, and a case fatality rate of approximately 3% have all been reported in Nigeria (Yinka-Ogunleye et al., 2018; Antunes et al., 2022). Thus, monkeypox is currently considered a threat not only to countries in West and Central Africa but to the entire world. This means that it is a significant public health concern worldwide.

In 2003, the United States of America became the first country outside of Africa to experience an outbreak of monkeypox (Sah et al., 2022). Prairie dogs kept as pets were identified as the source of this infection. These pets shared a cage with dormice and Gambian pouched rats illegally imported from Ghana. More than 70 cases of monkeypox have been identified in the United States as a result of this epidemic (Kabuga and El Zowalaty, 2019). Travelers from Nigeria have also been reported to have developed monkeypox in Israel in September 2018, in the United Kingdom in September 2018, in Singapore in December 2019, May 2021, and May 2022 (Adegboye et al., 2022), and in the United States of America in July and November 2021 (Kumar et al., 2022). In May 2022, multiple occurrences of monkeypox were detected in different countries. Since the beginning of May 2022, more than 3,000 cases of monkeypox virus infection have been documented in approximately 50 countries in five regions. These findings prompted the World Health Organization (WHO) to declare monkeypox an “emerging global public concern” on 23 June 2022 (Thornhill et al., 2022).

Although human babesiosis and monkeypox are both life-threatening conditions, there are currently no effective therapies or vaccines available to combat these diseases. Therefore, a viable drug to control human babesiosis and the monkeypox infection is urgently required to prevent another pandemic like SAR CoV-2 (Kumar et al., 2022; Nolasco et al., 2022). Thus, this study intends to explore potentially valuable drugs derived from natural sources. Since nature is regarded as a fantastic source of cures for all kinds of diseases (Jafari Porzani et al., 2022), in this case, *in silico* methods were chosen, and different drug design approaches were applied to establish them as potential candidates.

In silico strategies provide a framework for assessing the function of potential therapeutics against specific biological targets, which enables the selection of those with the best possible drug candidate for further *in vitro* and *in vivo* studies (Vougas et al., 2019). *In silico* techniques can also be used to monitor the function of existing therapeutics against biological targets. These techniques have reduced the time and cost of novel drug development by minimizing the use of resources in laboratory testing (Yu et al., 2022). The investigation of the effectiveness of natural product-derived drugs against monkeypox and MERS CoV-2 viral proteins is one of the areas where *in silico* approaches are beneficial. In the field of biomedicine, the development of *in silico* assays has proven to be quite beneficial, including molecular docking, molecular dynamics simulation, and ADMET analysis, which are both straightforward and reliable. Thus, this innovative research was conducted to investigate the efficacy of natural inhibitors against the monkeypox virus and *B. microti* (Tabti et al., 2022).

2. Materials and methods

2.1. Ligand preparation and molecular optimization

Calculation of thermodynamic, molecular orbital, and molecular electrostatic characteristics is essential to computational chemistry, and quantum mechanical techniques are often used in this field. Gaussian 09 software was used to refine and optimize the geometry of selected natural molecules (Gaussian et al., 2009). Then, the optimization process was carried out using DFT (3-21G) with Becke's and Lee, Yang, and Parr's (LYP) (B) functional theory. During the optimization process, water was used as the solvent medium. After optimization, all compounds were saved in SDF format for further computational work (Amin et al., 2022). Finally, the optimized structures were viewed in Material Studio 08, and the 3D structure was captured as shown in Figure 1.

2.2. Protein preparation and molecular docking study and visualization

The crystal structures of *Babesia microti* lactate dehydrogenase (PDB ID 6J9D), *Babesia microti* lactate dehydrogenase apo-form (PDB ID 6K12), monkeypox virus profilin-like protein (PDB ID 4QWO), and monkeypox virus DNA polymerase (PDB: 8HG1) were acquired from the RCSB Protein Data Bank (<https://www.rcsb.org/>; Burley et al., 2017). The three-dimensional protein structures were imported into the BIOVIA Discovery Studio Visualizer software to remove water molecules and unwanted heteroatoms (Figure 2). Water molecules often have no role in the substrate's ability to bind to the receptor. They were thus removed to accelerate computations and free up the binding site. Additionally, the intended active site of the receptor may be occupied by previously docked, unwanted heteroatoms. As a consequence, they were removed to free up the active site and speed up computations. Swiss PDB Viewer v.4.10 was employed for energy minimization (Sharma et al., 2019). PyRx version 0.8 virtual screening tools were used for docking with AutoDock Vina (Dallakyan and Olson, 2015). Finally, the docking results, complex structure, binding affinity, non-binding interactions, and binding pocket were visualized using the BIOVIA Discovery Studio Visualizer 4.5 software tools (Nath et al., 2021).

2.3. Determination of ADMET, Lipinski's rule, and pharmacokinetics

Drug-like and non-drug-like compounds were differentiated using Lipinski's rule of five. Drug-like criteria were used to more efficiently determine a molecule's drug-like qualities in its structural characteristics (Walters et al., 1999; Walters and Murcko, 2002). Important pharmacokinetic properties such as hydrogen bond acceptor, hydrogen bond donor, TPSA, bioavailability, molecular weight, and consensus log Po/w were estimated using SwissADME (<http://www.swissadme.ch/index.php>; Daina et al., 2017). Lipinski's Rule was also determined using the same web server.

ADMET stands for Absorption, Distribution, Metabolism, Excretion, and Toxicity. These pharmacological properties are determined for each drug candidate. Drug development significantly

depends on ADMET characteristics, with 50% of drugs failing because they violate these pharmacokinetic principles (Li, 2001). *In silico* ADMET studies were performed using an online web tool server called pkCSM (<https://biosig.lab.uq.edu.au/pkcsm/prediction>; Pires et al., 2015). Several pharmacokinetic parameters such as water solubility, coca-2 permeability, human intestinal absorption, Blood–Brain Barrier (BBB) penetration, cytochrome P450 inhibition and substrate, AMES toxicity, skin sensitization, and hepatotoxicity levels were calculated for selected natural compounds.

2.4. Calculation of QSAR and pIC₅₀

The abbreviation of QSAR is Quantitative Structure–Activity Relationship, which establishes a relationship between the chemical structure and the biological activity of chemical compounds (Miladiyah et al., 2018). QSAR is a quantum chemistry method that predicts the efficacy of compounds in drug discovery and development (Devillers, 1996). To perform QSAR and pIC₅₀ we used a freely available website called ChemDes (http://www.scbdd.com/chemopy_desc/index/; Kumer et al., 2019). These web services provided data such as Chiv5 molecular connectivity, bcutm1 mean burden descriptors, MRVSA9, MRVSA6, and PEOVSA5 are MOE type descriptors and GATSV4 indicating autocorrelation descriptors, with the last two parameters J and diameter suggesting topological descriptors of drug molecules for reported ligands. To determine and calculate the QSAR and pIC₅₀, the mentioned parameters were first collected from the ChemDes database, and after developing the multiple linear regression (MLR) in an Excel sheet and calculating the pIC₅₀ value, the mentioned MLR equations were applied for calculating the pIC50 values.

$$\begin{aligned} \text{pIC}_{50} \quad (\text{Activity}) = & -2.768483965 + 0.133928895 \times (\text{Chiv5}) + \\ & 1.59986423 \times (\text{bcutm1}) + (-0.02309681) \times (\text{MRVSA9}) + (-0.002946101) \\ & \times (\text{MRVSA6}) + (0.00671218) \times (\text{PEOVSA5}) + (-0.15963415) \times (\text{GATSV4}) \\ & + (0.207949857) \times (\text{J}) + (0.082568569) \times (\text{Diameter}) \quad (\text{Siddiquey et al., 2022}). \end{aligned}$$

2.5. Molecular dynamics simulations

Molecular dynamics simulations were performed using the GROMACS package on the docked protein-ligand complex to determine structural stability and protein properties. The simulation was carried out for 100 ns in water, and the AMBER force field was used. Trajectory and energy files were noted every 2 fs (Mir et al., 2022).

For solvation purposes, we used truncated cubic boxes that contained TIP3P water molecules. The box dimensions and vectors were set to $3.256 \times 3.061 \times 3.142$ nm and $5.7 \times 5.7 \times 5.7$ nm, respectively. To effectively comply with the minimum image convention, the protein was centered in the simulation box at a minimum distance of 1 nm from the box edge. The entire system contained 56,242 atoms, and the simulation was executed in 0.15M KCL by adding the necessary potassium and chloride ions. Energy minimization is a critical step to avoid static clashes. Therefore, energy minimization was performed using the steepest descent method for 5,000 steps. The process was terminated when the total maximum force of the system reached $<1,000$ (KJ mol⁻¹ nm⁻¹), followed by a brief 100 ps (50,000 steps) equilibration in the NVT ensemble and then 1,000 ps (1,000,000 steps) in the NPT ensemble.

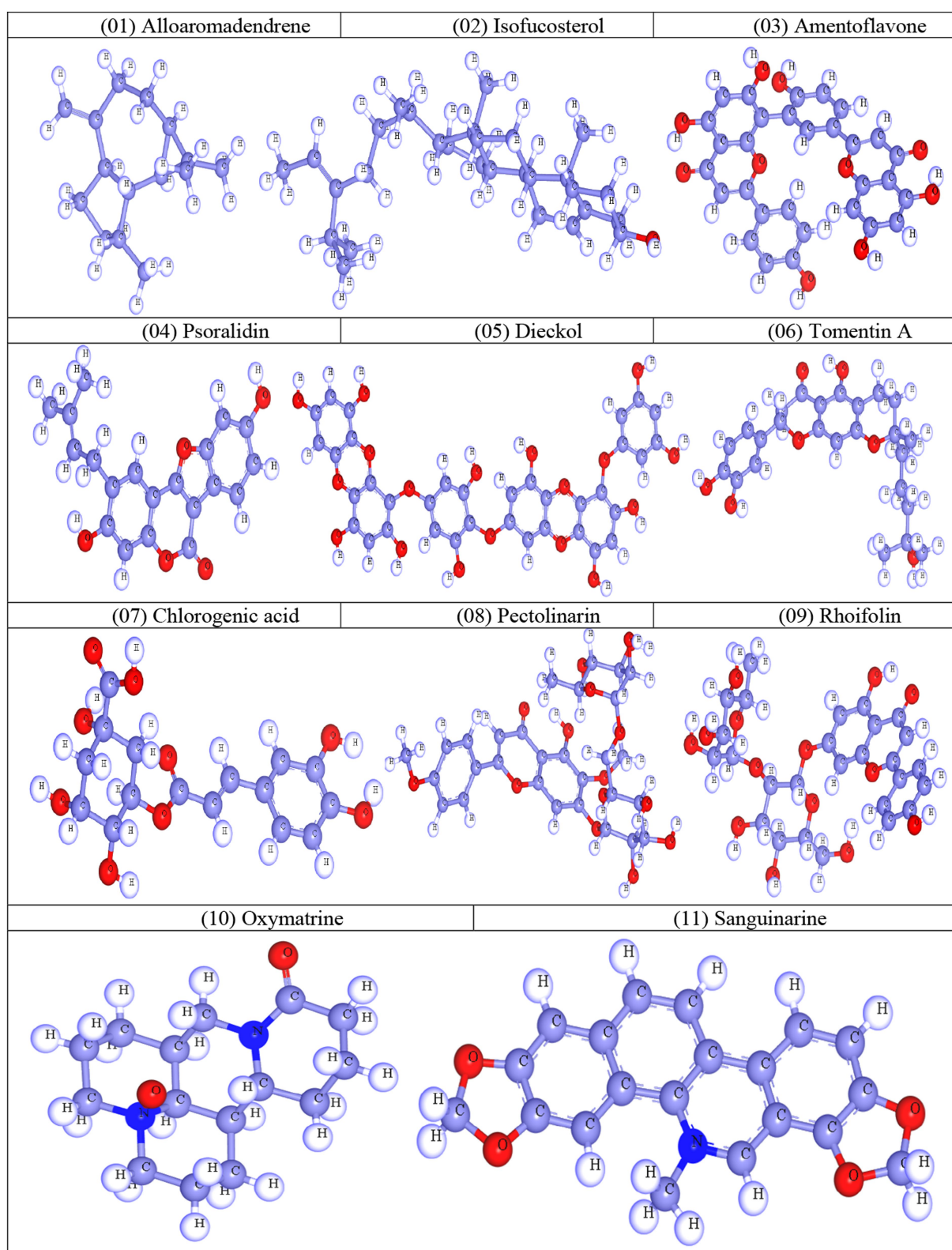


FIGURE 1
Optimized molecular structures of natural compounds.

Furthermore, a stable temperature and pressure of 300K and 1 atm were maintained using the Parrinello-Rahman algorithm for weak coupling velocity rescaling (modified Berendsen thermostat). The

relaxation intervals were set to $\tau_T = 0.2\text{ps}$ and $\tau_p = 1.0\text{ps}$. A Verlet scheme was applied to calculate non-bonded interactions. All x, y, and z directions used Periodic Boundary Conditions (PBC). Interactions

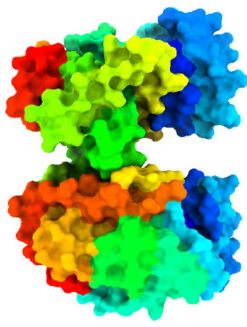
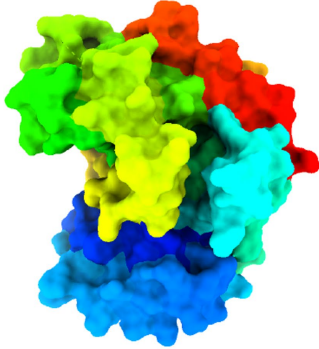


<i>Babesia microti</i> lactate dehydrogenase (PDB ID 6J9D)	<i>Babesia microti</i> lactate dehydrogenase apo-form (BmLDH) (PDB ID 6K12)	Monkeypox virus profilin-like protein (PDB ID 4QWO)	Monkeypox virus DNA polymerase (PDB: 8HG1)
			
Ref. [34]	Ref. [34]	Ref.[35]	Ref.[36]

FIGURE 2

Three-dimensional protein structure of the *monkeypox virus* (Minasov et al., 2014; Yu et al., 2019; Peng et al., 2022).

within a short-range threshold of 1.2 nm were calculated at each time step. The electrostatic interactions and forces for a homogeneous medium outside the long-range limit were calculated using Particle Mesh Ewald (PME). The trajectories generated during the 100 ns production run were utilized to calculate the radius of gyration (Rg), Root-Mean Square Deviation (RMSD), and Root-Mean Square Fluctuation (RMSF). All graphs were generated and visualized in Xmgrace.

2.6. Free energy calculations

The free binding was calculated using the traditional MM/PBSA method. The trajectories generated during the MD simulations were used for the free binding energy calculations. The ionic strength of the system was 0.150 M concentration with default grid dimensions. The non-polar solvation energy was calculated using the solvent-accessible surface area (SASA) model. The default values of solvent surface tension and SASA energy constant from previous MM/PBSA calculations of 0.0226778 kJ/mol Å² and 3.84982 kJ/mol, respectively, were also used. The average free binding energy of dieckol, amentoflavone, and the co-crystallized ligand cidofovir with monkeypox virus profilin-like protein and monkeypox virus DNA polymerase was determined using the MM/PBSA method (El-Barghouthi et al., 2009). The average free binding energy was calculated by the bootstrap method using MmPbSaStat.py. The decomposition energy of each amino acid contributing to the free energy binding was calculated using the MmPbSaDecomp.py module of *g_mmpbsa*. The ΔG_{bind} was calculated according to the following equation (1).

$${}^{\circ}G_{\text{binding}} = \Delta G_{\text{complex}} - (\Delta G_{\text{protein}}) + (\Delta G_{\text{ligand}}) \quad (1)$$

The above-mentioned method has been implemented by several authors to calculate the free binding energy of molecular scaffolds with template targets (Mir and Nayak, 2021; Mir et al., 2022, 2023).

2.7. DFT calculation

Optimizations of the selected phytochemicals were performed by use of functional B3LYP and basis set 6-311G++ of Gaussian 09v program (Kashyap et al., 2021; Kumer et al., 2022). The electronegative atom, oxygen, was assumed to be common to produce accurate results. Once the geometric optimization was done, molecular frontier orbital diagrams were identified: HOMO and LUMO. The HOMO and LUMO orbitals and their corresponding magnitudes were created using vibrational frequencies from the visualization interface of the Gaussian 09 program. The LUMO and HOMO were calculated from these frontier orbital energies, and the energy gap (E gap) was determined. Finally, the hardness (η), and softness (S) were observed through the DFT approach, which refers to the behavior of the molecule.

3. Analysis of the results

3.1. Lipinski's rule and pharmacokinetics

The drug-like assessment is a qualitative approach used to develop drugs or drug-like substances with respect to various parameters, such as bioavailability. In addition, the field of pharmacokinetics describes what happens to a chemical after it is absorbed by a living organism. Drug-likeness methods and Lipinski's rule of five help predict pharmacokinetic parameters based on the structure of the compound (Lipinski, 2004). The majority of the selected compounds in the current study, excluding dieckol, pectolinarin, and rhoifolin adheres to Lipinski's rule of five. These three molecules obeyed Lipinski's rule of five due to their higher molecular weights. Therefore, by ignoring the molecular weight, we continued with further computational studies. All the drug compounds in Table 1 have good bioavailability scores (most of them 0.55, or 55%), which indicate how much of the medication is absorbed into the bloodstream—lipophilicity, expressed by the logarithm of the octanol–water partition coefficient log P. In the development of new drugs, the assessment of lipophilicity values

should be significant to understand the affinity of any drug to a lipid environment (Pallicer et al., 2014).

3.2. Molecular docking analysis against the targeted receptor of the monkeypox virus

The molecular docking technique predicts the binding orientation of a ligand molecule to a specific biological macromolecule. This approach is also constructive for assessing how drug molecules can bind to biological target (Singh et al., 2022). Monkeypox virus profilin-like protein and monkeypox virus DNA polymerase were selected as receptor molecules to perform molecular docking. Blind docking was included for both receptors (Table 2). Dieckol and amentoflavone exhibited the most promising docking outcomes, with dieckol having a binding affinity of -10.1 Kcal/mol for profilin-like protein and -10.5 Kcal/mol for monkeypox virus DNA polymerase. This suggests that dieckol forms a stronger bond with the receptor macromolecules than cidofovir (the standard drug) and that this compound may play a critical role in the long-term stabilization of this protein. On the other hand, amentoflavone also showed strong binding affinity where the compound binds with profilin-like proteins with -9.5 Kcal/mol and amentoflavone-monkeypox DNA polymerase-bound complex exhibited binding affinity of -10.3 Kcal/mol. The docking profile for the rest of the compounds (except alloaromadendrene) was also promising, as these compounds also exhibited better binding affinity than cidofovir, but dieckol and amentoflavone were selected for further investigation with the most convincing binding affinity.

3.3. Molecular docking analysis against targeted human *Babesia microti*

Babesia microti lactate dehydrogenase and *Babesia microti* lactate dehydrogenase apo-form were docked against ligands selected to determine which ligands should be considered for molecular dynamics simulations. For *Babesia microti* lactate dehydrogenase, the standard drug diminazene expressed a binding affinity of -6.4 Kcal/mol. All the selected ligands showed better binding affinity than the standard drug, with amentoflavone (-10.2 Kcal/mol) and dieckol (-11.1 Kcal/mol) exhibiting the strongest binding affinity for the receptor. On the other hand, diminazene had a binding affinity of -6.2 Kcal/mol for the apo-form of *Babesia microti* lactate dehydrogenase. Amentoflavone (-10.0 Kcal/mol) and dieckol (-10.4 Kcal/mol) again expressed robust binding affinity toward receptors. Isofucosterol, psoralidin, tomentin A, chlorogenic acid, pectolinarin, rhoifolin, oxymatrine, and sanguinarine also exhibited promising docking scores (Table 3) against both receptors, with outstanding binding affinity but lower than amentoflavone and dieckol.

3.4. Molecular docking pose and interaction analysis

Molecular docking allows the prediction or investigation of the interactions that are the key components that substantially influence the affinity of a ligand for a receptor. According to molecular docking

studies, the receptor-ligand complex with the lowest amount of energy excreted is considered to have the highest binding affinity. From the previous step, we selected amentoflavone and dieckol as the most suitable ligands with the lowest binding energy and the highest binding affinity score for selected proteins. The receptor protein-ligand docked complexes are shown in Figure 3.

For the monkeypox virus profilin-like protein, amentoflavone formed a pi-sigma bond with ILE A:32 and ILE A:59; a pi-alkyl bond with PRO A:60; a conventional hydrogen bond with PHE A:41. On the other hand, dieckol formed a conventional hydrogen bond with ARG A:138 and GLU A:107; a pi-alkyl bond with VAL A:93 and ARG A:139; a pi-pi stacking with TYR A:142. There was no pi-sigma bond observed for dieckol.

Additionally, for *Babesia microti* lactate dehydrogenase, dieckol formed a conventional hydrogen bond with GLY A:32, GLY A:97, ASP A:195, and GLY A:246; a pi-alkyl bond with VAL A:31, ALA A:99, and ALA A:238; a Van der Waals bond with ALA A:30; a pi-sigma bond with GLY A:194; a pi-pi stacking with GLY A:29; a pi-pi T-shaped with TYR A:239. On top of that, dieckol with *Babesia microti* lactate dehydrogenase apo-form formed a conventional hydrogen bond with GLY A:29, GLY A:97, THR A:95, and GLN A:100; a pi-alkyl bond with VAL A:31, ALA A:96, and ALA A:238; a pi-sigma bond with GLY A:194; a pi-pi stacking with HIS A:193 and TYR A:239, and a pi-donor hydrogen bond was observed at residue ARG A:99.

3.5. ADMET data analysis

The pharmacokinetic properties of the studied compounds are shown in Table 4. We evaluated the absorption features of the compounds based on water solubility, Caco-2 permeability, and human intestinal absorption. Drug absorption depends on water solubility; a higher water-soluble compound implies higher absorption properties and ample bioavailability (Xu et al., 2013). Compound 02 is more water-soluble than compound 07, which is marginally more water-soluble. Caco-2 is a human colorectal adenocarcinoma cell line that has been immortalized and is primarily utilized as a reference model of the intestinal barrier (Lea, 2015). Compounds 11 and 05 show maximum (2.107) and minimum Caco-2 permeability (-0.967). The human intestinal absorption (HIA) rate is essential in predicting how well a medicine will be absorbed when administered orally (Arora and Khurana, 2022). Compound 09 has the lowest percentage (24.308%) of HIA, while compound 11 has the maximum percentage (100%) of Hof IA. Table 4 displays drug distribution characteristics such as the blood-brain barrier and volume of distribution (VD). A lower VD score (<-0.15) indicates that the therapeutic agent is more uniformly distributed in plasma as opposed to tissue, whereas a higher VD score (>0.45) reflects that the pharmaceutical molecule is more uniformly transported in tissues. Compound 03 has the lowest VD value, but compound 09 has a higher VD value. Additionally, the blood-brain barrier (BBB) prevents the substance from entering the brain and central nervous system. Only two of our reported biomolecules, alloaromadendrene, and isofucosterol, can cross the BBB.

Following drug distribution, the liver breaks down the compound through various enzymatic processes. The isoenzyme cytochrome P450 is in charge of the biotransformation and metabolism of drugs (Cresteil et al., 1994). The importance of drug metabolism by cytochrome P450

TABLE 1 Lipinski's rule of five data, pharmacokinetics.

PubChem CID	Ligand name	Molecular weight	Hydrogen bond acceptor	Hydrogen bond donor	Consensus log P_{ow}	Lipinski's rule		Bioavailability
						Result	Violation	
91,354	Alloaromadendrene	204.35	0	0	4.34	Yes	1	0.55
5,281,326	Isofucosterol	412.69	1	1	7.07	Yes	1	0.55
5,281,600	Amentoflavone	538.46	10	6	3.62	No	2	0.17
5,281,806	Psoralidin	336.34	5	2	3.98	Yes	0	0.55
3,008,868	Dieckol	742.55	18	11	3.39	No	3	0.17
71,659,627	Tomentin A	442.5	7	4	3.59	Yes	0	0.55
1,794,427	Chlorogenic acid	354.31	9	6	-0.38	Yes	1	0.11
168,849	Pectolinarin	622.57	15	7	-0.43	No	3	0.17
5,282,150	Rhoifolin	578.52	14	8	-0.66	No	3	0.17
114,850	Oxymatrine	264.36	2	0	0.41	Yes	0	0.55
5,154	Sanguinarine	332.33	4	0	2.88	Yes	0	0.55

can be attributed to concerns about medication toxicity and pharmacological effects (Ogu and Maxa, 2000). There are several routes for a drug molecule to be excreted from the body, namely through the liver, bile, and kidneys. A valuable piece of information for estimating drug excretion is the total clearance rate of the drug molecule. It indicates the amount of drug excreted per unit by the combination of the liver and kidney (Dowd et al., 2010). Organic cation transporter two, or OCT2 substrate, is a crucial excretion factor because it improves renal clearance. None of the compounds were predicted to act as OCT2 substrates. One of the leading causes of unsuccessful medication development is toxicity. None of our compounds showed any toxicity, such as skin sensitivity or liver damage, except compounds 04 and 11, which showed AMES toxicity.

3.6. Quantitative structure–activity relationship and pIC₅₀

The Quantitative Structure–Activity Relationship (QSAR) is a computer modeling approach that has been applied and used in the area of drug discovery and design to predict the biological activity of chemical compounds based on their molecular structures. This is accomplished and used mainly in the development of new drugs, especially in computer-aided drug design. It involves the development of mathematical models that connect the physicochemical descriptions or structural aspects of molecules to their biological action (Patel et al., 2014; Wang et al., 2015).

The standard ranges of QSAR are considered below 10. Any molecule below 10 is potential, according to the theory (Ahamed et al., 2023). In our current investigation, the overall values of QSAR and pIC₅₀ are positive (Table 5) and they are satisfied with standard ranges. The highest and lowest values of pIC₅₀ are 6.26 and 4.45, respectively. The outcome of PIC₅₀ suggests that the compound may have therapeutic efficacy against the targeted disease.

3.7. Molecular dynamics simulations for monkeypox complexes

Molecular dynamics simulations are essential to determining the structural conformations of protein-ligand complexes. Promising interactions of dieckol with PDB 4QWO further encourage us to perform MD simulations. Figure 4 shows the RMSD of dieckol, amentoflavone, and cidofovir with monkeypox virus profilin-like protein. It was observed that the RMSD of Dieckol remains between 0.1–0.2 nm throughout the simulations, but some jumps of dieckol have also been observed at the time scales of 40, 80, 83, and 85 ns of the simulations; this does not show any effect on the ligand binding with the binding site. Dieckol remained stable in the binding site throughout the simulation period. The RMSD of amentoflavone with monkeypox virus profilin-like protein remained intact throughout the simulation period. Also, the reference ligand, cidofovir, is a common drug used to treat the monkeypox virus. The MD simulations revealed that Cidofovir showed robust binding with the monkeypox protein for 45 ns of the time scale, then the cidofovir showed a higher change in conformations after 98 ns of the period, then unbinding of the cidofovir was observed after 98 ns of the period. This study demonstrated that dieckol is more stable than cidofovir and amentoflavone, as observed from the

TABLE 2 Binding affinity against the *monkeypox* virus.

Name of compound	Monkeypox virus profilin-like protein (PDB ID: 4QWO)	Monkeypox virus DNA polymerase (PDB: 8HG1)
	Binding affinity (Kcal/mol)	Binding affinity (Kcal/mol)
Alloaromadendrene	−6.3	−6.9
Isofucosterol	−8.3	−8.7
Amentoflavone	−9.5	−10.3
Psoralidin	−8.6	−8.5
Dieckol	−10.1	−10.5
Tomentin A	−9.1	−9.1
Chlorogenic acid	−8.0	−6.9
Pectolinarin	−8.4	−9.2
Rhoifolin	−8.5	−9.6
Oxymatrine	−7.0	−7.3
Sanguinarine	−8.3	−9.0
Standard Cidofovir	−6.5	−6.8

TABLE 3 Binding affinity against human *Babesia microti*.

Name	<i>Babesia microti</i> lactate dehydrogenase (PDB ID: 6J9D)	<i>Babesia microti</i> lactate dehydrogenase apo form (PDB ID: 6K12)
	Binding affinity (Kcal/mol)	Binding affinity (Kcal/mol)
Alloaromadendrene	−6.5	−6.2
Isofucosterol	−7.9	−9.1
Amentoflavone	−10.2	−10.0
Psoralidin	−8.2	−7.9
Dieckol	−11.1	−10.4
Tomentin A	−8.1	−8.2
Chlorogenic acid	−7.3	−7.5
Pectolinarin	−8.2	−8.9
Rhoifolin	−8.7	−8.7
Oxymatrine	−7.5	−6.8
Sanguinarine	−9.2	−8.4
Standard Diminaze	−6.4	−6.2

conformations of the molecules during simulations. Therefore, dieckol will exert more significant inhibition of the *monkeypox* virus than cidofovir. Next, each ligand was superposed with the respective ligand, and the conformations were calculated and represented in the form of root mean square deviations (RMSD) in Figures 4, 5.

The stability of the complexes was determined from the C α backbone of the monkeypox virus protein complexed with cidofovir, amentoflavone, and dieckol. The C α backbone is a crucial parameter in the MD simulation to determine the stability of the complexes. The *monkeypox* virus complexed with cidofovir, amentoflavone, and dieckol was compared with the apo complex. This analysis logically gives an in-depth idea of the effects on the structures due to the ligand binding. The RMSD of the apo complex (without ligand) ranged between 0.1–0.15 nm and reflects the stable trajectory

throughout the simulation period parallel to the amentoflavone-bound monkeypox virus. The complex reference cidofovir bound to monkeypox showed minor fluctuations for 10–20 ns, then became stable. Again, minor fluctuations were observed in the RMSD on a time scale of 60–70 ns. Finally, dieckol complexed with the monkeypox virus showed stability throughout the simulation period, and the RMSD was found to be 0.9–1.1 nm. This study shows that dieckol is the most stable and has no effect on the C α backbone of the *monkeypox* virus.

The root mean square fluctuations were found to be within an acceptable range; no higher fluctuations were observed; only the reference complex Cidofovir bound to monkeypox showed higher fluctuations at the terminal end. The limited fluctuations reveal the stability of the complexes in Figures 6, 7.

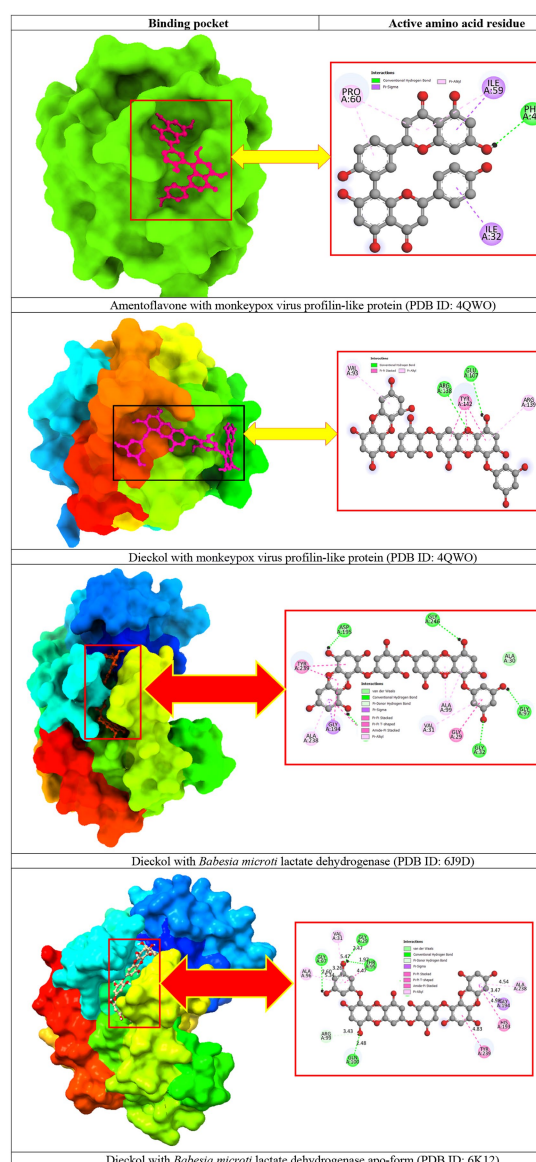


FIGURE 3

Docking interactions between the proposed compounds.

