

# Pharmacoinformatics: New developments and challenges in drug design

**Edited by**

David Ramírez, Rafael Peláez, Carlos F. Lagos  
and José L. Medina-Franco

**Published in**

Frontiers in Pharmacology



## 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 978-2-8325-5300-8  
DOI 10.3389/978-2-8325-5300-8

## 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](https://frontiersin.org/about/contact)

# Pharmacoinformatics: New developments and challenges in drug design

## Topic editors

David Ramírez — University of Concepcion, Chile

Rafael Peláez — University of Salamanca, Spain

Carlos F. Lagos — Universidad San Sebastian, Chile

José L. Medina-Franco — National Autonomous University of Mexico, Mexico

## Citation

Ramírez, D., Peláez, R., Lagos, C. F., Medina-Franco, J. L., eds. (2024).

*Pharmacoinformatics: New developments and challenges in drug design.*

Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5300-8

# Table of contents

- 05 Editorial: Pharmacoinformatics: new developments and challenges in drug design  
David Ramírez, Rafael Peláez, Carlos F. Lagos and José L. Medina-Franco
- 07 Deep learning-based multi-drug synergy prediction model for individually tailored anti-cancer therapies  
Shengnan She, Hengwei Chen, Wei Ji, Mengqiu Sun, Jiaxi Cheng, Mengjie Rui and Chunlai Feng
- 24 Network pharmacology-based approach to explore the underlying mechanism of sinomenine on sepsis-induced myocardial injury in rats  
Linggang Sun, Zhiyun Chen, Yunjie Ni and Zhengfei He
- 37 Computational study of the binding orientation and affinity of noncovalent inhibitors of the papain-like protease (PLpro) from SARS-CoV-1 considering the protein flexibility by using molecular dynamics and cross-docking  
Luis Castillo-Campos, José Luis Velázquez-Libera and Julio Caballero
- 51 Identifying target organ location of *Radix Achyranthis Bidentatae*: a bioinformatics approach on active compounds and genes  
Minh Nhat Tran, Su-Jin Baek, Hyeong Joon Jun and Sanghun Lee
- 64 MBC and ECBL libraries: outstanding tools for drug discovery  
Tiziana Ginex, Enrique Madruga, Ana Martinez and Carmen Gil
- 76 Unraveling the potential mechanisms of the anti-osteoporotic effects of the *Achyranthes bidentata*–*Dipsacus asper* herb pair: a network pharmacology and experimental study  
Tao Li, Wenzhao Li, Xiaoning Guo, Tingting Tan, Cheng Xiang and Zhengxiao Ouyang
- 91 Design of a multi-target focused library for antidiabetic targets using a comprehensive set of chemical transformation rules  
Fernanda I. Saldivar-González, Gabriel Navarrete-Vázquez and José L. Medina-Franco
- 107 Prediction of drug–disease associations based on reinforcement symmetric metric learning and graph convolution network  
Huimin Luo, Chunli Zhu, Jianlin Wang, Ge Zhang, Junwei Luo and Chaokun Yan



- 120 **Synthesis, biological and computational evaluation of novel cyanomethyl vinyl ether derivatives**  
Endika Martín-Encinas, María Fuertes, Samuel Delgado-Hernández, Fernando García-Tellado, David Tejedor and Concepción Alonso
- 132 **Repurposing the Open Global Health Library for the discovery of novel Mpro destabilizers with scope as broad-spectrum antivirals**  
Francisco Castillo, David Ramírez, María C. Ramos, Blanca Martínez-Arribas, Elisabeth Domingo-Contreras, Thomas A. Mackenzie, Carlos Peña-Varas, Sven Lindemann, Fernando Montero, Frederick Annang, Francisca Vicente, Olga Genilloud, Dolores González-Pacanowska and Rosario Fernandez-Godino



## OPEN ACCESS

EDITED AND REVIEWED BY  
Alastair George Stewart,  
The University of Melbourne, Australia

\*CORRESPONDENCE  
David Ramírez,  
✉ dramirezs@udec.cl

RECEIVED 09 July 2024  
ACCEPTED 15 July 2024  
PUBLISHED 30 July 2024

CITATION  
Ramírez D, Peláez R, Lagos CF and  
Medina-Franco JL (2024), Editorial:  
Pharmacoinformatics: new developments and  
challenges in drug design.  
*Front. Pharmacol.* 15:1462070.  
doi: 10.3389/fphar.2024.1462070

COPYRIGHT  
© 2024 Ramírez, Peláez, Lagos and Medina-  
Franco. 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: Pharmacoinformatics: new developments and challenges in drug design

David Ramírez<sup>1\*</sup>, Rafael Peláez<sup>2,3,4</sup>, Carlos F. Lagos<sup>5,6</sup> and  
José L. Medina-Franco<sup>7</sup>

<sup>1</sup>Departamento de Farmacología, Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción, Chile, <sup>2</sup>Laboratorio de Química Orgánica y Farmacéutica, Departamento de Ciencias Farmacéuticas, Facultad de Farmacia, Universidad de Salamanca, Salamanca, Spain, <sup>3</sup>Instituto de Investigación Biomédica de Salamanca (IBSAL), Facultad de Farmacia, Universidad de Salamanca, Salamanca, Spain, <sup>4</sup>Centro de Investigación de Enfermedades Tropicales de la Universidad de Salamanca (CIETUS), Facultad de Farmacia, Universidad de Salamanca, Salamanca, Spain, <sup>5</sup>Chemical Biology and Drug Discovery Lab, Escuela de Química y Farmacia, Facultad de Medicina y Ciencia, Universidad San Sebastián, Campus Los Leones, Santiago, Chile, <sup>6</sup>Centro Ciencia and Vida, Fundación Ciencia and Vida, Huechuraba, Santiago, Chile, <sup>7</sup>DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Mexico City, Mexico

## KEYWORDS

pharmacoinformatics, drug design, systems pharmacology, multi target directed ligand, pharmacology

## Editorial on the Research Topic

[Pharmacoinformatics: new developments and challenges in drug design](#)

Pharmacoinformatics represents a crucial nexus between computational sciences and pharmacology, aiming to leverage advanced data analysis and machine learning techniques to streamline drug discovery and development. The Research Topic, “Pharmacoinformatics: New Developments and Challenges in Drug Design,” brings together 10 original research articles contributed by more than 57 authors with over 18,000 views and downloads in all of the time until July 2024, that exemplify the transformative potential of computational approaches in modern pharmacology. The articles in this Research Topic underscore the diverse methodologies and applications of pharmacoinformatics, addressing a broad range of therapeutic areas and offering novel insights into drug design, efficacy, and mechanism of action.

One of the key contributions to this Research Topic is the study by [She et al.](#), titled “Deep learning-based multi-drug synergy prediction model for individually tailored anti-cancer therapies.” This paper presents a sophisticated deep learning model to predict synergistic multiple-drug combinations for cancer treatment. The model’s ability to tailor therapies to individual patients marks a significant step towards personalized medicine, demonstrating how machine learning can optimize therapeutic regimens and potentially improve clinical outcomes. The wide and appropriate use of artificial intelligence techniques such as machine learning techniques continues to revolutionize drug-disease association predictions, as demonstrated by [Luo et al.](#) in “Prediction of drug–disease associations based on reinforcement symmetric metric learning and graph convolution network.” The innovative use of reinforcement learning and graph convolution networks in this study highlights the evolving landscape of computational drug discovery, offering new avenues for identifying potential therapeutic applications of existing drugs.

Another compelling study, “Network pharmacology-based approach to explore the underlying mechanism of sinomenine on sepsis-induced myocardial injury in rats” by Sun et al., utilizes network pharmacology to unravel the mechanisms by which sinomenine exerts its protective effects against myocardial injury in sepsis. This research highlights the utility of network-based analyses in identifying key molecular interactions and pathways, providing a holistic understanding of drug action and facilitating the discovery of new therapeutic targets. The potential of network pharmacology is further exemplified by Li et al., where the authors studied the potential mechanisms of the anti-osteoporotic effects of the *Achyranthes bidentata*–*Dipsacus asper* herb pair. This research combines network pharmacology with experimental validation to decode the anti-osteoporotic mechanisms of a traditional herb pair, showcasing a robust approach to studying complex herbal formulations.

The investigation by Tran et al., “Identifying target organ location of *Radix Achyranthis Bidentatae*: a bioinformatics approach on active compounds and genes,” exemplifies the integration of bioinformatics in phytomedicine research. By identifying the target organs of herbal therapies this study not only advances our understanding of traditional medicinal herbs but also illustrates the broader applicability of pharmacoinformatics in natural product research.

In the field of antiviral drug discovery, Castillo-Campos et al. contribute a comprehensive computational analysis titled “Computational study of the binding orientation and affinity of noncovalent inhibitors of the papain-like protease (PLpro) from SARS-CoV-1 considering the protein flexibility by using molecular dynamics and cross-docking.” This work underscores the importance of accounting for protein flexibility in docking and molecular dynamics studies and provides valuable insights into the development of inhibitors against viral proteases, crucial for combating viral pathogens like SARS-CoV-1. In the same field, Castillo et al. discovered novel Mpro destabilizers with scope as broad-spectrum antivirals. They illustrate the innovative use of pharmacoinformatics for drug repurposing. By identifying Mpro destabilizers with potential broad-spectrum antiviral activity, this study contributes to the ongoing efforts to find effective treatments for viral infections.

Exploring the field of chemical libraries, Ginex et al. present two novel chemical libraries, MBC and ECBL as outstanding tools for drug discovery. Both libraries are designed to enhance the efficiency of drug discovery. The strategic design and diverse chemical space covered by these libraries are poised to accelerate the identification of promising drug candidates. The design of focused chemical libraries is further elaborated in Saldivar-González et al.’s work, “Design of a multi-target focused library for antidiabetic targets using a comprehensive set of chemical transformation rules.” By employing a multi-target approach, this study aims to address the multifactorial nature of diabetes, providing a valuable resource for the development of antidiabetic therapies.

Finally, in the synthesis and evaluation of novel compounds, Martín-Encinas et al.’s article, “Synthesis, biological and computational evaluation of novel cyanomethyl vinyl ether derivatives” stands out. This research integrates computational and experimental approaches to characterize new compounds, showcasing a comprehensive strategy for drug development.

Collectively, the contributions of the Research Topic, “Pharmacoinformatics: New Developments and Challenges in

Drug Design,” highlight the dynamic and interdisciplinary nature of pharmacoinformatics. The integration of advanced computational methods with pharmacological research not only accelerates drug discovery but also enhances our understanding of complex biological systems. As we continue to navigate the challenges and opportunities in this field, the studies presented in this Research Topic open new avenues for future innovations in drug design and therapeutic interventions.

We hope that this Research Topic of articles inspires further research and collaboration, ultimately leading to the development of more effective and personalized therapeutic strategies.

## Author contributions

DR: Writing–review and editing, Writing–original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal Analysis, Data curation, Conceptualization. RP: Writing–review and editing, Visualization, Investigation, Data curation, Conceptualization. CL: Writing–review and editing, Visualization, Investigation, Formal Analysis, Conceptualization. JM-F: Writing–review and editing, Visualization, Investigation, Formal Analysis, Conceptualization.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by Agencia Nacional de Investigación y Desarrollo (ANID) grant N° BASAL FB210008 and ACT210012, as well as Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) grant N° 1220656.

## Acknowledgments

The editors extend their gratitude to all the authors, reviewers, and Frontiers in Pharmacology Editorial team who have contributed to this Research Topic. Their dedication and insights are instrumental in advancing the field of pharmacoinformatics.

## 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.



## OPEN ACCESS

## EDITED BY

Linlin Shen,  
Shenzhen University, China

## REVIEWED BY

Qiang Yan,  
Taiyuan University of Technology, China  
Xinshao Wang,  
Zenith AI/University of Oxford,  
United Kingdom

## \*CORRESPONDENCE

Chunlai Feng,  
✉ feng@ujs.edu.cn  
Mengjie Rui,  
✉ mjru@ujs.edu.cn

<sup>†</sup>These authors have contributed equally to this work

## SPECIALTY SECTION

This article was submitted to  
Translational Pharmacology,  
a section of the journal  
Frontiers in Pharmacology

RECEIVED 31 August 2022

ACCEPTED 05 December 2022

PUBLISHED 15 December 2022

## CITATION

She S, Chen H, Ji W, Sun M, Cheng J,  
Rui M and Feng C (2022), Deep learning-  
based multi-drug synergy prediction  
model for individually tailored anti-  
cancer therapies.  
*Front. Pharmacol.* 13:1032875.  
doi: 10.3389/fphar.2022.1032875

## COPYRIGHT

© 2022 She, Chen, Ji, Sun, Cheng, Rui  
and Feng. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)  
(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.

# Deep learning-based multi-drug synergy prediction model for individually tailored anti-cancer therapies

Shengnan She<sup>†</sup>, Hengwei Chen<sup>†</sup>, Wei Ji, Mengqiu Sun,  
Jiaxi Cheng, Mengjie Rui\* and Chunlai Feng\*

Department of Pharmaceutics, School of Pharmacy, Jiangsu University, Zhenjiang, China

While synergistic drug combinations are more effective at fighting tumors with complex pathophysiology, preference compensating mechanisms, and drug resistance, the identification of novel synergistic drug combinations, especially complex higher-order combinations, remains challenging due to the size of combination space. Even though certain computational methods have been used to identify synergistic drug combinations *in lieu* of traditional *in vitro* and *in vivo* screening tests, the majority of previously published work has focused on predicting synergistic drug pairs for specific types of cancer and paid little attention to the sophisticated high-order combinations. The main objective of this study is to develop a deep learning-based approach that integrated multi-omics data to predict novel synergistic multi-drug combinations (DeepMDS) in a given cell line. To develop this approach, we firstly created a dataset comprising of gene expression profiles of cancer cell lines, target information of anti-cancer drugs, and drug response against a large variety of cancer cell lines. Based on the principle of a fully connected feed forward Deep Neural Network, the proposed model was constructed using this dataset, which achieved a high performance with a Mean Square Error (MSE) of 2.50 and a Root Mean Squared Error (RMSE) of 1.58 in the regression task, and gave the best classification accuracy of 0.94, an area under the Receiver Operating Characteristic curve (AUC) of 0.97, a sensitivity of 0.95, and a specificity of 0.93. Furthermore, we utilized three breast cancer cell subtypes (MCF-7, MDA-MD-468 and MDA-MB-231) and one lung cancer cell line A549 to validate the predicted results of our model, showing that the predicted top-ranked multi-drug combinations had superior anti-cancer effects to other combinations, particularly those that were widely used in clinical treatment. Our model has the potential to increase the practicality of expanding the drug combinational space and to leverage its capacity to prioritize the most effective multi-drug combinational therapy for precision oncology applications.