The compactness of the complexes was monitored, which primarily determines their rigidity. The higher the value of the radius of gyration (ROG), the more unstable the complex (Ghahremanian et al., 2022). Lower values result in greater stability in the complex. The apo complex, amentoflavone, and dieckol bound to the *monkeypox* virus showed ROG values of 1.3–1.33 nm throughout the simulation period, and similar compactness results in the stability of the complex. However, the reference complex Cidofovir bound with *monkeypox* showed higher ROG until 58 ns of the simulation period and then achieved ROG values that were comparatively similar to those of other complexes in the present study (Figure 8). Protein folding was calculated from the solvent-accessible surface area (SASA). In the reference complex, the SASA value is comparatively higher than in the other complexes in the present study. Similarly, the lower SASA value demonstrates a higher protein folding and is referred to as a stable complex,

whereas the higher protein folding results from the instability of the complexes. These results demonstrated that dieckol bound to the *monkeypox* virus showed linear trajectories and was correlated with the ROG (Figure 9).

3.7.1. Secondary structure analysis of proteins

The Definition Secondary Structure of Proteins (dssp) module in GROMACS was chosen to determine the structural changes of the *monkeypox* virus upon binding to amentoflavone and dieckol. The structural changes were correlated with the apo complex and the reference complex of cidofovir-bound *monkeypox*. Moreover, the changes in the apo complex and the reference complex (cidofovir-bound *monkeypox* protein) were observed from the dssp plot. The amino acids located between 44–56 ns showed transitions between α -helix and loop turns at position 55 ns of the time scale until the end of the simulations (Figure 10A).

TABLE 4 Thermotical ADME properties.

No	Water solubility Log S	Caco-2 permeability $\times 10^{-6}$	Human intestinal absorption (%)	VDss (human)	BBB permeability	CYP450 1A2 inhibitor	CYP450 2D6 substrate	Renal OCT2 substrate	AMES toxicity	Skin sensitization	Hepatotoxicity
01	−5.764	1.395	95.302	0.753	Yes	No	No	No	No	No	No
02	−6.715	1.212	94.642	0.179	Yes	No	No	No	No	No	No
03	−2.892	0.145	84.356	−1.066	No	No	No	No	No	No	No
04	−3.979	1.048	93.488	0.052	No	Yes	No	No	Yes	No	No
05	−2.892	−0.967	68.892	−0.218	No	No	No	No	No	No	No
06	−3.502	−0.366	86.57	0.615	No	No	No	No	No	No	No
07	−2.449	−0.84	36.377	0.518	No	No	No	No	No	No	No
08	−2.986	0.309	41.847	0.684	No	No	No	No	No	No	No
09	−2.862	−0.942	24.308	1.14	No	No	No	No	No	No	No
10	−3.58	1.269	96.121	0.404	No	No	No	No	No	No	No
11	−5.56	2.107	100.	0.298	No	Yes	No	No	Yes	No	No

TABLE 5 Data of QSAR calculation.

Ligand	Chiv5	bcutm1	(MRVSA9)	(MRVSA6)	(PEOEVSAS)	GATSv4	J	Diameter	PIC50
01	3.846	3.972	0.0	12.52	32.923	0.0	1.898	6.0	5.18
02	6.776	3.972	0.0	23.298	57.917	0.566	1.447	15	6.26
03	3.214	4.156	21.938	93.243	0.0	0.878	1.195	17	5.04
04	2.199	4.247	32.908	57.965	11.649	0.951	1.493	12	4.62
05	3.68	4.04	0.0	66.73	0.0	0.809	0.873	25	6.11
06	3.192	4.001	5.783	40.956	6.066	0.834	1.395	16	5.32
07	1.708	3.866	18.015	29.839	6.066	1.172	1.829	13	4.45
08	3.365	4.116	10.969	46.622	0.0	0.957	1.236	21	5.71
09	3.236	4.106	10.969	52.688	0.0	0.961	1.286	18	5.43
10	4.694	3.931	5.907	5.207	0.0	1.002	1.676	7	4.76
11	2.846	4.215	32.448	42.595	6.066	0.761	1.27	11	4.57

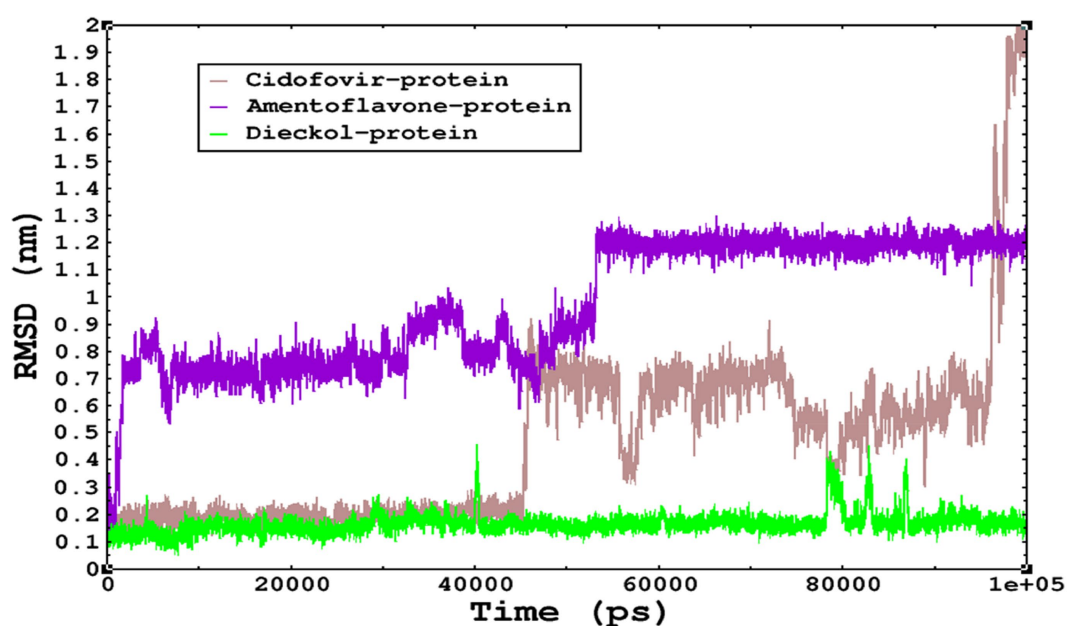


FIGURE 4

The conformations of dieckol, cidofovir, and amentoflavone with *monkeypox* virus during simulations in an aqueous medium. The trajectories are plotted, and the color represented is given in the legend box.

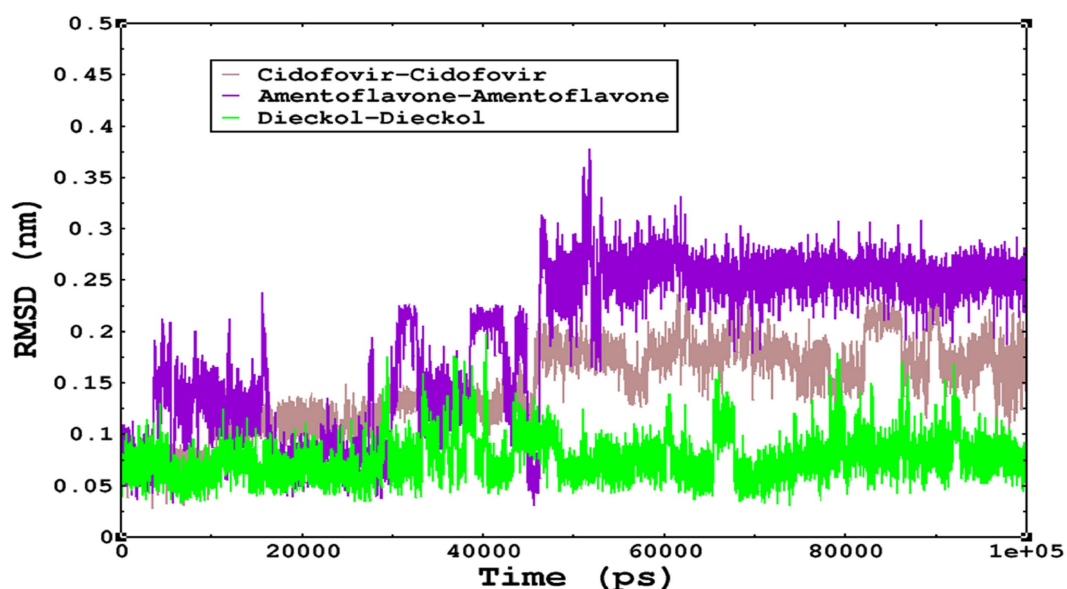


FIGURE 5

The RMSD of cidofovir, amentoflavone, and dieckol was calculated after superposing on the same molecular scaffold in the aqueous medium for 100ns of the simulation period.

Similarly, the reference complex showed bend changes into the loop turns, and the transitions occur among amino acids residing 44–56, and transitions occur between α -helix, and loop turns at 20, 41, 45, 70–80, and 90–100 ns of time scale during simulations (Figure 10D).

In the complex amentoflavone and dieckol bound to the monkeypox virus, the minor transitions occurred at amino acids residing at 44–56 ns, and transitions occurred between α -helix and

loop turns at varied simulation periods (Figures 10B,C). The overall understanding of the simulations is that amentoflavone and dieckol bound to the monkeypox virus are stable complexes and need further investigation.

3.7.2. Free binding energy calculations

The free binding energy was calculated using the MM/PBSA method. The 60–100 ns generated trajectory was used for the free binding energy

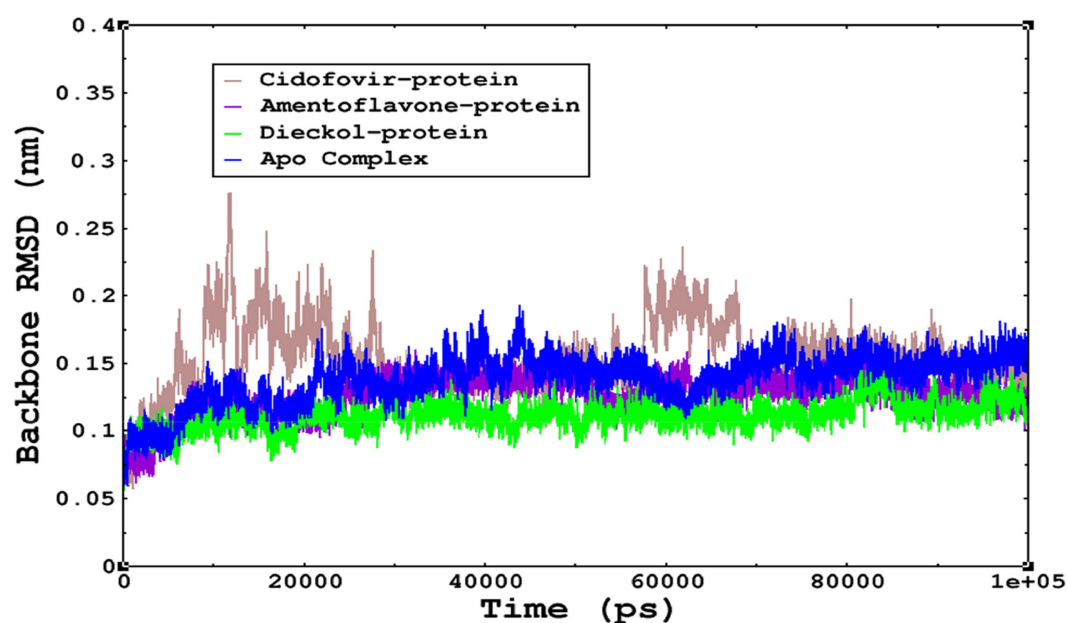


FIGURE 6

The stability of the complexes was monitored from the C α backbone during the simulations; all complexes were found to have a stable C α backbone; the most stable complex is dieckol.

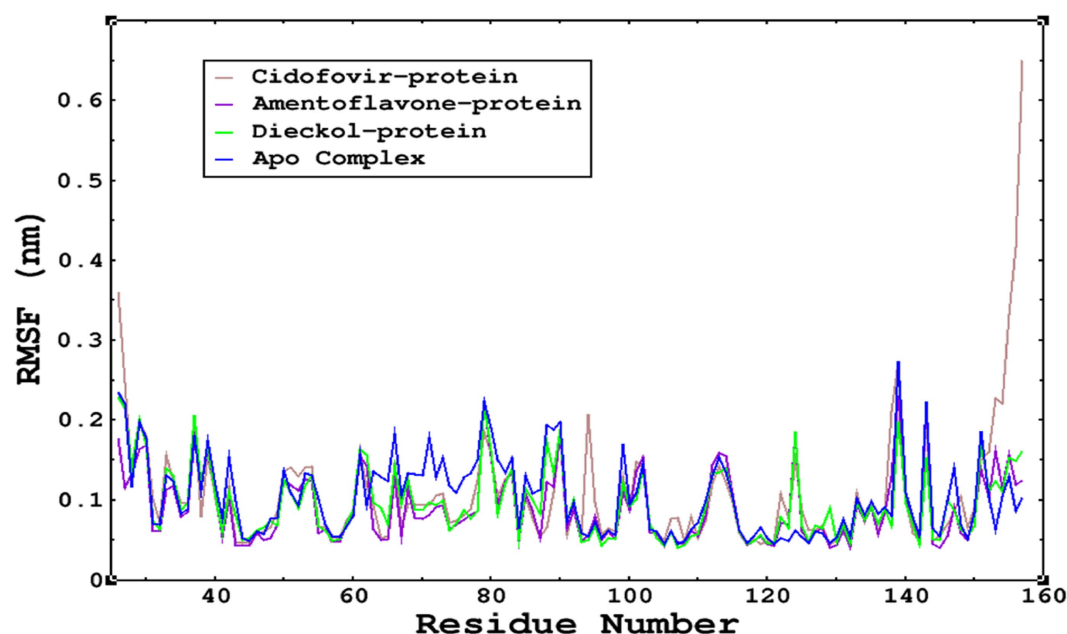


FIGURE 7

The RMSF was calculated by using the RMSF module of GROMACS; parallel fluctuations were observed between the Apo complex and the other complexes complexed with the ligands. This demonstrates no significant fluctuations between Apo and other complexes, except that the reference cidofovir bound to monkeypox shows higher fluctuations at the terminal end.

calculations (Table 6). All the molecules included in the present study showed lower contributions to hydrogen bonding. At the same time, a hydrophobic interaction was dominant in each complex. The binding energy exhibited by amentoflavone was found to be -115.610 ± 1.531

kJ/mol, and for the dieckol complex, the binding energy was -207.080 ± 1.797 kJ/mol. The reference complex cidofovir exhibited ΔG_{bind} 212.357 ± 2.766 kJ/mol, and the positive ΔG_{bind} obtained by cidofovir was due to the unbinding that occurred during the simulations.

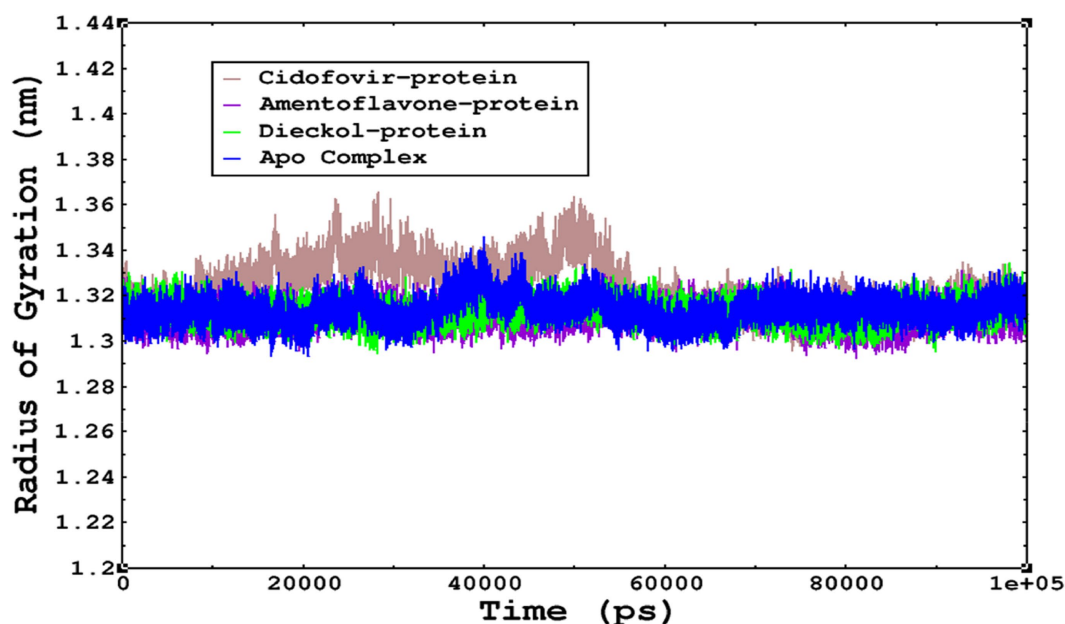


FIGURE 8

The radius of gyration refers to the compactness and rigidity of the complexes. The ROG trajectories are represented in the legend box for each complex.

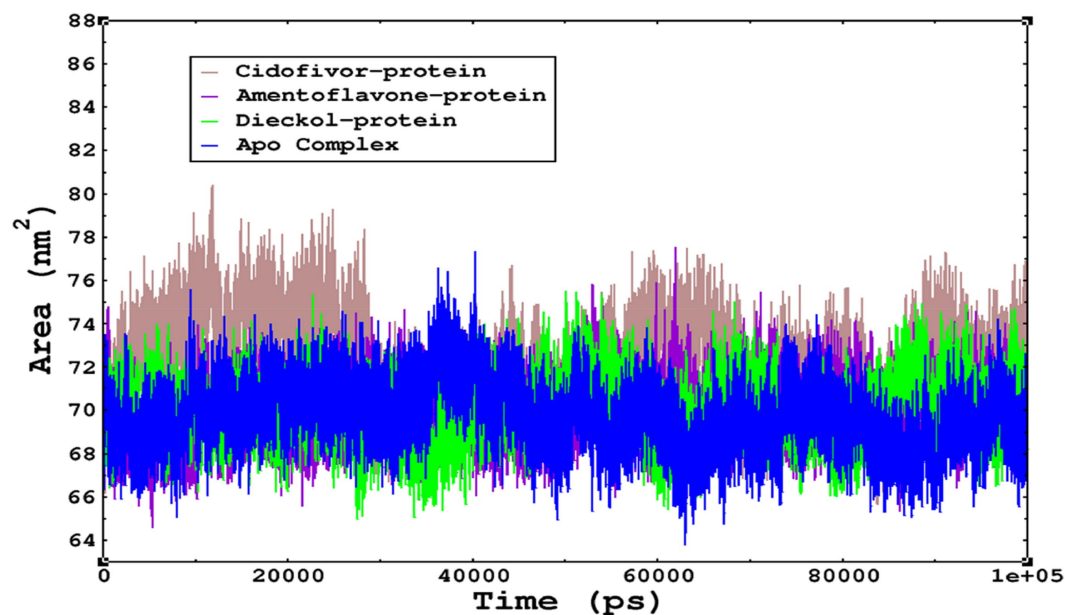


FIGURE 9

The SASA values of the complexes as calculated by gmx_sasa. The reflected SASA trajectories were plotted for each complex, and the complex representations are given in the legend box.

3.8. *Babesia microti* complexes MD simulations

We also conducted molecular dynamics simulations of amentoflavone, dieckol, and diminazene complexed with *Babesia microti* lactate dehydrogenase (PDB ID 6J9D). The RMSD of each

ligand was monitored. It was found that each ligand was stabilized in the binding site, and the orientations were represented in the form of RMSD. It was observed that the RMSD of amentoflavone ranged between 0.2–0.38 nm throughout the simulations. The RMSD of dieckol with protein (PDB 6J9D) remained intact with the protein throughout the simulation period. Moreover, the standard ligand

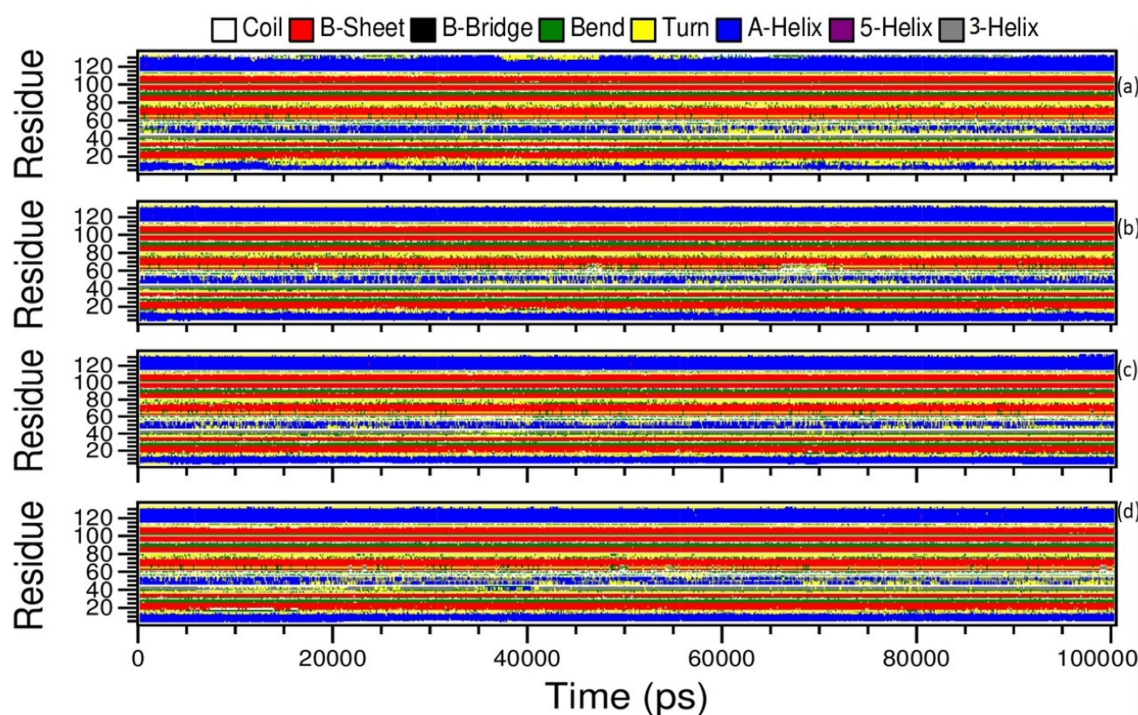


FIGURE 10

Definition Secondary Structure of Protein (DSSP) analysis of Amentoflavone, Dieckol, Cidofovir bound monkeypox, and Apo complex.

TABLE 6 Free binding energy data.

Ligand	ΔE_{vdw} kJ/mol	ΔE_{Elec} kJ/mol	$\Delta E_{\text{Polar solvation}}$ kJ/mol	SASA kJ/mol	$\Delta E_{\text{binding Energy}}$ kJ/mol
Amentoflavone	-168.239 ± 0.535	-20.276 ± 1.077	80.607 ± 1.282	-14.709 ± 0.038	-115.610 ± 1.531
Dieckol	-197.578 ± 1.083	-215.037 ± 2.576	226.411 ± 1.348	-20.845 ± 0.075	-207.080 ± 1.797
Cidofovir	-66.350 ± 0.922	250.383 ± 4.145	37.585 ± 2.008	-9.241 ± 0.094	212.357 ± 2.766

diminazene MD simulations revealed binding with the *Babesia microti* lactate dehydrogenase protein for 100 ns of the time scale, and the conformations showed stability from the start to the end of the simulations. This study demonstrated that amentoflavone, dieckol, and diminazene were found to be stable, as observed from the conformations of the molecules during simulations. Therefore, amentoflavone, dieckol, and diminazene will significantly inhibit *Babesia microti* lactate dehydrogenase. Next, each ligand was superposed with the respective ligand, and the conformations were calculated in the form of RMSDs (Figures 11, 12). The stability of the whole protein was determined by calculating the RMSD C α . The backbone C α revealed that each complex remained stable throughout the simulation period. The conformations of backbone C α were calculated in the form of RMSD C α and the data points were found to be parallel to each other. No major changes in RMSD were observed, with the RMSD C α ranging between 0.2–0.4 nm throughout the simulation period (Figure 13). The root mean square fluctuations of each amino acid present in the *Babesia microti* lactate dehydrogenase (PDB ID 6J9D) complexed with amentoflavone,

dieckol, and diminazene were investigated during the simulation period. Fewer fluctuations indicated protein stability. The binding site amino acids showed fewer fluctuations, confirming that the binding of amentoflavone, dieckol, and diminazene does not affect the protein stability (Figure 14). Moreover, the compactness of each protein was investigated by calculating the radius of gyration; higher values result in protein instability, whereas lower values of the radius of gyration result in a more compact protein structure. The protein complexed with amentoflavone exhibited a lower radius of gyration value than the other two complexes, i.e., 6J9D complexed with dieckol and diminazene (Figure 15). Next, the protein folding was determined by calculating the solvent-accessible surface area. The higher the protein folding, the higher the SASA values. From the SASA plots, it was found that amentoflavone, dieckol, and diminazene complexed with 6J9D were parallel, and the same protein folding was observed (Figure 16). Overall, these findings show that each complex of amentoflavone, dieckol, and diminazene complexed with *Babesia microti* lactate dehydrogenase was found to be stable. Further, *in vitro* experiments are required to validate this computational study.

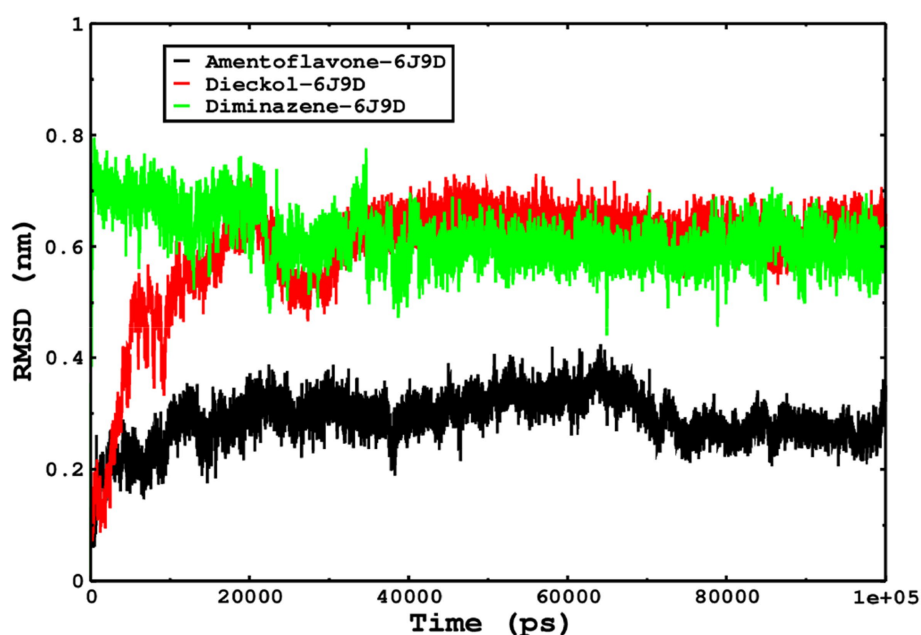


FIGURE 11

The ligand conformations were monitored in the aqueous medium for 100ns of simulations. The RMSD of each ligand was monitored by superposition on the protein 6J9D, and the plotted trajectories are represented in the legend box.

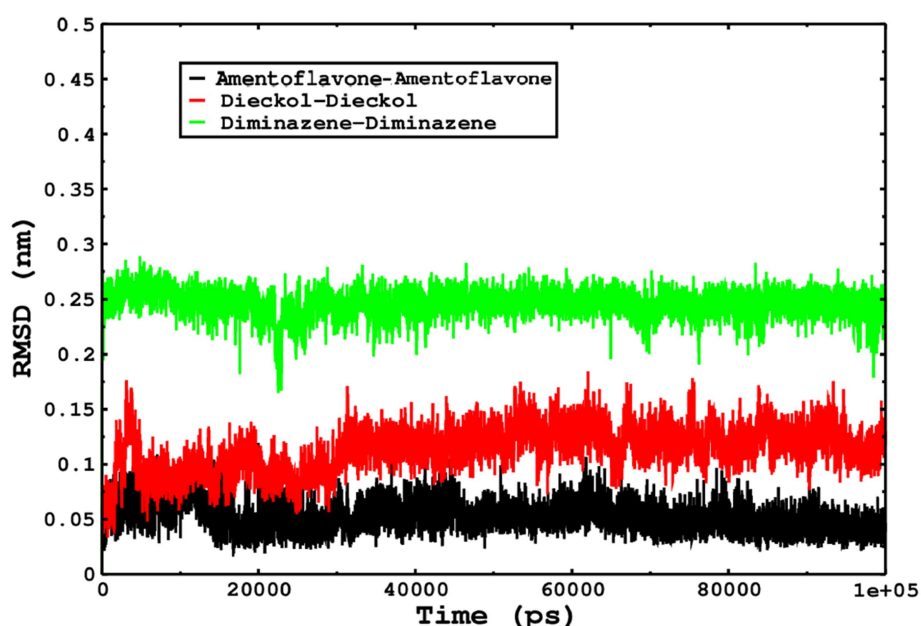


FIGURE 12

The conformations of only ligands were monitored by superposition on the respective ligand, and the conformations were plotted in the form of RMSD.

3.8.1. Free binding energy calculations of the *Babesia microti* lactate dehydrogenase complex

The free binding energy was calculated using the MM/PBSA method. The 60–100 ns generated trajectory was used for the free binding energy calculations (Table 7). All the molecules included in the present study showed lower contributions to hydrogen bonding. At the same

time, the hydrophobic interaction was dominant in each complex. The binding energy exhibited by amentoflavone was found to be -220.278 ± 2.302 kJ/mol, and for the dieckol complex, the binding energy was -252.554 ± 3.043 kJ/mol. The reference complex cidofovir had a ΔG_{bind} of -38.166 ± 6.111 kJ/mol; the positive ΔG_{bind} obtained by cidofovir was due to the unbinding that occurred during the simulations.

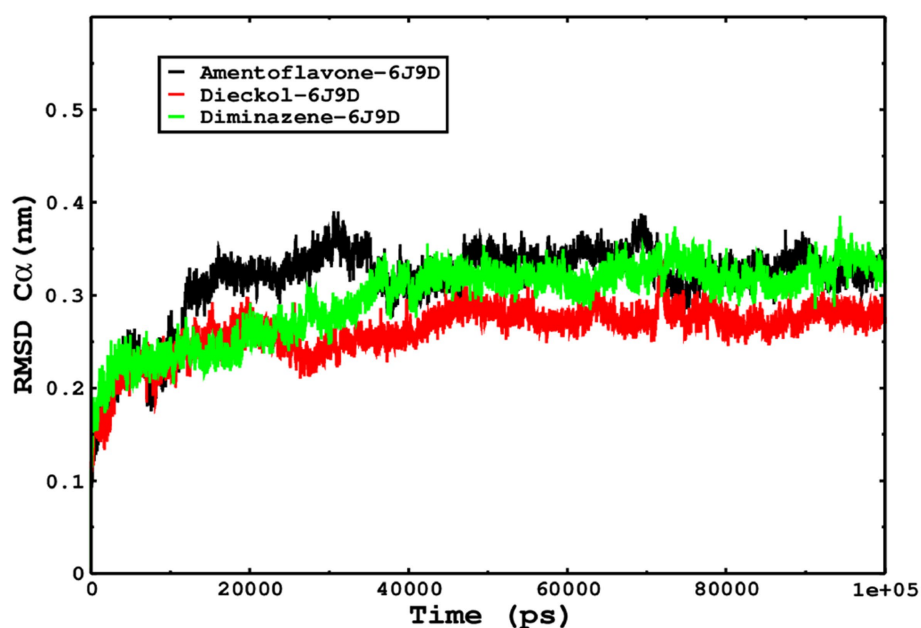


FIGURE 13

The protein stability of each complex simulated for 100ns was determined from the conformations of the C α , and the RMSD C α indicates that each complex was found stable.

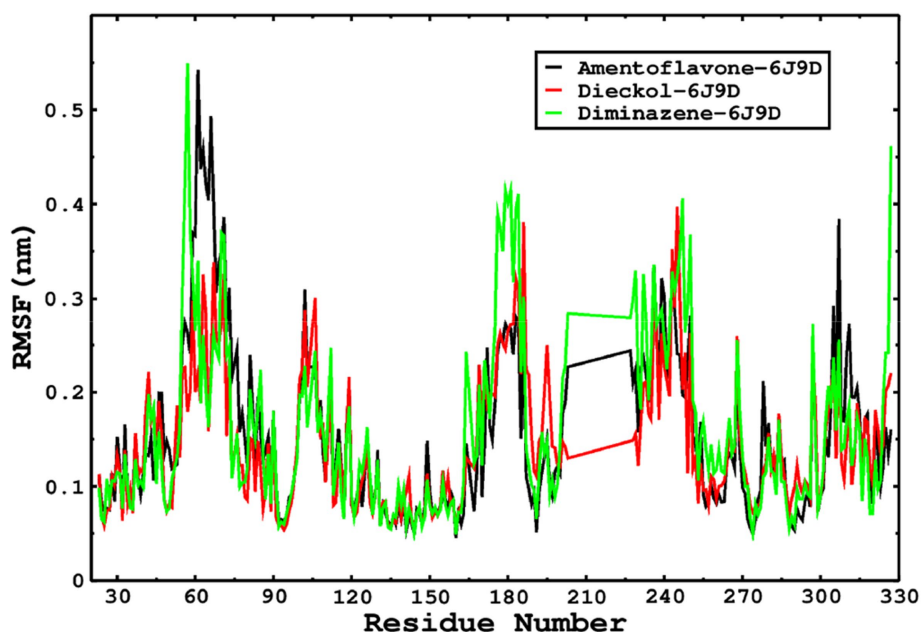


FIGURE 14

The RMSF was calculated by using the RMSF module of the GROMACS; the parallel fluctuations complexed with the ligands were found in parallel.

3.9. Frontier molecular orbitals and chemical reactivity descriptors

The following descriptors were used to determine the behavior of the phytochemicals selected in this study. The energy band gap between HOMO and LUMO was calculated, from which the Chemical

Reactivity Descriptors were determined. This parameter is essential to determining the stability of the molecule. A higher band gap between HOMO and LUMO results in a more stable molecule, whereas a lower energy gap exhibits higher reactivity and is a useful parameter in determining kinetic stability. The energy gap was found to range between 3 and 6 eV, which indicates that the chosen molecules are

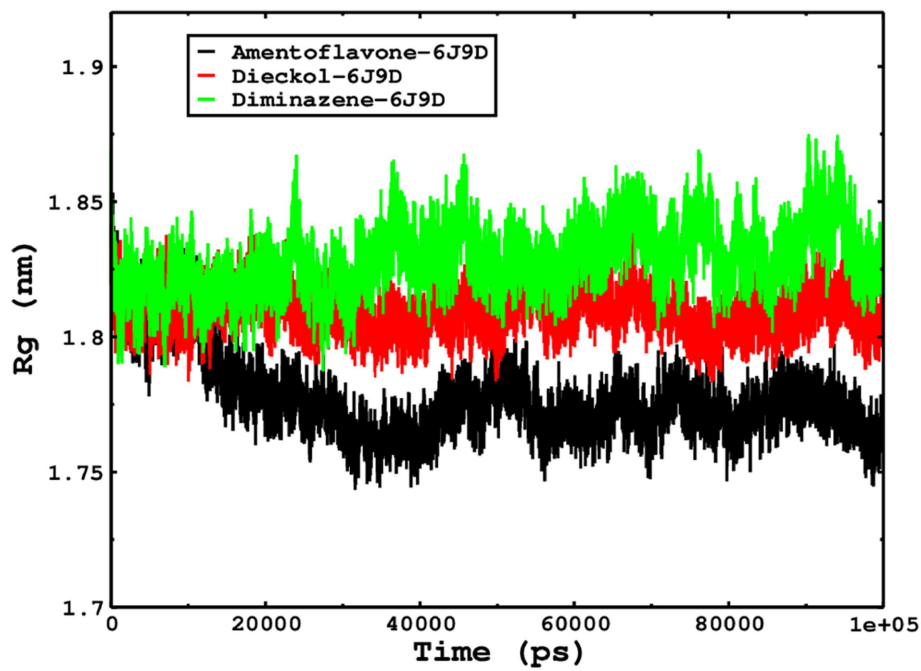


FIGURE 15
The radius of gyration is related to the compactness and rigidity of the complexes. The ROG trajectories are shown in the legend box for each complex.

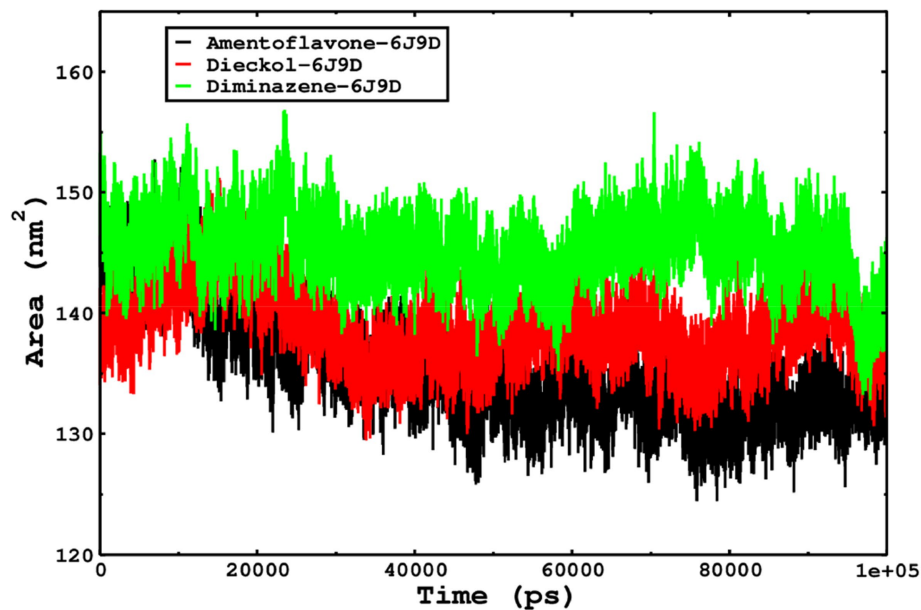


FIGURE 16
The SASA of the complexes was calculated by gmx_sasa. The reflected SASA trajectories were plotted for each complex, and the complex representations are shown in the legend box.

TABLE 7 Free binding energy data.

Ligand	ΔE_{vdw} kJ/mol	ΔE_{Elec} kJ/mol	$\Delta E_{Polar\ solvation}$ kJ/mol	SASA kJ/mol	$\Delta E_{binding\ Energy}$ kJ/mol
Amentoflavone	-245.713 ± 1.353	-94.143 ± 2.019	141.932 ± 1.144	-22.344 ± 0.122	-220.278 ± 2.302
Dieckol	-269.596 ± 1.589	-162.381 ± 3.689	210.245 ± 1.678	-30.788 ± 0.118	-252.554 ± 3.043
Diminazene	-126.488 ± 1.637	-247.210 ± 10.249	352.501 ± 4.680	-17.069 ± 0.128	-38.166 ± 6.111

TABLE 8 Chemical reactivity descriptors.

Drug Name	LUMO	HOMO	Gap	I=-HOMO	A=-LUMO	Hardness	Softness
Alloaromadendrene	-0.3235	-6.5184	6.1949	6.5184	0.3235	3.4209	0.3228
Isoucosterol	-0.3899	-6.3731	5.9832	6.3731	0.3899	2.9916	0.3343
Amentoflavone	-2.4852	-6.4969	4.0118	6.4969	2.4852	2.0059	0.4956
Psoralidin	-2.0989	-6.0303	3.9314	6.0303	2.0989	1.9656	0.5088
Dieckol	-0.5659	-4.8836	4.3176	4.8836	0.5659	2.1599	0.4632
Tomentin A	-2.0173	-6.3302	4.2784	6.3302	2.0173	2.1394	0.4675
Pectolinarin	-2.2677	-5.9919	3.7242	5.9919	2.2677	1.8620	0.5370
Rhoifolin	-2.5518	-6.6515	4.0996	6.6515	2.5518	2.0498	0.4878

stable and reactive (Table 8). The HOMO and LUMO data calculations and frontier molecular orbital images are represented in Table 7; Supplementary Figure S1. The stability order of the molecules is as follows: alloaromadendrene, isoucosterol, dieckol, tomentin a, rhoifolin, amentoflavone, psoralidin, and pectolinarin. Their reactivity is in the reverse order of the above-represented molecules. Hardness determines how hard a compound is, and softness determines how quickly a compound dissolves or breaks down when in contact with the liquid phase. Hardness and softness are opposites of each other, and it is seen that the softness value for all compounds is much lower in comparison to hardness. Thus, it is estimated that these compounds will dissolve quickly (Rahman et al., 2022).

4. Discussion

The monkeypox virus and the *Babesia microti* parasite are highly pathogenic infectious diseases that have emerged as a growing public health concern in recent years. There is currently no specific treatment or vaccine available for these diseases, and they can cause severe illness and even death in some cases. There are significant differences between these two pathogens in their biology, physiological characteristics, and pathogenesis. The primary objective of this investigation is to inhibit the monkeypox virus profilin-like protein and *Babesia microti* lactate dehydrogenase. If this is possible, it would likely prevent and limit their ability to cause disease. In the case of monkeypox, inhibiting this target protein would likely prevent the virus from replicating in infected cells, reducing the severity of the infection and potentially preventing the spread of the virus to other individuals. On the other hand, inhibiting the target protein of *Babesia microti* would prevent the parasite from multiplying inside red blood cells, which is essential for its survival and ability to cause disease. This would likely limit the severity of the infection and reduce the risk of complications such as anemia that can result from the destruction of infected red blood cells.

Second, nature is an excellent source for novel drug development. So, the mentioned phytochemicals have been selected for having many pharmacological activities against different infectious diseases, and based on the literature, this study has been designed to investigate the dual function of some of these compounds against the mentioned virus and parasite using advanced computational and drug design approaches such as molecular docking, molecular dynamics simulation, DFT, ADMET, drug-likeness, etc. The molecular docking score was reported as outstanding against both pathogens. Then, free binding energy, molecular dynamics simulation, DFT, and other related studies are performed step by step, and all the levels of computational experiments are satisfied by the mentioned phytochemicals. Finally, this section should conclude that these phytocompounds could be further studied in a wet lab to investigate their actual performance and validate the computational result.

5. Conclusion

The application of advanced computational strategies and combined drug design approaches, such as ADMET evaluation, ligand drug-likeness quantification, and molecular docking analysis, has led to the identification and characterization of potential inhibitors of the viral pathogens *Babesia microti* and monkeypox. Through these

methods, a total of 11 promising lead compounds, including Alloaromadendrene, Isofucosterol, Amentoflavone, Psoralidin, Dieckol, Tomentin A, Chlorogenic acid, Pectolinarin, Rhoifolin, Oxymatrine, and Sanguinarine, have been demonstrated to have high potency against the active catalytic sites of the target enzymes, outstanding drug-like properties, and no toxic effects. Moreover, the binding affinities of the selected natural biomolecules were measured using the AutoDock Vina tool, resulting in ranges of -6.5 kcal/mol to -11.1 kcal/mol against *Babesia microti* lactate dehydrogenase (PDB ID 6J9D), -6.2 kcal/mol to -10.4 kcal/mol against *Babesia microti* lactate dehydrogenase apo form (PDB ID 6K12), -6.3 kcal/mol to -10.1 kcal/mol for monkeypox virus profilin-like protein (PDB ID 4QWO), and -6.9 kcal/mol to -10.5 kcal/mol for monkeypox virus DNA polymerase (PDB ID 8HG1). Notably, Dieckol and Amentoflavone exhibited higher reactivity and better affinity for both the *Babesia microti* and monkeypox-targeted proteins, with high predicted affinities. Dieckol, in particular, demonstrated effective and potent binding ability against monkeypox and remained stable within the binding site during MD simulations. The MM/PBSA method calculated the highest negative free energy, with Amentoflavone and Dieckol showing free binding energies of -115.610 ± 1.531 kJ/mol and -207.080 ± 1.797 kJ/mol, respectively, while Cidofovir showed a free binding energy of 212.357 ± 2.766 kJ/mol. This research focuses on the inhibition of monkeypox and *Babesia microti* using phytochemicals, and among them, the multifaceted role of Dieckol and Amentoflavone has been discovered, as they surprisingly bind and suppress both monkeypox and *Babesia microti* pathogens effectively.

6. Prospects of the study

In conclusion, the prospects of this study include conducting *in vitro* and *in vivo* validation of the lead compounds, particularly Dieckol and Amentoflavone, to assess their efficacy against monkeypox and *Babesia microti*. Structural optimization and combination therapy approaches hold promise for enhancing their potency and broadening their spectrum of activity. Additionally, clinical trials may be conducted to evaluate the safety and effectiveness of these compounds in humans. These future directions will contribute to advancing our understanding and potential treatments for inhibiting monkeypox and *Babesia microti* infections.

7. Limitations of the study

Despite the significant findings and promising results obtained in this study, several limitations should be acknowledged. First, it is important to note that this investigation is purely theoretical in nature, relying on computational methods and simulations. While these approaches provide valuable insights and predictions, further validation through extensive *in vitro* and *in vivo* experiments is required. The practical value of these phytochemicals can only be determined through extensive preclinical and clinical studies. Additionally, this work focused on a specific set of target proteins associated with *Babesia microti* and monkeypox, and further investigation should expand the scope to include a broader range of potential targets. Furthermore, the

study primarily explored the binding affinities and drug-like properties of the identified lead compounds, but factors such as pharmacokinetics, bioavailability, and potential side effects should be thoroughly investigated to ensure the development of newer and safer drugs from natural sources. Therefore, to fully validate the findings of this investigation and unlock the potential therapeutic applications of these compounds, it is imperative to conduct comprehensive experimental studies that include computational, preclinical, and clinical trials.

Data availability statement

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

Author contributions

SA, SAM, NM, and SH: conceptualization. SA, SM, SAM, and NM: methodology. SA and SH: validation. SA, SH, SM, NM, and BN: formal analysis. H-AN, YB, AM, MB, and SA: data curation. SA, H-AN, YB, AM, MB, and SAM: writing—original draft preparation. BN: writing—review and editing and supervision. All authors have contributed to the article and approved the submitted version.

Acknowledgments

The authors would like to extend their sincere appreciation to the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia for funding this work through the project number (RSP2023R457).

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.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1206816/full#supplementary-material>

References

- Adegboye, O. A., Eugenia Castellanos, M., Alele, F. O., Pak, A., Ezechukwu, H. C., Hou, K., et al. (2022). Travel-related monkeypox outbreaks in the era of COVID-19 pandemic: are we prepared? *Viruses* 14:1283. doi: 10.3390/v14061283
- Ahamed, F. M., Chinnam, S., Challa, M., Kariyanna, G., Kumer, A., Jadoun, S., et al. (2023). Molecular dynamics simulation, QSAR, DFT, molecular docking, ADMET, and synthesis of ethyl 3-((5-Bromopyridin-2-yl) Imino) butanoate analogues as potential inhibitors of SARS-CoV-2. *Polycycl. Aromat. Compd.*, 1–19. doi: 10.1080/10406638.2023.2173618
- Amin, M. R., Yasmin, F., Dey, S., Mahmud, S., Saleh, M. A., Emran, T. B., et al. (2022). Methyl β -D-galactopyranoside esters as potential inhibitors for SARS-CoV-2 protease enzyme: synthesis, antimicrobial, PASS, molecular docking, molecular dynamics simulations and quantum computations. *Glycoconj. J.* 39, 261–290. doi: 10.1007/s10719-021-10039-3
- Antunes, F., Cordeiro, R., and Virgolino, A. (2022). Monkeypox: from a neglected tropical disease to a public health threat. *Infect. Dis. Rep.* 14, 772–783. doi: 10.3390/idr14050079
- Arora, D., and Khurana, B. (2022, 189). "Computer-aided biopharmaceutical characterization: gastrointestinal absorption simulation and in silico computational modeling" in *Computer aided pharmaceuticals and drug delivery* (London: Springer), –215.
- Beer, E. M., and Rao, V. B. (2019). A systematic review of the epidemiology of human monkeypox outbreaks and implications for outbreak strategy. *PLoS Negl. Trop. Dis.* 13:e0007791. doi: 10.1371/journal.pntd.0007791
- Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., and Velankar, S. (2017). Protein data Bank (PDB): the single global macromolecular structure archive. *Protein Crystallogr.*, 627–641. doi: 10.1007/978-1-4939-7000-1_26
- Chand, M., Choi, J.-Y., Pal, A. C., Singh, P., Kumari, V., Thekkiniath, J., et al. (2022). Epitope profiling of monoclonal antibodies to the immunodominant antigen BmGP12 of the human pathogen *Babesia microti*. *Front. Cell. Infect. Microbiol.* 12:1754. doi: 10.3389/fcimb.2022.1039197
- Chiu, J. E., Renard, I., Pal, A. C., Singh, P., Vidyam, P., Thekkiniath, J., et al. (2021). Effective therapy targeting cytochrome bc 1 prevents babesia erythrocytic development and protects from lethal infection. *Antimicrob. Agents Chemother.* 65, e00662–e00621. doi: 10.1128/AAC.00662-21
- Cresteil, T., Monsarrat, B., Alvinerie, P., Tréluyer, J. M., Vieira, I., and Wright, M. (1994). Taxol metabolism by human liver microsomes: identification of cytochrome P450 isozymes involved in its biotransformation. *Cancer Res.* 54, 386–392.
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 1–13. doi: 10.1038/srep42717
- Dallakyan, S., and Olson, A. J. (2015). "Small-rPmolecule library screening by docking with PyRx" in *Chemical biology* (London: Springer), 243–250.
- Devillers, J., *Neural networks in QSAR and drug design* Cambridge: Academic Press (1996).
- Djokic, V., Primus, S., Akoole, L., Chakraborti, M., and Parveen, N. (2018). Age-related differential stimulation of immune response by *Babesia microti* and *Borrelia burgdorferi* during the acute phase of infection affects disease severity. *Front. Immunol.* 9:2891. doi: 10.3389/fimmu.2018.02891
- Dowd, F. J., Yagiela, J. A., Johnson, B., Mariotti, A., and Neidle, E. A., *Pharmacology and therapeutics for dentistry-E-book* Amsterdam: Elsevier Health Sciences (2010).
- El-Barghouti, M., Jaime, C., Akielah, R., Al-Sakhen, N., Masoud, N., Issa, A., et al. (2009). Free energy perturbation and MM/PBSA studies on inclusion complexes of some structurally related compounds with β -cyclodextrin. *Supramol. Chem.* 21, 603–610. doi: 10.1080/10610270802613562
- Farasani, A. (2022). Monkeypox virus: future role in human population. *J. Infect. Public Health* 15, 1270–1275. doi: 10.1016/j.jiph.2022.10.002
- "Gaussian, R.A., Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., et al. *Gaussian 09, revision a.02* Wallingford, CT, USA: Gaussian Inc (2009)."
- Ghahremanian, S., Rashidi, M. M., Raeisi, K., and Toghraie, D. (2022). Molecular dynamics simulation approach for discovering potential inhibitors against SARS-CoV-2: a structural review. *J. Mol. Liq.* 354:118901. doi: 10.1016/j.molliq.2022.118901
- Hunfeld, K.-P., Hildebrandt, A., and Gray, J. (2008). Babesiosis: recent insights into an ancient disease. *Int. J. Parasitol.* 38, 1219–1237. doi: 10.1016/j.ijpara.2008.03.001
- Jafari Porzani, S., Konur, O., and Nowruzi, B. (2022). Cyanobacterial natural products as sources for antiviral drug discovery against COVID-19. *J. Biomol. Struct. Dyn.* 40, 7629–7644. doi: 10.1080/07391102.2021.1899050
- Kabuga, A. I., and El Zowalaty, M. E. (2019). A review of the monkeypox virus and a recent outbreak of skin rash disease in Nigeria. *J. Med. Virol.* 91, 533–540. doi: 10.1002/jmv.25348
- Kashyap, D., Jakhmola, S., Tiwari, D., Kumar, R., Moorthy, N. H. N., Elangovan, M., et al. (2021). Plant derived active compounds as potential anti SARS-CoV-2 agents: an in-silico study. *J. Biomol. Struct. Dyn.* 40, 10629–10650. doi: 10.1080/07391102.2021.1947384
- Kumar, N., Acharya, A., Gendelman, H. E., and Byrareddy, S. N. (2022). The 2022 outbreak and the pathobiology of the monkeypox virus. *J. Autoimmun.*:102855. doi: 10.1016/j.jaut.2022.102855
- Kumar, A., O'Bryan, J., and Krause, P. J. (2021). The global emergence of human babesiosis. *Pathogens* 10:1447. doi: 10.3390/pathogens10111447
- Kumer, A., Chakma, U., Chandro, A., Howlader, D., Akash, S., Kobir, M., et al. (2022). Modified D-glucofuranose computationally screening for inhibitor of breast cancer and triple breast cancer: chemical descriptor, molecular docking, molecular dynamics and QSAR. *J. Chil. Chem. Soc.* 67, 5623–5635. doi: 10.407/S0717-97072022000305623
- Kumer, A., Sarker, N., Paul, S., and Zannat, A. (2019). The theoretical prediction of thermophysical properties, HOMO, LUMO, QSAR and biological indic of cannabinoids (CBD) and tetrahydrocannabinol (THC) by computational chemistry. *Adv. J. Chem.* 2, 190–202. doi: 10.33945/SAMI/AJCA.2019.2.190202
- Lea, T., "Caco-2 cell line," The impact of food bioactives on health pp. 103–111 (2015).
- Li, A. P. (2001). Screening for human ADME/tox drug properties in drug discovery. *Drug Discov. Today* 6, 357–366. doi: 10.1016/S1359-6446(01)01712-3
- Lipinski, C. A. (2004). Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* 1, 337–341. doi: 10.1016/j.ddtec.2004.11.007
- Miladiyah, I., Jumina, J., Haryana, S. M., and Mustofa, M. (2018). Biological activity, quantitative structure–activity relationship analysis, and molecular docking of xanthone derivatives as anticancer drugs. *Drug Des. Devel. Ther.* 12, 149–158. doi: 10.2147/DDDT.S149973
- Minasov, G., Shuvalova, L., Dubrovskaya, I., Flores, K., Grimshaw, S., Kwon, K., et al. (2014). "1.52 angstrom crystal structure of A42R profilin-like protein from monkeypox virus Zaire-96-I-16" Center for Structural Genomics of Infectious Diseases (CSGID).
- Mir, S. A., Meher, R. K., Baitharu, I., and Nayak, B. (2022). Molecular dynamic simulation, free binding energy calculation of Thiazolo-[2, 3-b] quinazolinone derivatives against EGFR-TKD and their anticancer activity. *Results Chem.* 4:100418. doi: 10.1016/j.rechem.2022.100418
- Mir, S. A., Muhammad, A., Padhiary, A., Ekka, N. J., Baitharu, I., Naik, P. K., et al. (2023). Identification of potent EGFR-TKD inhibitors from NPACT database through combined computational approaches. *J. Biomol. Struct. Dyn.*, 1–14. doi: 10.1080/07391102.2023.2218932
- Mir, S., and Nayak, B., Molecular dynamic simulations and binding free energy evaluations of Thiazolo-[2, 3-b] quinazolinone derivatives with wtEGFR-TKD and TMLR mutant EGFR-TKD (2021).
- Mohr, A., Lobetti, R., and Van der Lugt, J. (2000). Acute pancreatitis: a newly recognised potential complication of canine babesiosis. *J. S. Afr. Vet. Assoc.* 71, 232–239. doi: 10.4102/jsava.v71i4.721
- Nath, A., Kumer, A., Zaben, F., and Khan, M. (2021). Investigating the binding affinity, molecular dynamics, and ADMET properties of 2, 3-dihydrobenzofuran derivatives as an inhibitor of fungi, bacteria, and virus protein. *Beni-Suef Univ. J. Basic Appl. Sci.* 10, 1–13. doi: 10.1186/s43088-021-00117-8
- Nolasco, S., Vitale, F., Geremia, A., Tramuto, F., Maida, C. M., Sciuto, A., et al. (2022). First case of monkeypox virus, SARS-CoV-2 and HIV co-infection. *J. Infect.* 86, e21–e23. doi: 10.1016/j.jinf.2022.08.014
- Ogu, C. C., and Maxa, J. L., "Drug interactions due to cytochrome P450," Waco: Baylor University Medical Center Proceedings (2000).
- Pallicer, J. M., Rosés, M., Ràfols, C., Bosch, E., Pascual, R., and Port, A. (2014). Evaluation of log Po/w values of drugs from some molecular structure calculation software. *ADMET DMPK* 2, 107–114. doi: 10.5599/admet.2.2.45
- Parker, S., Nuara, A., Buller, R. M. L., and Schultz, D. A. (2007). *Human monkeypox: an emerging zoonotic disease*, vol. 2, 17–34. doi: 10.2217/17460913.2.1.17
- Patel, H. M., Noolvi, M. N., Sharma, P., Jaiswal, V., Bansal, S., Lohan, S., et al. (2014). Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Med. Chem. Res.* 23, 4991–5007. doi: 10.1007/s00044-014-1072-3
- Peng, Q., Xie, Y., Kuai, L., Wang, H., Qi, J., Gao, G. F., et al. (2022). Structure of monkeypox virus DNA polymerase holoenzyme. *Science* eade6360. doi: 10.1126/science.ade6360
- Pires, D. E., Blundell, T. L., and Ascher, D. B. (2015). pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* 58, 4066–4072. doi: 10.1021/acs.jmedchem.5b00104
- Rahman, M. A., Chakma, U., Kumer, A., Rahman, M. R., and Matin, M. M. (2022). Uridine-derived 4-aminophenyl 1-thioglucoisides: DFT optimized FMO, ADME, and antiviral activities study. *Biointerface Res. Appl. Chem.* 13, 1–15. doi: 10.33263/BRIAC131.052
- Rosner, F., Zarrabi, M. H., Benach, J. L., and Habicht, G. S. (1984). Babesiosis in splenectomized adults: review of 22 reported cases. *Am. J. Med.* 76, 696–701. doi: 10.1016/0002-9343(84)90298-5
- Sah, R., Padhi, B. K., Siddiq, A., Abdelal, A., Reda, A., Ismail Lashin, B., et al. (2022). Public health emergency of international concern declared by the World Health

- Organization for monkeypox. *Glob. Secur. Health Sci. Policy* 7, 51–56. doi: 10.1080/23779497.2022.2124185
- Sharma, S., Kumar, P., and Chandra, R. (2019). Applications of BIOVIA materials studio, LAMMPS, and GROMACS in various fields of science and engineering molecular dynamics simulation of nanocomposites using BIOVIA materials studio. *Lammps Gromacs*, 329–341. doi: 10.1016/B978-0-12-816954-4.00007-3
- Siddiquey, F., Roni, M., Kumer, A., Chakma, U., and Matin, M. (2022). Computational investigation of Betalain derivatives as natural inhibitor against food borne bacteria. *Curr. Chem. Lett.* 11, 309–320. doi: 10.5267/j.ccl.2022.3.003
- Singh, S., Baker, Q. B., and Singh, D. B. (2022). “Molecular docking and molecular dynamics simulation” in *Bioinformatics* (London: Elsevier), 291–304.
- Spielman, A. (1994). The emergence of Lyme disease and human babesiosis in a changing environment. *Ann. N. Y. Acad. Sci.* 740, 146–156. doi: 10.1111/j.1749-6632.1994.tb19865.x
- Tabti, K., Elmchichi, L., Sbai, A., Maghat, H., Bouachrine, M., Lakhli, T., et al. (2022). In silico design of novel PIN1 inhibitors by combined of 3D-QSAR, molecular docking, molecular dynamic simulation and ADMET studies. *J. Mol. Struct.* 1253:132291. doi: 10.1016/j.molstruc.2021.132291
- Tanowitz, H. B., and Weiss, L. M., *Parasitic diseases, diagnostic approaches, and therapies*, (2009). Wiley: Hindawi.
- Thornhill, J. P., Barkati, S., Walmsley, S., Rockstroh, J., Antinori, A., Harrison, L. B., et al. (2022). Monkeypox virus infection in humans across 16 countries—April–June 2022. *N. Engl. J. Med.* 387, 679–691. doi: 10.1056/NEJMoa2207323
- U.S.NEWS (2023) Tick-borne illness babesiosis spreads to new U.S. states. (U.S. News) Available at: <https://www.usnews.com/news/health-news/articles/2023-03-17/tick-borne-illness-babesiosis-spreads-to-new-u-s-states#:~:text=March2023ata.m.&text=FRIDAY20March2023U.S.healthofficialsreported%20Friday> (accessed March 17, 2023).
- Vannier, E., and Gelfand, J. A., *Babesia species, principles and practice of infectious diseases*. Philadelphia, PA: Elsevier, 3400–3409 (2020).
- Vougas, K., Sakellariopoulos, T., Kotsinas, A., Foukas, G.-R. P., Ntargaras, A., Koinis, F., et al. (2019). Machine learning and data mining frameworks for predicting drug response in cancer: an overview and a novel in silico screening process based on association rule mining. *Pharmacol. Ther.* 203:107395. doi: 10.1016/j.pharmthera.2019.107395
- Walters, W. P., and Murcko, M. A. (2002). Prediction of ‘drug-likeness’. *Adv. Drug Deliv. Rev.* 54, 255–271. doi: 10.1016/S0169-409X(02)00003-0
- Walters, W. P., Murcko, A. A., and Murcko, M. A. (1999). Recognizing molecules with drug-like properties. *Curr. Opin. Chem. Biol.* 3, 384–387. doi: 10.1016/S1367-5931(99)80058-1
- Wang, T., Wu, M.-B., Lin, J.-P., and Yang, L.-R. (2015). Quantitative structure–activity relationship: promising advances in drug discovery platforms. *Expert Opin. Drug Discovery* 10, 1283–1300. doi: 10.1517/17460441.2015.1083006
- Xu, W., Ling, P., and Zhang, T. (2013). Polymeric micelles, a promising drug delivery system to enhance bioavailability of poorly water-soluble drugs. *J. Drug Delivery* 2013, 1–15. doi: 10.1155/2013/340315
- Yannielli, L., and Alcamo, I. E., *Lyme disease* (2009) New York: Infobase publishing.
- Yinka-Ogunleye, A., Aruna, O., Ogoina, D., Aworabhi, N., Eteng, W., Badaru, S., et al. (2018). Reemergence of human monkeypox in Nigeria, 2017. *Emerg. Infect. Dis.* 24, 1149–1151. doi: 10.3201/eid2406.180017
- Yu, L., Shen, Z., Liu, Q., Zhan, X., Luo, X., An, X., et al. (2019). Crystal structures of *Babesia microti* lactate dehydrogenase BmLDH reveal a critical role for Arg99 in catalysis. *FASEB J.* 33, 13669–13682. doi: 10.1096/fj.201901259R
- Yu, D., Wang, L., and Wang, Y. (2022). Recent advances in application of computer-aided drug Design in Anti-Influenza a Virus Drug Discovery. *Int. J. Mol. Sci.* 23:4738. doi: 10.3390/ijms23094738



OPEN ACCESS

EDITED BY

Sinosh Skariyachan,
St. Pius X College, India

REVIEWED BY

Anand Anbarasu,
VIT University, India
Milad Zandi,
Tehran University of Medical Sciences, Iran

*CORRESPONDENCE

Yunjiao He
✉ heyj@sustech.edu.cn
Peng George Wang
✉ wangp6@sustech.edu.cn

†These authors have contributed equally to this work

RECEIVED 10 April 2023

ACCEPTED 03 July 2023

PUBLISHED 21 July 2023

CITATION

Rcheulishvili N, Mao J, Papukashvili D, Feng S, Liu C, Wang X, He Y and Wang PG (2023) Design, evaluation, and immune simulation of potentially universal multi-epitope mpox vaccine candidate: focus on DNA vaccine. *Front. Microbiol.* 14:1203355. doi: 10.3389/fmicb.2023.1203355

COPYRIGHT

© 2023 Rcheulishvili, Mao, Papukashvili, Feng, Liu, Wang, He and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Design, evaluation, and immune simulation of potentially universal multi-epitope mpox vaccine candidate: focus on DNA vaccine

Nino Rcheulishvili[†], Jiawei Mao[†], Dimitri Papukashvili, Shunping Feng, Cong Liu, Xingyun Wang, Yunjiao He^{*} and Peng George Wang^{*}

Department of Pharmacology, School of Medicine, Southern University of Science and Technology, Shenzhen, China

Monkeypox (mpox) is a zoonotic infectious disease caused by the mpox virus. Mpox symptoms are similar to smallpox with less severity and lower mortality. As yet mpox virus is not characterized by as high transmissibility as some severe acute respiratory syndrome 2 (SARS-CoV-2) variants, still, it is spreading, especially among men who have sex with men (MSM). Thus, taking preventive measures, such as vaccination, is highly recommended. While the smallpox vaccine has demonstrated considerable efficacy against the mpox virus due to the antigenic similarities, the development of a universal anti-mpox vaccine remains a necessary pursuit. Recently, nucleic acid vaccines have garnered special attention owing to their numerous advantages compared to traditional vaccines. Importantly, DNA vaccines have certain advantages over mRNA vaccines. In this study, a potentially universal DNA vaccine candidate against mpox based on conserved epitopes was designed and its efficacy was evaluated *via* an immunoinformatics approach. The vaccine candidate demonstrated potent humoral and cellular immune responses *in silico*, indicating the potential efficacy *in vivo* and the need for further research.