## KEYWORDS

anti-cancer combination therapy, high-order drug combinations, cancer cell subtype-specific models, deep learning framework, precision oncology

## 1 Introduction

Various carcinogenic factors and pathogenesis have been linked to cancer, which has been identified as a collection of complex diseases (Tolomeo and Simoni, 2002). This complicates the application of a single treatment for a single target, as it activates redundant activities in cancer cells such as various downstream factors and parallel pathways due to compensatory mechanisms (Alexander and Friedl, 2012). Inter-tumor and intra-tumor heterogeneity are a major contributor to drug resistance and disease progression in clinical cancer treatment, ultimately leading to disease relapse (Holohan et al., 2013). Combination therapy has been shown to be a well-established and superior solution to these problems because of its improved clinical efficacy and lack of development of drug resistance. Since the dose of each drug is smaller than what is used in monotherapy, it is possible that the side effects will be minimized (Mahase, 2019). So far, significant efforts have been undertaken to systematically evaluate the synergistic combinations from a large pool of chemical compounds (MacGowan et al., 1990; Wiesner et al., 2002; Sopirala et al., 2010). Finding successful drug combinations is still incredibly difficult, especially with today's high-throughput screening technologies (Sun et al., 2013). Furthermore, high-order combinations have the potential to regulate biological systems more powerfully than drug pairs because they favor compensatory mechanisms, which tumors greatly exploit; however, the number of experiments run to identify promising high-order combinations would explode by several orders of magnitude, which is far beyond the current exploration ability. There is a pressing need for systemic methodologies, and an urgent need to make it feasible to find new therapeutic combinations of more than two agents, including synthetic chemicals, biological molecules and natural products.

An extensive range of computational methods spanning a large area of methodologies (Bansal et al., 2014; Gayvert et al., 2017; Chen et al., 2018; Huang et al., 2019) has tremendously aided research into anti-cancer drug combinations in the recent years. Different machine learning models and the burgeoning field of deep learning are examples of possible approaches. A machine learning based classification model could extract features from multiple drug profiles including drug targeted proteins and Anatomical Therapeutic Chemical Classification System (ATC) codes, and, as a result, it enabled the prediction of potential synergistic drug pairs (Iwata et al., 2015). But ATC code is available only for marketed drugs, suggesting that the processing of uncharacterized drugs or new candidate compounds is considerably beyond the power of this approach. In another method, two machine learning algorithms, random forest (RF) and extreme gradient boosting (XGBoost), were applied to establish models for drug combination prediction, indicating that XGBoost resulted in a better perform than the RF model (Sidorov et al., 2019). As

trained on a pre-cell line, these two models should be rebuilt when applying for another cell line. Recent impressive breakthroughs of deep neural networks, which profit from the explosion of big data and the ability to automatically extract key features, have produced greatly enhanced performance in biomedical research. A deep learning approach, DeepSynergy, proposed by Preuer et al., integrated the chemical descriptors of drugs and genomic data of cell lines of interest for predicting synergistic drug combinations (Preuer et al., 2018). Following this, numerous techniques based on deep learning framework, such as AuDNNsynergy (Zhang T. et al., 2021), MatchMaker (Kuru et al., 2021) and Deep Signaling Synergy (Zhang H. et al., 2021), have been suggested with multi-omics data to prioritize drug combinations, revealing their benefits on the prediction of paired drug combination. However, the existing deep learning models mainly focused on predicting drug pairs which might not be efficient to inhibit the aggressive growth of tumors driven by complex mechanisms (Holohan et al., 2013; Dry et al., 2016).

With the approval of multi-drug combinations for a variety of diseases such as cancers and *tuberculosis* (Gotwals et al., 2017; Davies et al., 2019), the focus of the search for combinational therapies has shifted partially away from pairwise combinations and toward high-order ones containing three or more drugs. Yet there are limited tools to predict multi-drug synergy in diseases. A recent web application, Synergy Finder 2.0, is developed to analyze the drug combination screen data and provide the best multi-drug synergy patterns (Ianevski et al., 2020). However, this tool is based on the dose-response data collected by a huge number of multi-drug screening activities, which make it infeasible to find prospective high-order combinations in a labor- and time-saving manner. So far, we lack deep learning-based approaches to predict the synergy of high-order combinations by integrating multi-omics data, and this is a problem.

The methodological advances of deep learning-based models have made it easier to investigate the best possible high-order combinations within the defined disease module. In this study, we developed a deep learning-based model for the prediction of synergistic multi-drug combinations (DeepMDS) through using a large-scale dataset that integrated by targets information, drug response data and large-scale genomic profile of cancer cell lines from varied tissues. DeepMDS can generate predicted pseudo-IC50 values, which can be used to quantify and, by extension, rank the synergistic anti-cancer effect of drug combinations. As a comparison, we used some of the most advanced machine learning algorithms as reference models, including K Nearest Neighbor (KNN), Random Forest (RF), Support Vector Machine (SVM) and Gradient Boosting Machine (GBM). These algorithms have all been succeeded in modeling drug synergy and were among the top winning methods of the 2019 AstraZeneca-Sanger drug combination prediction DREAM Challenge (Menden et al., 2019). More importantly, the performance of our DeepMDS were further extensively

validated by published literatures and rigorous studies based on biologically heterogeneous breast cancer cell subtypes (MCF-7, MDA-MD-468 and MDA-MB-231) as well as lung cancer cell line A549.

## 2 Materials and methods

### 2.1 Data collection

In this work, we collected, pre-processed, and combined gene expression profiles of cancer cell lines and target information of anti-cancer drugs to generate modeling dataset. Then, data on drug response against a large variety of cancer cell lines were also collected for the purpose of labeling modeling samples. Herein, the precise process of datasets construction was described in detail in this section.

#### 2.1.1 Gene expression features

Based on Affymetrix Human Genome U219 Array plates, basal gene expression profiles of 1,000 human cancer cell lines were measured and identified utilizing a wide variety of anti-cancer therapeutics in the Genomics of Drug Sensitivity in Cancer (GDSC) project (Iorio et al., 2016). The gene expression data of cancer cell lines were demonstrated to be useful information, which faithfully recapitulated cancer-driven alterations in 11,289 tumors from 29 tissues. Meanwhile, many of the genomic information were highly associated with drug sensitivity or resistance and thus it could be efficiently applied to predict drug response as sample features. The public available transcriptional profiles of 1,000 human cancer cell lines were carefully retrieved from the ArrayExpress database (Parkinson et al., 2005) and then the data pre-processing was conducted based on the platform R v3.5.0. To begin, oligo-package was applied to convert the downloaded raw data (CEL files) into standard genomic profiles. Then missing and invalid values were filled and replaced using the impute 1.52.0 package from Bioconductor Library (Gentleman et al., 2004). In further, Robust Multichip Average (RMA) algorithm was used to normalize the refilled datasets, preventing erroneous results generated by maxima and minima as well as decreasing computing burden. Next, based on the annotation file of gene chip, each probe ID was matched with its corresponding gene symbol and the mean expression value of the multiple probe IDs matched the same official gene symbol was computed to reflect the expression intensity. A phenomenon known as the “curse of dimensionality” may cause prediction models to perform poorly due to the large number of genes covered by the expression profiles (Aliper et al., 2016). To avoid this difficulty, genes in cancer-related pathways were selected to lower the size of gene expression features. In practice, 14 gene sets, which were defined by cBioPortal, consisted of cancer-related pathways (Cerami et al., 2012), such as DNA damage response or RTK signaling

pathways (Jeon et al., 2018). Finally, a total number of 215 genes were selected as genomic features and their corresponding gene expression data were used as the feature representations of cancer cell lines (Supplementary Data Sheet S1).

#### 2.1.2 Target information

Along with gene expression features, this study gathered information on the targets of anti-cancer drugs. To begin, we obtained target information for 265 chemical compounds from DrugBank (Wishart et al., 2018) and PubChem (Wang et al., 2009). This information was merged with determined drug sensitivity of cancer cell lines from the GDSC project. On the other hand, 1,574 naturally occurring anticancer compounds were obtained from the Naturally occurring Plant based Anticancerous Compound-Activity-Target DataBase (NPACT), and the related target information for each compound was retrieved from TCMSP (Ru et al., 2014), DrugBank and PubChem. Finally, a total of 1,093 targets were obtained as target features of compounds. The target information of each compound was used to generate the feature representation of the compound. More specifically, the target feature values corresponding to the targets of the compound were encoded as “1” and the others were encoded as “0” (Supplementary Data Sheet S2).

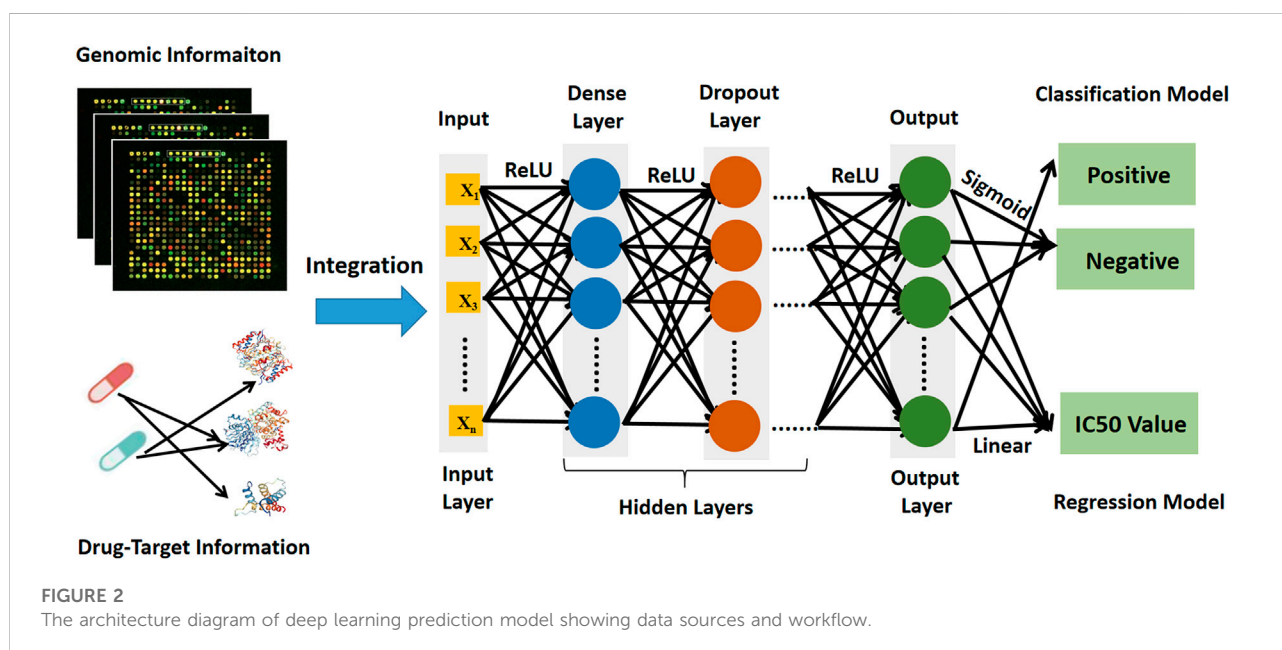
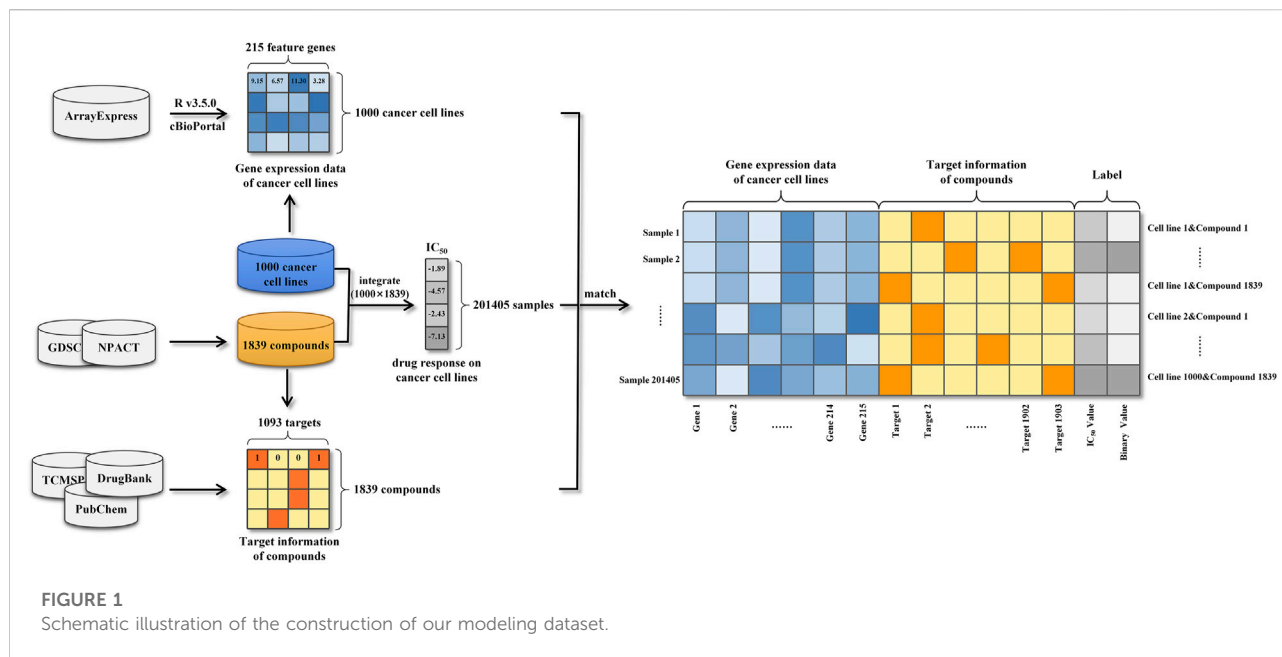
#### 2.1.3 Drug response information

Drug response information, also called as monotherapy information, assessed drug effects on cell lines and was used in this study to label samples. The GDSC project experimentally determined and quantified the drug responses of over 265 chemical compounds to 1,000 cancer cell lines using the half maximum inhibitory concentration (IC50) (Iorio et al., 2016). Additionally, we gathered equivalent data for 1,574 natural chemicals in response to distinct cell lines from NPACT, PubChem and related literatures. In total, the drug responses of 201,405 drug-cancer cell line pairs were collected and used as the labels (IC50 in the regression task and binary value in the classification task) (Supplementary Data Sheet S3).

#### 2.1.4 Data integration

Gene expression profiles of cancer cell lines, target information of anti-cancer compounds and drug responses against a large variety of cancer cell lines were integrated into 201405 modeling samples (Figure 1). Specifically, each sample was represented as a vector consisting of a 215-dimensional genomic feature representation of cancer cell line and a 1093-dimensional target feature representation of compound. Following that, the drug response was used to label the sample. Due to the considerable dimension disparity between gene expression features and target information, all samples' data were adjusted using zero-centered processing and normalized square deviation.