KEYWORDS

monkeypox, mpox virus, universal vaccine, DNA vaccine, immunoinformatics, *in silico*, epidemics, pandemics

1. Introduction

The history of pandemics has shown us that not a single virus outbreak should be neglected as sooner or later the re-emergence of the same or even more virulent strain may occur with severe outcomes. The spread of the monkeypox (mpox) virus was considered limited to Central and West African countries until May 2022 when it crossed the continental borders and gained global health emergency status by World Health Organization (WHO) on 23 July 2022 (Papukashvili et al., 2022b; World Health Organization, 2022b). Mpox is a zoonotic disease with similar clinical manifestation as the smallpox virus which also belongs to the genus *Orthopoxvirus* with an additional symptom of lymph node enlargement. Mpox disease is characterized by a rash appearing 1–3 days following the onset of fever and lymphadenopathy. The rash is usually distributed on the whole body, especially on extremities (Sklenovská and Van Ranst, 2018), genitals, and anus (Kmieć and Kirchhoff, 2022; Mileto et al., 2022; Moschese et al., 2022; Sah et al., 2022; World Health Organization, 2022a). The mpox virus is a large, ~280 nm X ~220, brick/oval-shaped virus with a large linear double-stranded ~197 kb long DNA genome (Papukashvili et al., 2022b) with more

than 190 open reading frames (ORFs) (Shchelkunov et al., 2002; Kugelman et al., 2014; Kmiec and Kirchhoff, 2022; Vandenberg et al., 2022; Zhu et al., 2022). The mpox virus produces two infectious forms of virions from the infected cells—an intracellular mature virion (IMV) and an extracellular enveloped virion (EEV) with an extra envelope. The structure of the mpox virus is complex, and many of the viral antigens are not well-studied. Infection by EEV is more efficient than by IMV (Locker et al., 2000).

One of the main reasons for the recent mpox multi-country outbreaks is the cessation of smallpox vaccination in 1980 which seems reasonable as most of the current mpox cases are detected in smallpox-unvaccinated people (Papukashvili et al., 2022b). Although the smallpox vaccine which is based on the vaccinia virus, another representative of *Orthopoxvirus*, is effective against the mpox virus (Zandi et al., 2023), there is no specific mpox vaccine available till now. The modified vaccinia virus Ankara-Bavarian Nordic (MVA-BN), also known as Imvamune, Jynneos, or Imvanex, is a third-generation authorized smallpox vaccine and is also used as the mpox vaccine (Zaack et al., 2022). Although the currently available vaccine is effective for the mpox virus (CDC, 2022; Kandeel et al., 2023), Zaack et al. have demonstrated that immunization series with the two-shot MVA-BN vaccine, in non-primed individuals, yields relatively low levels of mpox virus-neutralizing antibodies (Zaack et al., 2022). The line of preclinical studies has made efforts in the development of the mpox vaccine. Hooper et al. have shown that the DNA vaccine encoding the antigens L1, A27, A33, and B5R protected rhesus macaques from the lethal challenge of the mpox virus (Hooper et al., 2004). Other studies have provided evidence of protection from the mpox virus challenge in non-human primates after immunization with A27, A33, B5, and L1 proteins (Buchman et al., 2010; Heraud et al., 2022). However, these vaccines were strain-specific. Importantly, recently worldwide spread mpox virus is of the West African clade with a mortality rate of approximately 1–3.6% (Kmiec and Kirchhoff, 2022; Kozlov, 2022; Yang, 2022), while the mpox of Central African clade is deadlier with a mortality rate of up to 11% (Jezek et al., 1987; Shafaati and Zandi, 2022; Yang, 2022). Considering the possible severe outcomes of whether the mpox virus of the Central African clade is spread, developing a new universal vaccine against mpox is urgently needed. Nucleic acid vaccines have gained enormous attention in recent years due to their efficacy, safety, cost-effectiveness, and time-saving features compared with conventional approaches. Apparently, the available mRNA vaccines for the current coronavirus disease 2019 (COVID-19) have high efficacy and safety profiles (Pardi et al., 2018; Corbett et al., 2020; Walsh et al., 2020; Haas et al., 2021; Kowalik et al., 2021; Liu et al., 2022b). On the other hand, studies on DNA vaccines demonstrated no less advantageous features than mRNA vaccines including the requirement of even less production time and high stability (Flingai et al., 2013; Williams, 2013; Chavda et al., 2021b; Xia et al., 2021; Wang et al., 2022). More importantly, there is already an available DNA vaccine for COVID-19 approved in India (Sheridan, 2021). According to the current situation of available vaccines and the general epidemiological picture, developing a universal next-generation effective anti-mpox vaccine is necessary. To design a potentially universal DNA vaccine contender against mpox virus, in the present research, according to their functions and immunogenicity (Shchelkunov et al., 2002; Hooper et al.,

2004; Buchman et al., 2010; Hirao et al., 2011; Meng et al., 2018; Zhang et al., 2023), sequences of four antigen proteins (A5L, A15L, A35R, and B6R) were retrieved and aligned, and the conserved sequences were selected to predict B-cell, MHC-I, and MHC-II binding epitopes. Except for the importance of immune response and protection against viral challenges, another rationale behind choosing these representative antigens was to assess the effectiveness of the novel antigen combination, aiming to propose another candidate for an anti-mpox vaccine. After the design of the vaccine, the final construct was optimized, and the structure and various characteristics of the vaccine were predicted. Immune simulation analyses demonstrated the strong humoral and cellular responses that warrant the contribution of the vaccine candidate in potentially universal multi-epitope DNA vaccine development.

2. Materials and methods

2.1. Selection of antigens, collection of data, and selection of conserved sequences

Four antigen proteins, namely, A5L, A15L, A35R, and B6R, were chosen according to their function in the life cycle and immunogenicity of the mpox virus. This selection also aimed to assess the immune responses generated by the vaccine against this combination of antigens. They are the orthologs of vaccinia virus glycoproteins—A4L, A14L, A33R, and B5R, respectively (Shchelkunov et al., 2002). These antigens are highly conserved among the mpox, vaccinia, and variola viruses. The protein sequences of mpox virus antigens reported on the National Center for Biotechnology Information (NCBI) database were retrieved. The number of downloaded sequences for A5L, A15L, A35R, and B6R was 1819, 355, 349, and 376, respectively. All the sequences of A5L were used for further analyses as they met the selection criteria such as the known length of the protein sequence. For the rest of the three proteins, the redundant sequences were removed according to their length and quality—179 sequences of A15L, 178 sequences of A35R, and 178 sequences of B6R were used for further analyses. The software Jalview 2.11.1.4 was used for the alignment and conservancy analysis (Waterhouse et al., 2009). The threshold filter of conservation was adjusted to 10 below the threshold, and the consensus sequence was used for B-cell and T-cell (MHC-I and MHC-II binding) epitope prediction.

2.2. Epitope prediction and selection

NetMHCpan EL 4.1 server was applied for the prediction of cytotoxic T lymphocyte (CTL) epitopes (<https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1>) (Reynisson et al., 2020). This server predicts the binding of peptides to any MHC-I molecule of the known sequence using artificial neural networks (ANNs) that are a collection of simple interconnected algorithms processing information according to the external input. The globally most prevalent alleles HLA-A*01:01, HLA-B*07:02, and HLA-B*08:01 were selected from the “Allele Frequency Net Database” and used for the prediction of the MHC-I binding epitope. The default parameters of the server were used: the

threshold for strong binders was set to 0.5 (% rank), while for weak binders it was 2%. The predicted peptides were identified as strong binders when the % rank was below the set threshold (0.5%), while the weak binders were considered when the peptide rank was above 0.5% and below 2%. For the prediction of helper T lymphocyte (HTL) epitopes, the NetMHCIIpan-4.0 server which performs the prediction of peptide binding to any MHC-II molecule of the known sequence using ANNs was used. The worldwide most prevalent MHC-II alleles DRB1*07:01 and DRB1*15:01 were selected for the prediction of MHC-II binding peptides of the mpox virus conserved sequences. According to the default parameters, the predicted peptides were identified as strong binders when the % rank was below the set threshold (1%), while the weak binders were considered when the peptide rank was above 1% and below 2%. The epitopes with the strongest binding capacity to MHC-I and MHC-II alleles were selected. Both NetMHCpan-4.1 and NetMHCIIpan-4.0 use the NNAlign_MA machine learning framework, which has ANNs at its core. NetMHCpan-4.1, a computational tool, has been trained using a comprehensive dataset comprising over 850,000 peptides with quantitative binding affinity (BA) and mass spectrometry-eluted ligand (EL) information. The BA data utilized in the training of NetMHCpan-4.1 encompass 170 MHC molecules from various species—humans (HLA-A, HLA-B, HLA-C, and HLA-E), mice (H-2), cattle (BoLA), primates (Patr, Mamu, and Gogo), equine (Eqca), and swine (SLA). The EL data used in the analysis include 177 MHC molecules from a range of species, such as humans (HLA-A, B, C, E), mice (H-2), cattle (BoLA), primates (Patr, Mamu, and Gogo), swine (SLA), equine (Eqca), and dogs (DLA). NetMHCIIpan-4.0 has been trained using a comprehensive dataset comprising more than 500,000 measurements of BA and EL. This dataset covers the three human MHC class II isotypes, namely HLA-DR, HLA-DQ, and HLA-DP, along with the mouse molecules (H-2). The inclusion of EL data expands the coverage of MHC-II molecules, as the BA data cover 59 molecules, while the EL data cover 74 molecules. Both NetMHCpan-4.1 and NetMHCIIpan-4.0 are the currently recommended algorithms in Immune Epitope Database (IEDB). They demonstrate cutting-edge performance and surpass other algorithms in benchmark experiments (Reynisson et al., 2020).

For predicting the linear B-cell epitopes, the IEDB with the method of Bepipred linear epitope prediction tool (v2.0) was utilized (Jespersen et al., 2017). The BepiPred-2.0 server performs the prediction of B-cell epitopes from a protein sequence *via* a random forest algorithm trained on epitopes and non-epitope amino acids determined from crystal structures. The threshold was set at 0.6 meaning that the residues scoring above the threshold were predicted to be part of the epitopes. The B-cell epitopes were then selected.

2.3. Multi-epitope DNA vaccine design

After the prediction of B-cell, MHC-I, and MHC-II epitopes, the epitopes with the most optimal features were selected. Multi-peptide vaccine construct was designed and epitopes of A5L, A15L, A35R, and B6R were fused *via* the appropriate linkers. The linkers play a crucial role in multi-epitope vaccine development

as their reasonable selection allows to enhance structural stability, flexibility, and proper folding, as well as to increase the immune response by allowing to include multiple epitopes in a single vaccine construct. Thus, the well-described linkers KK, GPGPG, and EAAAK were used to connect different peptide components. EAAAK is a rigid linker that provides an alpha helix-forming structure between the domains that enhances the stability and maintains the constant spacing between the domains of the fusion protein as well as preserves their functions (Chen et al., 2013). B-cell epitopes were linked with a flexible KK linker (Rahmani et al., 2021; Tarrahimofrad et al., 2021). The KK linker allows the presentation of each linked epitope to antibodies while avoiding antibody induction against the whole joined sequences resulting in the antibody reactivity to each of the B-cell epitopes (Yano et al., 2005). MHC-I and MHC-II epitopes were conjugated together with another flexible linker GPGPG which is a glycine-rich linker, in addition enhancing the solubility of vaccine construct, and also provides high accessibility and flexibility for the adjacent epitopes (Dong et al., 2020; Khan et al., 2021; Martinelli, 2022). Except for the epitopes, to rationally optimize the fused multi-peptide construct for the optimal immunogenic outcome, adjuvants were also fused. Particularly, 45 amino acid long human β -defensin 3 (hBD3) with immunomodulatory effect (Hoover et al., 2003; Ali et al., 2017; Qamar et al., 2020) was adjoined to the N-terminal of the vaccine after the leading sequence of tissue plasminogen activator which also facilitates the antigen presentation (Ahammad and Lira, 2020). hBD3 is one of the top five adjuvants utilized for COVID-19 subunit vaccines (Mekonnen et al., 2022). Furthermore, defensins play a crucial role in defending against pathogen infections as they effectively bridge the innate and adaptive immune responses through leukocytes such as dendritic cells (DCs) and T cells (Bellamkonda et al., 2022). The rigid linker EAAAK (Nezafat et al., 2014; Tarrahimofrad et al., 2021) was used to conjugate the hBD3 with the pan-HLA-DR-binding epitope (PADRE) which triggers antigen-specific CD4⁺ T cells (Alexander et al., 2000). EAAAK was used to connect the PADRE sequence to the rest of the construct as well. For future *in vitro* validation of the expression of the multi-epitope construct, the C-terminal polyhistidine tag—6xHis—was also adjoined (Boyoglu-Barnum et al., 2021). The untranslated regions (UTRs) were included for ensuring the proper stability and translation of the multi-epitope construct when it is transcribed into mRNA and for allowing the translation efficiency and stability of mRNA while validating the protein expression *in vitro* and *in vivo*. The following UTRs were used: cytomegalovirus (HCMV) immediate early (IE) partial sequence as 5'UTR, and human growth hormone partial sequence as a 3'UTR as they have demonstrated the effective stimulation of the protein production *in vivo* (Rybakova et al., 2019).

2.4. Vaccine properties—*In silico* assessment

The vaccine should have wide human population coverage. Thus, the IEDB population coverage analysis tool was employed to evaluate the worldwide human population coverage of the designed DNA vaccine (Bui et al., 2006). The allergenicity of

the designed vaccine was predicted *via* the AllerCatPro server which predicts the allergenic proteins based on the similarity of both their amino acid sequences and 3D structures compared with the comprehensive dataset of already known allergens from the WHO/International Union of Immunological Societies, Comprehensive Protein Allergen Resource, Food Allergy Research and Resource Program, UniProtKB, and Allergome (Maurer-Stroh et al., 2019; Nguyen et al., 2022). VaxiJen v2.0 was applied to predict the antigenicity of the multi-epitope protein construct (Doytchinova and Flower, 2007a,b). The threshold was set as 0.4, and default parameters were employed (Ghafouri et al., 2022).

The multi-epitope vaccine physico-chemical properties were assessed by ProtParam (Gasteiger et al., 2005). Molecular weight (MW), the composition of atoms and amino acids, instability, estimated half-life, theoretical isoelectric point (pI), aliphatic indexes, and grand average of hydropathicity (GRAVY) of the vaccine construct were evaluated. The MW was determined by summing amino acids' average isotopic masses and the average isotopic mass of a single water molecule. pI was calculated by amino acids' pKa value. Half-life estimation describes the necessary time from protein synthesis until the disappearance of its half-amount from the cell. The instability index was used to determine how stable is the protein in a test tube. The value <40 estimates the protein as stable. GRAVY calculates the hydropathicity index of all amino acids divided by the length of the sequence. The larger number denotes more hydrophobicity of amino acids (Gasteiger et al., 2005). An aliphatic index is used to characterize the relative volume of protein which is occupied by aliphatic side chains and is considered a positive factor in increasing thermostability. The solubility of the vaccine protein was predicted *via* DeepSoluE. This server is a deep-learning predictor, and its prediction algorithm outperforms the other servers for the prediction of protein solubility (Wang and Zou, 2023).

2.5. Prediction and quality assessment of the vaccine's 3D structure

The 3D structure of the designed vaccine was constructed *via* RoseTTAFold (Baek et al., 2021). This tool uses a three-track network: integrating protein sequence patterns, amino acid interactions, and 3D structure to accurately predict protein structure and interaction (Baek et al., 2021). After the 3D structure of the vaccine protein was predicted, it was optimized with GalaxyRefine and FG-MD which is a molecular dynamics (MD)-based algorithm for protein structure refinement at the atomic level (Zhang et al., 2011; Heo et al., 2013). This immunoinformatics approach uses molecular dynamics simulation and relaxes and refines the protein structure (Heo et al., 2013). Predicting and knowing the potential structure is important as it determines the function of the protein that allows its application. As the predicted 3D structure provides the understanding of its interactions with the immune system, it is then used for further analyses, e.g., prediction of B-cell discontinuous epitopes and molecular docking. The tertiary structure quality of the 3D protein was then verified with the Ramachandran plot and ERRAT. The Ramachandran plot illustrates energetically permissible regions for the backbone

dihedral angles ψ and ϕ of amino acid residues within a protein structure allowing the assessment of 3D protein structure quality (Nelson et al., 2021). ERRAT also represents the quality but for non-bonded interactions, and a higher score of ERRAT indicates a higher quality of the tertiary structure of the protein.

2.6. Presence of conformational B-cell epitopes in the vaccine

To predict the discontinuous B-cell epitopes, ElliPro was employed. This tool is the recourse of the immune epitope database (IEDB). It allows for predicting the conformational B-cell epitopes according to the antigen's tertiary structure. ElliPro associates predicted discontinuous epitopes with protrusion index (PI) which is averaged over epitopes' residues. The approximation of the antigen's tertiary structure is achieved *via* the number of ellipsoids. The PI score is defined based on the residue's mass center which is lying outside the largest ellipsoid. A higher value suggests higher solvent accessibility which, on the other hand, is crucial in protein stability and folding (Ponomarenko et al., 2008).

2.7. *In silico* immunization

The C-ImmSim server was utilized to analyze the simulated humoral and cellular immune responses elicited against the multi-epitope vaccine protein. After submitting the sequence of vaccines in the FASTA format, the server predicts the immune responses *via* a position-specific scoring matrix and machine learning algorithms (Rapin et al., 2011). Two HLA-A (01:01 and 02:01), two HLA-B (07:02 and 08:01), and two DRB (07:01 and 15:01) were selected through Allele Frequency Net Database (AFND). The prime, second, and booster doses of the vaccine were injected with 4 weeks gap. The volume and steps of the simulation were 10 and 270, respectively. The vaccine did not contain LPS. The random seed was 12,345, and the injection time periods were 1, 85, and 169 (Rapin et al., 2010).

2.8. Docking of vaccine antigen with host receptors

To evaluate the binding capacity of the vaccine antigen to its recognition receptors of the host, molecular docking analysis was conducted by the server of ClusPro. ClusPro determines the molecular docking of two proteins providing an outcome *via* presenting putative protein complexes in a ranked list. Ligand conformation, orientation, and position along with the assessment of binding affinity are the major properties that determine molecular docking. Ultimately, the electrostatic-favored protein-protein docked complexes with favorable desolvation-free energies are picked (Kozakov et al., 2017). Analysis of vaccine antigen's docking with MHC-I (HLA-A*02:01) (6TDS) and MHC-II (HLA-DRB1*01:01) (1AQD) host receptors was performed. The MHC molecules were used to dock the following epitopes in the vaccine construct:

MHC-I ligands—ETNDLVTNVY (A5L), SPTRTWKVL (A15L), DSDVSQEVRY (A35R), and VSDYVSELY (B6R); MHC-II ligands—APILLPSSTAPVLKP (A5L) and HSDYKSFEDAKANCA (A35R). The tertiary structure of these epitopes was predicted with AlphaFold2 (Jumper et al., 2021; Mirdita et al., 2022). The protein databank (PDB) files 6TDS (MHC-I) and 1AQD (MHC-II) were edited and cleaned to remove heteroatoms, bound peptides, and water molecules. The server of ClusPro performs the docking of rigid body *via* sampling billions of conformations, clusters 1000 lowest energy structures to find the most favorable models, and refines selected structures *via* minimization of energy (Kozakov et al., 2017).

3. Results

3.1. Design of the final construct

All the predicted B-cell, MHC-I, and MHC-II epitopes are given in [Supplementary materials 1–12](#). In total, 30 epitopes were used for the final vaccine construct design. The open reading frame (ORF) of the vaccine construct is shown in [Supplementary material](#). The selected four antigens of the mpox virus are shown in [Figure 1](#). The location, functions, and immune properties of the selected antigens are listed in [Table 1](#). The strategy of *in silico* design and evaluation of a potentially universal DNA vaccine for mpox are shown in [Figure 2](#). All the epitopes included in the proposed multi-epitope vaccine construct along with the DNA vaccine are given in [Figure 3](#).

The population coverage analyses demonstrated that the 5 MHC-I alleles (HLA-A*01:01, HLA-A*02:01, HLA-B*07:02, HLA-B*08:01, and HLA-A*02:12) with IC50 binding affinity <500 nM with the MHC-I epitopes and the 13 MHC-II alleles (HLA-DRB1*07:01, HLA-DRB1*01:01, HLA-DRB1*09:01, HLA-DRB1*10:01, HLA-DRB1*04:01, HLA-DRB1*15:01, HLA-DRB1*14:01, HLA-DRB1*13:01, HLA-DRB1*08:01, HLA-DRB1*03:01, HLA-DRB1*11:01, HLA-DRB1*12:01, and HLA-DRB1*16:01) with IC50 binding affinity <500 nM with MHC-II epitopes cover 95.21% of the human population, which means that these epitopes could cover approximately 95% of the population from different regions of the world ([Supplementary Figure 1](#) and [Supplementary Table 1](#)).

3.2. Analyses of protein structure and assessment of physico-chemical characteristics

The predicted 3D structure of the vaccine construct along with the quality evaluation is given in [Figure 4](#). The distribution of amino acid residues in the Ramachandran plot was as follows: refined structure had 279 residues in the most favored region, 93 residues in the additionally allowed region, 13 in the generously allowed region, and only 7 residues in the disallowed region. The Ramachandran plot allows us to visualize the energetically allowed and disallowed regions for the dihedral angles. For example, in poor quality homology models, many dihedral angles can be found in the forbidden regions of the plot which indicates the problems

with the structure. The favored regions in the Ramachandran plot correspond to the regular secondary structures. The ERRAT score was 88.7218 meaning that all these parameters indicate the favorable overall quality of the vaccine antigen protein.

The instability index of the proposed vaccine construct was 29.12, which classifies the vaccine as stable. The aliphatic index was 60.22 which indicates the high thermostability of the vaccine protein. GRAVY was -0.678 , stipulating hydrophilicity. The half-life *in vitro* was estimated to be 30 h, while the *in vivo* in the yeast and *E. coli* were estimated to be over 20 h and over 10 h, respectively. The number of amino acids was 542, MW was 56.55076 kDa, and the theoretical pI was 9.28. The vaccine construct contains 46 negatively charged amino acid residues (Asp+Glu) and 64 positively charged amino acid residues (Arg+Lys). The probability of the vaccine solubility was 0.7963, indicating the solubility of the vaccine protein (Wang and Zou, 2023).

3.3. Presence of discontinuous B-cell epitopes

All three conformational epitopes along with the scores are listed in [Table 2](#). The locations of each discontinuous epitope in the tertiary structure of the antigen protein are illustrated in [Figure 5](#).

3.4. Assessment of allergenicity, antigenicity, and molecular docking

A vaccine protein construct did not show evidence of allergenicity allowing us to presume that the multi-epitope construct can be considered with low allergenic potential. On the other hand, the multi-peptide was demonstrated to be a probable antigen with a score of 0.4248.

The molecular docking with MHC molecules demonstrated stable binding. ClusPro identifies the most probable complex models by determining the largest clusters available. The interaction positions of the docked complex of vaccine epitopes and MHC molecules are shown in [Figure 6](#). The interaction energy is given in [Table 3](#) where the cluster size is the size of the largest cluster found, while the lowest energy is the Gibbs free energy of the best complex conformation (Kozakov et al., 2017).

3.5. Evaluation of immune response induced by immune simulation

In silico analyses demonstrated that the immune response triggered by the proposed vaccine was compatible with the responses generally induced *via in vivo* immunization. The immune responses after the additional two booster doses of vaccine were stronger compared to the prime immunization. The immunization lowered the antigen level, while high levels of antibodies (IgG and IgM) were produced. Immunoglobulin and antigen levels varied over time. The antigen abundance peaked at each injection time point ([Figure 7A](#)). The humoral

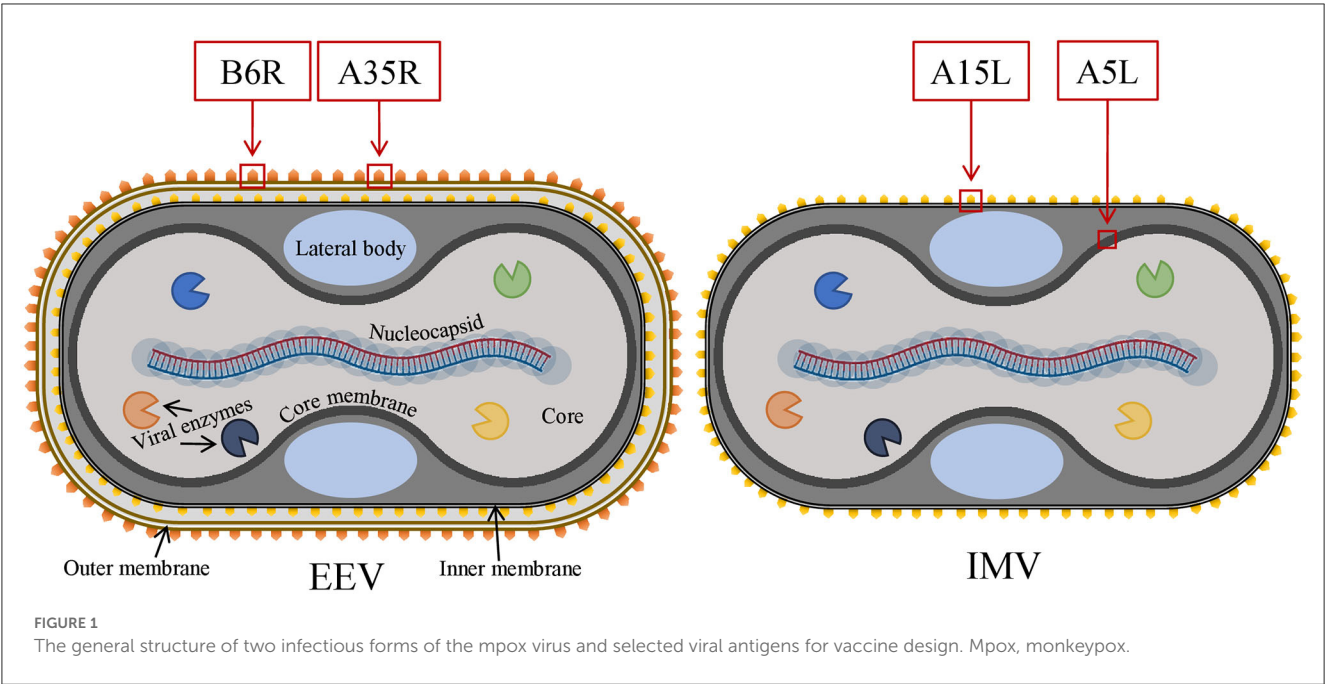


TABLE 1 Selected antigens of the mpox virus and their functions.

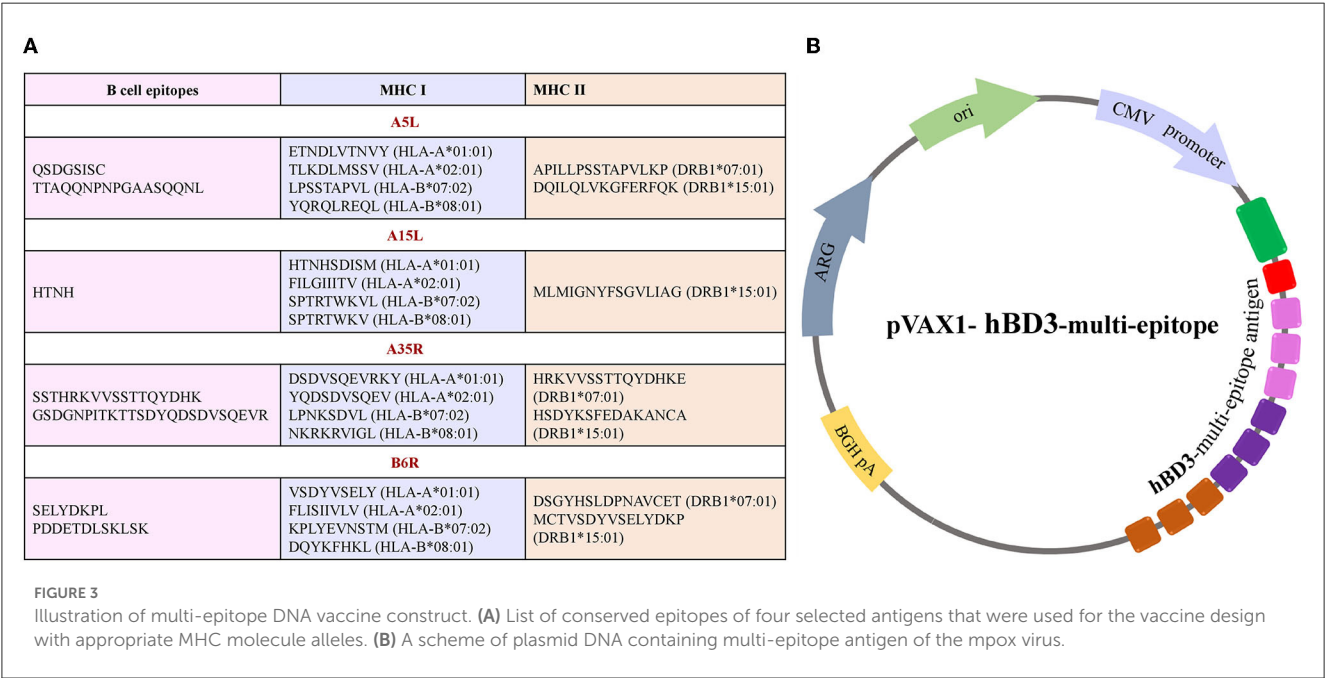
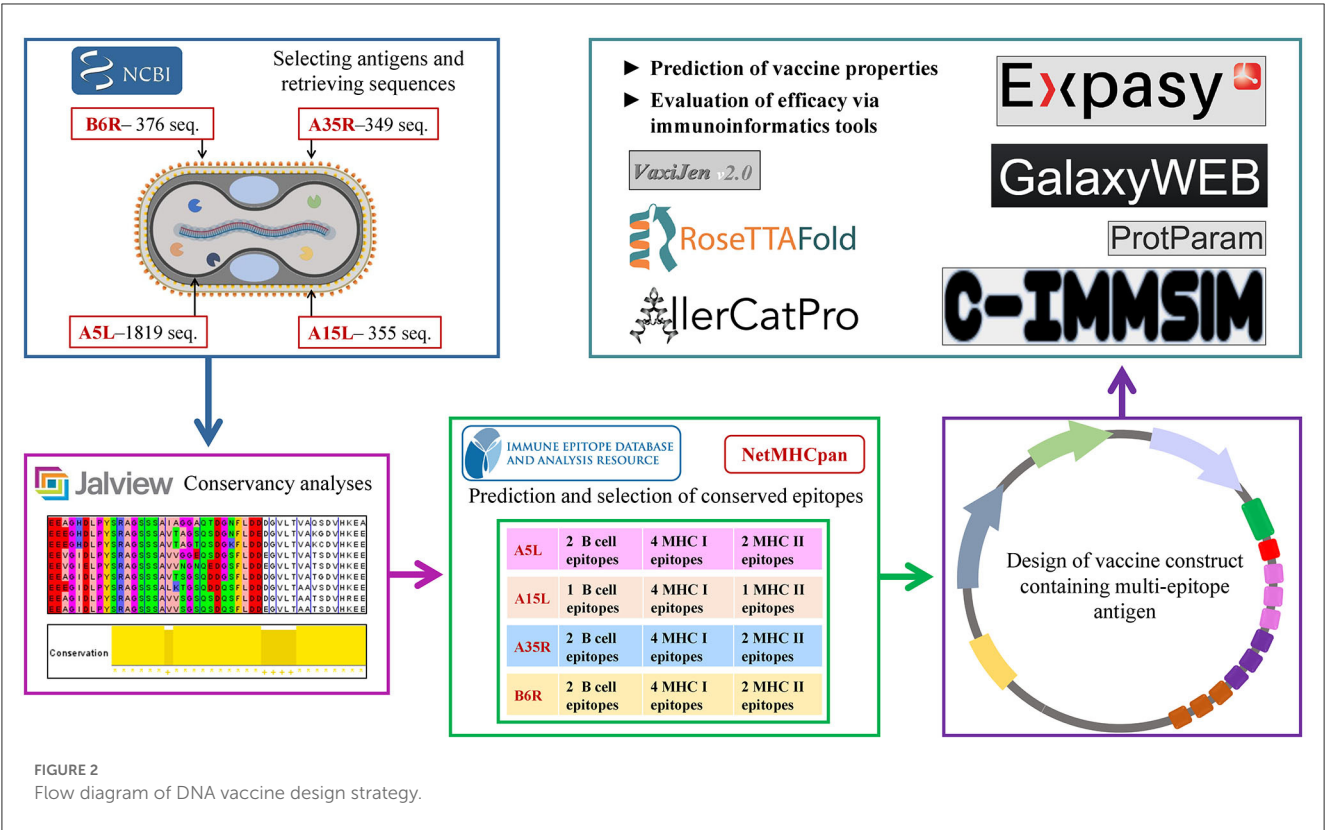
Antigen name	Location	Function/properties	Ref.
A5L	IMV	Immunodominant virion core protein (281 aa) needed for the assembly and disassembly of virion; A highly antigenic protein of the viral core; Enhances the effect of cytotoxic T lymphocytes in the mpox challenge model; Similar to VACV A4L protein.	(Maa and Esteban, 1987; Shchelkunov et al., 2002; Hirao et al., 2011)
A15L	IMV	Essential inner membrane protein (90 aa) of IMV; Immunodominant antigen in smallpox vaccine; Similar to VACV A14L protein.	(Shchelkunov et al., 2002; Meng et al., 2018)
A35R	EEV	Envelope glycoprotein (181 aa); Cell-to-cell spread of virion; Neutralizing antibody target; Similar to VACV A33R protein	(Shchelkunov et al., 2002; Hirao et al., 2011)
B6R	EEV	Palmitoylated glycoprotein (317 aa); Required for efficient cell spread; Complement control; Similar to VACV B5R protein	(Shchelkunov et al., 2002; Hirao et al., 2011)

Mpox, monkeypox; IMV, intracellular mature virus; EEV, extracellular enveloped virus.

response was induced after each shot, and the antibody levels remained elevated during the weeks after the last vaccine injection. The immunization stimulated the B cells (Figures 7B, C). The ignorable number of anergic cells and activation of CTLs was observed (Figure 7D). The counts of HTLs demonstrate that the duplication phase starts immediately after each injection (Figure 7E). Days after each injection and upon successful interaction with antigen-presenting cells, HTLs start to duplicate and differentiate into memory cells. HTLs also foster the secretion of cytokines and humoral response. High levels of IFN- γ , TGF- β , and IL-10 were induced upon each injection (Figure 7F). The other parameters of immune responses are shown in Supplementary Figure 2.

4. Discussion

Since the mpox virus crossed African borders and caused a multi-country outbreak (Supplementary Figure 3) with increased cases of human-to-human transmission, the global concern has increased; thus, there is a strong need for the development of specific drugs and vaccines. Although the available smallpox vaccine is effective for mpox prevention (Shafaati and Zandi, 2023), except for the painful immunization procedure, adverse effects were also reported (Maurer et al., 2003). Certain groups of the population are vulnerable to vaccination as serious side effects including pericarditis and myocarditis have been noted. A risk of recombination of genes between the mpox virus and attenuated or



live poxvirus-based vaccines exists (Lum et al., 2022). In addition, after immunization with the smallpox vaccine, a manifestation of mpox disease still has been observed (Meyer et al., 2002). All the abovementioned disadvantages of the smallpox vaccine can be potentially overcome by the new DNA-based universal mpox vaccine candidate as, on the one hand, DNA vaccines are safe and simple to manufacture compared to their conventional counterparts. On the other hand, the mpox vaccine designed in this study has the potential to be effective against various strains of mpox viruses and smallpox virus since these two species share a high level of similarity. Noteworthy, the majority of the cases occur in men who have sex with men (MSM). Close physical contact is indeed crucial in transmission (Heskin et al., 2022; Martínez et al., 2022; Orviz et al., 2022). The sex-related infection

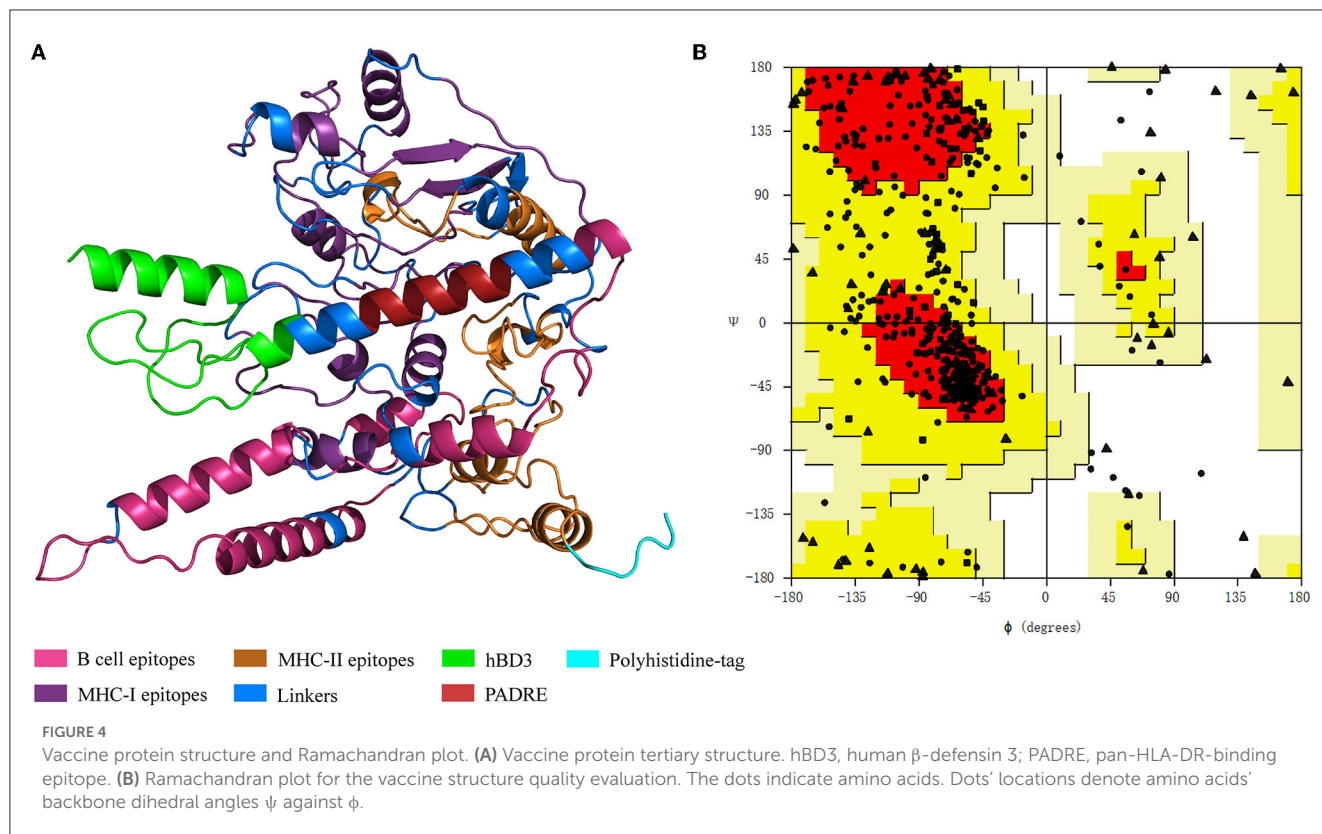


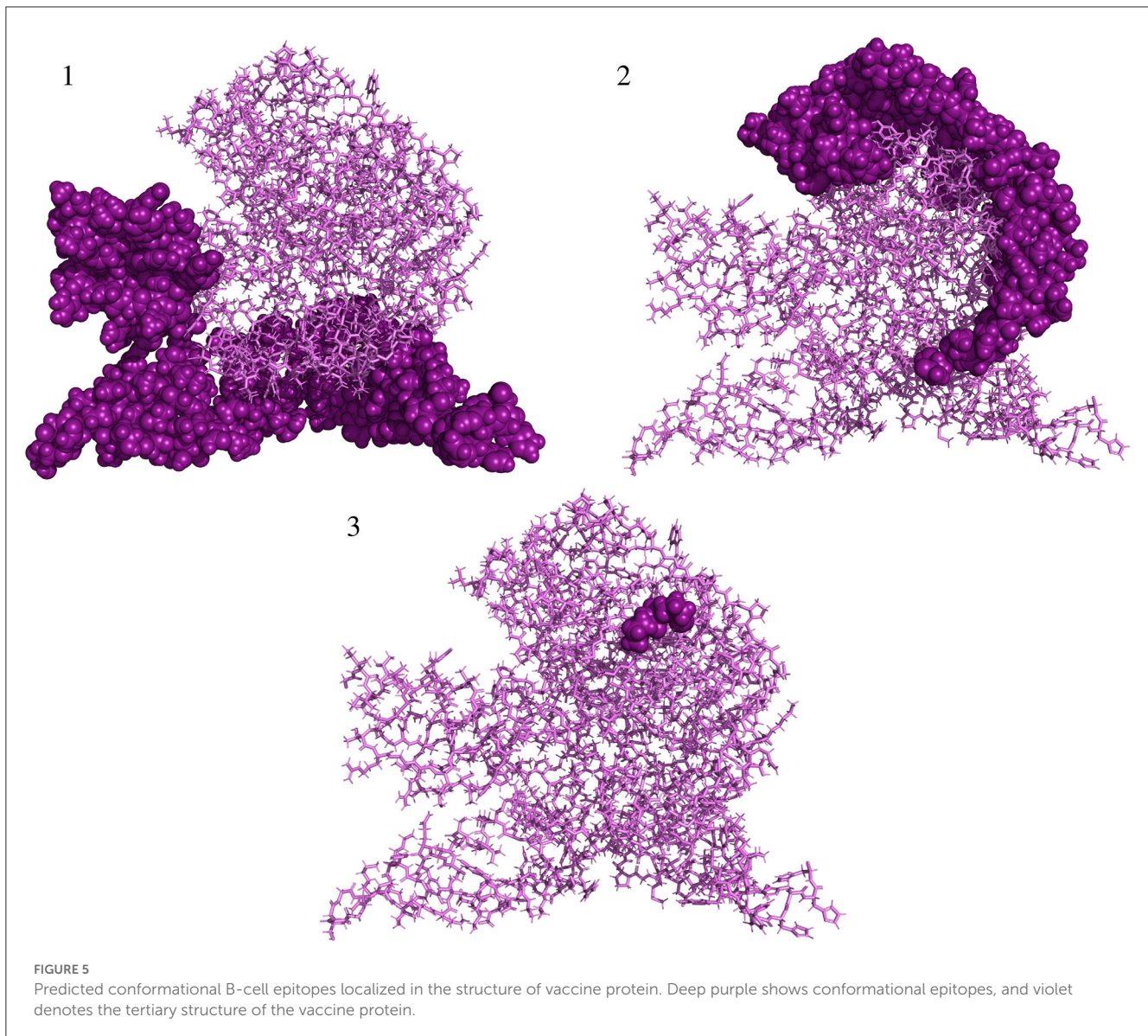
TABLE 2 Predicted discontinuous B-cell epitopes.

#	Epitopes	Residue number	Score
1	A:G1, A:I2, A:I3, A:N4, A:T5, A:L6, A:Q7, A:K8, A:Y9, A:Y10, A:C11, A:V13, A:R14, A:G16, A:R17, A:C18, A:A19, A:V20, A:L21, A:S22, A:C23, A:L24, A:E28, A:Q29, A:I30, A:G31, A:K32, A:C33, A:S34, A:T35, A:R36, A:G37, A:R38, A:K39, A:C40, A:C41, A:R42, A:K109, A:V110, A:V111, A:S112, A:S113, A:T114, A:T115, A:Q116, A:Y117, A:D118, A:H119, A:K120, A:K121, A:K122, A:G123, A:S124, A:D125, A:G126, A:N127, A:P128, A:I129, A:T130, A:K131, A:T132, A:T133, A:S134, A:D135, A:Y136, A:Q137, A:D138, A:S139, A:D140, A:V141, A:S142, A:Q143, A:E144, A:V145, A:R146, A:K147, A:K148, A:S149, A:E150, A:L151, A:Y152, A:D153, A:K154, A:P155, A:L156, A:K157, A:K158, A:P159, A:D160, A:D161, A:E162, A:T163, A:T164, A:L165, A:K167, A:L168, A:K170, A:G171, A:P172, A:G173, A:P174, A:G175, A:D179, A:P203, A:G204, A:Y472, A:H474, A:K475, A:E476, A:G477, A:Y485, A:S487, A:F488, A:G499, A:P500, A:G501, A:D502, A:S503, A:G504, A:Y505, A:H506, A:S507, A:L508, A:D509, A:P510, A:N511, A:A512, A:V513, A:C514, A:E515, A:T516, A:G517, A:P518, A:G519, A:P520, A:G521, A:M522, A:C523, A:T524, A:V525, A:S526, A:D527, A:Y528, A:V529, A:S530, A:E531, A:L532, A:Y533, A:D534, A:K535, A:P536, A:H537, A:H538, A:H539, A:H540, A:H541, A:H542	157	0.746
2	A:A67, A:K68, A:Q69, A:S70, A:D71, A:G72, A:S73, A:I74, A:S75, A:C76, A:K78, A:T79, A:T80, A:A81, A:Q82, A:Q83, A:N84, A:P85, A:N86, A:P87, A:G88, A:A89, A:Q92, A:G272, A:P273, A:G274, A:S275, A:P276, A:T277, A:R278, A:T279, A:W280, A:K281, A:V282, A:G283, A:P284, A:G285, A:P286, A:G287, A:D288, A:S289, A:D290, A:V291, A:S292, A:Q293, A:E294, A:V295, A:R296, A:K297, A:Y298, A:G299, A:P300, A:G301, A:P302, A:G303, A:Y304, A:Q305, A:D306, A:S307, A:D308, A:V309, A:S310, A:Q311, A:E312, A:V313, A:G314, A:P315, A:G316, A:P317, A:G318, A:L319, A:P320, A:N321, A:K322, A:S323, A:D324, A:V325, A:L326, A:G327, A:P328, A:G329, A:P330, A:G331, A:N332, A:K333, A:R334, A:K335, A:R336, A:I338, A:G339, A:L340, A:G341, A:Y354, A:G355, A:P356, A:G357, A:V368, A:G369, A:P370, A:G371, A:P372, A:G373, A:K374, A:P375, A:G384, A:P385, A:G386, A:P387, A:G388, A:H394, A:K395, A:L396, A:G397, A:P398, A:G399, A:P400, A:G401, A:P403, A:D422, A:K429, A:E432, A:R433, A:Q435, A:K436, A:G437, A:P438, A:G439, A:P440, A:G446, A:G456, A:G457, A:P458, A:G459, A:P460	134	0.663
3	A:P416, A:G417, A:P418	3	0.504

tendency was always reported before the current multi-country outbreak (Sklenovská and Van Ranst, 2018).

The report of human-to-dog transmission of the mpox virus also complicates combating the outbreak (Seang et al., 2022). Additionally, the re-emergence risk of smallpox (mortality rate 10–75%) (Javier et al., 2022) or risk of spreading to the Central African clade of the mpox virus, which is much deadlier compared with

the currently spread West African clade (Bunge et al., 2022), is very real and has a huge potential to cause severe outcomes. The recent multi-country outbreak also demonstrated that the mpox virus which belongs to the West African clade can be life-threatening (Meyer et al., 2002). All the abovementioned conditions indicate the urgency of designing a new, specific next-generation mpox vaccine that will be effective, safe, and easy to develop.



In general, nucleic acid-based approaches find ways through various applications such as vaccines against bacterial (Wang et al., 2023) and viral (Hobernik and Bros, 2018) infections, protein replacement (Papukashvili et al., 2022a; Vavilis et al., 2023), and cancer treatment (Hobernik and Bros, 2018; Liu et al., 2022a). *In silico*-designed and evaluated multi-peptide vaccines are thermodynamically stable, effective, specific, and easily and inexpensively developed compared with conventional vaccines.

In this study, four antigens of the mpox virus that are the target of immunity were selected according to their functions and role in immune response (Table 1). A5L is an immunodominant core protein that is necessary for the assembly and disassembly of the virion (Maa and Esteban, 1987; Shchelkunov et al., 2002). A4L, which is the ortholog of the A5L protein in the vaccinia virus, has been used as a part of the DNA vaccine along with the orthologs of A35R and B6R, among the other antigens and induced strong immune response in non-human primates, providing protection from the mpox virus (Hirao et al., 2011). A15L (similar to A14L

in the vaccinia virus) is an inner membrane immunodominant protein which is a target of antibodies (Shchelkunov et al., 2002; Meng et al., 2018). In addition to the aforementioned rationale for selecting these antigens, another reason was to make a vaccine by combining these four antigens to assess the immune response. This approach aimed to generate an additional vaccine candidate as having multiple vaccine candidates would greatly enhance the effectiveness of pandemic prevention efforts. After the conservancy analyses of each protein sequence downloaded from the NCBI database, B- and T-cell conserved epitopes were predicted to design a DNA vaccine. DNA vaccines are a type of nucleic acid vaccine with a number of advantages over conventional immunization strategies (Supplementary Figure 4).

DNA vaccines instruct the host cells to produce antigens similar to the virus and the body becomes ready for the future possible infection, which is capable to stop the spreading of the virus and, thus, a manifestation of clinical symptoms (Chavda et al., 2021a; Qin et al., 2021). After the delivery into the host cell, the

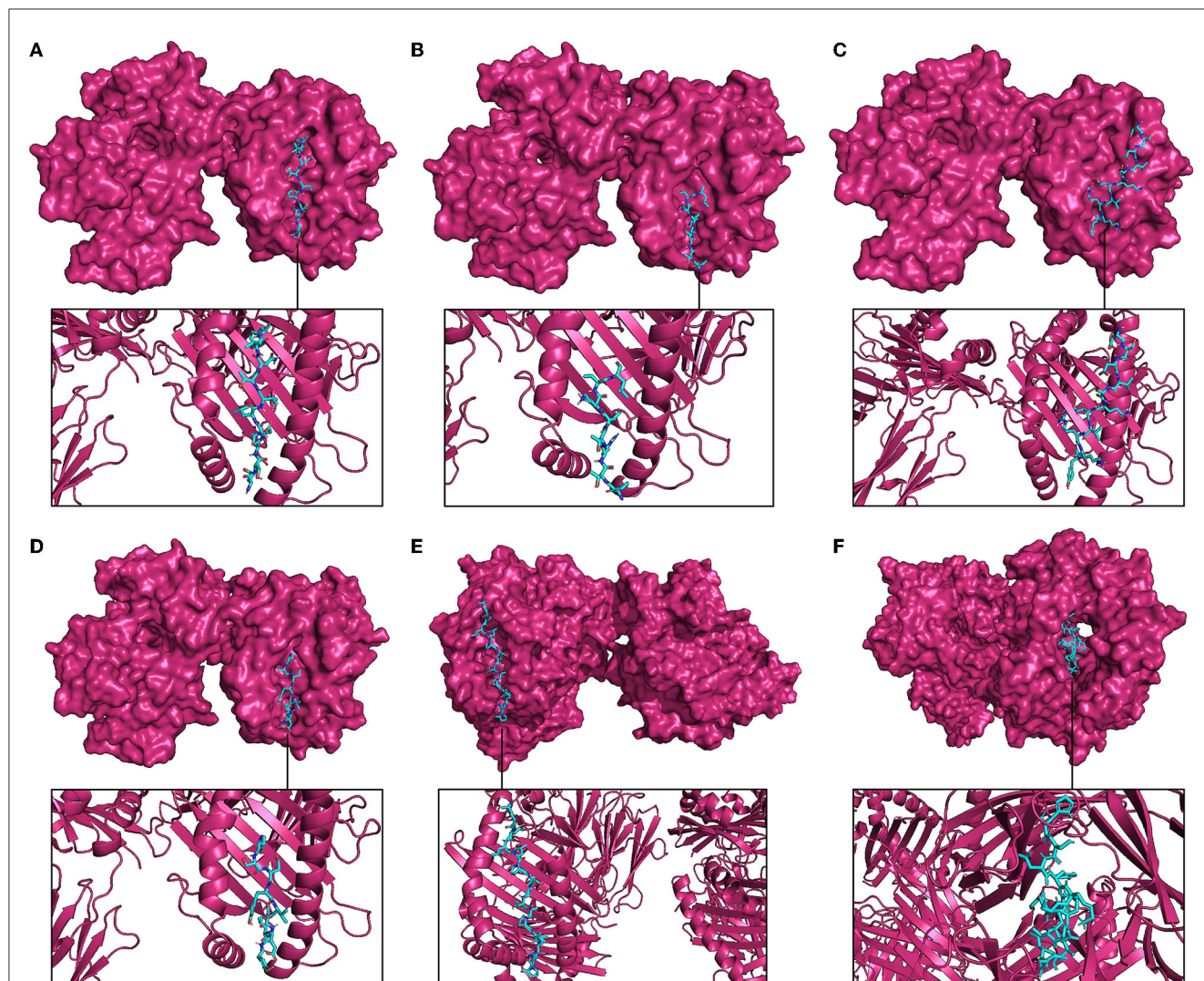


FIGURE 6

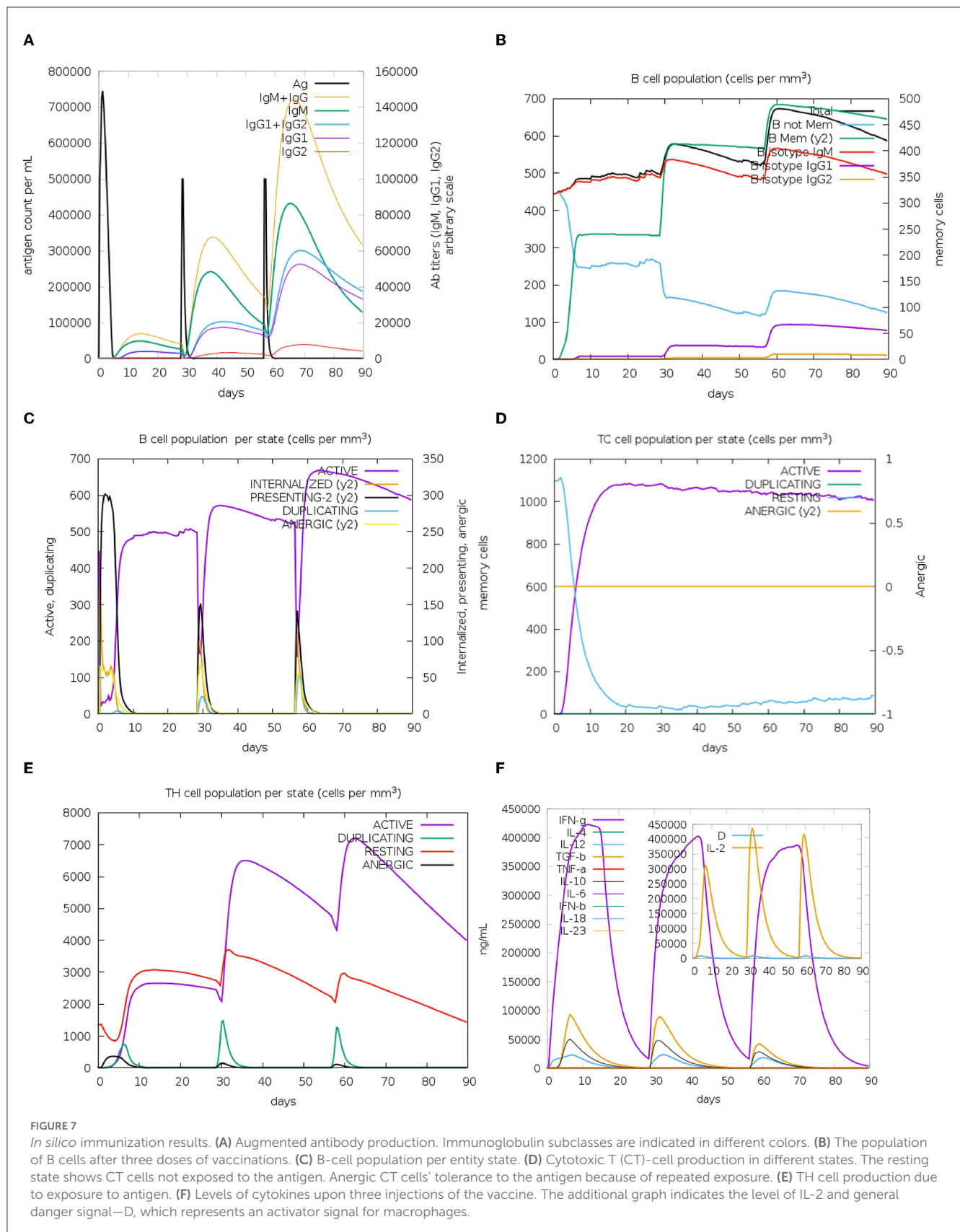
Molecular docking between representative epitopes (cyan) of vaccine construct and MHC molecules (magenta). (A) Docked complex of epitope ETNDLVTNVY (A5L) and MHC-I molecule. (B) Docked complex of epitope SPTRTWKVL (A15L) and MHC-I molecule. (C) Docked complex of epitope DSDVSQEVRY (A35R) and MHC-I molecule. (D) Docked complex of epitope VSDYVSELY (B6R) and MHC-I molecule. (E) Docked complex of epitope APILLPSSTAPVLKP (A5L) and MHC-II molecule. (F) Docked complex of epitope HSDYKSFEDAKANCA (A35R) and MHC-II molecule.

TABLE 3 Molecular docking of selected vaccine epitopes and MHC molecules.

Peptide	Cluster size	Center weighted score	Lowest energy
Molecular docking of epitopes with 6TDS (MHC-I)			
ETNDLVTNVY	252	−909.7	−1,089.0
SPTRTWKVL	396	−1,083.7	−1,239.3
DSDVSQEVRY	217	−647.4	−794.9
VSDYVSELY	344	−1,194.4	−1,194.4
Molecular docking of epitopes with 1AQD (MHC-II)			
APILLPSSTAPVLKP	190	−1,296.4	−1,296.4
HSDYKSFEDAKANCA	244	−845.1	−1,023.0

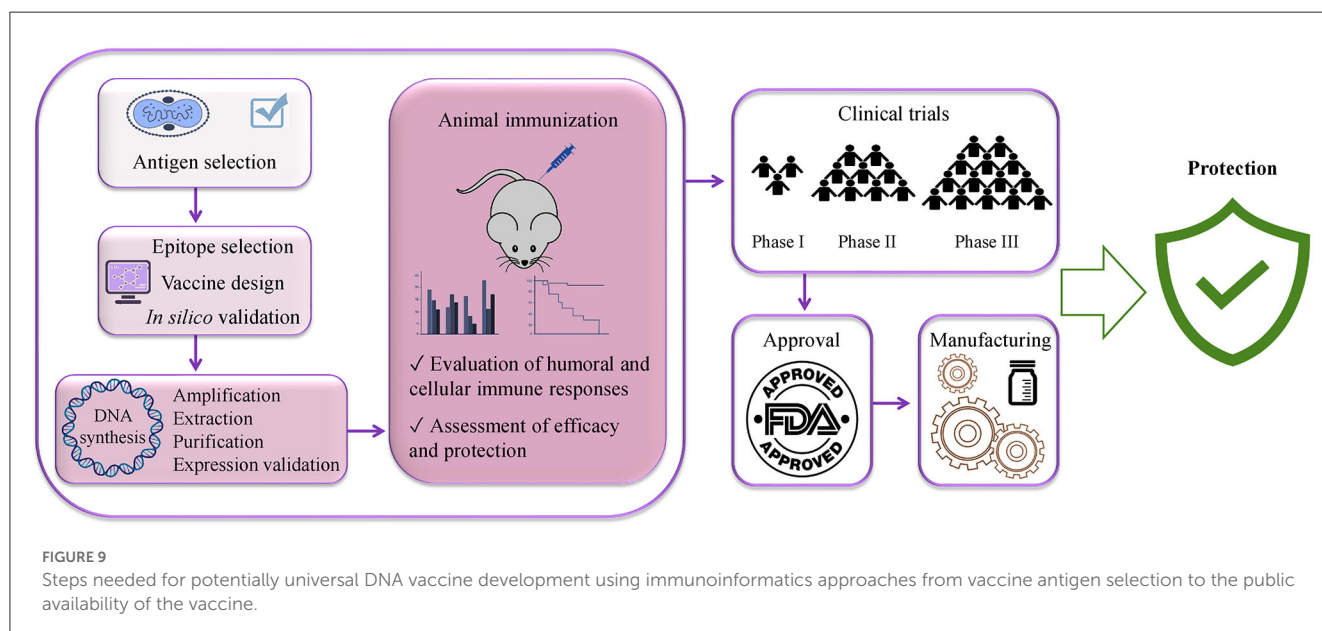
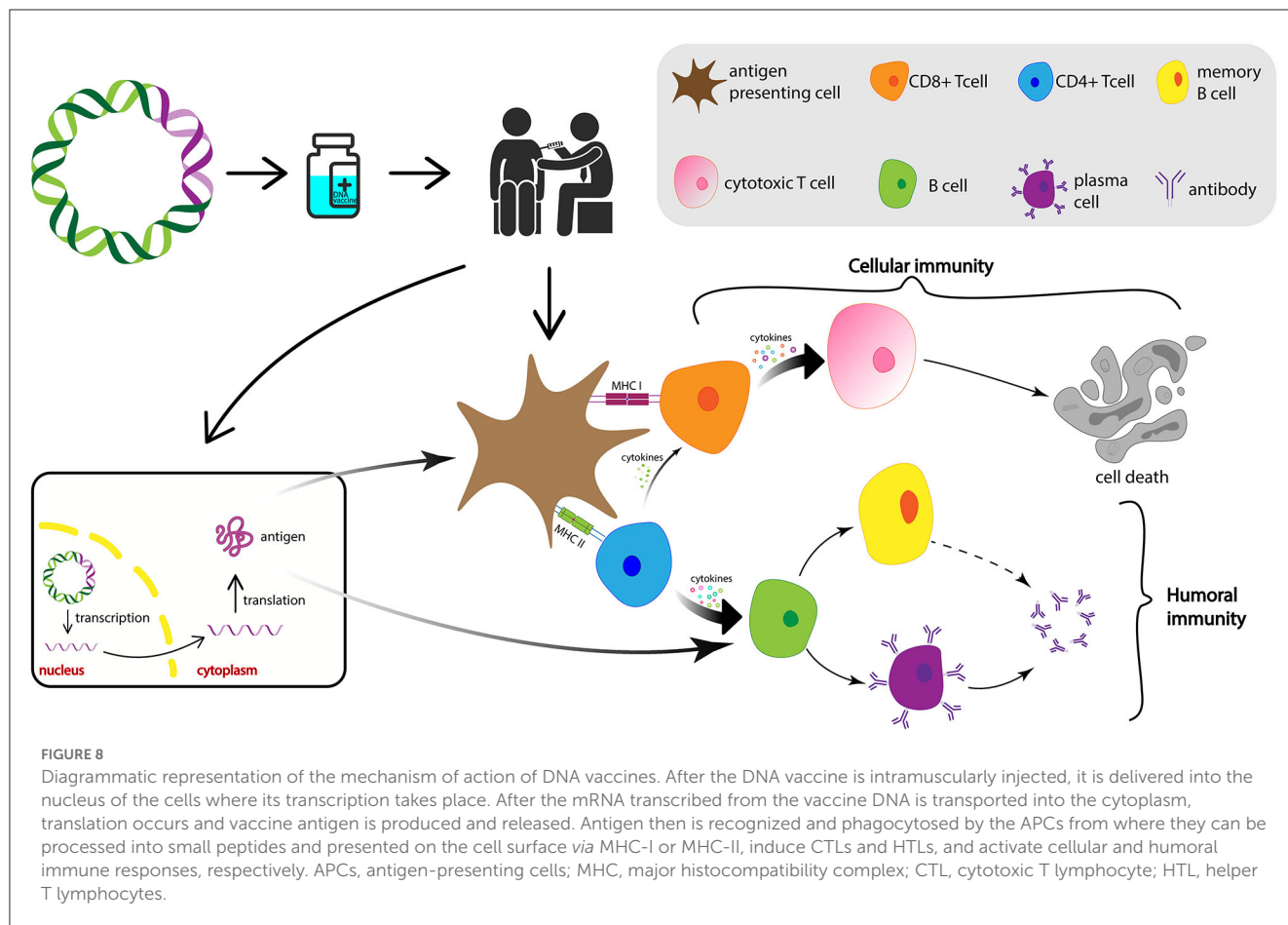
DNA is transported to the nucleus where transcription takes place. Subsequently, the mRNA is conveyed to the cytoplasm where it undergoes translation to produce the antigens for the vaccine. Figure 8 depicts the schematic representation of the mechanism of action for DNA vaccines upon delivery into the host cell.