## 2.2 Model construction

Among the processed datasets, 80% (161,124) of samples were randomly chosen for the training dataset, while 20% (40,281) were used as the test dataset. Then, using the training and test datasets, a deep learning prediction model and other models based on various machine learning algorithms were constructed and optimized, and their performances were compared.

### 2.2.1 Deep learning prediction model

The deep learning prediction model (DeepMDS) was built sequentially in Python (version 3.6) using the Keras platform, which is a high-level neural networks API running on top of Theano (Feng et al., 2019). The basic architecture of deep learning models was illustrated in Figure 2. To begin, gene expression data from cell lines and target information of drugs as input were loaded in the nodes (also called neurons) of the input layer. Then the loaded information from input layer

was propagated through the neighboring hidden layers, including the dense layer and the dropout layer. Finally, the output layer could provide the predicted IC50 values for each sample. To address sophisticated regression problems, each layer among the deep learning architecture was followed by non-linear activation functions (Feng et al., 2019). The Rectified Linear Unit (ReLU) activation function was used to activate the input layer and hidden layer in this study because it has the capacity to reduce the vanishing gradient problem and has a rapid computing speed (Eq. 1) (Feng et al., 2019). Then for the output layer, a linear activation function was applied in the regression model to fit the distribution of predicted IC50 values better (Eq. 2). Meanwhile, the classification model was developed using the deep learning architecture, which enables a similar assessment of model performance. The construction of classification model constructed in the same manner as stated previously, except that the Sigmoid activation function (Eq. 3) was applied to produce the classification labels in the output layer (Eq. 3). Here, the samples labeled with IC50 values  $\leq 10$  nM were considered positive samples, whereas those labeled with IC50 values  $> 10$  nM were considered negative samples.

$$y = \text{ReLU}(Wx + b) \quad (1)$$

where  $y$  was the activation value of the hidden layer,  $x$  was the input data,  $W$  was weight matrix and  $b$  was bias.

$$z = \text{linear}(W'y + b') \quad (2)$$

where  $z$  was the predicted IC50 values,  $y$  was the activation value of the hidden layer,  $W'$  was transposed weight matrix and  $b'$  was transposed bias.

$$z = \text{sigmoid}(W'y + b') \quad (3)$$

where  $z$  was the classification labels,  $y$  was the activation value of the hidden layer,  $W'$  was transposed weight matrix and  $b'$  was transposed bias.

In order to train the model, the loss functions of MSE (mean square error) and binary cross-entropy were used to estimate performance of regression and classification models, respectively, by comparing the difference between the actual label of input data in input layer ( $x$ ) and the predicted label of output layer ( $z$ ), where SGD (stochastic gradient descent) was applied to search the optimal parameters (Eq. 4).

$$L_H(x, z) = -\sum_{k=1}^d [x_k \log z_k + (1 - x_k) \log (1 - z_k)] \quad (4)$$

where  $x$  was the actual value of input data in input layer,  $z$  was the predicted value of output layer,  $d$  was the epoch number.

In addition, Adam (adaptive moment estimation) and RMSprop (Root Mean Square prop) were selected as optimization functions for the construction of regression and classification models, respectively. Throughout the training process, the aforementioned processes were repeated in order to update the weights and bias until the optimal weight matrix  $W$  and bias  $b$  were obtained.

### 2.2.1.1 Optimization of deep learning prediction model

The performance of a deep learning prediction model is determined not only by its architecture of deep learning but also by its hyper parameters. Traditionally, the ideal parameter combination for a deep learning model was established by human experience, which was neither accurate nor objective. To obtain the optimal DeepMDS, a grid search algorithm was used to find the best combination from a parameter space including epoch number, batch size, learning rate, dropout rate and hidden units of hidden layers. Finally, using the same datasets, 5,625 ( $5 \times 5 \times 3 \times 3 \times 5 \times 5$ ) regression and classification models were developed individually to seek their own optimal parameter combinations using 10-fold cross validation. According to the optimization results, the conic architecture with two hidden layers having 200 nodes in the first layer and 100 nodes in the second layer was the optimal regression and classification model.

Also, a big dropout rate of 0.5 followed behind each dense layer to avoid the overfitting problems. Furthermore, a smaller learning rate of  $10^{-5}$ , a batch size of 128 and an epoch number of 200 were set up to constitute the optimal regression model. Meanwhile, a learning rate of  $10^{-3}$ , a batch size of 32 and an epoch number of 500 were chosen for the best classification model (Supplementary Table S2, Supplementary Data Sheet D4).

## 2.2.2 Model evaluation and comparison

To compare the performance of deep learning model to that of other models based on state-of-the-art machine learning algorithms, the same datasets were used to develop a  $k$  nearest neighbor (KNN) model, a random forest (RF) model, a support vector machine (SVM) model and a gradient boosting machine (GBM) model. Also, each model was allowed to optimize hyper parameters using a grid search algorithm and cross validation.

### 2.2.2.1 K nearest neighbor model

The variable selection  $k$  nearest neighbor (KNN) algorithm was applied to develop the prediction model based on Python (version 3.6). Regarding hyper parameter setting, number of neighbors, types of weight functions and algorithms were tuned to achieve the optimal KNN model. Following a grid search in a value space of considered parameters, the optimal parameters for the KNN regression model were 6 neighbors, a “uniform” weight function and a “auto” algorithm. In addition, for the KNN classification model, the optimal model consisted of 5 neighbors, a “uniform” weight function, and a “auto” algorithm (Supplementary Table S3, Supplementary Data Sheet S4).

### 2.2.2.2 Random forests model

Based on random forest (RF) algorithm and Bagging architecture, Random Forest Regressor and Random Forest Classifier functions were used to develop RF regression and classification models using Python (version 3.6) respectively.



In terms of hyper parameter setting, the number of features considered in each split, the number of estimators (trees), and the minimal number of leaved samples were all adjusted. As a consequence, the RF regression model's optimized parameters were 200 estimators, 'auto' for features considered, and a min\_samples\_leaf of 50. The best settings for the RF classification model were set at 100 estimators, 'auto' for features considered, and a min\_samples\_leaf of 10 (Supplementary Table S4, Supplementary Data Sheet S4).

### 2.2.2.3 Support vector machine model

Based on Support Vector Machine (SVM) algorithm, Support Vector Regression (SVR) and Support Vector Classification (SVC) functions were applied to develop SVM regression and classification models using Python (version 3.6) respectively. During the process of hyper parameter setting, the type of kernel function, penalty factor C and gamma were tuned to achieve the optimal SVM model. According to the optimization results, the optimal SVM regression model was determined to be the RBF kernel function, a penalty factor C of 10 and a gamma of 0.01. Then, for the SVM classification model, the optimal parameters were determined to be the RBF kernel function, a penalty factor C of 1 and a gamma of 0.1 (Supplementary Material S5, Supplementary Data Sheet S4).

### 2.2.2.4 Gradient boosting machine model

Based on Gradient Boosting Machine (GBM) algorithm and Boosting architecture, Gradient Boosting Regressor and Gradient Boosting Classifier functions were applied to construct GBM regression and classification models *via* Python (version 3.6), respectively. When setting the hyper parameters, number of trees, learning rates, number of features in each split, min\_samples\_split and min\_samples\_leaf were took into consideration. According to the optimization results, the optimal GBM regression model consisted of 500 estimators, a min\_samples\_split of 1,000, a learning rate of 0.01, and a min\_samples\_leaf of 60. Also, the optimal parameters for the GBM classification model were then adjusted as 200 estimators, a min\_samples\_split of 600, a learning rate of 0.01, and a min\_samples\_leaf of 60 (Supplementary Material Table S6, Supplementary Material Data D4).

## 2.2.3 Performance metrics

In order to assess and compare the performances of above optimized prediction models, the mean square error (MSE, Eq. 5), the root mean square error (RMSE, Eq. 6) and R-Square ( $R^2$  score, Eq. 7) were used as metrics to evaluate their ability to predict IC50 values of drug combinations in the regression task. Meanwhile, the standard criteria for classification work including Sensitivity (SEN, Eq. 8), Specificity (SPE, Eq. 9), Accuracy (ACC, Eq. 10) and Matthews correlation coefficient (MCC, Eq. 11) were also applied to evaluate model performance for the classification task.

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_{true}^{(i)} - y_{pre}^{(i)})^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_{true}^{(i)} - y_{pre}^{(i)})^2} \quad (6)$$

$$R^2\_score = 1 - \frac{\frac{1}{m} \sum_{i=1}^m (y_{true}^{(i)} - y_{pre}^{(i)})^2}{\frac{1}{m} \sum_{i=1}^m (y_{true}^{(i)} - \bar{y})^2} \quad (7)$$

where  $y_{true}$  was the actual values of samples,  $y_{pre}$  was the predicted values of samples,  $m$  was the number of samples.

$$SEN = \frac{TP}{TP + FN} \quad (8)$$

$$SPE = \frac{TN}{FP + TN} \quad (9)$$

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \quad (10)$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FN) \cdot (TP + FP) \cdot (TN + FN) \cdot (TN + FP)}} \quad (11)$$

where TP meant true positive; TN meant true negative; FP meant false positive; FN meant false negative.

Furthermore, the area under the Receiver Operating Characteristic (ROC) curve (AUC) was also used to evaluate the model performance for the classification task. Specifically, the best possible prediction was 100% sensitivity and 100% specificity with area under the curve (AUC) of 1, while an AUC value of  $\leq 0.5$  represented random selection.

## 2.2 Prediction and validation with literature synergy data

To further verify the performance of constructed DeepMDS model built above, literature validation was carried out. Sun's work (Sun et al., 2015) rated 17 drug pairs comprised of 12 single agents (sorafenib, erlotinib, gefitinib, tamoxifen, everolimus, dasatinib, sunitinib, BIBW-2992, thalidomide, PD98059, flavopiridol and toremifene) based on their RACS model-predicted synergy. Meanwhile, to confirm the predicted results, each drug pair was experimentally tested at four different concentration ratios (4:1, 3:2, 2:3, and 1:4) using MCF-7 cell line. The synergistic effect of these 17 drug pairs was also predicted and ranked by DeepMDS using target information and gene expression data of MCF-7 cell line. DeepMDS's predicted results were then compared to experimental results from the literature to determine the model's performance.

## 2.3 Prediction and validation by *in vitro* cellular experiments

To further evaluate the capability of DeepMDS to predict the synergy effect of multi-drug combinations, seven recommended chemotherapy drugs (docetaxel, paclitaxel, doxorubicin,

epirubicin, gemcitabine, 5-fluorouracil, and methotrexate) from breast cancer clinical treatment guidelines (Telli and Carlson, 2009) were randomly grouped to generate drug combinations, including drug pairs and high-order combinations. Following that, the synergy effect of drug combinations was then predicted using DeepMDS and evaluated using an *in vitro* cell viability assay. In brief, 120 drug combinations were constructed using seven chemotherapeutic agents (II2-II28 indicated two-drug combinations, III10-III56 indicated three-drug combinations, IV16-IV70 indicated four-drug combinations, V21-V56 indicated five-drug combinations, VI16-VI28 indicated six-drug combinations, VII7 indicated seven-drug combinations).

Following the collection of target information for each medication from GDSC, PubChem, and DrugBank, the datasets were pre-processed to construct prediction samples. To examine the synergistic effect of the aforementioned drug combinations, three distinct subtypes of breast cancer cell lines were used: MCF-7, MDA-MB-468, and MDA-MB-231. Furthermore, to validate DeepMDS's robustness and applicability, this model was used to predict another cancer cell line A549 from lung tissue. Each cell line's gene expression data were analyzed and then utilized to construct prediction samples. Finally, DeepMDS was used to predict the sample datasets. For each cell line, the optimized DeepMDS model predicted and ranked the IC<sub>50</sub> values of 120 drug combinations.

The corresponding validation experiments were carried out *in vitro*. MCF-7, MDA-MB-468, MDA-MB-231, and A549 cell lines were obtained from the Cell Bank of Type Culture Collection of Chinese Academy of Sciences (CBTCCAS). Four cancer cells were cultured in DMEM medium supplemented with 10% fetal bovine serum, and kept at 37°C and 5% CO<sub>2</sub> in a humidified incubator. Docetaxel, paclitaxel, doxorubicin, epirubicin, gemcitabine, 5-Fluorouracil and methotrexate were purchased from Meryer (Shanghai, China), and the purity of each drug (compound) is above 98%. Each drug (compound) was dissolved in DMEM medium and then used alone or in combination with other drugs at various concentration ratios so that we could ensure each drug attained its best synergistical ratio throughout a wide concentration range (Table 1). Then, exponentially growing cells were seeded in 96-well plates at a density of 5×10<sup>3</sup> per well and cultured for 24 h.

Afterward, the cells were then treated for 72 h with a variety of single drugs or multi-drug combinations at a series of diluted concentrations. There are three replicates for each measurement, and the cytotoxicities of individual drugs or combinations were determined using the cell counting kit-8 (CCK-8) assay. IC<sub>50</sub> values for each sample was calculated in line with the manufacturer's instructions. In addition, the combination index (CI) (Chou and Talalay, 1984) was calculated using the CompuSyn software (Chou and Martin, 2007), and then CI

values were applied to define and quantify the synergistic effect of each drug combination. In general, a drug combination is synergistic if the CI value is less than 0.9, additive if the CI value is between 0.9 and 1.1, and antagonistic if the CI value is greater than 1.1 (Sun et al., 2016). In this study, a drug combination was considered synergistic if the CI values for all concentration ratios were all less than 0.9.

## 2.4 Pathway enrichment analysis of drug combinations

To explore the synergistic mechanism of predicted combinations in given cell lines, KEGG pathway enrichment analysis was performed on the specific feature genes of cancer cell lines and the target information of drug combinations, and the pathways of synergistic combinations against different cancer cell lines were compared.

## 3 Results

### 3.1 Overview of DeepMDS model

Here, we present DeepMDS, a Deep Neural Network (DNN)-based methodology for the prediction of the pseudo-IC<sub>50</sub> values of a series of drug combinations in a given cell line. Figure 2 illustrates the framework of the DeepMDS, which contains two main features: 1) identification of top-ranked drug combinations from a pool of drug pairs and combinations of three or more compounds, that is, high-order combinations, and 2) cancer cell line-specific prediction by integrating gene expression profile, target information of drugs, and drug responses. In other words, DeepMDS not only allows us to predict the most potent combination, but it also allows us to deliver the best prospective combination susceptible to a specific molecular subtype of cancer cells, which mimics the way that precision medicine is utilized in clinical trials.

### 3.2 Model comparison

We first validated our DeepMDS using the test dataset and compared it to four other machine learning-based methods (Table 2 and Table 3). In terms of performance metrics, regardless of whether the regression task is used to predict the pseudo-IC<sub>50</sub> values or the classification task is used to identify positive results, it is clear that our deep learning model outperformed those developed using traditional machine learning algorithms. In specific, DeepMDS achieved a test MSE of 2.50 in the regression task, while GBM, SVM, RF and KNN models performed poorly with MSEs of 5.75, 8.66,

**TABLE 1 The settings of concentration ratios for different drug combinations.**

The number of drugs in a combination	I	II	III	IV	V	VI
Two	1:1	2:1	1:2	—	—	—
Three	1:1:1	2:1:1	1:2:1	1:1:2	—	—
Four	1:1:1:1	2:1:1:1	1:2:1:1	1:1:2:1	1:1:1:2	—
Five	1:1:1:1:1	2:1:1:1:1	1:2:1:1:1	1:1:2:1:1	1:1:1:2:1	1:1:1:1:2

Note: roman numerals, including I, II, III, IV, V, and VI, indicated different drug molar ratios in a drug combination.

**TABLE 2 Model performances of prediction models for regression task.**

Model	MSE	RMSE	$R^2\_score$
DeepMDS	2.50	1.58	0.86
GBM	5.75	2.40	0.81
SVM	8.66	2.94	0.75
RF	13.11	3.62	0.72
KNN	16.73	4.09	0.67
DeepSynergy	255.49	15.91	0.73

Note: The columns showed mean square error (MSE), root mean square error (RMSE) and R-Square ( $R^2\_score$ ).

**TABLE 3 Model performances of prediction models for classification task.**

Model	SEN	SPE	MCC	ACC	AUC
DeepMDS	0.95	0.93	0.88	0.94	0.97
GBM	0.87	0.85	0.72	0.86	0.92
SVM	0.81	0.85	0.66	0.83	0.89
RF	0.74	0.82	0.56	0.78	0.83
KNN	0.75	0.71	0.46	0.73	0.76
DeepSynergy	0.57	0.95	NA	0.92	0.90

Note: The columns showed sensitivity (SEN), specificity (SPE), Matthews correlation coefficient (MCC), accuracy (ACC), and the performance measures area under ROC, curve (AUC). "NA" indicated that no MCC, data was provided in literature.

13.11 and 16.73, respectively. Along with MSE, two more evaluation metrics, RMSE and  $R^2\_score$ , showed a similar trend. It is worth mentioning that the square root of  $R^2\_score$  equals the Pearson correlation coefficient in this case, as  $R^2\_score$  was used to determine the linear correlation between predicted and actual values in this regression task. In the classification challenge, DeepMDS also outperformed the competition, increasing the ACC to 0.94 and the AUC to 0.97, while the second-best approach, the GBM model, achieved an ACC of 0.86 and an AUC of 0.92.

Additionally, we compared the performance of DeepMDS to that of DeepSynergy, a deep learning-based model for predicting synergy in a given cell line. DeepSynergy achieved an ACC of 0.92 and an AUC of 0.90 for classification, and an

MSE of 255.49 and an RMSE of 15.91 for regression. As shown in Table 2 and 3, our DeepMDS still performed well. Also, DeepMDS achieved a SEN of 0.95 and a SPE of 0.93 for the classification task, compared to 0.57 and 0.95 for DeepSynergy. Moreover, we compared the performance of DeepMDS against other deep learning-based methods. DeepMDS predictions showed a significant correlation with actual combination viabilities (Pearson's  $r = 0.93$ , Supplementary Table S1), outperforming other four models developed in the last 2 years. These findings demonstrated that the strength of our deep learning-based model, which was able to achieve steady and robust model performance in both regression and classification tasks, as well as superior accuracy in drug synergy prediction.

RACS Ranking	Drug 1	Drug 2	The combination index of drug combinations				DeepMDS Ranking
			4:1	3:2	2:3	1:4	
10	Sorafenib	Dasatinib	0.62±0.04	0.54±0.02	0.76±0.06	0.75±0.01	1
3	Gefitinib	Tamoxifen	0.68±0.03	0.22±0.05	0.25±0.03	0.86±0.01	2
6	Sorafenib	Tamoxifen	0.71±0.04	0.67±0.04	0.30±0.07	0.22±0.04	3
7	Gefitinib	Toremifene	0.89±0.01	0.55±0.12	0.75±0.07	0.76±0.02	4
8	Erlotinib	Sorafenib	0.61±0.09	0.55±0.05	0.28±0.01	0.21±0.03	5
17	Erlotinib	Sunitinib	0.54±0.02	0.77±0.05	1.09±0.16	0.73±0.08	6
1	Gefitinib	Everolimus	0.67±0.03	0.73±0.02	0.74±0.05	0.66±0.04	7
9	Gefitinib	BIBW-2992	1.04±0.02	0.76±0.01	0.89±0.05	0.93±0.03	8
4	Erlotinib	Tamoxifen	0.36±0.08	0.23±0.06	0.45±0.04	0.63±0.02	9
2	Gefitinib	Thalidomide	0.82±0.01	0.89±0.03	0.77±0.05	0.86±0.02	10
12	Gefitinib	PD98059	0.61±0.05	0.73±0.01	0.80±0.04	0.89±0.02	11
14	Tamoxifen	Flavopiridol	3.32±0.15	0.86±0.02	2.60±0.05	0.92±0.01	12
16	Gefitinib	Erlotinib	1.18±0.04	0.97±0.02	0.96±0.05	0.96±0.02	13
5	Gefitinib	Sorafenib	1.09±0.05	0.44±0.14	0.85±0.08	0.65±0.06	14
13	Everolimus	BIBW-2992	3.47±0.07	0.78±0.07	1.31±0.02	0.94±0.04	15
11	Sorafenib	Everolimus	0.82±0.02	1.05±0.10	0.98±0.04	1.23±0.09	16
15	Erlotinib	Flavopiridol	0.95±0.02	0.93±0.04	0.84±0.06	1.46±0.02	17

**FIGURE 3**  
The comparison results between DeepMDS and literature. The synergy effect of each drug pair was retrieved from literature (Sun et al., 2015), and described using combination index (CI). The left ranking was predicted using RACS model, validated by *in vitro* experiments on MCF-7 (Sun et al., 2015). Dark green indicated strong synergy (CI < 0.3); pale green indicated synergy (0.3 < CI < 0.9); yellow indicated additive (0.9 < CI < 1.1); and red indicated antagonism (CI > 1.1). The different CI values of each drug pair were calculated at four dual-drug ratios, including 4:1, 3:2, 2:3, and 1:4.

### 3.3 Literature validation

To verify our DeepMDS’s predictive power, we first focused on previously published drug combinations, the majority of which were paired combinations. Seventeen drug pairs and twelve single agents were predicted using DeepMDS and were shown to be consistent with published literature (Sun et al., 2015) (Supplementary Data Sheet S5). Notably, the output layer of the DeepMDS was the predicted pseudo-IC50 value for each combination, which did not represent the actual therapeutic efficacy but was used to rank the therapeutic efficacies of multi-drug combinations. Here we confined the predicted outcomes to pairwise drug combinations and ranked 17 drug pairs according to their increasing pseudo-IC50 values, followed by a comparison to experimental data from the literature (Figure 3).

Four of the seventeen drug pairs in the reference data were validated as having significant synergistic antitumor effects at optimal dose ratios (Sun et al., 2015), as seen by the dark green coloration in Figure 3. DeepMDS re-ranked these drug combinations, revealing that three highly synergistic couples were correctly predicted in the top five combinations. Sorafenib and dasatinib and gefitinib and toremifene, the next two most effective medication combinations, also revealed

synergistic mechanisms at all drug ratios. Notably, the bottom-ranked combination was verified to exhibit additive or even antagonistic effects, as predicted by DeepMDS. Collectively, the ranking of pairwise combinations predicted by DeepMDS was largely comparable with experimental data from the literature (Sun et al., 2015), demonstrating our model’s adeptness at filtering and enriching synergistic medication combinations.

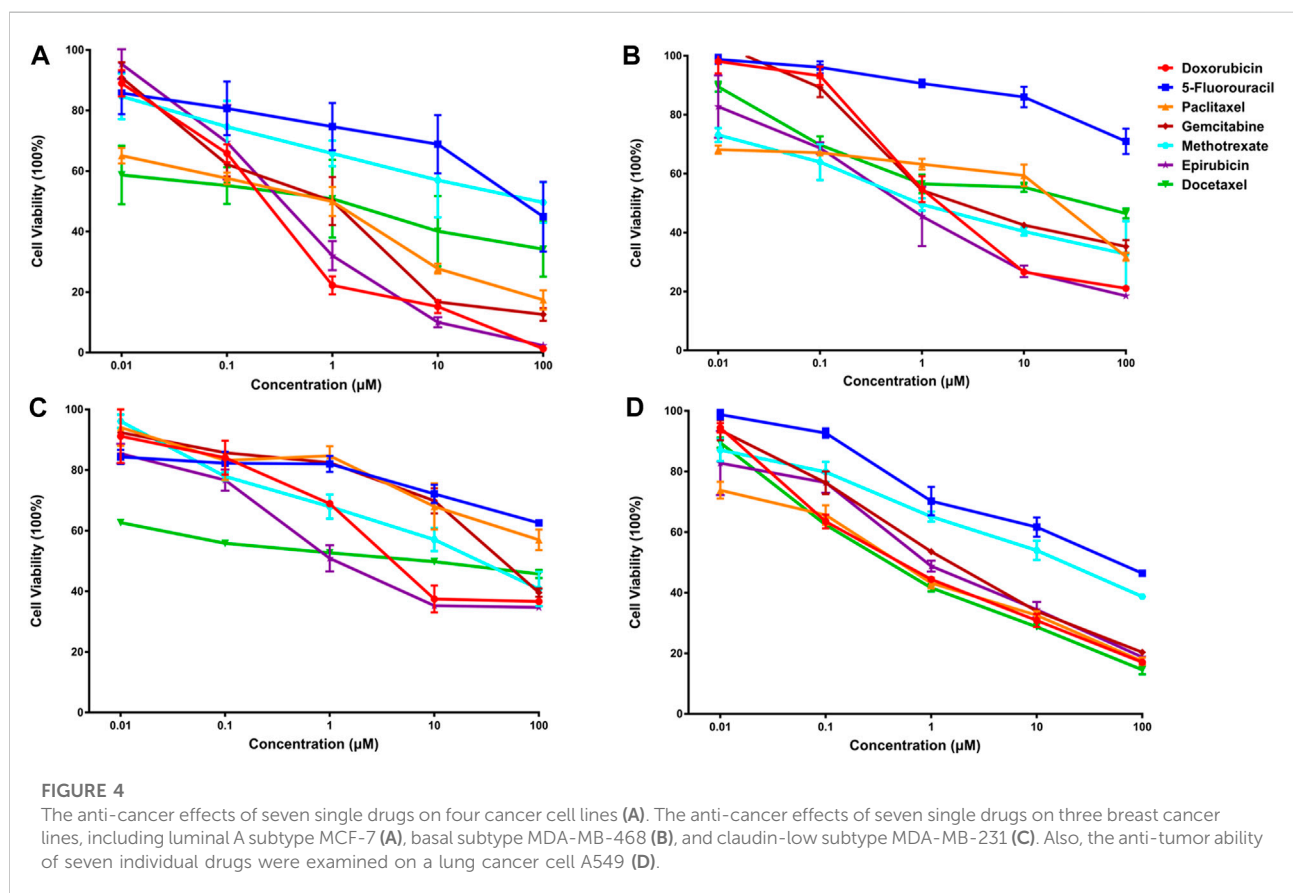
### 3.4 De novo prediction of multi-drug combinations for specific cancer cell lines

To further explore DeepMDS’s ability to predict novel high-order combinations, we chose seven anticancer drugs that have been approved by the FDA for breast cancer (National Comprehensive Cancer Network, 2021). These drugs were randomly assigned into 120 combinations, ranging from simple drug pairs to more sophisticated three- or more-drug combinations. The anticancer activity of these combinations was then predicted using our DeepMDS on four cancer cell lines, followed by in-house experimental validation.

Regarding the heterogeneous biological markers of breast cancer cell lines, we chose three representative subtypes: MCF-7

TABLE 4 The top three predicted combinations for a variety of cancer cell lines.

Predicted ranking	MCF-7	MDA-MB-468	MDA-MB-231	A549
1	III12 (doxorubicin, docetaxel, and gemcitabine)	III12 (doxorubicin, docetaxel and gemcitabine)	IV33 (doxorubicin, gemcitabine, methotrexate, and paclitaxel)	II28 (epirubicin and paclitaxel)
2	III7 (doxorubicin, 5-Fluorouracil, and docetaxel)	II3 (doxorubicin and docetaxel)	V32 (doxorubicin, docetaxel, gemcitabine, methotrexate, and paclitaxel)	III12 (doxorubicin, epirubicin, and paclitaxel)
3	III18 (doxorubicin, gemcitabine, and paclitaxel)	III7 (doxorubicin, 5-Fluorouracil and docetaxel)	IV59 (5-Fluorouracil, docetaxel, methotrexate and epirubicin)	III52 (docetaxel, epirubicin, and paclitaxel)



for luminal A subtype (ER<sup>+</sup>, PR<sup>+/−</sup>, HER2<sup>−</sup>), MDA-MB-468 for basal subtype (ER<sup>−</sup>, PR<sup>−</sup>, HER2<sup>−</sup>), and MDA-MB-231 for claudin-low subtype (ER<sup>−</sup>, PR<sup>−</sup>, HER2<sup>−</sup>), the latter two of which were also referred to as triple-negative cell lines (Holliday and Speirs, 2011). For the sake of comparison, one lung cancer cell line A549 was chosen to assess the prediction ability of DeepMDS. Then, 120 drug combinations were ranked according to their predicted pseudo-IC<sub>50</sub> values for each cell line.

According to the findings (Table 4), the top three synergistic combinations for MCF-7 and MDA-MB-468 shared commonalities, including III12 and III7. When compared to MDA-MB-468, the top three choices for another triple-negative

MDA-MB-231 had no similar result. The top three regimens for MDA-MB-231 were combinations of more than three drugs, including IV33, V32 and IV59. On another A549 lung cancer cell line, II28, III12 and III52 were the top three.

### 3.5 Experimental validation of predicted synergistic combinations

Subsequently, an *in vitro* cell viability study was undertaken on each cancer cell line to evaluate the predicted findings. IC<sub>50</sub> values for individual drugs were first obtained for each



TABLE 5 The anti-cancer effects of nine combinations of three drugs on MCF-7 cells.