Remarkably, although DNA vaccine enters the host nucleus, which is the main concern related to DNA vaccines, the likelihood of its integration into the host genome is extremely low (Williams, 2013). Indeed, there is a number of clinical trials on the application of DNA vaccines for the prevention of various infectious diseases: NCT04591184, NCT01498718, NCT01487876, and NCT04445389. Furthermore, India has designated the emergency use of the first DNA vaccine against COVID-19 in 2021 (Sheridan, 2021) and there are a number of other COVID-19 DNA vaccine candidates in clinical (Sheridan, 2021; Silveira et al., 2021) and preclinical development (Shafaati et al.,



2022; Wang et al., 2022). More importantly, DNA vaccines have shown protective efficacy against the mpox virus in non-human primates (Hooper et al., 2004; Hirao et al., 2011). The steps

of developing a DNA-based vaccine from the antigen selection to the commercial availability are vaccine design, optimization with proper linkers and addition of immunomodulator sequences,



synthesis, transformation into competent bacterial cells for its amplification, extraction, purification, animal immunization, assessment of immune responses, clinical trials, approval, and manufacturing (Papukashvili et al., 2022b; Rcheulishvili et al., 2022). Steps that are needed for developing a potentially

universal DNA-based anti-mpox multi-epitope vaccine are given in Figure 9.

Designing a vaccine employing immunoinformatics approaches significantly shortens the time of vaccine development and lays the groundwork for the rational design of the effective

vaccine, while in the case of conventional techniques, the development of a vaccine may take decades of laboratory work (Rappuoli and Aderem, 2011). Leveraging immunoinformatics enables us to predict the potential outcome of the vaccine. Upon achieving favorable results with the designed vaccine, it becomes more rational to proceed with *in vitro* and *in vivo* screening of the multi-epitope construct.

Leveraging immunoinformatics tools enables us to forecast the potential efficacy of the vaccine. Upon achieving favorable outcomes with the designed vaccine, it becomes more logical to proceed with *in vitro* and *in vivo* screening of the multi-epitope construct.

The links of the employed servers and tools are given in [Supplementary Table 2](#). Apparently, there are a number of *in silico* studies which designed the vaccine and computationally evaluated its efficacy, stability, etc., e.g., vaccines for COVID-19 (Dong et al., 2020; Oliveira et al., 2020), influenza (Behbahani et al., 2021; Sharma et al., 2021; Rcheulishvili et al., 2023), mpox (Hirao et al., 2011; Akhtar et al., 2022), and other viruses (Ali et al., 2019; Mahmudul et al., 2020; Ros-Lucas et al., 2020).

In this study, the vaccine construct designed here was found to have favorable physico-chemical properties and induce strong cellular and humoral immune responses. The immune simulation analysis shows that immunization with the multi-epitope vaccine candidate induces the production of HTLs and CTLs and stimulates the B-cell population, antibodies, and cytokines. The computationally designed vaccine in this study requires to be validated with *in vitro* and *in vivo* studies to confirm the outcomes obtained in this study.

5. Conclusion

In summary, the development of new, next-generation, specific vaccine candidates against the mpox virus is undoubtedly essential. The available immunoinformatics approaches allow for the rational design of the potentially protective vaccine and facilitate the process of vaccine development. The results obtained in this study demonstrate that the multi-epitope vaccine designed here may be suggested as an auspicious vaccine candidate which has the prospect of eliciting strong immune responses and providing protection against the mpox virus. In addition, the strategy of developing a universal multi-epitope DNA vaccine used in this study may have a positive impact on the development of a potentially universal vaccine against the mpox virus and other viruses and, thus, will aid in averting future outbreaks or pandemics.

References

- Ahammad, I., and Lira, S. S. (2020). Designing a novel mRNA vaccine against SARS-CoV-2: an immunoinformatics approach. *Int. J. Biol. Macromol.* 162, 820–837. doi: 10.1016/j.ijbiomac.2020.06.213
- Akhtar, N., Kaushik, V., Grewal, R. K., and Wani, A. K. (2022). Immunoinformatics-Aided Design of a Peptide Based Multiepitope Vaccine Targeting Glycoproteins and Membrane Proteins against Monkeypox Virus. *Viruses* 14, 2374. doi: 10.3390/v14112374
- Alexander, J., del Guercio, M.-F., Maewal, A., Qiao, L., Fikes, J., Chesnut, R. W., et al. (2000). Linear PADRE T helper epitope and carbohydrate B cell epitope conjugates induce specific high titer IgG antibody responses. *J. Immunol.* 164, 1625–1633. doi: 10.4049/jimmunol.164.3.1625
- Ali, A., Khan, A., Kaushik, A. C., Wang, Y., Ali, S. S., Junaid, M., et al. (2019). Immunoinformatic and systems biology approaches to predict and

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 authors.

Author contributions

NR conceived and designed the study and wrote the original draft. NR and JM contributed to the methodology and investigation. NR, DP, and SF contributed to the visualization. CL and XW revised the manuscript. YH supervised the study. PW supervised and administered the project. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Shenzhen Science and Technology Innovation Program (Grant No. KQTD20200909113758004).

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.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1203355/full#supplementary-material>

- validate peptide vaccines against Epstein-Barr virus (EBV). *Sci. Rep.* 9, 720. doi: 10.1038/s41598-018-37070-z
- Ali, M., Pandey, R. K., Khaton, N., Narula, A., Mishra, A., and Prajapati, V. K. (2017). Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci. Rep.* 7, 9232. doi: 10.1038/s41598-017-09199-w
- Baek, M., Dimaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a 3-track neural network. *Science* 373, 871–876. doi: 10.1126/science.abj8754
- Behbahani, M., Moradi, M., and Mohabatkari, H. (2021). In silico design of a multi-epitope peptide construct as a potential vaccine candidate for Influenza A based on neuraminidase protein. *Silico Pharmacol.* 9, 36. doi: 10.1007/s40203-021-00095-w
- Bellamkonda, N., Lambe, U. P., Sawant, S., Nandi, S. S., Chakraborty, C., and Shukla, D. (2022). Immune response to SARS-CoV-2 vaccines. *Biomedicine* 10, 1464. doi: 10.3390/biomedicine10071464
- Boyoglu-Barnum, S., Ellis, D., Gillespie, R. A., Hutchinson, G. B., Park, Y., Moin, S. M., et al. (2021). Quadrivalent influenza nanoparticle vaccines induce broad protection. *Nature* 592, 623–628. doi: 10.1038/s41586-021-03365-x
- Buchman, G. W., Cohen, M. E., Xiao, Y., Richardson-Harman, N., Silvera, P., DeTolla, L. J., et al. (2010). A protein-based smallpox vaccine protects non-human primates from a lethal monkeypox virus challenge. *Vaccine* 28, 6627–6636. doi: 10.1016/j.vaccine.2010.07.030
- Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., and Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 7, 153. doi: 10.1186/1471-2105-7-153
- Bunge, E. M., Hoet, B., Chen, L., Lienert, F., Weidenthaler, H., Baer, L. R., et al. (2022). The changing epidemiology of human monkeypox—a potential threat? A systematic review. *PLoS Negl. Trop. Dis.* 16, e0010141. doi: 10.1371/journal.pntd.0010141
- CDC (2022). JYNNEOS Vaccine. Available online at: <https://www.cdc.gov/poxvirus/mpox/interim-considerations/jynneos-vaccine.html> (accessed June 30, 2023).
- Chavda, V. P., Hossain, K., Beladiya, J., and Apostolopoulos, V. (2021a). Nucleic acid vaccines for COVID-19: a paradigm shift in the vaccine development arena. *Biologics* 1, 337–356. doi: 10.3390/biologics1030020
- Chavda, V. P., Pandya, R., and Apostolopoulos, V. (2021b). DNA vaccines for SARS-CoV-2: toward third-generation vaccination era. *Expert Rev. Vaccines* 20, 1549–1560. doi: 10.1080/14760584.2021.1987223
- Chen, X., Zaro, J. L., and Shen, W. C. (2013). Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.* 65, 1357–1369. doi: 10.1016/j.addr.2012.09.039
- Corbett, K. S., Flynn, B., Foulds, K. E., Francica, J. R., Boyoglu-Barnum, S., Werner, A. P., et al. (2020). Evaluation of the mRNA-1273 vaccine against SARS-CoV-2 in nonhuman primates. *N. Engl. J. Med.* 383, 1544–1555. doi: 10.1056/NEJMoa2024671
- Dong, R., Chu, Z., Yu, F., and Zha, Y. (2020). Contriving multi-epitope subunit of vaccine for COVID-19: immunoinformatics approaches. *Front. Immunol.* 11, 1784. doi: 10.3389/fimmu.2020.01784
- Doytchinova, I. A., and Flower, D. R. (2007a). Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* 25, 856–866. doi: 10.1016/j.vaccine.2006.09.032
- Doytchinova, I. A., and Flower, D. R. (2007b). Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* 8, 4. doi: 10.1186/1471-2105-8-4
- Flingai, S., Czerwonko, M., Goodman, J., Kudchodkar, S. B., Muthumani, K., and Weiner, D. B. (2013). Synthetic DNA vaccines: improved vaccine potency by electroporation and co-delivered genetic adjuvants. *Front. Immunol.* 4, 354. doi: 10.3389/fimmu.2013.00354
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). Protein identification and analysis tools on the ExPASy server; the proteomics protocols handbook. *Humana Press* 112, 531–552. doi: 10.1385/1-59259-890-0:571
- Ghafari, F., Cohan, R. A., Samimi, H., Hosseini, A., Naderi, M., Noorbakhsh, F., et al. (2022). Development of a multipitope vaccine against SARS-CoV-2: immunoinformatics study. *JMIR Bioinforma. Biotechnol.* 3, e36100. doi: 10.2196/36100
- Haas, E. J., Angulo, F. J., McLaughlin, J. M., Anis, E., Singer, S. R., Khan, F., et al. (2021). Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet* 397, 1819–1829. doi: 10.1016/S0140-6736(21)00947-8
- Heo, L., Park, H., and Seok, C. (2013). GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 41, 384–388. doi: 10.1093/nar/gkt458
- Heraud, J., Edghill-smith, Y., Ayala, V., Kalisz, I., Parrino, J., Kalyanaraman, V. S., et al. (2022). Subunit recombinant vaccine protects against monkeypox. *J. Immunol.* 177, 2552–2564. doi: 10.4049/jimmunol.177.4.2552
- Heskin, J., Belfield, A., Milne, C., Brown, N., Walters, Y., Scott, C., et al. (2022). Transmission of monkeypox virus through sexual contact – a novel route of infection. *J. Infect.* 85, 334–363. doi: 10.1016/j.jinf.2022.05.028
- Hirao, L. A., Draghia-akli, R., Prigge, J. T., Yang, M., Satishchandran, A., Wu, L., et al. (2011). Multivalent smallpox DNA vaccine delivered by intradermal electroporation drives protective immunity in nonhuman primates against lethal monkeypox challenge. *J. Infect. Dis.* 203, 95–102. doi: 10.1093/infdis/jiq017
- Hobernik, D., and Bros, M. (2018). DNA vaccines—how far from clinical use? *Int. J. Mol. Sci.* 19, 3605. doi: 10.3390/ijms19113605
- Hooper, J. W., Thompson, E., Wilhelmsen, C., Zimmerman, M., Ichou, M. A., Steffen, S. E., et al. (2004). Smallpox DNA vaccine protects nonhuman primates against lethal monkeypox. *J. Virol.* 78, 4433–4443. doi: 10.1128/JVI.78.9.4433-4443.2004
- Hoover, D. M., Wu, Z., Tucker, K., Lu, W., and Lubkowski, J. (2003). Antimicrobial characterization of human β -defensin 3 derivatives. *Antimicrob. Agents Chemother.* 47, 2804–2809. doi: 10.1128/AAC.47.9.2804-2809.2003
- Javier, F., Sánchez, M., Martínez-sellés, M., Ruiz-galiana, J., and Ramos, P. D. L. (2022). Monkeypox in humans: a new outbreak. *Off. J. Spanish Soc. Chemother.* 6, 2022.
- Jespersen, M. C., Peters, B., Nielsen, M., and Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* 45, W24–W29. doi: 10.1093/nar/gkx346
- Jezek, Z., Szczeniowski, M., and Paluku, K. M. (1987). Human monkeypox: clinical features of 282 patients. *J. Infect. Dis.* 156, 293–298. doi: 10.1093/infdis/156.2.293
- 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
- Kandeel, M., Morsy, M. A., El-lateef, H. M. A., Marzok, M., El-Beltagi, H. S., Khodair, K. M., et al. (2023). Efficacy of the modified vaccinia Ankara virus vaccine and the replication-competent vaccine ACAM2000 in monkeypox prevention Mahmoud. *Int. Immunopharmacol.* 119, 110206. doi: 10.1016/j.intimp.2023.110206
- Khan, M. T., Islam, R., Jerin, T. J., Mahmud, A., Khatun, S., Kobir, A., et al. (2021). Immunoinformatics and molecular dynamics approaches: next generation vaccine design against West Nile virus. *PLoS ONE* 16, e0253393. doi: 10.1371/journal.pone.0253393
- Kmieć, D., and Kirchhoff, F. (2022). Monkeypox: a new threat? *Int. J. Mol. Sci.* 23, 7866. doi: 10.3390/ijms23147866
- Kowalik, F., Schreiner, D., Jensen, C., Teschner, D., Gehring, S., and Zepp, F. (2021). mRNA-based vaccines. *Vaccines* 9, 390. doi: 10.3390/vaccines9040390
- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., et al. (2017). The ClusPro web server for protein-protein docking. *Nat. Protoc.* 12, 255–278. doi: 10.1038/nprot.2016.169
- Kozlov, M. (2022). Monkeypox outbreaks: 4 key questions researchers have. *Nature* 606, 238–239. doi: 10.1038/d41586-022-01493-6
- Kugelman, J. R., Johnston, S. C., Mulembakani, P. M., Kitalu, N., Lee, M. S., Koroleva, G., et al. (2014). Genomic variability of monkeypox virus among humans, democratic Republic of the Congo. *Emerg. Infect. Dis.* 20, 232–239. doi: 10.3201/eid2002.130118
- Liu, C., Papukashvili, D., Dong, Y., Wang, X., Hu, X., Yang, N., et al. (2022a). Identification of tumor antigens and design of mRNA vaccine for colorectal cancer based on the immune subtype. *Front. Cell Dev. Biol.* 9, 783527. doi: 10.3389/fcell.2021.783527
- Liu, C., Rcheulishvili, N., Shen, Z., Papukashvili, D., Xie, F., Wang, Z., et al. (2022b). Development of an LNP-Encapsulated mRNA-RBD vaccine against SARS-CoV-2 and its variants. *Pharmaceutics* 14, 1101. doi: 10.3390/pharmaceutics14051101
- Locker, J. K., Kuehn, A., Schleich, S., Rutter, G., Hohenberg, H., Wepf, R., et al. (2000). Entry of the two infectious forms of vaccinia virus at the plasma membrane is signaling-dependent for the IMV but not the EEV. *Mol. Biol. Cell* 11, 2497–2511. doi: 10.1091/mbc.11.7.2497
- Lum, F.-M., Ruesta, A. T., Tay, M. Z., Lin, R. T. P., Lye, D. C., Réna, L., et al. (2022). Monkeypox: disease epidemiology, host immunity and clinical interventions. *Nat. Rev. Immunol.* 22, 597–613. doi: 10.1038/s41577-022-00775-4
- Maa, J. S., and Esteban, M. (1987). Structural and functional studies of a 39,000-Mr immunodominant protein of vaccinia virus. *J. Virol.* 61, 3910–3919. doi: 10.1128/jvi.61.12.3910-3919.1987
- Mahmud, H., Shiful, I., Sourav, C., Abu, H. M., Kazi, F. A., Ziaul, F. J., et al. (2020). Contriving a chimeric polyvalent vaccine to prevent infections caused by herpes simplex virus (Type-1 and Type-2): an exploratory immunoinformatic approach. *J. Biomol. Struct. Dyn.* 38, 2898–2915. doi: 10.1080/07391102.2019.1647286
- Martinelli, D. D. (2022). In silico vaccine design: a tutorial in immunoinformatics. *Healthc. Anal.* 2, 100044. doi: 10.1016/j.health.2022.100044
- Martínez, J. I., Montalbán, E. G., Bueno, S. J., Martínez, F. M., Juliá, A. N., Díaz, S., et al. (2022). Monkeypox outbreak predominantly affecting men who have sex with men, Madrid, Spain, 26 April to 16 June 2022. *Eurosurveillance* 27, 2200471. doi: 10.2807/1560-7917.ES.2022.27.22.2200471

- Maurer, D. M., Harrington, B., Army, D., Hospital, C., Hood, F., and Lane, J. M. (2003). Smallpox vaccine: contraindications, administration, and adverse reactions. *Am. Fam. Physician* 68, 889–896.
- Maurer-Stroh, S., Krutz, N. L., Kern, P. S., Gunalan, V., Nguyen, M. N., Limviphuvadh, V., et al. (2019). AllerCatPro-prediction of protein allergenicity potential from the protein sequence. *Bioinformatics* 35, 3020–3027. doi: 10.1093/bioinformatics/btz029
- Mekonnen, D., Mengist, H. M., and Jin, T. (2022). SARS-CoV-2 subunit vaccine adjuvants and their signaling pathways. *Expert Rev. Vaccines* 21, 69–81. doi: 10.1080/14760584.2021.1991794
- Meng, X., Kaever, T., Yan, B., Traktman, P., Zajonc, D. M., Peters, B., et al. (2018). Characterization of murine antibody responses to vaccinia virus envelope protein A14 reveals an immunodominant antigen lacking of effective neutralization targets. *Virology* 518, 284–292. doi: 10.1016/j.virol.2018.03.005
- Meyer, H., Perrichot, M., Stemmler, M., Emmerich, P., Schmitz, H., Varaine, F., et al. (2002). Outbreaks of disease suspected of being due to human monkeypox virus infection in the Democratic Republic of Congo in 2001. *J. Clin. Microbiol.* 40, 2919–2921. doi: 10.1128/JCM.40.8.2919-2921.2002
- Mileto, D., Riva, A., Cutrera, M., Moschese, D., Mancon, A., Meroni, L., et al. (2022). New challenges in human monkeypox outside Africa: a review and case report from Italy. *Travel Med. Infect. Dis.* 49, 102386. doi: 10.1016/j.tmaid.2022.102386
- Mirdita, M., Schütze, K., Moriawaki, Y., and Heo, L. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi: 10.1038/s41592-022-01488-1
- Moschese, D., Giacomelli, A., Beltrami, M., Pozza, G., Mileto, D., Reato, S., et al. (2022). Hospitalisation for monkeypox in Milan, Italy. *Travel Med. Infect. Dis.* 49, 102417. doi: 10.1016/j.tmaid.2022.102417
- Nelson, D. L., Cox, M. M., and Hoskins, A. A. (2021). *Lehninger Principles of Biochemistry*. 8th ed. Austin: Macmillan Learning.
- Nezafat, N., Ghasemi, Y., Javadi, G., Khoshnoud, M. J., and Omidinia, E. (2014). A novel multi-epitope peptide vaccine against cancer: an in silico approach. *J. Theor. Biol.* 349, 121–134. doi: 10.1016/j.jtbi.2014.01.018
- Nguyen, M. N., Krutz, N. L., Limviphuvadh, V., Lopata, A. L., Gerberick, G. F., and Maurer-Stroh, S. (2022). AllerCatPro 2.0: a web server for predicting protein allergenicity potential. *Nucleic Acids Res.* 50, W36–W43. doi: 10.1093/nar/gkac446
- Oliveira, S. C., Magalhães, M. T. Q., De, and Homan, E. J. (2020). Immunoinformatic analysis of SARS-CoV-2 nucleocapsid protein and identification of COVID-19 vaccine targets. *Front. Immunol.* 11, 587615. doi: 10.3389/fimmu.2020.587615
- Orviz, E., Negro, A., Ayerdi, O., Vázquez, A., Muñoz-Gomez, A., Monzón, S., et al. (2022). Monkeypox outbreak in Madrid (Spain): clinical and virological aspects. *J. Infect.* 85, 412–417. doi: 10.1016/j.jinf.2022.07.005
- Papukashvili, D., Rcheulishvili, N., Liu, C., Ji, Y., He, Y., and Wang, P. G. (2022a). Self-amplifying RNA approach for protein replacement therapy. *Int. J. Mol. Sci.* 23, 12884. doi: 10.3390/ijms232112884
- Papukashvili, D., Rcheulishvili, N., Liu, C., Wang, X., He, Y., and Wang, P. G. (2022b). Strategy of developing nucleic acid-based universal monkeypox vaccine candidates. *Front. Immunol.* 13, 1050309. doi: 10.3389/fimmu.2022.1050309
- Pardi, N., Hogan, M. J., Porter, F. W., and Weissman, D. (2018). mRNA vaccines—a new era in vaccinology. *Nat. Rev. Drug Discov.* 17, 261–279. doi: 10.1038/nrd.2017.243
- Ponomarenko, J., Bui, H. H., Li, W., Füsseder, N., Bourne, P. E., Sette, A., et al. (2008). ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* 9, 514. doi: 10.1186/1471-2105-9-514
- Qamar, M. T. U., Shokat, Z., Muneer, I., Ashfaq, U. A., Javed, H., Anwar, F., et al. (2020). Multiepitope-based subunit vaccine design and evaluation against respiratory syncytial virus using reverse vaccinology approach. *Vaccines* 8, 288. doi: 10.3390/vaccines8020288
- Qin, F., Xia, F., Chen, H., Cui, B., Feng, Y., Zhang, P., et al. (2021). A guide to nucleic acid vaccines in the prevention and treatment of infectious diseases and cancers: from basic principles to current applications. *Front. Cell Dev. Biol.* 9, 633776. doi: 10.3389/fcell.2021.633776
- Rahmani, A., Bae, M., Saleki, K., Moradi, S., and Nouri, H. R. (2021). Applying high throughput and comprehensive immunoinformatics approaches to design a trivalent subunit vaccine for induction of immune response against emerging human coronaviruses SARS-CoV, MERS-CoV and SARS-CoV-2. *J. Biomol. Struct. Dyn.* 40, 6097–6113. doi: 10.21203/rs.3.rs-92515/v1
- Rapin, N., Lund, O., Bernaschi, M., and Castiglione, F. (2010). Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE* 5, e9862. doi: 10.1371/journal.pone.0009862
- Rapin, N., Lund, O., and Castiglione, F. (2011). Immune system simulation online. *Bioinformatics* 27, 2013–2014. doi: 10.1093/bioinformatics/btr335
- Rappuoli, R., and Aderem, A. (2011). A 2020 vision for vaccines against HIV, tuberculosis and malaria. *Nature* 473, 463–469. doi: 10.1038/nature10124
- Rcheulishvili, N., Mao, J., Papukashvili, D., Liu, C., Wang, Z., Zhao, J., et al. (2023). Designing multi-epitope mRNA construct as a universal influenza vaccine candidate for future epidemic/pandemic preparedness. *Int. J. Biol. Macromol.* 226, 885–899. doi: 10.1016/j.ijbiomac.2022.12.066
- Rcheulishvili, N., Papukashvili, D., Liu, C., Ji, Y., He, Y., and Wang, P. G. (2022). Promising strategy for developing mRNA-based universal influenza virus vaccine for human population, poultry, and pigs—focus on the bigger picture. *Front. Immunol.* 13, 1025884. doi: 10.3389/fimmu.2022.1025884
- Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 48, W449–W454. doi: 10.1093/nar/gkaa379
- Ros-Lucas, A., Correa-Fiz, F., Bosch-Camós, L., Rodríguez, F., and Alonso-Padilla, J. (2020). Computational analysis of african swine fever virus protein space for the design of an epitope-based vaccine ensemble. *Pathogens* 9, 1078.
- Rybakova, Y., Kowalski, P. S., Huang, Y., Gonzalez, J. T., Heartlein, M. W., Derosa, F., et al. (2019). mRNA delivery for therapeutic Anti-HER2 antibody expression in vivo. *Mol. Ther.* 27, 1415–1423. doi: 10.1016/j.ymthe.2019.05.012
- Sah, R., Abdelaal, A., Reda, A., Katamesh, B. E., and Manirambona, E. (2022). Monkeypox and its possible sexual transmission: where are we now with its evidence? *Pathogens* 11, 924. doi: 10.3390/pathogens11080924
- Seang, S., Burrell, S., Todesco, E., Leducq, V., Monsel, G., Pluart, D., et al. (2022). Evidence of human-to-dog transmission of monkeypox virus. *Lancet* 400, 658–659. doi: 10.1016/S0140-6736(22)01487-8
- Shafaati, M., Mirzaei, R., Saidijam, M., Amirheidari, B., Soleimani, M., Mahaki, H., et al. (2022). A brief review on DNA vaccines in the era of COVID-19. *Future Virol.* 17, 49–66. doi: 10.2217/fvl-2021-0170
- Shafaati, M., and Zandi, M. (2022). State-of-the-art on monkeypox virus: an emerging zoonotic disease. *Infection* 50, 1425–1430. doi: 10.1007/s15010-022-01935-3
- Shafaati, M., and Zandi, M. (2023). Human monkeypox (hMPXV) re-emergence: host immunity status and current vaccines landscape. *J. Med. Virol.* 95, e28251. doi: 10.1002/jmv.28251
- Sharma, S., Kumari, V., Vasant, B., Mukherjee, A., Pandey, R., and Kondabagil, K. (2021). Immunoinformatics approach for a novel multi-epitope subunit vaccine design against various subtypes of Influenza A virus. *Immunobiology* 226, 152053. doi: 10.1016/j.imbio.2021.152053
- Shchelkunov, S. N., Totmenin, A. V., Safronov, P. F., Mikheev, M. V., Gutorov, V. V., Ryazankina, O. I., et al. (2002). Analysis of the monkeypox virus genome. *Virology* 197, 172–194. doi: 10.1006/viro.2002.1446
- Sheridan, C. (2021). First COVID-19 DNA vaccine approved, others in hot pursuit. *Nat. Biotechnol.* 39, 1479–1482. doi: 10.1038/d41587-021-00023-5
- Silveira, M. M., Moreira, G. M. S. G., and Mendonça, M. (2021). DNA vaccines against COVID-19: Perspectives and challenges. *Life Sci.* 267, 118919. doi: 10.1016/j.lfs.2020.118919
- Skenovská, N., and Van Ranst, M. (2018). Emergence of monkeypox as the most important orthopoxvirus infection in humans. *Front. Public Heal.* 6, 241. doi: 10.3389/fpubh.2018.00241
- Tarrahimofrad, H., Rahimnahl, S., Zamani, J., Jahangirian, E., and Aminzadeh, S. (2021). Designing a multi-epitope vaccine to provoke the robust immune response against influenza A H7N9. *Sci. Rep.* 11, 24485. doi: 10.1038/s41598-021-03932-2
- Vandenbogaert, M., Kwasiborski, A., Gonofio, E., Descorps, S., Selekon, B., Andy, A., et al. (2022). Nanopore sequencing of a monkeypox virus strain isolated from a pustular lesion in the Central African Republic. *Sci. Rep.* 12, 10768. doi: 10.1038/s41598-022-15073-1
- Vavilis, T., Stamoula, E., Aintzoglou, A., Sachinidis, A., Lamprinou, M., Dardalas, I., et al. (2023). mRNA in the context of protein replacement therapy. *Pharmaceutics* 15, 166. doi: 10.3390/pharmaceutics15010166
- Walsh, E. E., Frenck, R. W., Falsey, A. R., Kitchin, N., Absalon, J., Gurtman, A., et al. (2020). Safety and immunogenicity of Two RNA-based Covid-19 vaccine candidates. *N. Engl. J. Med.* 383, 2439–2450. doi: 10.1056/NEJMoa2027906
- Wang, C., and Zou, Q. (2023). Prediction of protein solubility based on sequence physicochemical patterns and distributed representation information with DeepSoluE. *BMC Biol.* 21, 12. doi: 10.1186/s12915-023-01510-8
- Wang, X., Liu, C., Rcheulishvili, N., Papukashvili, D., Xie, F., Zhao, J., et al. (2023). Strong immune responses and protection of PcrV and OprF-I mRNA vaccine candidates against *Pseudomonas aeruginosa*. *npj Vaccines* 8, 76. doi: 10.1038/s41541-023-00672-4
- Wang, X., Rcheulishvili, N., Cai, J., Liu, C., Xie, F., Hu, X., et al. (2022). Development of DNA vaccine candidate against SARS-CoV-2. *Viruses* 14, 1049. doi: 10.3390/v14051049
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Williams, J. (2013). Vector design for improved DNA vaccine efficacy, safety and production. *Vaccines* 1, 225–249. doi: 10.3390/vaccines1030225

- World Health Organization (2022a). *Monkeypox*. Available online at: <https://www.who.int/news-room/fact-sheets/detail/monkeypox> (accessed September 16, 2022).
- World Health Organization (2022b). *WHO Director-General Declares the Ongoing Monkeypox Outbreak a Public Health Emergency of International Concern*. Geneva: WHO.
- Xia, D., Jin, R., Byagathalli, G., Yu, H., Ye, L., Lu, C. Y., et al. (2021). An ultra-low-cost electroporator with microneedle electrodes (ePatch) for SARS-CoV-2 vaccination. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2110817118. doi: 10.1073/pnas.2110817118
- Yang, Z. (2022). Monkeypox: a potential global threat? *J. Med. Virol.* 94, 4034–4036. doi: 10.1002/jmv.27884
- Yano, A., Onozuka, A., Asahi-Ozaki, Y., Imai, S., Hanada, N., Miwa, Y., et al. (2005). An ingenious design for peptide vaccines. *Vaccine* 23, 2322–2326. doi: 10.1016/j.vaccine.2005.01.031
- Zaeck, L. M., Lamers, M. M., Verstrepen, B. E., Bestebroer, T. M., Royen, M. E., van, Götz, H., et al. (2022). Low levels of monkeypox virus neutralizing antibodies after MVA-BN vaccination in healthy individuals. *Nat. Med.* 29, 270–278. doi: 10.1038/s41591-022-02090-w
- Zandi, M., Shafaati, M., and Hosseini, F. (2023). Mechanisms of immune evasion of monkeypox virus. *Front. Microbiol.* 14, 1106247. doi: 10.3389/fmicb.2023.1106247
- Zhang, J., Liang, Y., and Zhang, Y. (2011). Atomic-level protein structure refinement using fragment guided molecular dynamics conformation sampling. *Structure* 19, 1784–1795. doi: 10.1016/j.str.2011.09.022
- Zhang, R. R., Wang, Z. J., Zhu, Y. L., Tang, W., Zhou, C., Zhao, S. Q., et al. (2023). Rational development of multicomponent mRNA vaccine candidates against mpox. *Emerg. Microbes Infect.* 12, 2192815. doi: 10.1080/22221751.2023.2192815
- Zhu, M., Ji, J., Shi, D., Lu, X., Wang, B., Wu, N., et al. (2022). Unusual global outbreak of monkeypox: what should we do? *Front. Med.* 16, 507–517. doi: 10.1007/s11684-022-0952-z



OPEN ACCESS

EDITED BY

Sinosh Skariyachan,
St. Pius X College, India

REVIEWED BY

Milad Zandi,
Tehran University of Medical Sciences, Iran
Md Tajmul,
National Institute of Diabetes and Digestive and
Kidney Diseases (NIH), United States
Fnu Wahiduzzaman,
St. Jude Children's Research Hospital,
United States

*CORRESPONDENCE

Yao Sun
✉ sunyao0819@hit.edu.cn

[†]These authors have contributed equally to this work

RECEIVED 05 June 2023

ACCEPTED 16 August 2023

PUBLISHED 20 September 2023

CITATION

Jiao Y, Shi C and Sun Y (2023) The use of
Xuanbai Chengqi decoction on monkeypox
disease through the estrone-target AR
interaction.
Front. Microbiol. 14:1234817.
doi: 10.3389/fmicb.2023.1234817

COPYRIGHT

© 2023 Jiao, Shi and Sun. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The use of Xuanbai Chengqi decoction on monkeypox disease through the estrone-target AR interaction

Yanqi Jiao^{1†}, Chengcheng Shi^{2†} and Yao Sun^{1*}

¹School of Science, Harbin Institute of Technology (Shenzhen), Shenzhen, China, ²School of Science/State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology (Shenzhen), Shenzhen, China

Introduction: After COVID-19, there was an outbreak of a new infectious disease caused by monkeypox virus. So far, no specific drug has been found to treat it. Xuanbai Chengqi decoction (XBCQD) has shown effects against a variety of viruses in China.