Predicted ranking	Group number	The IC <sub>50</sub> (nM) of drug combinations				The best IC <sub>50</sub>
		I	II	III	IV	
1	III12	90.43	31.37	30.88	42.94	30.88
2	III7	290.99	101.66	56.52	85.66	56.52
3	III18	102.14	128.37	88.62	168.34	88.62
62	III40	2268.13	535.42	5823.46	308.97	308.97
63	III43	640.90	6386.59	4144.85	598.10	598.10
65	III47	10584.30	827.17	18532.30	10219.80	827.17
118	III9	23145.90	11,675.80	2121.31	2784.08	2121.31
119	III44	2606.11	2197.47	1499.70	1523.97	1499.70
125	III41	6022.00	354825.00	84020.00	24170.00	6022.00

Note: the predicted ranking included 120 drug combinations and individual drugs themselves.

cancer cell line (Figure 4). Then, in a similar fashion, the synergistic effects of predicted drug combinations were measured for each cell line.

### 3.5.1 Synergistic effects of predicted combinations on luminal a breast cancer cell line

For MCF-7 cell line, the results indicated that realistic IC<sub>50</sub> values for single drugs ranged from 138.3 nM to 97.25  $\mu$ M, with docetaxel exhibiting the best anti-cancer ability and 5-fluorouracil exhibiting the least (Figure 4A). In light of the ranked combinations, several combinations containing the top three (III12, III7 and III18), the middle level ones (III40, III43 and III47), and the bottom three (III9, III44 and III41) were examined on MCF-7 cell line using the defined drug ratios listed in Table 1.

As a result, the lowest IC<sub>50</sub> value for each combination across all drug ratios was considered the best experimental result and was used to rank the synergistic effect (Table 5). Except for the bottom combinations III9 and III44, the rest of experimental results were identical to the predicted order. With respect to the combinations III9 and III44, the actual IC<sub>50</sub> values reversed their ranking, which could be explained in part by the fact that both combinations elicited strong antagonistic responses on MCF-7 cell line, and DeepMDS may be insensitive to negative examples with additive or antagonistic effects. Additionally, the associated CI values of drug combinations were also calculated (Supplementary Figure S1). The top three combinations had a clear synergy impact on MCF-7 cell line ( $0.3 < CI < 0.9$ ), with III12 exhibiting the strongest synergy effect ( $CI < 0.3$ ). By contrast, the middle three and the bottom three demonstrated antagonism effect ( $CI > 1.1$ ).

To further evaluate DeepMDS's accuracy and robustness, the best synergistic combination, III12, was compared to different combinations including either two or all three drugs from III12 at the optimal drug ratio. For example, we

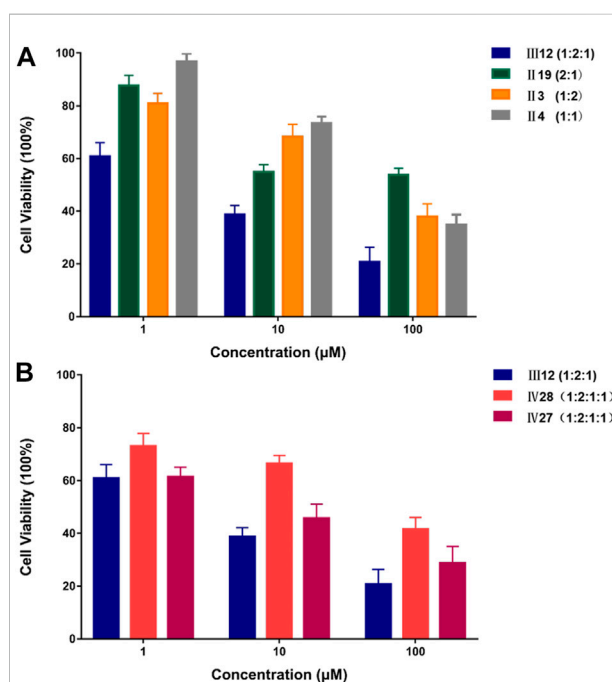


FIGURE 5

The comparison of anti-cancer effect on MCF-7 cell line between III12 and related combinations. (A). The comparison of anti-cancer effect on MCF-7 cells between III12 and related pairwise combinations that were extracted from III12. (B). The comparison of anti-cancer effect on MCF-7 cells between III12 and four-drug combinations which included the entire III12 composition.

chose II19 (docetaxel/gemcitabine, 2:1), II3 (doxorubicin/docetaxel, 1:2), and II4 (doxorubicin/gemcitabine, 1:1) as components of III12 (Figure 5A); and other groups, IV27 and IV28, contained the whole combination setting of

**TABLE 6** The comparison results of anti-cancer effect between clinically used combinations and III12 on MCF-7 cell line.

Predicted ranking	Group number	The IC50 (nM) of drug combinations				The best IC50
		I	II	III	IV	
1	III12	90.43	31.37	30.88	42.94	30.88
60	II3	570.00	440.00	377.18	—	377.18
92	III55	3695.00	7789.00	9444.00	814.10	814.10

Note: the predicted ranking included 120 drug combinations and individual drugs themselves.

**TABLE 7** The anti-cancer effects of drug combinations on MDA-MB-468 cells.

Predicted ranking	Group number	The IC50 (nM) of drug combinations				The best IC50
		I	II	III	IV	
1	III12	137.80	115.50	410.50	468.50	115.50
2	II3	353.90	207.60	443.30	—	207.60
3	III7	3497.00	522.90	1561.00	2767.00	522.90
27	III55	3095.00	774.20	1056.00	1613.00	774.20

Note: the predicted ranking included 120 drug combinations and individual drugs themselves.

III12 (Figure 5B). In addition, the commonly used clinical combinations (III55: gemcitabine, epirubicin and paclitaxel, and II3: doxorubicin and docetaxel) (Telli and Carlson, 2009) were assessed under the same circumstance as combination III12. Not unexpectedly, *in vitro* cellular experimental results indicated that III12 continues to exhibit the best anti-cancer synergistic activity when compared to any other combination (Table 6 and Supplementary Figure S2). Taking all the above validation data into account, the predicted III12 (doxorubicin, docetaxel and gemcitabine) was the most synergistic combination for the MCF-7 cell line.

### 3.5.2 Synergistic effects of predicted combinations on triple-negative breast cancer cell line

Additionally, for the MDA-MB-468 cell line (triple-negative basal subtype), IC50 values for various drugs ranged from 881.6 nM to 536.2  $\mu$ M, with paclitaxel exerting the greatest anti-cancer ability (Figure 4B). Then, the synergy impact of the top three combinations at various drug ratios was evaluated on MDA-MB-468. Similarly, III12 (doxorubicin, docetaxel, and gemcitabine) achieved the best IC50 value of 115.5 nM when used in a 2:1:1 M ratio (Table 7). Additionally, the clinically used drug combinations III55 (gemcitabine, epirubicin, and paclitaxel) (Telli and Carlson, 2009) was evaluated, and its best IC50 value was 774.2 nM, ranking 27th in the predicted results.

However, another triple-negative claudin-low subtype, MDA-MB-231, showed different drug responses. For example, monotherapy demonstrated that epirubicin had the lowest IC50 value of 2.81  $\mu$ M while 5-fluorouracil remained the worst one (Figure 4C). Experiments indicated that IV33, a four-drug combination, was the best of the predicted top three. Two commonly used drug combinations (III55 and II3) in clinical treatment were also compared, and it was discovered that III55, which was ranked 27th, and II3, which was ranked 70th, had significantly higher IC50 values and inferior anticancer activity (Table 8).

In addition, the synergistic mechanisms of all combinations were calculated for MDA-MB-468 and MDA-MB-231, respectively. With regards to MDA-MB-468, all three top combinations indicated strong synergy at each drug ratio, with CI values smaller than 0.3. Besides, the clinically used III55 showed strong synergy (CI = 0.27) and modest synergy (CI = 0.74) at 2:1:1 and 1:2:1 ratios, respectively; however, this regime had additive effect and antagonistic effect at the ratio of 1:1:2 (CI > 0.9) and 1:1:1 (CI > 1.1), respectively (Supplementary Figure S3). For the MDA-MB-231 cell line, the top two, IV33 and V32, exhibited strong synergistic effect at all drug ratios. However, the ranked third combination IV59 would exhibit some antagonistic activity at the ratio of 1:1:1:1 and 1:2:1:1, while still presenting strong synergy at other ratios. By contrast, II3 and III55, both of which have been used in clinical practice, had at least modest synergistic effects at each drug ratio (Supplementary Figure S4).

**TABLE 8** The anti-cancer effects of drug combinations on MDA-MB-231 cells.

Predicted ranking	Group number	The IC50 (nM) of drug combinations						The best IC50
		I	II	III	IV	V	VI	
1	IV33	177	52	139	2152	167	—	52
2	V32	1087	743	868	852	219	552	219
3	IV59	147,600	1952	19,260	244	850	—	244
27	III55	1506	5757	1793	5019	—	—	1506
70	II3	7983	9414	4567	—	—	—	4567

Note: the predicted ranking included 120 drug combinations and individual drugs themselves.

**TABLE 9** The anti-cancer effect of III12 and IV33 on MCF-7, MDA-MB-231 and MDA-MB-468 cells.

Group number	Cell line	The IC50 (nM) of drug combinations					The best IC50
		I	II	III	IV	V	
III12	MCF-7	90.4	31.3	30.8	42.9	—	30.8
	MDA-MB-468	137.8	115.5	410.5	468.5	—	115.5
	MDA-MB-231	15,720	8044.0	887.0	4666.0	—	887.0
IV33	MCF-7	8090.0	787.6	4355.0	788.9	1296.0	787.6
	MDA-MB-468	2803.0	1275.0	1988.0	1451.0	2665.0	1275.0
	MDA-MB-231	177.5	52.3	139.8	2152.0	167.5	52.3

### 3.5.3 Validation of predictive specificity for various breast cancer subtypes

The predicted result's specificity for cancer cell lines was further confirmed. IV33, the best drug combination for MDA-MB-231, was evaluated using other subtypes of breast cancer cell lines such as MCF-7 and MDA-MB-468. Rather than that, III12, which shown the greatest anticancer activity against MCF-7 and MDA-MB-468, was evaluated in a similar manner against MDA-MB-231. IV33 was predicted to rank 66th for MCF-7 and 33rd for MDA-MB-468, respectively, and III12 was predicted to rank 10th for the MDA-MB-231 (Table 9). Experiments proved the anticancer abilities of various combinations predicted for each specific subtype. And regardless of the drug ratio, IV33 exhibited antagonistic activity against MCF-7; however, IV33 had synergistic anti-cancer effects on MDA-MB-468 at all but 1:1:1:1 and 1:1:1:2. Compared with the outcomes of IV33 on 2 cell lines, III12, which performed slightly better on MDA-MB-231, exhibited synergy effect at all ratios.

To identify the potential synergistic mechanism of III12 and IV33 on various breast cancer subtypes, KEGG pathway enrichment analysis was carried out with a *p*-value cutoff of 0.01 (Supplementary Figure S7). The enrichment analysis showed that pathway in cancer, PI3K-Akt signaling pathway

and notch signaling pathway were the common pathways of III12 and IV33 on three breast cancer subtypes. More importantly, MAPK signaling pathway may be a special mechanism for the synergistic anti-cancer effect of III12 on MCF7 and MDA-MB-468, and Rap1 signaling pathway may be another important mechanism for III12 on MCF7. In addition, MAPK signaling pathway was the common pathway of IV33 on three breast cancer subtypes. Further analysis revealed that Gap junction may not contribute significantly to the synergistic anti-cancer effect of IV33. Collectively, each subtype of breast cancer cell lines had its own best synergized drug combinations, indicating an excellent specificity of DeepMDS for cell line subtypes that reflect dramatic genetic and epigenetic changes during the development of cancer.

### 3.5.4 Synergistic effects of predicted combinations on lung cancer cell line

An additional lung cancer cell line, A549, was used to examine the applicability of DeepMDS. It first predicted the synergistic anticancer effects of 120 drug combinations using DeepMDS (Supplementary Data Sheet D6), and then six combinations that ranked at the top two (II28 and III21), middle level (III12 and III47), or bottom two (III41 and III42) were examined in terms of cellular



TABLE 10 The anti-cancer effects of drug combinations on A549 cells.

Predicted ranking	Group number	The IC50 (nM) of drug combinations				The best IC50
		I	II	III	IV	
1	II28	181.5	140.9	73.8	—	73.8
2	III21	240.6	194.2	240.3	234.1	194.2
30	III12	215.1	135.4	128.4	191.1	128.4
41	III47	1353.0	696.8	750.7	3213.0	696.8
119	III41	9304	15,100	4739	15,110	4739
125	III42	1353	3807	1112	1223	1112

Note: the predicted ranking included 120 drug combinations and individual drugs themselves.

toxicity. IC50 values of individual drug ranged from 0.44  $\mu$ M to 50.23  $\mu$ M, with doxorubicin exerting the best anti-cancer ability and 5-fluorouracil still having the weakest efficacy (Figure 4D). It was shown that the anti-cancer capacity of various combinations at their optimal drug ratios was compatible with the predicted results of DeepMDS, thereby proving the reliability of this model for predicting potential multi-drug combinations (Table 10).

When we looked at the mechanisms of action for these combinations, it was clear that the synergy effect at every drug ratio was the most noticeable benefit of the top two combinations. Combinations in the middle (III47), as well as those at the bottom, showed additive or antagonistic effects at the majority of the ratios, but the 30th-ranked combination (III12) also demonstrated a strong synergistic mechanism in some cases (Supplementary Figure S6). It is possible that some underlying drug-target interactions, which were part of the overall synergistic mechanism, were not collected in the current training dataset because the predicted rank for III12 and III42 differed from the experimental results. Big biomedical data, which is becoming more widely available, could help us better predict the best combination for a given cancer cell line.

## 4 Discussion

Although it is now well established that combination therapies are significantly more effective at treating complicated disorders, experimentally assessing novel combinations is difficult due to the huge number of possible drug combinations. In this study, a deep learning-based model (DeepMDS) was successfully built to expedite the development of novel synergistic multi-drug combinations for clinical cancer treatment. DeepMDS enabled the ranking of all multi-drug combinations constructed randomly from a pool of medications using large-scale integrated features taken from gene expression profiles of human cancer cell lines, the multiscale interactome, and drug response data. Also, the predicted ranking of drug combinations revealed the likely mechanisms of action; for example, the higher ranked combination had a significantly greater

synergy impact, whereas the lower ranked combination would have an antagonistic effect. DeepMDS performed admirably in terms of accuracy. For the classification job, it earned an ACC of 0.94, an AUC of 0.97, a SEN of 0.95, and a SPE of 0.93. When facing a regression task, this model achieved a MSE of 2.50 and a RMSE of 1.58. A lack of experimental validation for some deep learning-based models may result in erroneous and/or unprofitable predictions when evaluating combinations of unknown druggable chemicals, natural products, and/or new cell lines. So, an *in vitro* cell experiment with seven clinically used anti-cancer drugs was used to test the ranked drug combinations predicted by DeepMDS in this work. In comparison to other drug combinations, it is clear that all of the predicted optimal synergistic combinations had a significant synergistic anti-cancer effect on each individual cell line.

As per the knowledge of authors, one of the biggest advantages of our model is to accurately predict the most promising three- or more-drug combinations for a certain cancer cell line. High-order combinations of drugs, as opposed to simple drug pairs, can regulate many anti-cancer networks simultaneously, hence improving tumor growth inhibition efficacy while avoiding drug resistance. Results indicated that DeepMDS leveraged its ability to rank high-order combinations which were randomly formed in the training space and so far untested. In addition, another advantage of DeepMDS is to predict synergistic combinations specific to a cancer cell line and even to a subtype of cell line. Experiments demonstrated that DeepMDS consistently gained high prediction performance across various subtypes of breast cancer cells and tissue-specific cancer cell lines. For example, III12 (doxorubicin, docetaxel and gemcitabine) had the best synergistic anti-cancer activity on hormone-responsive breast cancer cell line MCF-7, but IV33 (doxorubicin, gemcitabine, methotrexate and paclitaxel) and III12 were the most effective combinations against claudin-low MDA-MB-231 and basal MDA-MB-468, respectively, for triple-negative breast cancers. A549, a lung cancer cell line, was also used to evaluate the cell line specificity of DeepMDS, and one drug pair (II28) was found to be the best regimen. Therefore, it doesn't matter what the multi-drug combinations are, DeepMDS was able to accurately predict and rank synergistic combinations against the

cell line of interest, showing a wide range of applications. DeepMDS, in particular, makes it easier in the future to give targeted multi-drug combinations when taking into account the heterogeneity in genomic data of each patient.

To further validate the prediction power of DeepMDS for unseen combinations, clinically used breast cancer drug combinations, including II3 (doxorubicin and docetaxel) and III55 (gemcitabine, epirubicin, and paclitaxel), were tested in three subtypes of breast cancer cell line and, by extension, compared with the predicted best combinations. As a result, the IC50 values of II3 and III55 against MCF-7 increased about tenfold to twentyfold when compared to the predicted best combination (III12). Also, for triple-negative MDA-MB-468, this triple-drug combination (III12) was predicted to be the best, with an IC50 value of approximately 55.6 percent of II3's and 14.9 percent of III55's, respectively. In another triple-negative MDA-MB-231, we observed that the IC50 values of II3 leaped by about 88-fold, of III55 by 28.9-fold, when compared to the best combination IV33. Thus, it is possible to apply the novel synergistic drug combinations predicted by DeepMDS for breast cancer clinical trials, especially with regard to the triple negative breast cancer.

During the construction of machine learning and deep learning models, data are of critical relevance. In some cases, low quality predictive performance was mainly due to the incomplete dataset. For example, DeepSynergy was unable to accurately predict the response of novel medications and novel cell lines; more specifically, DeepSynergy indicated MSEs between 414 and 500 for novel drugs, and MSEs between 387 and 461 for novel cell lines. Because there were only 38 training instances of chemical compounds and cell lines, the authors speculated that the low prediction performance was due to a lack of training data (39 examples). In this case, the larger-scale integrated modeling datasets (201,405), which include 1,000 human cancer cell lines and 1839 chemicals, could substantially improve the performance and increase the accuracy of ranking the combinations. Another characteristic is the incorporation of drug-target data into modeling data. Rather than relying on descriptors of chemical structures to compare the structure similarity of two drugs, the drug-target information drives our prediction model to produce more accurate results in a biomedical context, which is beneficial for elucidating the underlying mechanisms of synergy action.

However, one limitation of our suggested strategy is that the modeling data contains insufficient information on drug targets. As a result, in some situations, a portion of a drug's target information may be omitted from the existing features, resulting in a discrepancy between predicted and actual outcomes. Notably, we did not feel that this constraint would eliminate the clinical use of our DeepMDS. By updating experimental drug targets data or adding predicted drug targets, this problem can be solved and the prediction accuracy of DeepMDS can be further enhanced. In addition, drug concentration ratio is also important for the synergistic effect of drug combination. Due to the lack of available data on drug concentration ratios of drug

combinations and the corresponding synergies, in this model, it was assumed that the drug concentration of each drug was sufficient to act on their targets and produce efficacy. We will continue to develop a computational method to predict the optimal drug concentration ratio for drug combination in future studies. Despite the limitations, our prediction model was able to translate monotherapy data into clinically useful predictions and expand the Universe of possible synergistic medication combinations, prioritizing promising multi-drug combinations for distinct types of cancer.

## 5 Conclusion

In this study, we developed a deep learning-based model that could aid in the discovery of the probable best combinations for a certain cell line or cell subtype. With regard to the high-cost experimental screening of drug combinations, our DeepMDS would significantly simplify the process of prioritizing the most promising multi-drug combinations for future pre-clinical studies. More importantly, our experimental validation proved that high-order combinations including three or more drugs, in most of cases, consistently outperformed drug pairs typically utilized in clinical treatment. Also, precise and robust prediction of drug combinations could identify the possible targeted combinations for personalized medicine, thereby expediting the development of combination therapy to combat against drug resistance and to improve efficacy.

## 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

SS and HC contributed equally to this study as the first authors. CF conceptualized the research. SS, HC, and MS established the model. SS, HC, WJ, MS, and JC performed the experiments and participated in analyzing the data. MR and CF supervised the project. HC, WJ, and MR wrote the manuscript. All authors reviewed and approved the manuscript.

## Funding

This work was supported, in part, by the National Natural Science Foundation of China (No.82074286, 81373897),

Natural Science Foundation of Jiangsu Province (No. BK20191428, BK20181445), Six Talent Peak Project from Government of Jiangsu Province (No. SWYY-013), the Science and Technology Innovation Fund of Zhenjiang-International cooperation projects (GJ2021012), and the Science and Technology Project of Jiangsu Administration of Traditional Chinese Medicine (2020ZX21).

## 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.

## References

- Alexander, S., and Friedl, P. (2012). Cancer invasion and resistance: Interconnected processes of disease progression and therapy failure. *Trends Mol. Med.* 18 (1), 13–26. doi:10.1016/j.molmed.2011.11.003
- Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., and Zhavoronkov, A. (2016). Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* 13 (7), 2524–2530. doi:10.1021/acs.molpharmaceut.6b00248
- Bansal, M., Yang, J., Karan, C., Menden, M. P., Costello, J. C., Tang, H., et al. (2014). A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* 32 (12), 1213–1222. doi:10.1038/nbt.3052
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2 (5), 401–404. doi:10.1158/2159-8290.CD-12-0095
- Chen, G., Tsoi, A., Xu, H., and Zheng, W. J. (2018). Predict effective drug combination by deep belief network and ontology fingerprints. *J. Biomed. Inf.* 85, 149–154. doi:10.1016/j.jbi.2018.07.024
- Chou, T.-C., and Martin, N. (2007). strong>The mass-action law-based new computer software, CompuSyn, for automated simulation of synergism and antagonism in drug combination studies</strong>. *Cancer Res.* 67 (9), 637.
- Chou, T. C., and Talalay, P. (1984). Quantitative analysis of dose-effect relationships: The combined effects of multiple drugs or enzyme inhibitors. *Adv. Enzyme Regul.* 22, 27–55. doi:10.1016/0065-2571(84)90007-4
- Davies, G., Boeree, M., Hermann, D., and Hoelscher, M. (2019). Accelerating the transition of new tuberculosis drug combinations from Phase II to Phase III trials: New technologies and innovative designs. *PLoS Med.* 16 (7), e1002851. doi:10.1371/journal.pmed.1002851
- Dry, J. R., Yang, M., and Saez-Rodriguez, J. (2016). Looking beyond the cancer cell for effective drug combinations. *Genome Med.* 8 (1), 125. doi:10.1186/s13073-016-0379-8
- Feng, C., Chen, H., Yuan, X., Sun, M., Chu, K., Liu, H., et al. (2019). Gene expression data based deep learning model for accurate prediction of drug-induced liver injury in advance. *J. Chem. Inf. Model.* 59 (7), 3240–3250. doi:10.1021/acs.jcim.9b00143
- Gayvert, K. M., Aly, O., Platt, J., Bosenberg, M. W., Stern, D. F., and Elemento, O. (2017). A computational approach for identifying synergistic drug combinations. *PLoS Comput. Biol.* 13 (1), e1005308. doi:10.1371/journal.pcbi.1005308
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* 5 (10), R80. doi:10.1186/gb-2004-5-10-r80
- Gotwals, P., Cameron, S., Cipolletta, D., Cremasco, V., Crystal, A., Hewes, B., et al. (2017). Prospects for combining targeted and conventional cancer therapy with immunotherapy. *Nat. Rev. Cancer* 17 (5), 286–301. doi:10.1038/nrc.2017.17
- Holliday, D. L., and Speirs, V. (2011). Choosing the right cell line for breast cancer research. *Breast Cancer Res.* 13 (4), 215. doi:10.1186/bcr2889
- Holohan, C., Van Schaeybroeck, S., Longley, D. B., and Johnston, P. G. (2013). Cancer drug resistance: An evolving paradigm. *Nat. Rev. Cancer* 13 (10), 714–726. doi:10.1038/nrc3599
- Huang, L., Brunell, D., Stephan, C., Mancuso, J., Yu, X., He, B., et al. (2019). Driver network as a biomarker: Systematic integration and network modeling of multi-omics data to derive driver signaling pathways for drug combination prediction. *Bioinformatics* 35 (19), 3709–3717. doi:10.1093/bioinformatics/btz109
- Ianevski, A., Giri, A. K., and Aittokallio, T. (2020). SynergyFinder 2.0: Visual analytics of multi-drug combination synergies. *Nucleic Acids Res.* 48 (1), W488–W493. doi:10.1093/nar/gkaa216
- Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., et al. (2016). A landscape of pharmacogenomic interactions in cancer. *Cell* 166 (3), 740–754. doi:10.1016/j.cell.2016.06.017
- Iwata, H., Sawada, R., Mizutani, S., Kotera, M., and Yamanishi, Y. (2015). Large-scale prediction of beneficial drug combinations using drug efficacy and target profiles. *J. Chem. Inf. Model.* 55 (12), 2705–2716. doi:10.1021/acs.jcim.5b00444
- Jeon, M., Kim, S., Park, S., Lee, H., and Kang, J. (2018). *In silico* drug combination discovery for personalized cancer therapy. *BMC Syst. Biol.* 12 (2), 16. doi:10.1186/s12918-018-0546-1
- Kuru, H. b., Tastan, O., and Cicek, E. (2021). MatchMaker: A deep learning framework for drug synergy prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 19, 2334–2344. doi:10.1109/TCBB.2021.3086702
- MacGowan, A. P., Holt, H. A., and Reeves, D. S. (1990). *In-vitro* synergy testing of nine antimicrobial combinations against *Listeria monocytogenes*. *J. Antimicrob. Chemother.* 25 (4), 561–566. doi:10.1093/jac/25.4.561
- Mahase, E. (2019). Breast cancer: NICE approves new drug combination treatment. *Bmj* 366, l4727. doi:10.1136/bmj.l4727
- Menden, M. P., Wang, D., Mason, M. J., Szalai, B., Bulusu, K. C., Guan, Y., et al. (2019). Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat. Commun.* 10 (1), 2674. doi:10.1038/s41467-019-09799-2
- National Comprehensive Cancer Network (2021). Breast cancer. Version 5.2021. Available: [www.nccn.org/professionals/physician\\_gls/pdf/breast.pdf](http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf) (Accessed June 14, 2021).
- Parkinson, H., Sarkans, U., Shojatalab, M., Abeygunawardena, N., Contrino, S., Coulson, R., et al. (2005). ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 33, D553–D555. Database issue. doi:10.1093/nar/gki056
- Preuer, K., Lewis, R. P. I., Hochreiter, S., Bender, A., Bulusu, K. C., and Klambauer, G. (2018). DeepSynergy: Predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 34 (9), 1538–1546. doi:10.1093/bioinformatics/btx806
- Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., et al. (2014). Tcmisp: A database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminform.* 6, 13. doi:10.1186/1758-2946-6-13

## 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/fphar.2022.1032875/full#supplementary-material>

- Sidorov, P., Naulaerts, S., Arieu-Bonnet, J., Pasquier, E., and Ballester, P. J. (2019). Predicting synergism of cancer drug combinations using NCI-ALMANAC data. *Front. Chem.* 7, 509. doi:10.3389/fchem.2019.00509
- Sopirala, M. M., Mangino, J. E., Gebreyes, W. A., Biller, B., Bannerman, T., Balada-Llasat, J. M., et al. (2010). Synergy testing by Etest, microdilution checkerboard, and time-kill methods for pan-drug-resistant *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54 (11), 4678–4683. doi:10.1128/aac.00497-10
- Sun, W., Sanderson, P. E., and Zheng, W. (2016). Drug combination therapy increases successful drug repositioning. *Drug Discov. Today* 21 (7), 1189–1195. doi:10.1016/j.drudis.2016.05.015
- Sun, X., Vilar, S., and Tatonetti, N. P. (2013). High-throughput methods for combinatorial drug discovery. *Sci. Transl. Med.* 5 (205), 205rv1. 205rv201. doi:10.1126/scitranslmed.3006667
- Sun, Y., Sheng, Z., Ma, C., Tang, K., Zhu, R., Wu, Z., et al. (2015). Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. *Nat. Commun.* 6, 8481. doi:10.1038/ncomms9481
- Telli, M. L., and Carlson, R. W. (2009). First-line chemotherapy for metastatic breast cancer. *Clin. Breast Cancer* 9 (2), S66–S72. doi:10.3816/CBC.2009.s.007
- Tolomeo, M., and Simoni, D. (2002). Drug resistance and apoptosis in cancer treatment: Development of new apoptosis-inducing agents active in drug resistant malignancies. *Curr. Med. Chem. Anticancer. Agents* 2 (3), 387–401. doi:10.2174/1568011024606361
- Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., and Bryant, S. H. (2009). PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* 37, W623–W633. doi:10.1093/nar/gkp456
- Wiesner, J., Henschker, D., Hutchinson, D. B., Beck, E., and Jomaa, H. (2002). *In vitro* and *in vivo* synergy of fosmidomycin, a novel antimalarial drug, with clindamycin. *Antimicrob. Agents Chemother.* 46 (9), 2889–2894. doi:10.1128/aac.46.9.2889-2894.2002
- 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 (D1), D1074–d1082. doi:10.1093/nar/gkx1037
- Zhang, H., Feng, J., Zeng, A., Payne, P., and Li, F. (2021a). Predicting tumor cell response to synergistic drug combinations using a novel simplified deep learning model. *AMIA Annu. Symp. Proc.* 2020, 1364–1372.
- Zhang, T., Zhang, L., Payne, P. R. O., and Li, F. (2021b). Synergistic drug combination prediction by integrating Multiomics data in deep learning models. *Methods Mol. Biol.* 2194, 223–238. doi:10.1007/978-1-0716-0849-4\_12



## OPEN ACCESS

## EDITED BY

Yusof Kamisah,  
Faculty of Medicine Universiti Kebangsaan  
Malaysia, Malaysia

## REVIEWED BY

Xiaohui Wang,  
East Tennessee State University,  
United States  
Song Hui Jia,  
St. Michael's Hospital, Canada

## \*CORRESPONDENCE

Zhengfei He,  
✉ hzf\_19750101@163.com

RECEIVED 18 January 2023

ACCEPTED 05 June 2023

PUBLISHED 14 June 2023

## CITATION

Sun L, Chen Z, Ni Y and He Z (2023),  
Network pharmacology-based approach  
to explore the underlying mechanism of  
sinomenine on sepsis-induced  
myocardial injury in rats.  
*Front. Pharmacol.* 14:1138858.  
doi: 10.3389/fphar.2023.1138858

## COPYRIGHT

© 2023 Sun, Chen, Ni and He. 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.