Methods: We searched for the active compounds and potential targets for XBCQD from multiple open databases and literature. Monkeypox related targets were searched out from the OMIM and GeneCards databases. After determining the assumed targets of XBCQD for monkeypox treatment, we built the PPI network and used R for GO enrichment and KEGG pathway analysis. The interactions between the active compounds and the hub targets were investigated by molecular docking and molecular dynamics (MD) simulations.

Results: In total, 5 active compounds and 10 hub targets of XBCQD were screened out. GO enrichment and KEGG analysis demonstrated that XBCQD plays a therapeutic role in monkeypox mainly by regulating signaling pathways related to viral infection and inflammatory response. The main active compound estrone binding to target AR was confirmed to be the best therapy choice for monkeypox.

Discussion: This study systematically explored the interactions between the bioactive compounds of XBCQD and the monkeypox-specific XBCQD targets using network pharmacological methods, bioinformatics analyses and molecular simulations, suggesting that XBCQD could have a beneficial therapeutic effect on monkeypox by reducing the inflammatory damage and viral replication via multiple pathways. The use of XBCQD on monkeypox disease was confirmed to be best worked through the estrone-target AR interaction. Our work could provide evidence and guidance for further research on the treatment of monkeypox disease.

KEYWORDS

Xuanbai Chengqi decoction, monkeypox, network pharmacology, molecular docking, molecular dynamics simulation

1. Background

Monkeypox, a virus that used to be common around rainforests in Central and West Africa, has spread out globally (Zardi and Chello, 2022). In June 2022, the World Health Organization (WHO) reported that more than 550 confirmed cases of monkeypox were identified in 30 countries and territories worldwide (Hasan and Saeed, 2022). In July 2022, WHO sounded the

alarm again, declaring the monkeypox outbreak to be a Public Health Emergency of International Concern (PHEIC) (Shafaati and Zandi, 2023). It was the seventh PHEIC the WHO had declared by far, and was the highest alert level the agency could give (Wilder-Smith and Osman, 2020; Shafaati et al., 2022).

Monkeypox disease is a rare zoonotic disease caused by the monkeypox virus (Dou et al., 2023). It is endemic in central and west Africa, with the greatest concentration in the Democratic Republic of the Congo. Although it was firstly found in captive monkeys, available data suggested that the African rodents were natural hosts. For instance, squirrels, rats, mice, monkeys, marmots and humans were all infected (Chen et al., 2022). There were two genetically distinct branches identified (Hutson et al., 2010), and the Congo Basin (central Africa) branch was reported more frequently than the west African branch. The current outbreak of Monkeypox in 2022 involved multiple countries on different continents, mainly in men who had sex with men (MSM), and its manifestations were related to genital lesions (Abu-Hammad et al., 2023). Up to 22 June 2022, 99% of the 508 confirmed cases of monkeypox in Madrid region of Spain belonged to the MSM population, and the lesions affected the genital, perineum, or perianal area (Iñigo Martínez et al., 2022). It was also identified swollen inguinal lymph nodes as the main features, indicating that sexual transmission was the primary mode of transmission (Letafati and Sakhavarz, 2023). On 6 July 2022, Germany reported 1,304 confirmed cases, which were also concentrated in the MSM population (Patel and Patel, 2023). Sequencing data from countries indicated that the 2022 outbreak was caused by the west African branch of the monkeypox virus. However, the up-to-date research suggested that there were two distinct lineages of the monkeypox virus with separate sources found in the US.

The monkeypox virus is a double-stranded DNA virus that is related to the variola virus (Kaler et al., 2022; Shafaati and Zandi, 2022). The clinical manifestations of human monkeypox are similar to those of smallpox, which often cause rash, fever, chills, and muscle soreness (Reynolds et al., 2019). The fatality rate of monkeypox is about 3%~6%, but relatively higher in children, young adults, and immunodeficient individuals (Ligon, 2004). When monkeypox is complicated with septicemia, meningitis, osteomyelitis, and other diseases, the death rate could be as high as 10% (Patel et al., 2022; Rizk et al., 2022). Up to now, some antiviral drugs and vaccines initially developed in smallpox have been approved for the treatment and prevention of monkeypox (Hung et al., 2022). But the effects of the treatment and prevention are still being investigated (Shamim et al., 2023). In fact, no specific drugs against monkeypox virus have been developed by far (Zovi et al., 2022). In view of the immunomodulatory and antiviral effects of traditional Chinese medicine (TCM) and its long history of clinical applications, this study aims to explore the potential treatment of monkeypox with TCM.

Xuanbai Chengqi decoction (XBCQD) is a TCM consisting of four Chinese herbs, including mineral-based gypsum, herbal rhubarb, bitter almond, and trichosanthes (Hanzlicek et al., 2014). It was reported to improve the disease symptoms effectively and prognosis of acute lung injury (ALI) patients with fewer adverse reactions by protecting lung function, alleviating excessive inflammatory reactions and tissue damage (Zhu et al., 2021). In addition, XBCQD was found to be an alternative treatment for severe infectious lung diseases caused by influenza and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Huang et al., 2021; Zhu et al., 2021).

In fact, XBCQD is a representative Chinese medicine prescription in the Differentiation of Febrile Disease written by Wu Jutong in Qing dynasty. It has been widely used in China for the treatment of lung injury, pulmonary fibrosis, chronic obstructive pulmonary and other common respiratory diseases. It can effectively reduce phlegm, heat, cough, wheezing, chest tightness, and has less adverse reactions (Wang et al., 2021; Huo et al., 2022). With the spread of monkeypox worldwide, the China National Health Commission and the National Administration of TCM issued the guidelines for the diagnosis and treatment of monkeypox in June 2022, in which XBCQD and other TCM were recommended to treat monkeypox patients with different symptoms (Desai et al., 2022; Warner et al., 2022).

Network pharmacology is a powerful tool to uncover the effective compounds in TCM from a systematical molecular way, integrating multiple open databases and bioinformatics techniques to construct a comprehensive drug-target-disease network (Zhang et al., 2013). Such a multi-component, multi-target network could reveal the mechanism behind the action of the drug (Noor et al., 2022). Molecular docking offers a possible way to reveal the *in vivo* binding patterns of ligand-receptor (Salmaso and Moro, 2018), which could be used to further explore the ligand-receptor relationship (Meng et al., 2011). Molecular dynamics (MD) simulation allows for the study of various ligand-receptor motions based on Newtonian mechanics to assess their long-time stabilities and flexibilities (Hollingsworth and Dror, 2018).

In this work, we first utilized network pharmacology to screen active drug compounds for the monkeypox virus and explored the potential biological mechanism behind the treatment of monkeypox by XBCQD from a systematic and molecular perspective. In total, 5 active compounds and 10 hub targets of XBCQD were screened out. Subsequently, the biological functional network of XBCQD was constructed to elucidate the regulatory way of XBCQD. The results demonstrated that XBCQD played a therapeutic role in monkeypox mainly by regulating signaling pathways related to viral infection and inflammatory response. Molecular docking was used to predict the binding energy scores and patterns between the hub targets and the potential therapeutic compounds. Finally, MD simulation was adopted to simulate the interaction dynamics and calculate the change of binding free energy of the target-compound complex, so as to provide theoretical foundation for the future clinical applications. The main active compound estrone binding to target AR was finally confirmed to be the best therapy choice for monkeypox.

2. Materials and methods

2.1. Active compounds in XBCQD and the corresponding targets

The TCMSP platform¹ was adopted to screen out the active compounds in XBCQD under the following standard criteria: the oral availability (OB) $\geq 30\%$ and drug-like likeness (DL) ≥ 0.18 , followed by a target search for active compounds by their MOLIDs. The standard criteria is the general standard for screening Chinese medicine according to the relevant literature (Ru et al., 2014). The

¹ <https://tcmsp-e.com/>

active compounds of *Trichosanthes pericarpium* were collected from the SYMmap database.² The PubChem database³ was used to derive the Simplified Molecular Input Line Entry System (SMILES) of the active compounds. The SMILES were then input into the SwissTargetPrediction structural similarity forecast target database⁴ to predict the valid targets. Unmatched names of targets were supplemented by literature review (Richards et al., 2015; Wang et al., 2021). Finally, the Uniprot database⁵ was used to annotate the relevant targets.

2.2. The compound-target network

The Cytoscape 3.9.2 software was adopted to prepare the compound-target network file and type file, as well as to conduct the network topology analysis. To construct the compound-target network map, the target map, color, transparency, and size were adjusted according to the connectivity (degree) of the targets.

2.3. The targets of monkeypox disease

The targets of monkeypox disease were searched by using OMIM⁶ and GeneCards⁷ databases. Through the search in GeneCards database, “monkeypox” and “monkeypox virus” were set as keywords to get monkeypox related targets. Monkeypox related targets were also obtained by searching “monkeypox” as a keyword in the OMIM database (see footnote 6). The targets symbols information corresponding to monkeypox disease was downloaded. Targets and functions were set to “human” and “VLOOKUP” to match target gene names.

2.4. The cross targets

Using the Venny online database,⁸ a reflection of intersections of drug target genes and disease genes, that is, the crossover between potential targets for XBCQD and monkeypox related targets could be identified. These identified crossover targets were considered as potential anti-monkeypox hub targets.

2.5. PPI network and cluster analysis

The String⁹ platform was used to construct the PPI network. Potential anti-monkeypox key targets (identified in the “The cross targets” section) were evaluated using the String platform with PPI highest confidence and species limited to 0.900 and homo sapiens.

The string PPI analysis results were then uploaded to Cytoscape 3.9.2 software to identify potential hub anti-monkeypox targets. In addition, the Simple Text Data Format (.tsv) files of PPI results were imported into Cytoscape 3.9.2 software to visualize the PPI network. As the number of nodes in the PPI network decreases, its color changes from red to yellow. Nodes that met the requirement of degree centrality were retrieved and identified as the hub targets of monkeypox. In the PPI network, the degree of the node represents the number of edges between it and other nodes. And the number of connections with other nodes indicates the significance of the node (Yu et al., 2022).

2.6. Go and KEGG analyses

GO function and KEGG path enrichment were analyzed using Bioconductor¹⁰ platform in R language. GO analysis of drug therapy gene functions was annotated in terms of biological process (BP), cellular component (CC), and molecular function (MF). KEGG is mainly a pathway analysis, aiming to elucidate the major signaling pathways for drug therapy. After installation of the R package, we introduced the disease target genes with specific parameters. The “Selection identifier” and “List type” were set to “official gene symbol” and “gene list.” The species were defined as “Homo,” “background” and “*Homo sapiens*” in “List.” The target, the minimum overlap, the *p*-value and the minimum concentration were set to “human,” 3, 0.05 and 1.5, respectively. Then we ran the gene ontology function and KEGG pathway on the target genes. We screened and preserved the results of the most important BP, CC, MF, and KEGG pathways. Finally, the corresponding bubble diagram was derived.

2.7. Molecular docking

To reveal the binding patterns between the active compounds and the targets, AutoDock Vina software, Discovery Studio 4.5 Client and PyMOL software were used. The 3D structures of the central targets (receptors) were obtained from the RCSB PDB¹¹ database and saved into PDB format. The obtained 3D structures (in PDB format) were further processed using PyMOL software (version 2.2.0) to remove water molecules (“solvent removal” command) and small ligands (“organics removal” command). The active compounds in 2D structures in SDF format were downloaded from PubChem website (see footnote 3). The files were then converted into PDB format using the Open Babel software (version 2.4.1). Hydrogen and Gasteiger charges were added to the above receptors and ligands using AutoDock Vina software, and then saved into PDBQT format. The AutoGrid tool of AutoDock Vina software was used to set the interfacing frame parameters, including the grid box that contained the entire system. The parameter was set to Lamarck Genetic Algorithm (LGA), which generated 10 docking results for each ligand and corresponding receptor. All the docking results were visualized by PyMOL software. Finally, the optimal docked structure could

² <http://www.symmap.org/>

³ <https://pubchem.ncbi.nlm.nih.gov>

⁴ <http://www.swisstargetprediction.ch>

⁵ <https://www.uniprot.org>

⁶ <https://www.omim.org/>

⁷ <https://www.genecards.org/>

⁸ <https://bioinfogp.cnb.csic.es/tools/venny/>

⁹ <https://string-db.org/>

¹⁰ <http://www.bioconductor.org/>

¹¹ <https://www.rcsb.org/>

be obtained based on the docking scores of all the possible docked structures.

2.8. MD simulation

The long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method. The target was placed in the center of simulation box filled with water TIP3P molecules with distance of 1.2 nm from the box boundary. There were 33 Na⁺ ions introduced in the water box to neutralize the charge of the whole system. The system was firstly balanced with energy minimization process which ran up to 100,000 steps using the steepest descent algorithm. Then the system was equilibrated to 310 K with v-rescale (velocity rescaling) method and backbone restrained. Subsequently, the system was further equilibrated at constant pressure (1 bar) and constant temperature (310 K). Finally, 200 ns MD test was conducted after all the restraints released. The simulation results were analyzed by Gromacs 5.1.2 built-in tools and our in-house scripts.

3. Results

3.1. Active compounds in XBCQD and the corresponding targets

In total, 285 compounds were screened out by the TCMSP database, among which 16, 19 and 11 compounds were from herbal rhubarb, bitter almond, and trichosanthes, respectively. And 43 active compounds met the OB \geq 30% and DL \geq 0.18 standards, which were selected for further analysis after removal of the duplicates. The TCMSP database was used to obtain the corresponding targets. All targets were then entered into the Uniprot database and normalized by removing the repeated ones. Finally, 116 potential targets were obtained.

3.2. Targets of monkeypox virus

The GeneCards database (Supplementary Table S1) and the OMIM database (Supplementary Table S2) yielded a total of 36 and 95 targets, respectively. After removal of duplicates, the number of related targets was 95. The 116 XBCQD-related targets and the 95 monkeypox virus gene targets were mapped to each other using the online tool Venny 2.1.0 software.¹² 36 XBCQD-monkeypox virus intersection targets were obtained (Figure 1A). All of the intersection targets were located between the differentially expressed genes in the monkeypox virus dataset.

3.3. PPI network analysis

According to the PPI network analysis (Figure 1C), the 36 predicted destinations were imported into String. If a node's degree,

betweenness, and proximity meet certain criteria, it can be designated as a hub node (Wan et al., 2019). The network centrality was used, individually or collectively, to define the network properties of the compounds (degree centrality, betweenness centrality, and closeness centrality) and to judge the importance of the nodes (Valente et al., 2008; Oldham et al., 2019). Nodes with higher levels (larger sizes) were considered to play a more critical role in the network (Chen et al., 2013).

The ranking of the nodes of the most important active compounds in XBCQD was summarized in Table 1, including beta-sitosterol, Stigmasterol, Gamma-Aminobutyric Acid, Phytol, estrone, Machiline, l-SPD, aloe-emodin, Glabridin, and Licochalcone B, etc. The top 10 targets were ESR1 (Estrogen Receptor 1, degree = 19), TP53 (tumor protein p53, degree = 19), CASP3 (cysteine-aspartic acid protease 3, degree = 17), JUN (transcription factor Jun, degree = 17), TNF (Tumor Necrosis Factor, degree = 17), MYC (Cellular myelocytomatosis oncogene, degree = 17), CDKN1A (Cyclin-dependent kinase inhibitor 1A, degree = 13), IL1B (Interleukin 1 Beta, degree = 13), CCNB1 (G2/mitotic-specific cyclin-B1, degree = 12), and AR (androgen receptor, degree = 11) (Figure 1B). Based on the above compound and target information, a compound-target pathway network was constructed to explain the mechanism of XBCQD against monkeypox virus, as shown in Figure 1D. The yellow V-shaped, light purple hexagon, light pink hexagon, light green hexagon and blue diamond shape represent herb names, KXR targets, DH targets, GLP targets and intersection core targets, respectively.

3.4. GO and KEGG pathway enrichment analyses

The 36 XBCQD-monkeypox virus intersection targets were imported into the Metascape platform. GO functional enrichment analysis was performed on the targets of the active ingredients in the treatment of monkey pox virus from the levels of BP, CC, and MF (Figure 2A). The top 10 items were selected for visual analysis (Figure 2B). In total, there were 2,455 BP items (Supplementary Table S3). The size of the circles indicates the number of targets, and the darker of the circle indicates the larger log *p* value of the BP item.

It can be found that BP was related to rhythmic process, positive regulation of cell cycle, response to drug, epithelial cell proliferation, DNA-templated transcription, initiation, response to steroid hormone, regulation of epithelial cell proliferation, negative regulation of protein phosphorylation, regulation of epithelial cell proliferation, and transcription initiation from RNA polymerase II promoter. There were 155 items in the CC analysis (Supplementary Table S4), and the top 10 items were selected for visual analysis (Supplementary Figure S1), including postsynaptic membrane, membrane raft, membrane microdomain, presynaptic membrane components, presynaptic membrane components, nuclear periphery, integral component of postsynaptic membrane, integral component of synaptic membrane and intrinsic component of synaptic membrane. In addition, there were a total of 260 items in MF molecular function analysis (Supplementary Table S5), with top 10 entries also visualized (Supplementary Figure S2), which were related to drug binding, nuclear receptor activity, ligand-activated transcription factor activity, protein kinase C activity,

¹² <http://bioinfo.genotoul.fr/jvenn>

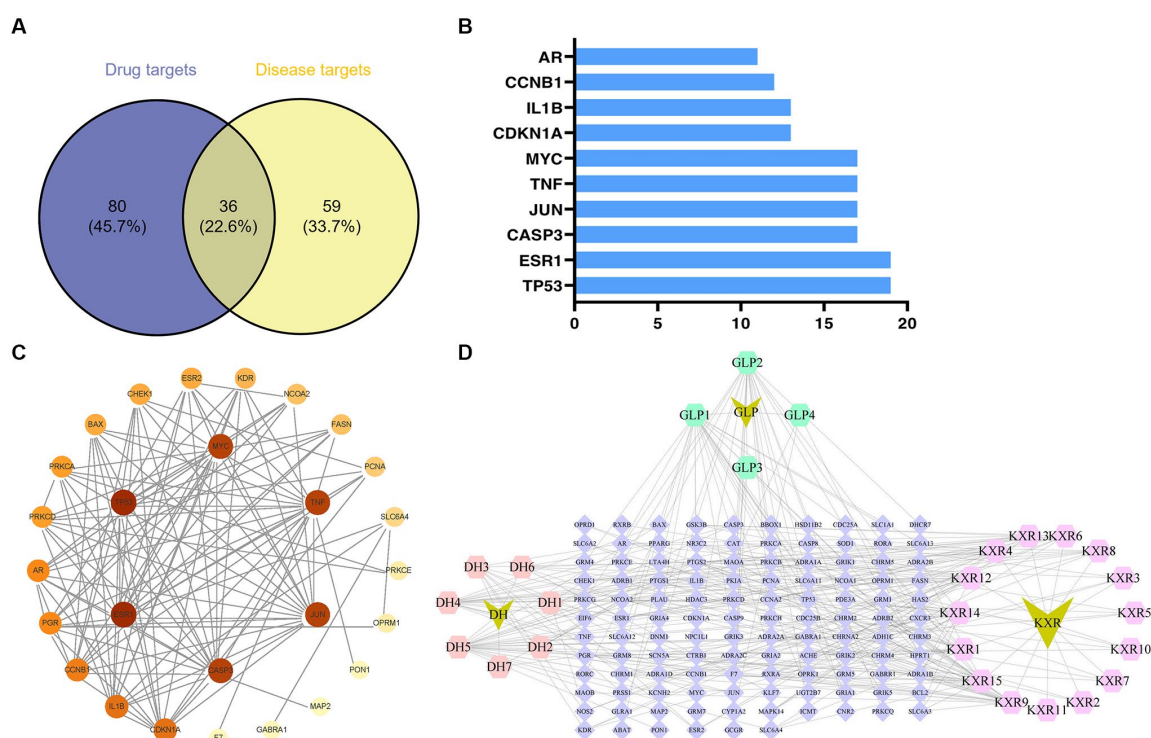


FIGURE 1

Potential targets of XBCQD and monkeypox, and PPI network. **(A)** Venn diagram of potential gene targets. **(B)** The top 10 monkeypox targets by degree. **(C)** The PPI network of XBCQD for the treatment of monkeypox. **(D)** The compounds-targets network showing potential mechanism of XBCQD for the treatment of monkeypox.

calcium-dependent protein kinase C activity, histone kinase activity, RNA polymerase II general transcription initiation factor binding, postsynaptic neurotransmitter receptor activity, acetylcholine receptor activity, and calcium-dependent protein serine/threonine kinase activity. According to the log p value in Figure 2D, 10 signaling pathways with high probability were screened according to the enrichment factor value and the number of genes involved in each pathway, which were closely related to the therapeutic mechanism of monkeypox virus. The size of the circle indicates the number of targets, and the darker of the circle indicates the larger log p value of the path.

To analyze the significance of hub targets in pathways involved in monkeypox virus treatment, the top 10 pathways determined according to gene counts and adjusted p values from the KEGG enrichment analysis and related targets were used to construct the KEGG key pathway network (Figure 2C). According to Figure 2C and Supplementary Table S6, XBCQD in the treatment of monkeypox virus could be mainly related to Hepatitis B (Supplementary Figure S3), AGE-RAGE signal pathway of diabetes complications, Breast cancer, Proteoglycans in cancer, Human cytomegalovirus infection, MAPK signaling pathway, Estrogen signaling pathway, Epstein-Barr virus infection, Human immunodeficiency virus 1 infection and Chemical carcinogenesis-receptor activation. Therefore, XBCQD could target multiple functional and biological factors in the treatment of monkeypox virus, and its effect was mainly reflected in affecting the process of cell proliferation and apoptosis. The components in XBCQD had direct or indirect regulatory effects on the inflammatory response

with regard to pruritic inflammation. However, the effects and far-reaching impact are still needed to be further verified.

3.5. Molecular docking of compound-target

Based on the monkeypox-related targets and selected compounds from the PPI network, molecular docking was performed. The interactions between the potential active compounds and the hub targets were analyzed using the AutoDock Vina, Discovery Studio 4.5 Client and PyMOL software applications. The selected top 5 active compounds included beta-sitosterol (MOL000087), Stigmasterol (MOL003035), Gamma-Aminobutyric Acid (MOL000388), Phytol (MOL001442) and estrone (MOL008204). The protein structures of hub targets were acquired online from RCSB PDB, including AR (PDB ID: 2QPY), CASP3 (PDB ID: 6BDV), CCNB1 (PDB ID: 2B9R), CDKN1A (PDB ID: 3TS8), ESR1 (PDB ID: 1A52), IL1B (PDB ID: 4DEP), JUN (PDB ID: 1FOS), MYC (PDB ID: 7C36), TP53 (PDB ID: 4MZI) and TNF (PDB ID: 1TNF). The values of affinity energy were obtained by docking analyses (Table 2). Notably, the lower of the affinity energy indicates the stronger of the binding capacity and more stable of the binding conformation. In general, the energy less than -5 kcal/mol indicates that the receptor has some binding ability to the ligand. Our results showed that estrone-AR (-12 kcal/mol), estrone-ESR1 (-11.1 kcal/mol), Stigmasterol-CASP3 (-10.4 kcal/mol), and beta-sitosterol-CASP3 (-9.9 kcal/mol) exhibited stronger binding affinities than the other moieties. The docking patterns of all molecules were shown in Figure 3.

3.6. Structural stability and interaction energy by MD simulation

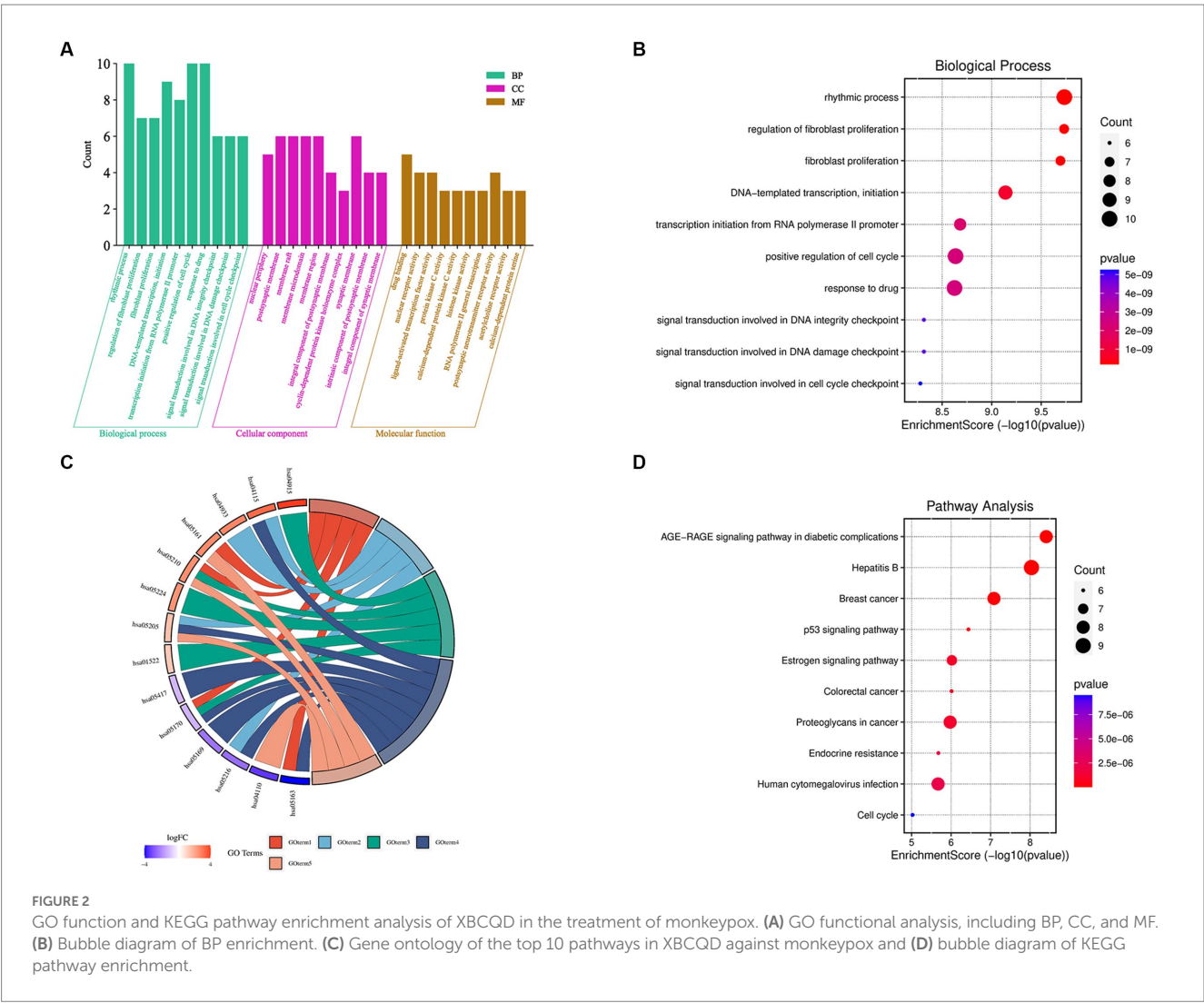
The two best dockings between the compounds and the targets (estrone-AR and estrone-ESR1) were selected to perform the MD

TABLE 1 Node ranking of the main active compounds.

ID	Name	Degree
MOL000087	beta-sitosterol	28
MOL003035	Stigmasterol	26
MOL000388	Gamma-Aminobutyric Acid	24
MOL001442	Phytol	23
MOL008204	estrone	20
MOL007207	Machiline	20
MOL012922	l-SPD	20
MOL000471	aloe-emodin	19
MOL004908	Glabridin	19
MOL004841	Licochalcone B	13

simulations. After 200 ns of MD simulations, the dynamic evolutions of estrone-AR and estrone-ESR1 complexes were studied. The conformations of estrone-AR and estrone-ESR1 complexes as well as their contact residues were shown in [Figures 4A,B](#). The interaction energy, radius of gyration (Rg), distance distribution, number of hydrogen bonds, root-mean-square fluctuation (RMSF) and root-mean-square deviation (RMSD) were analyzed accordingly.

As shown in [Figure 4A](#), the RMSD of the last 50 ns for estrone-ESR1 was 0.37 ± 0.06 nm, which was higher than the value of 0.20 ± 0.03 nm for estrone-AR, demonstrating a higher structural flexibility of estrone-ESR1 complex. Subsequently, we analyzed the change of protein cyclotron radius during the 200 ns simulation, which could characterize the compactness of the protein structure. From [Figure 4B](#), it can be observed that the radius of gyration (Rg) of AR basically remained at 1.85 ± 0.02 nm during the whole simulation process, whereas the Rg of ESR1 decreased from 1.97 ± 0.04 nm in the initial 50 ns to 1.89 ± 0.03 nm in the last 50 ns. The evolutions of Rg values were consistent with those of RMSD, indicating that the presence of estrone led to a tighter structure of ESR1. According to the changes of interaction energy during the 200 ns simulations shown in [Figure 4C](#), the estrone-AR complex showed lower interaction



energy of -186 ± 27 kJ/mol than that of estrone-ESR1 complex (-157 ± 25 kJ/mol) during the last 50 ns. The radial distribution functions from estrone to ESR1 and AR were plotted in Figure 4D. It can be observed that estrone had a slightly closer contact with AR than ESR1.

From Figures 5A,B, it can be observed that amino acid Glu 353 of AR and Asn 705 of ESR1 could form hydrogen bonds with phenol hydroxyl of estrone. Comparing Figures 5C,D, it can be clearly seen that AR-estrone were more likely to generate hydrogen bonds and more stable than the ESR1-estrone. It can be seen from Figures 5E,F that the RMSF values of amino acids in AR (residues 676–913: 0.11 ± 0.04 nm) were generally lower than those in ESR1 (residues 312–521: 0.15 ± 0.05 nm). Based on the results of MD simulations, the estrone-AR complex was proposed to possess better interaction stability and binding ability than the estrone-ESR1 complex.