# Network pharmacology-based approach to explore the underlying mechanism of sinomenine on sepsis-induced myocardial injury in rats

Linggong Sun, Zhiyun Chen, Yunjie Ni and Zhengfei He\*

Department of Cardiology, The First People's Hospital of Fuyang Hangzhou, Hangzhou, Zhejiang, China

**Background:** Sepsis, a systemic disease, usually induces myocardial injury (MI), and sepsis-induced MI has become a significant contributor to sepsis-related deaths in the intensive care unit. The objective of this study is to investigate the role of sinomenine (SIN) on sepsis-induced MI and clarify the underlying mechanism based on the techniques of network pharmacology.

**Methods:** Cecum ligation and puncture (CLP) was adopted to induce sepsis in male Sprague-Dawley (SD) rats. Serum indicators, echocardiographic cardiac parameters, and hematoxylin and eosin (H&E) staining were conducted to gauge the severity of cardiac damage. The candidate targets and potential mechanism of SIN against sepsis-induced MI were analyzed via network pharmacology. Enzyme-linked immunosorbent assay was performed for detecting the serum concentration of inflammatory cytokines. Western blot was applied for evaluating the levels of protein expression. Terminal deoxynucleotidyl transferase-mediated dUTP biotin nick end labeling assay was applied to assess cardiomyocyte apoptosis.

**Results:** SIN significantly improved the cardiac functions, and attenuated myocardial structural damage of rats as compared with the CLP group. In total, 178 targets of SIN and 945 sepsis-related genes were identified, and 33 overlapped targets were considered as candidate targets of SIN against sepsis. Enrichment analysis results demonstrated that these putative targets were significantly associated with the Interleukin 17 (IL-17) signal pathway, inflammatory response, cytokines-mediated signal pathway, and Janus Kinase-Signal Transducers and Activators of Transcription (JAK-STAT) pathway. Molecular docking suggested that SIN had favorable binding affinities with Mitogen-Activated Protein Kinase 8 (MAPK8), Janus Kinase 1 (JAK1), Janus Kinase 2 (JAK2), Signal Transducer and Activator of Transcription 3 (STAT3), and nuclear factor kappa-B (NF- $\kappa$ B). SIN significantly reduced the serum concentration of Tumor Necrosis Factor- $\alpha$  (TNF- $\alpha$ ), Interleukin 1 Beta (IL-1 $\beta$ ), Interleukin 6 (IL-6), Interferon gamma (IFN- $\gamma$ ), and C-X-C Motif Chemokine Ligand 8 (CXCL8), lowered the protein expression of phosphorylated c-Jun N-terminal kinase 1 (JNK1), JAK1, JAK2, STAT3, NF- $\kappa$ B, and decreased the proportion of cleaved-caspase3/caspase3. In addition, SIN also significantly inhibited the apoptosis of cardiomyocytes as compared with the CLP group.



**Conclusion:** Based on network pharmacology analysis and corresponding experiments, it was concluded that SIN could mediate related targets and pathways to protect against sepsis-induced MI.

#### KEYWORDS

sinomenine, sepsis, myocardial injury, network pharmacology, pathway

## 1 Introduction

Sepsis is a potentially fatal organ failure resulted from aberrant or dysfunctional host response to infection (Salomao et al., 2019). According to epidemiological investigations, 270,000 of the 1.7 million sepsis patients in the U.S. died in 2014, and sepsis has emerged as a prominent contributor to mortality among patients admitted to hospitals (McLaughlin et al., 2020). The incidence of sepsis is rising in China at a rate of 1.5% each year. Besides, the aging population and the wide application of invasive surgery contribute to the increase of morbidity and mortality of severe sepsis year by year (Corrales et al., 2022). Clinical and fundamental research showed that sepsis affects the cardiovascular system of patients, with myocardial injury (MI) occurring in about 40%–50% of them and leading to a death rate of 70%–90% (Xie et al., 2021). Sepsis-induced MI is first characterized by dysfunctions including decreased myocardial contractility and biventricular dilatation with decreased left ventricular ejection fraction (LVEF), followed by morphological changes such as myocardial cell degeneration, focal necrosis, and blurred myocardial striated lines. However, no targeted treatments are now available for sepsis-induced MI. Currently, multiple studies have linked the elevated inflammatory cytokines with sepsis-induced MI, and inflammation inhibition has been regarded as a promising therapeutic strategy for sepsis-induced MI.

Recently, the application of traditional Chinese Medicine (TCM) in treating sepsis-induced MI is attracting increasing attention. For instance, ShenFu injection showed favorable efficacy in treating sepsis-induced MI by reducing mitochondrial apoptosis (Xu et al., 2020). Sinomenine (SIN), as an alkaloid isolated from the root and stem of *Sinomenium acutum* (Thunb.) Rehder et Wilson or *S. acutum* var. *cinereum*, has been utilized extensively in the treatment of rheumatic diseases and arrhythmia (Liu et al., 2018). Accumulating evidence revealed that SIN exhibits diverse pharmacological effects, for instance, it is anti-inflammatory (Zeng and Tong, 2020), anti-cancer (Song L. et al., 2021), and analgesic (Jiang et al., 2020). Recently, SIN was reported to ameliorate lung injury in sepsis. The Nuclear factor erythroid 2-related factor 2–Kelch Like ECH Associated Protein 1 (Nrf2–Keap1) axis (Wang et al., 2020) or altering intestinal homeostasis through the aryl hydrocarbon receptor/Nrf2 axis (Song W. et al., 2021) are two potential mechanisms by which SIN could reduce septic acute lung damage in rats. Liu et al. claimed that SIN could improve lipopolysaccharide (LPS)-induced cardiomyocyte injury *in vitro* (Liu et al., 2021). Therefore, it is necessary to investigate the potential application and underlying mechanisms of SIN in treating sepsis-induced MI.

Network pharmacology has been developed by integrating biochemistry, bioinformatics, and system biology for studying the complex mechanism of TCM and discovering potential targets and

mechanisms associated with various TCM monomers, such as artemisinin (Lin et al., 2021), melatonin (Song W. et al., 2021), and SIN (Li et al., 2021). In this study, the effects of SIN on cardiac dysfunctions was evaluated based on a rat model of sepsis. Network pharmacology was employed to identify SIN's possible targets and pathways against sepsis-induced MI, and the binding affinity between SIN and corresponding candidate targets was simulated by molecular docking. Furthermore, the candidate targets and pathways were also validated *in vivo*.

## 2 Materials and methods

### 2.1 Prediction of SIN targets

The structure of SIN (CID: 5459308) was downloaded from the database of PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). The candidate targets of SIN were predicted by the online tools of PharmMapper database (<http://www.lilab-ecust.cn/pharmmapper/>), Swiss Target Prediction (<http://www.swisstargetprediction.ch/>), HERB database (<http://herb.ac.cn/>), and TCM potential target database (TCM-PTD, <http://tcm.zju.edu.cn/>). After merging the results obtained from the four databases and excluding non-human genes, the rest genes were regarded as potential targets of SIN.

### 2.2 Screening sepsis-related genes

Sepsis-related genes were screened out from GeneCards (<https://www.genecards.org/>), Online Mendelian Inheritance in Man (OMIM, <https://www.omim.org/>), and Comparative Toxicogenomics Database (CTD, <http://ctdbase.org/>). “Sepsis” was used as a keyword for searching. After removing duplicates, the rest were identified as sepsis-related genes.

### 2.3 Network and enrichment analysis

The protein-protein interaction (PPI) of the shared targets between SIN and sepsis were retrieved from STRING (<https://string-db.org/>) with a medium confidence, and the PPI network was constructed with the use of Cytoscape software (<https://cytoscape.org/>). The topological parameters were calculated by the “Network Analyzer” plug-in. The “clusterprofiler” R package was used to analyze the enrichment of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The default threshold was set at a Bonferroni-corrected *p*-value of  $\leq 0.05$ . The SIN-pathway-gene network was then established using Cytoscape software.

## 2.4 Molecular docking

Molecular docking analysis of SIN and its related targets was conducted with the use of AutoDock Vina software (version 1.1.2) to anticipate the strength of their interaction. The 2D structure of SIN was obtained from PubChem. The crystal structure of SIN-targets was obtained from RCSB protein data bank (RCSB PDB: <https://www.rcsb.org/>), including Mitogen-Activated Protein Kinase 8 (MAPK8, 3elj), nuclear factor kappa-B (NF- $\kappa$ B, 7RG5), Janus Kinase 1 (JAK1, 4ei4), Janus Kinase 2 (JAK2, 7f7w), and Signal Transducer And Activator of Transcription 3 (STAT3, 6nuq). Before the docking, PyMOL (version 4.5.0) software was used for protein preparation by removing water molecules, solvent molecules, and other protein chains. Then, the software of AutoDock Tools 1.5.6 was used to add nonpolar hydrogens and calculate Gasteiger charges of protein structures. The Lamarckian Genetic algorithm was used to perform the conformational search and generate 100 conformations. The conformation with the best affinity was selected as the final docking conformation. The 2D diagrams of the SIN-targets complex were generated using LigPlus (version 2.24), and the 3D complex was visualized by PyMOL.

## 2.5 Animal model and SIN treatment

The Institutional Animal Care and Use Committee of Zhejiang Center of Laboratory Animals approved the animal procedures and experimental protocols (Approval Number: ZJCLA-IACUC-20020101). Male Sprague-Dawley (SD) rats (180–200 g) were obtained from Hangzhou Medical College Laboratory Animal Center, and maintained in a specified pathogen-free environment with unlimited availability of food and water on a 12-h day and night cycle. The sepsis model was induced by cecum ligation and puncture (CLP), following the previously described procedures (Mishra and Choudhury, 2018). The four groups—sham, CLP, CLP + LSIN, and CLP + HSIN groups—each containing ten experimental rats, were randomly allocated. The exposed cecum of the rats was sutured with 3–0 silk suture 1.2 cm to its distal end and punctured twice with a 22-gauge needle to create sepsis model. Following the surgical procedure, all rats received a subcutaneous injection of 50 mL/kg compound sodium chloride. The rats in the sham group underwent identical procedures as described above, excluding the CLP treatment. In the CLP + LSIN and CLP + HSIN groups, rats received SIN administration via tail vein injection at doses of 50 mg/kg and 100 mg/kg, respectively, 15 min prior to sepsis induction. After 24 h of sepsis, the hearts were extracted and a total of 500 mL blood was collected to obtain serum.

## 2.6 Determination of serum biochemical parameters

After blood collection from each group of rats, serum obtained by subjecting the blood samples to centrifugation at 3,000 g/min for 30 min. An automated analyzer (Modular DPP H7600; Roche Diagnostics, Basel, Switzerland) was employed to assess the

serum concentration of lactate dehydrogenase (LDH) and creatine kinase and its MB isoenzyme (CK-MB). Enzyme linked immunosorbent assay (ELISA)-based assay was performed for the concentration detection of cardiac troponin I (cTnI, mlbio, China, Shanghai), cardiac myosin light chain-1 (cMLC1, EK-Bioscience, China, Shanghai), as well as several inflammatory cytokines such as Tumor Necrosis Factor- $\alpha$  (TNF- $\alpha$ , Applygen, China, Beijing), Interleukin 1 Beta (IL-1 $\beta$ , wksbio, China, Shanghai), Interleukin 6 (IL-6, Thermo Fisher Scientific, USA, Massachusetts), C-X-C Motif Chemokine Ligand 8 (CXCL8, Shanghai yiyao biotechnology Co. Ltd., China, Shanghai), and Interferon gamma (IFN- $\gamma$ , Sino Biological, China, Beijing) in plasma.

## 2.7 Echocardiography

The cardiac parameters were evaluated after 12 h of CLP by echocardiography according to the methodology reported previously (Bayer et al., 2021). The rats were anesthetized with a mixture of 2% isoflurane and 0.5 L/min 100% O<sub>2</sub> before they were positioned on a warming pad (37 °C). A Vevo 2,100 Imaging System was used to take echocardiographic measures (FUJIFILM VisualSonics, Inc., Toronto, Ontario, Canada). To evaluate heart functioning, the ejection fraction (EF) and left ventricular interior dimension (LVID) were calculated. All measurements were conducted by a cardiologist who was unaware of the experimental details, ensuring a blinded assessment.

## 2.8 Hematoxylin and eosin (H&E) staining

The rat myocardial tissues were collected from every group 24 h after CLP and immediately fixed with 4% paraformaldehyde overnight at room temperature to facilitate subsequent histological analysis. The materials were then divided into 4  $\mu$ m-thick slices and embedded in paraffin. After that, the slices underwent H&E staining and observed under a light microscope at a magnification of  $\times 400$ .