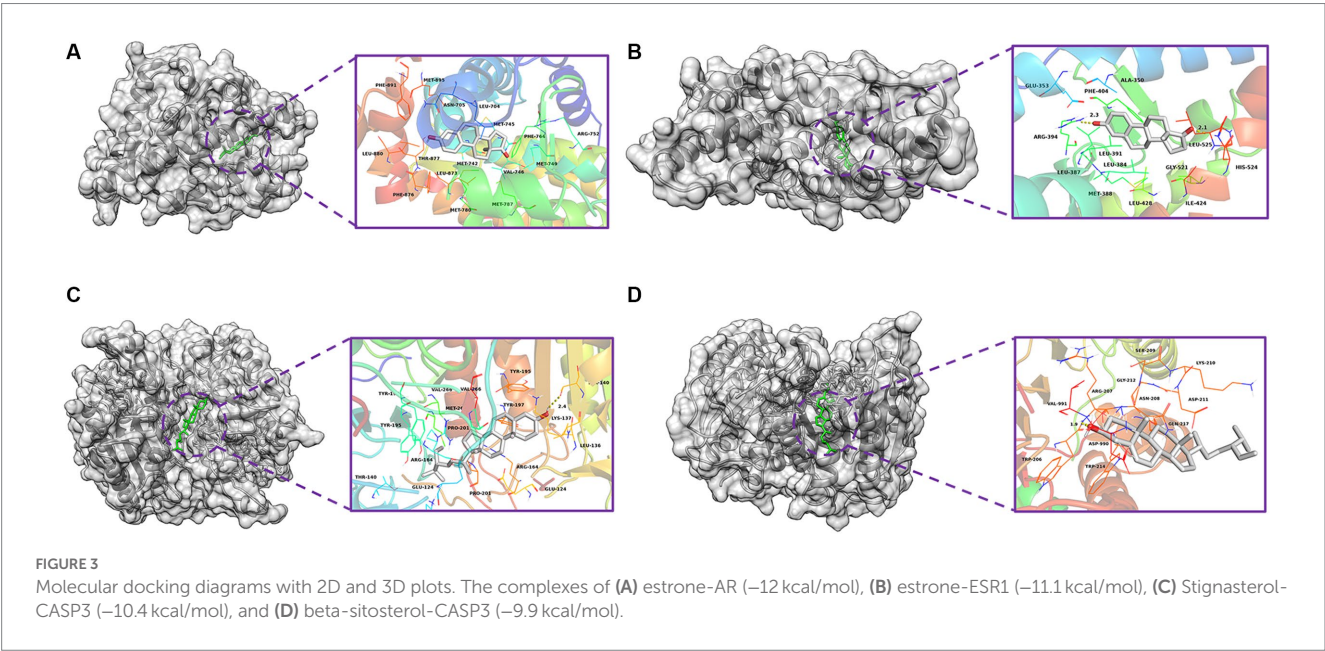
4. Discussion

TCM has accumulated a long history of clinical experience and efficacious prescriptions in preventing and treating diseases (Xing and Liu, 2021). It has been demonstrated to inhibit viral replications (Magden et al., 2005), however, the molecular mechanisms behind its effects have not been fully elucidated (Estep et al., 2011). To explore the potential pharmacological and molecular mechanism of XBCQD against monkeypox, we first employed network pharmacology in this study. A total of 36 potential targets associated with monkeypox were identified. Many of the targets were found to be hit by more than one compound. The results indicated that the active compounds of XBCQD could regulate more than one target and could have a synergistic effect on these targets.

The PPI analysis of the 36 targets showed that the top 10 hub targets, including ESR1, TP53, CASP3, JUN, TNF, MYC, CDKN1A, IL1B,

TABLE 2 The docking energy scores of the potential active compounds and hub targets.

ID	MOL000087	MOL008204	MOL000388	MOL001442	MOL003035
Name	Beta-sitosterol	Estrone	Gamma-Aminobutyric Acid	Phytol	Stigmasterol
AR	−8	−12	−4.3	−6.7	−8.4
CASP3	−9.9	−8.3	−4	−6.1	−10.4
CCNB1	−7.1	−7.3	−4.2	−4.8	−7.3
CDKN1A	−7.3	−7.2	−3.8	−5	−7.5
ESR1	−7	−11.1	−4.3	−5.8	−7.3
IL1B	−6.8	−7	−4.3	−4.4	−7.3
JUN	−8.1	−8.2	−4.7	−5.7	−8.2
MYC	−6.8	−7.6	−4.4	−5.1	−7.1
TNF	−6.6	−6.8	−4.2	−4.3	−6.8
TP53	−7.8	−7.9	−3.8	−5.3	−8.1



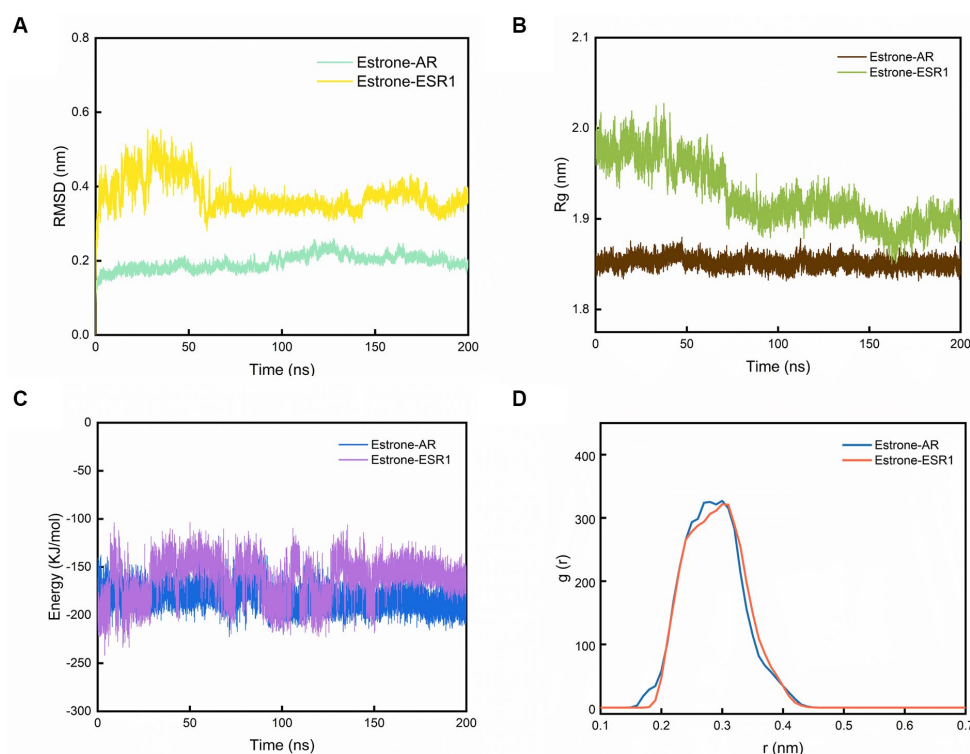


FIGURE 4

The results of MD simulations. (A) Evolutions of the RMSD, (B) Rg, (C) interaction energy values over 200 ns for the estrone-AR and estrone-ESR1, and (D) radial distribution function from estrone to AR and ESR1.

CCNB1, and AR may be the key targets of the treatment of monkeypox virus. The top 5 beta-sitosterol, Stigmasterol, Gamma-Aminobutyric Acid, Phytol, estrone and the top 10 targets were selected in this study. These bioactive compounds were found to effectively bind to the 10 targets according to the results of molecular docking. We further performed GO enrichment analysis on the 36 selected targets to better understand the multiple effects of XBCQD against monkeypox virus from a systematic perspective. The top 10 GO functional categories indicated that XBCQD may exert its effects through the involvement in the BP, MF, and CC. Among these targets, TP53 and ESR1 ranked highest. It was reported that TP53 could be one of the key enzymes in prostaglandin biosynthesis, which had an association with inflammation and mitosis (Costa et al., 2002). In addition, the inhibition of HIV-1, HCV and HSV by the serine protease inhibitor antithrombin III (ATIII) might be the result of TP53-mediated downstream synthesis of arachidonic acids, including prostaglandins (Smeets et al., 2014; He et al., 2015; Mussbacher et al., 2019). There was another study suggesting that the inhibition of ESR1 might be a potential therapeutic strategy for SARS-CoV-2 infection (Stilhan et al., 2020; Zhou et al., 2020; Li et al., 2022). Other identified hub targets were reported to be related to immune response and cytokine secretion (Lei et al., 2021). For example, JUN was found to be a stress response gene that altered cell structure during human development (Provençal et al., 2020). IL-6 was one of the classical pro-inflammatory cytokines that directly or indirectly activated a range of different cell types, further causing the secretion of cytokines (Turner et al., 2014; Hirano, 2021).

The concentration of GO terms in hub targets showed that XBCQD treatment of monkeypox was mainly involved the rhythmic

process, positive regulation of cell cycle, response to drug, epithelial cell proliferation and DNA-templated transcription, etc. KEGG pathway enrichment analysis of these targets indicated that they were involved in Hepatitis B, AGE-RAGE signal pathway of diabetes complications, Breast cancer, Proteoglycans in cancer, Human cytomegalovirus infection, etc. The AGE-RAGE signaling pathway was reported to have a regulatory role in diabetes (Ramamany et al., 2011; Kay et al., 2016). Some studies reported that IL-17 regulated viral infections (Ma et al., 2019; Ge et al., 2020; Mills, 2023). As a key cytokine in the pathogenesis of inflammation, TNF was involved in viral infection (Seo and Webster, 2002; Tuazon Kels et al., 2020; Darif et al., 2021). The Hepatitis B pathway was reported as major cellular signaling pathway activated by a variety of viruses (Branda and Wands, 2006; Nguyen et al., 2008; Herrscher et al., 2020). Moreover, estrogen signaling pathway was found to be enriched in both monkeypox infected monkeys and human models (Chadwick et al., 2005; Falcinelli et al., 2016; Xuan et al., 2022). GO and KEGG enrichment analysis suggested that XBCQD may have a positive effect on monkeypox by inhibiting inflammation and viral replication *via* these pathways.

The results of network pharmacology were validated by molecular docking of the top 10 targets and 5 active compounds. The two complexes with lowest binding energy scores were found to be estrone-AR and estrone-ESR1. Furthermore, MD simulations showed that the average value of interaction energy of estrone-AR was -186 kcal/mol (energy drift: 27 kcal/mol), while was lower than the value of -157 kcal/mol (energy drift: 25 kcal/mol) of the estrone-ESR1 complex. Therefore, the interaction of estrone-AR complex was more stable comparing to estrone-ESR1 complex.

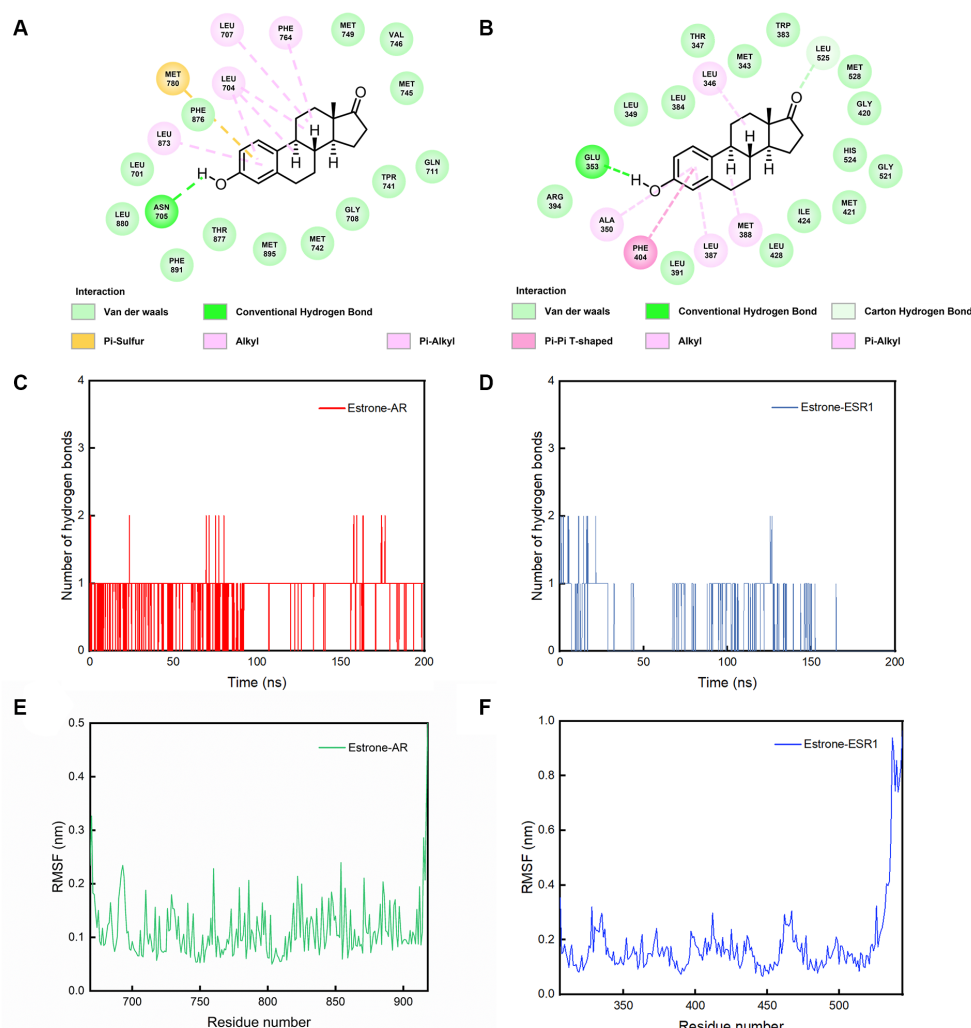


FIGURE 5

The conformations of contact residues for (A) estrone-AR and (B) estrone-ESR1 complexes, number of hydrogen bonds between the (C) estrone-AR and (D) estrone-ESR1 complexes, and (E) RMSF curves for AR and (F) ESR1 during the 200 ns simulations.

In summary, this study systematically explored the interactions between the bioactive compounds of XBCQD and the monkeypox-specific XBCQD targets using network pharmacological methods, bioinformatics analyses and molecular simulations, suggesting that XBCQD could have a beneficial therapeutic effect on monkeypox by reducing the inflammatory damage and viral replication *via* multiple pathways. In addition, the use of XBCQD on monkeypox disease was confirmed to be best worked through the estrone-target AR interaction. By using a series of computational approaches, our study established the drug screening of XBCQD on monkeypox disease for the first time, hopefully could provide some guidance for future drug development.

5. Conclusion

By combining network pharmacology, molecular docking and MD simulation, the molecular mechanism of XBCQD in the

treatment of monkeypox virus was systematically investigated. According to our results, the top 5 compounds (beta-sitosterol, Stigmasterol, Gamma-Aminobutyric Acid, Phytol, estrone) and the top 10 targets (ESR1, TP53, CASP3, JUN, TNF, MYC, CDKN1A, IL1B, CCNB1, and AR) were identified. We also found that XBCQD in the treatment of monkeypox virus could be mainly related to Hepatitis B, AGE-RAGE signal pathway of diabetes complications, Breast cancer, Proteoglycans in cancer, Human cytomegalovirus infection, MAPK signaling pathway, Estrogen signaling pathway, Epstein-Barr virus infection, Human immunodeficiency virus 1 infection and Chemical carcinogenesis-receptor activation. Molecular docking showed that estrone-AR and estrone-ESR1 exhibited stronger binding affinities than the other moieties. And the estrone-AR possessed higher structural and interaction stabilities than the estrone-ESR1 with the extension of molecular simulation time. Our study provides a comprehensive explanation of the multi-component, multi-target and multi-pathway intervention

mechanism of XBCQD in the treatment of monkeypox, which is expected to provide a basis and new insights for further pharmacological research.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary materials](#), further inquiries can be directed to the corresponding author.

Author contributions

YS designed the simulations. YJ and CS performed the simulations and analyzed the data. YS, YJ, and CS wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by the National Natural Science Foundation of China (Ref: 12102113) and the Major program of the National Natural Science Foundation of China (T2293720/T2293722).

References

- Abu-Hammad, O., Abu-Hammad, A., Jaber, A.-R., Jaber, A. R., and Dar-Odeh, N. (2023). Factors associated with geographic variations in the 2022 monkeypox outbreak: a systematic review. *New Microbes New Infect.* 51:101078. doi: 10.1016/j.nmni.2022.101078
- Branda, M., and Wands, J. R. (2006). Signal transduction cascades and hepatitis B and C related hepatocellular carcinoma. *Hepatology* 43, 891–902. doi: 10.1002/hep.211196
- Chadwick, C. C., Chippari, S., Matelan, E., Borges-Marcucci, L., Eckert, A. M., Keith, J. C., et al. (2005). Identification of pathway-selective estrogen receptor ligands that inhibit NF- κ B transcriptional activity. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2543–2548. doi: 10.1073/pnas.0405841102
- Chen, J. M., Chen, R. X., Gong, H. Y., Zhao, M. M., Ji, Y. F., Sun, M. H., et al. (2022). Epidemiology-based analysis of the risks and elimination strategies of the monkeypox outbreak in 2022. *Front. Vet. Sci.* 9:1064766. doi: 10.3389/fvets.2022.1064766
- Chen, D.-B., Gao, H., Lü, L., and Zhou, T. (2013). Identifying influential nodes in large-scale directed networks: the role of clustering. *PLoS One* 8:e77455. doi: 10.1371/journal.pone.0077455
- Costa, C., Soares, R., Reis-Filho, J. S., Leitão, D., Amendoeira, I., and Schmitt, F. C. (2002). Cyclo-oxygenase 2 expression is associated with angiogenesis and lymph node metastasis in human breast cancer. *J. Clin. Pathol.* 55, 429–434. doi: 10.1136/jcp.55.6.429
- Darif, D., Hammi, I., Kihel, A., El Idriissi Saik, I., Guessous, F., and Akarid, K. (2021). The pro-inflammatory cytokines in COVID-19 pathogenesis: what goes wrong? *Microb. Pathog.* 153:104799. doi: 10.1016/j.micpath.2021.104799
- Desai, A. N., Thompson, G. R., Iii Neumeister, S. M., Arutyunova, A. M., Trigg, K., and Cohen, S. H. (2022). Compassionate use of tecovirimat for the treatment of monkeypox infection. *JAMA* 328, 1348–1350. doi: 10.1001/jama.2022.15336
- Dou, Y. M., Yuan, H., and Tian, H. W. (2023). Monkeypox virus: past and present. *World J. Pediatr.* 19, 224–230. doi: 10.1007/s12519-022-00618-1
- Estep, R. D., Messaoudi, I., O'connor, M. A., Li, H., Sprague, J., Barron, A., et al. (2011). Deletion of the monkeypox virus inhibitor of complement enzymes locus impacts the adaptive immune response to monkeypox virus in a nonhuman primate model of infection. *J. Virol.* 85, 9527–9542. doi: 10.1128/jvi.00199-11
- Falcinelli, S. D., Chertow, D. S., and Kindrachuk, J. (2016). Integration of global analyses of host molecular responses with clinical data to evaluate pathogenesis and advance therapies for emerging and re-emerging viral infections. *ACS Infect. Dis.* 2, 787–799. doi: 10.1021/acsinfecdis.6b00104
- Ge, Y., Huang, M., and Yao, Y.-M. (2020). Biology of interleukin-17 and its pathophysiological significance in sepsis. *Front. Immunol.* 11:1558. doi: 10.3389/fimmu.2020.01558
- Hanzlicek, A. S., Roof, C. J., Sanderson, M. W., and Grauer, G. F. (2014). The effect of Chinese rhubarb, *rheum officinale*, with and without benazepril on the progression of naturally occurring chronic kidney disease in cats. *J. Vet. Intern. Med.* 28, 1221–1228. doi: 10.1111/jvim.12365
- Hasan, S., and Saeed, S. (2022). Monkeypox disease: an emerging public health concern in the shadow of COVID-19 pandemic: an update. *Trop. Med. Infect. Dis.* 7:283. doi: 10.3390/tropicalmed7100283
- He, W., Shi, F., Zhou, Z. W., Li, B., Zhang, K., Zhang, X., et al. (2015). A bioinformatic and mechanistic study elicits the antifibrotic effect of ursolic acid through the attenuation of oxidative stress with the involvement of ERK, PI3K/Akt, and p38 MAPK signaling pathways in human hepatic stellate cells and rat liver. *Drug Des. Devel. Ther.* 9, 3989–4104. doi: 10.2147/dddt.S85426
- Herrscher, C., Roingeard, P., and Blanchard, E. (2020). Hepatitis B virus entry into cells. *Cells* 9:1486. doi: 10.3390/cells9061486
- Hirano, T. (2021). IL-6 in inflammation, autoimmunity and cancer. *Int. Immunol.* 33, 127–148. doi: 10.1093/intimm/dxaa078
- Hollingsworth, S. A., and Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron* 99, 1129–1143. doi: 10.1016/j.neuron.2018.08.011
- Huang, K., Zhang, P., Zhang, Z., Youn, J. Y., Wang, C., Zhang, H., et al. (2021). Traditional Chinese medicine (TCM) in the treatment of COVID-19 and other viral infections: efficacies and mechanisms. *Pharmacol. Ther.* 225:107843. doi: 10.1016/j.pharmthera.2021.107843
- Hung, Y.-P., Lee, C.-C., Lee, J.-C., Chiu, C.-W., Hsueh, P.-R., and Ko, W.-C. (2022). A brief on new waves of monkeypox and vaccines and antiviral drugs for monkeypox. *J. Microbiol. Immunol. Infect.* 55, 795–802. doi: 10.1016/j.jmii.2022.08.016
- Huo, J., Wang, T., Wei, B., Shi, X., Yang, A., Chen, D., et al. (2022). Integrated network pharmacology and intestinal flora analysis to determine the protective effect of Xuanbai-Chengqi decoction on lung and gut injuries in influenza virus-infected mice. *J. Ethnopharmacol.* 298:115649. doi: 10.1016/j.jep.2022.115649
- Hutson, C. L., Abel, J. A., Carroll, D. S., Olson, V. A., Braden, Z. H., Hughes, C. M., et al. (2010). Comparison of west African and Congo Basin monkeypox viruses in BALB/c and C57BL/6 mice. *PLoS One* 5:e8912. doi: 10.1371/journal.pone.0008912

Acknowledgments

The authors thank all of the participants and peer workers.

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.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1234817/full#supplementary-material>

- Iñigo Martínez, J., Gil Montalbán, E., Jiménez Bueno, S., Martín Martínez, F., Nieto Juliá, A., Sánchez Díaz, J., et al. (2022). Monkeypox outbreak predominantly affecting men who have sex with men, Madrid, Spain, 26 April to 16 June 2022. *Euro Surveill.* 27:2200471. doi: 10.2807/1560-7917.Es.2022.27.27.2200471
- Kaler, J., Hussain, A., Flores, G., Kheiri, S., and Desrosiers, D. (2022). Monkeypox: a comprehensive review of transmission, pathogenesis, and manifestation. *Cureus* 14:e26531. doi: 10.7759/cureus.26531
- Kay, A. M., Simpson, C. L., and Stewart, J. A. Jr. (2016). The role of AGE/RAGE signaling in diabetes-mediated vascular calcification. *J. Diabetes Res.* 2016:6809703. doi: 10.1155/2016/6809703
- Lei, T.-Y., Ye, Y.-Z., Zhu, X.-Q., Smerin, D., Gu, L.-J., Xiong, X.-X., et al. (2021). The immune response of T cells and therapeutic targets related to regulating the levels of T helper cells after ischaemic stroke. *J. Neuroinflammation* 18:25. doi: 10.1186/s12974-020-02057-z
- Letafati, A., and Sakthavaz, T. (2023). Monkeypox virus: a review. *Microb. Pathog.* 176:106027. doi: 10.1016/j.micpath.2023.106027
- Li, F., Boon, A. C. M., Michelson, A. P., Foraker, R. E., Zhan, M., and Payne, P. R. O. (2022). Estrogen hormone is an essential sex factor inhibiting inflammation and immune response in COVID-19. *Sci. Rep.* 12:9462. doi: 10.21203/rs.3.rs-936900/v1
- Ligon, B. L. (2004). Monkeypox: a review of the history and emergence in the western hemisphere. *Semin. Pediatr. Infect. Dis.* 15, 280–287. doi: 10.1053/j.spid.2004.09.001
- Ma, W. T., Yao, X. T., Peng, Q., and Chen, D. K. (2019). The protective and pathogenic roles of IL-17 in viral infections: friend or foe? *Open Biol.* 9:190109. doi: 10.1098/rsob.190109
- Magden, J., Kääriäinen, L., and Ahola, T. (2005). Inhibitors of virus replication: recent developments and prospects. *Appl. Microbiol. Biotechnol.* 66, 612–621. doi: 10.1007/s00253-004-1783-3
- Meng, X. Y., Zhang, H. X., Mezei, M., and Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* 7, 146–157. doi: 10.2174/157340911795677602
- Mills, K. H. G. (2023). IL-17 and IL-17-producing cells in protection versus pathology. *Nat. Rev. Immunol.* 23, 38–54. doi: 10.1038/s41577-022-00746-9
- Mussbacher, M., Salzmann, M., Brostjan, C., Hoesel, B., Schoergenhofer, C., Datler, H., et al. (2019). Cell type-specific roles of NF- κ B linking inflammation and thrombosis. *Front. Immunol.* 10:85. doi: 10.3389/fimmu.2019.00085
- Nguyen, D. H., Ludgate, L., and Hu, J. (2008). Hepatitis B virus-cell interactions and pathogenesis. *J. Cell. Physiol.* 216, 289–294. doi: 10.1002/jcp.21416
- Noor, F., Tahir ul Qamar, M., Ashfaq, U. A., Albutti, A., Alwashmi, A. S. S., and Aljasir, M. A. (2022). Network pharmacology approach for medicinal plants: review and assessment. *Pharmaceuticals* 15:572. doi: 10.3390/ph15050572
- Oldham, S., Fulcher, B., Parkes, L., Arnatkevičiūtė, A., Suo, C., and Fornito, A. (2019). Consistency and differences between centrality measures across distinct classes of networks. *PLoS One* 14:e0220061. doi: 10.1371/journal.pone.0220061
- Patel, A., Bilinska, J., Tam, J. C. H., da Silva Fontoura, D., Mason, C. Y., Daunt, A., et al. (2022). Clinical features and novel presentations of human monkeypox in a Central London Centre during the 2022 outbreak: descriptive case series. *BMJ* 378:e072410. doi: 10.1136/bmj-2022-072410
- Patel, V. M., and Patel, S. V. (2023). Epidemiological review on monkeypox. *Cureus* 15:e34653. doi: 10.7759/cureus.34653
- Provençal, N., Arloth, J., Cattaneo, A., Anacker, C., Cattane, N., Wiechmann, T., et al. (2020). Glucocorticoid exposure during hippocampal neurogenesis primes future stress response by inducing changes in DNA methylation. *Proc. Natl. Acad. Sci. U. S. A.* 117, 23280–23285. doi: 10.1073/pnas.1820842116
- Ramasamy, R., Yan, S. F., and Schmidt, A. M. (2011). Receptor for AGE (RAGE): signaling mechanisms in the pathogenesis of diabetes and its complications. *Ann. N. Y. Acad. Sci.* 1243, 88–102. doi: 10.1111/j.1749-6632.2011.06320.x
- Reynolds, M. G., Doty, J. B., Mccollum, A. M., Olson, V. A., and Nakazawa, Y. (2019). Monkeypox re-emergence in Africa: a call to expand the concept and practice of one health. *Expert Rev. Anti-Infect. Ther.* 17, 129–139. doi: 10.1080/14787210.2019.1567330
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–424. doi: 10.1038/gim.2015.30
- Rizk, J. G., Lippi, G., Henry, B. M., Forthal, D. N., and Rizk, Y. (2022). Prevention and treatment of monkeypox. *Drugs* 82, 957–963. doi: 10.1007/s40265-022-01742-y
- Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., et al. (2014). TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminform.* 6:13. doi: 10.1186/1758-2946-6-13
- Salmaso, V., and Moro, S. (2018). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. *Front. Pharmacol.* 9:923. doi: 10.3389/fphar.2018.00923
- Seo, S. H., and Webster, R. G. (2002). Tumor necrosis factor alpha exerts powerful anti-influenza virus effects in lung epithelial cells. *J. Virol.* 76, 1071–1076. doi: 10.1128/jvi.76.3.1071-1076.2002
- Shafaati, M., and Zandi, M. (2022). State-of-the-art on monkeypox virus: an emerging zoonotic disease. *Infection* 50, 1425–1430. doi: 10.1007/s15010-022-01935-3
- Shafaati, M., and Zandi, M. (2023). Human monkeypox (hMPXV) re-emergence: host immunity status and current vaccines landscape. *J. Med. Virol.* 95:e28251. doi: 10.1002/jmv.28251
- Shafaati, M., Zandi, M., and Choudhary, O. P. (2022). Monkeypox virus crosstalk with HIV; where do we stand now? *Int. J. Surg.* 105:106897. doi: 10.1016/j.ijssu.2022.106897
- Shamim, M. A., Padhi, B. K., Satapathy, P., Veeramachaneni, S. D., Chatterjee, C., Tripathy, S., et al. (2023). The use of antivirals in the treatment of human monkeypox outbreaks: a systematic review. *Int. J. Infect. Dis.* 127, 150–161. doi: 10.1016/j.ijid.2022.11.040
- Smee, D. F., Hurst, B. L., Day, C. W., and Geiben-Lynn, R. (2014). Influenza virus H1N1 inhibition by serine protease inhibitor (serpin) antithrombin III. *Int. Trends Immun.* 2, 83–86.
- Stilhano, R. S., Costa, A. J., Nishino, M. S., Shams, S., Bartolomeo, C. S., Breithaupt-Faloppa, A. C., et al. (2020). SARS-CoV-2 and the possible connection to ERs, ACE2, and RAGE: focus on susceptibility factors. *FASEB J.* 34, 14103–14119. doi: 10.1096/fj.202001394RR
- Tuazon Kels, M. J., Ng, E., Al Rumaih, Z., Pandey, P., Ruuls, S. R., Korner, H., et al. (2020). TNF deficiency dysregulates inflammatory cytokine production, leading to lung pathology and death during respiratory poxvirus infection. *Proc. Natl. Acad. Sci. U. S. A.* 117, 15935–15946. doi: 10.1073/pnas.2004615117
- Turner, M. D., Nedjai, B., Hurst, T., and Pennington, D. J. (2014). Cytokines and chemokines: at the crossroads of cell signalling and inflammatory disease. *Biochim. Biophys. Acta, Mol. Cell Res.* 1843, 2563–2582. doi: 10.1016/j.bbamcr.2014.05.014
- Valente, T. W., Coronges, K., Lakon, C., and Costenbader, E. (2008). How correlated are network centrality measures? *Connect (Tor)* 28, 16–26.
- Wan, Y., Xu, L., Liu, Z., Yang, M., Jiang, X., Zhang, Q., et al. (2019). Utilising network pharmacology to explore the underlying mechanism of Wumei pill in treating pancreatic neoplasms. *BMC Complement. Altern. Med.* 19:158. doi: 10.1186/s12906-019-2580-y
- Wang, Y., Li, N., Li, Q., Liu, Z., Li, Y., Kong, J., et al. (2021). Xuanbai Chengqi decoction ameliorates pulmonary inflammation via reshaping gut microbiota and rectifying Th17/Treg imbalance in a murine model of chronic obstructive pulmonary disease. *Int. J. Chron. Obstruct. Pulmon. Dis.* 16, 3317–3335. doi: 10.2147/COPD.S337181
- Warner, B. M., Klassen, L., Sloan, A., Deschambault, Y., Soule, G., Banadyga, L., et al. (2022). In vitro and in vivo efficacy of tecovirimat against a recently emerged 2022 monkeypox virus isolate. *Sci. Transl. Med.* 14:eade7646. doi: 10.1126/scitranslmed.ade7646
- Wilder-Smith, A., and Osman, S. (2020). Public health emergencies of international concern: a historic overview. *J. Travel Med.* 27:taaa227. doi: 10.1093/jtm/taaa227
- Xing, D., and Liu, Z. (2021). Effectiveness and safety of traditional Chinese medicine in treating COVID-19: clinical evidence from China. *Aging Dis.* 12, 1850–1856. doi: 10.14339/ad.2021.0906
- Xuan, D. T. M., Yeh, I. J., Wu, C. C., Su, C. Y., Liu, H. L., Chiao, C. C., et al. (2022). Comparison of transcriptomic signatures between monkeypox-infected monkey and human cell lines. *J. Immunol Res* 2022:3883822. doi: 10.1155/2022/3883822
- Yu, Y., Zhou, B., Chen, L., Gao, T., and Liu, J. (2022). Identifying important nodes in complex networks based on node propagation entropy. *Entropy (Basel)* 24:275. doi: 10.3390/e24020275
- Zardi, E. M., and Chello, C. (2022). Human monkeypox-a global public health emergency. *Int. J. Environ. Res. Public Health* 19:16781. doi: 10.3390/ijerph192416781
- Zhang, G. B., Li, Q. Y., Chen, Q. L., and Su, S. B. (2013). Network pharmacology: a new approach for Chinese herbal medicine research. *Evid. Based Complement. Alternat. Med.* 2013:621423. doi: 10.1155/2013/621423
- Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., and Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* 6:14. doi: 10.1038/s41421-020-0153-3
- Zhu, H., Wang, S., Shan, C., Li, X., Tan, B., Chen, Q., et al. (2021). Mechanism of protective effect of xuan-bai-cheng-qi decoction on LPS-induced acute lung injury based on an integrated network pharmacology and RNA-sequencing approach. *Respir. Res.* 22:188. doi: 10.1186/s12931-021-01781-1
- Zovi, A., Ferrara, F., Langella, R., and Vitiello, A. (2022). Pharmacological agents with antiviral activity against monkeypox infection. *Int. J. Mol. Sci.* 23:15941. doi: 10.3390/ijms232415941

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