## 2.9 Western blot

The myocardial tissues were homogenized using Radio Immunoprecipitation Assay (RIPA, Thermo Fisher Scientific, USA, Massachusetts) lysis buffer, and subsequently centrifuged at 13,200 g at 4°C for 30 min. To determine the protein concentrations, the Bradford assay was conducted, and the supernatant was collected for total protein analysis. Subsequently, the extracted proteins (25  $\mu$ g) were subjected to Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) for separation, blotted and probed with the following antibodies: anti-c-Jun N-terminal kinase 1 (JNK1, ab199380, Abcam, UK, Cambridge, 1/2,500), anti-phospho-JNK1 (ab215208, Abcam, UK, Cambridge, 1/1,000), anti-NF- $\kappa$ B (ab16502, Abcam, UK, Cambridge, 1/1,000), anti-phospho-NF- $\kappa$ B (ab76302, Abcam, UK, Cambridge, 1/1,000), anti-JAK1 (ab133666, Abcam, UK, Cambridge, 1/1,000), anti-phospho-JAK1 (ab215338, Abcam, UK, Cambridge, 1/5,000), anti-JAK2

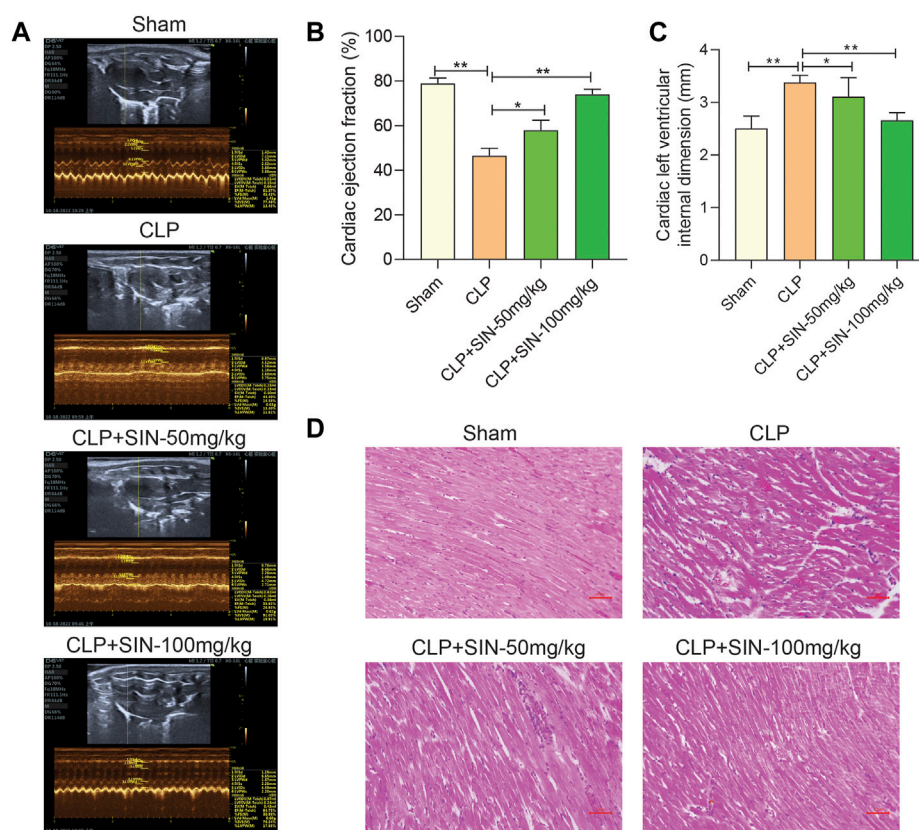


FIGURE 1

SIN increased the survival of rats with sepsis brought on by CLP and reduced the severity of the cardiac dysfunction those animals experienced. (A) Representative images of echocardiography for each group. Ejection fraction (B) and left ventricular internal dimension (C) were assessed using echocardiography as indicators of cardiac function. (D) Rat cardiac tissues' histological alterations at 12 h after CLP (H&E staining,  $\times 40$ ).  $n = 6$ .  $*p < 0.05$ ,  $**p < 0.01$ .

(ab108596, Abcam, UK, Cambridge, 1/1,000); anti-phospho-JAK2 (ab32101, Abcam, UK, Cambridge, 1/1,000); anti-STAT3 (ab68153, Abcam, UK, Cambridge, 1/2,000); anti-phospho-STAT3 (ab76315, Abcam, UK, Cambridge, 1/1,000); anti-Caspase-3 (9662S, Cell Signaling Technology, USA, Massachusetts 1/1,000); anti-Cleaved Caspase-3 (9664S, Cell Signaling Technology, USA, Massachusetts). For loading control, the blots were probed with antibody for GAPDH (ab8245, Abcam, UK, Cambridge, 1/500). The blots were measured by the chemiluminescence system (Millipore, Billerica, MA, USA), and the signals were quantified by densitometry. Use ImageJ 1.8.0 (<https://imagej.nih.gov/ij/>) to read the density of the bands.

## 2.10 Terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) staining

TUNEL test was used to ascertain the rate of cardiomyocyte apoptosis in the heart tissues of rats according to relevant instructions (Roche, USA). The heart tissue sections were fixed and permeated, followed by co-staining of TUNEL and 4',6-diamidino-2-phenylindole (DAPI).

## 2.11 Statistical analysis

Statistical analyses were performed using SPSS 25.0 software, and the results were reported as the mean  $\pm$  standard deviation. A  $p$ -value  $\leq 0.05$  is considered to be statistically significant.

## 3 Results

### 3.1 SIN improved the cardiac function of septic rats

To determine whether SIN has cardioprotective effects, the rats that developed sepsis via CLP were examined using echocardiography. SIN demonstrated a dose-dependent effect in significantly increasing the low EF in rats with CLP-induced sepsis, whereas significantly reduced the elevated LVID of septic rats in the CLP group. These findings suggested that SIN ameliorated cardiac function in septic rats (Figures 1A–C). Moreover, histopathological changes in the rat myocardial tissues were observed via H&E staining to assess the beneficial effect of SIN on MI in septic rats. As shown in Figure 1D, there was no degeneration, necrosis, or aberrant alterations in the myocardial interstitium of rats in the



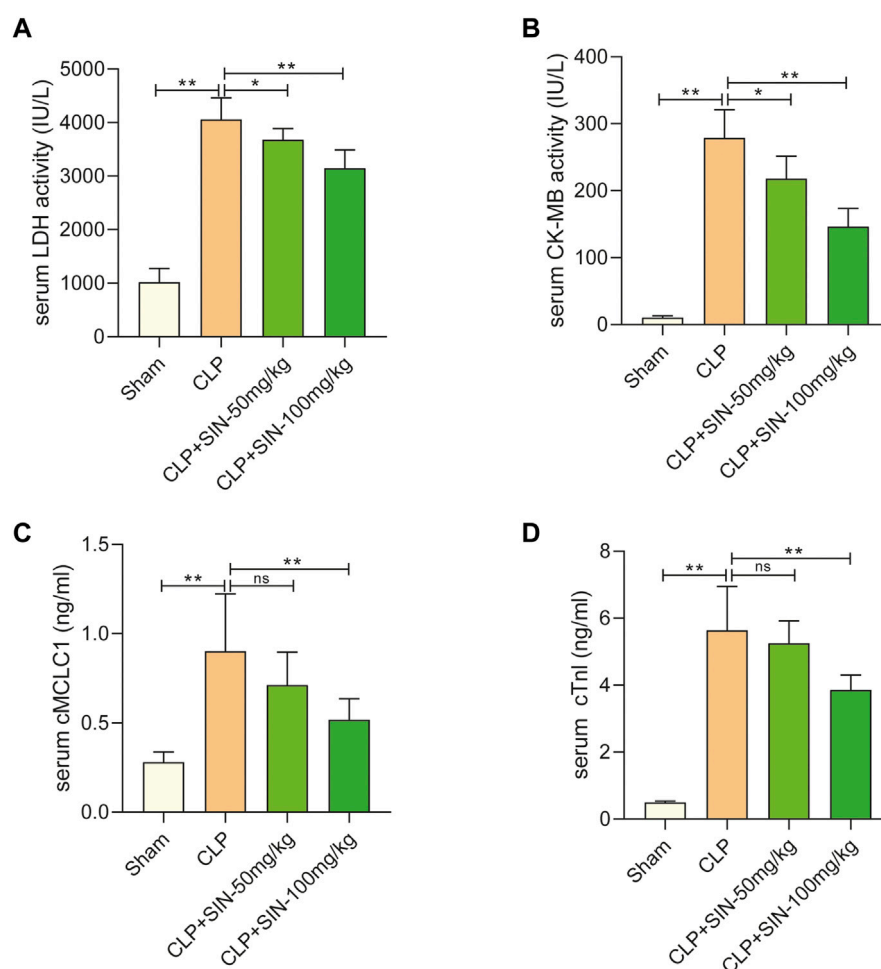


FIGURE 2

SIN reduced the cardiac dysfunction brought on by CLP. (A) Rat serum LDH response to SIN. (B) Rat serum CK-MB response to SIN. (C) Rat serum cMLC1 response to SIN. (D) Rat serum cTnI response to SIN.  $n = 6$ . \* $p < 0.05$ , \*\* $p < 0.01$ .

sham group, although they did have obvious transverse stripes of myocardial fibers. Rats in the CLP group showed significant pathological changes in their myocardial tissues, including myocardial fiber partial rupture and breakdown, myocardial stripe blur partial disappearance, and interstitial edema. Notably, administration of 50 and 100 mg/kg SIN considerably reduced these pathogenic alterations.

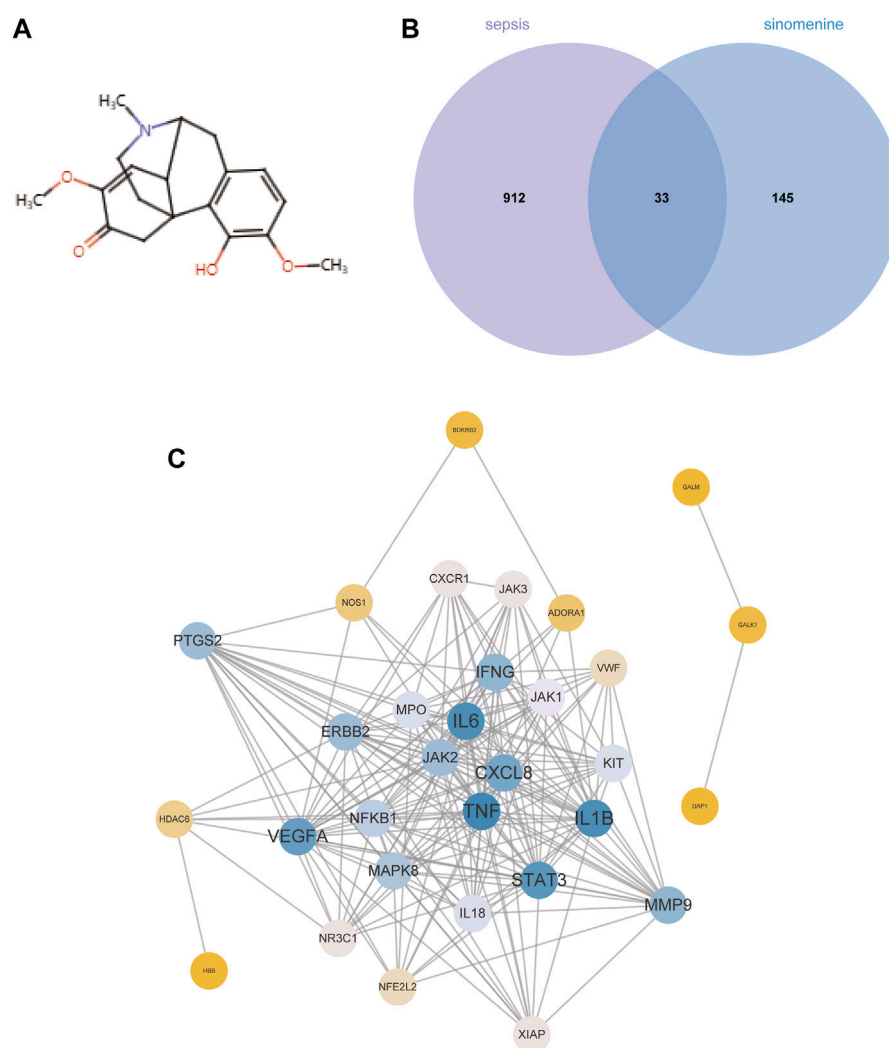
### 3.2 SIN attenuated the myocardial injuries of septic rats

Four serum biomarkers were used in this study for evaluation of the myocardial injuries of rats: LDH, CK-MB, cMLC1, and cTnI (Shiroorkar et al., 2020). Twelve hours after CLP surgery, the concentrations of LDH, CK-MB, cTnI, and cMLC1 in the serum were measured in each group. As shown in Figure 2, the concentration of LDH, CK-MB, cMLC1, and cTnI in the serum of septic rats was obviously elevated as compared with the normal rats, indicating the presence of CLP-induced MI. Importantly, SIN demonstrated a dose-dependent effect in significantly decreasing the

forementioned injury in rats with CLP-induced sepsis. These data suggested that SIN can attenuate MI and ameliorate myocardial function.

### 3.3 Potential targets and pathways of SIN against sepsis-induced MI

Figure 3A depicts the chemical structure of SIN. In total, 178 targets of SIN were identified (Supplementary Table S1), including 55 in SwissTargetPrediction, 103 in PharmMapper, 16 in HERB, and 10 in TCM-PTD. Meanwhile, 945 sepsis-related genes were screened out from GeneCard (902), OMIM (77), and Therapeutic Target Database (TTD) (67) (Supplementary Table S2). There were 33 shared targets between the targets of SIN and sepsis-related genes, as shown in Figure 3B. These proteins were identified as candidate targets of SIN against sepsis. Figure 3C shows the PPI network of these common targets, and the top 4 nodes with a greater degree were TNF, IL6, IL1 $\beta$ , and STAT3, respectively. These data indicated that the four targets play more crucial roles in the treatment of sepsis-induced MI by SIN.

**FIGURE 3**

The candidate targets of SIN against sepsis-induced MI. **(A)** The structure of SIN. **(B)** Venn diagram of the genes associated with sepsis and SIN targets. **(C)** The PPI network of the 33 candidate targets was generated by Cytoscape.

To reveal the potential functions and pathways of the candidate targets, functional enrichment analyses were conducted accordingly, and the findings are summarized in [Supplementary Tables S3, S4](#). [Figure 4A](#) displays the top 10 GO terms for biological process, cellular component, and molecular function. It was demonstrated that the candidate targets were primarily associated with cytokines mediation, inflammatory response, and phosphorylation of STAT protein. The enriched KEGG pathways primarily included Interleukin 17 (IL-17) signaling pathway, nucleotide-binding oligomerization domain (NOD)-like receptor signaling pathway, as well as various pathways related to viral or inflammatory diseases ([Figure 4B](#)). To better understand the correlations between candidate targets and enriched pathways, a gene-pathway network was established. [Figure 4C](#) illustrates that the larger the node font is, the more pathways the corresponding target are involved in.

### 3.4 Binding affinity of SIN with target proteins

The candidate targets' possible binding modes with SIN were evaluated using molecular docking analysis ([Figure 5](#)). The autodock scores were summarized in [Table 1](#), and a lower score indicated a better binding affinity between SIN and the proteins. Our data indicated that SIN was most tightly bound to JAK1 and loosely bound to JNK1. Additionally, hydrogen bonds were observed in all SIN-target complexes and the O3 in SIN was prone to form hydrogen bonds with residues. The main groups that bind residues with H donor moieties at the terminal are believed to be the carbonyl, methoxy, and hydroxy groups. The 2D and 3D docking images showed that SIN interacted with JAK1, JAK2, STAT3, JNK1, and NF- $\kappa$ B via hydrogen bonds and hydrophobic contacts. These data indicated that SIN can interact with JAK1, JAK2, STAT3, JNK1, and NF- $\kappa$ B to form compact complexes.

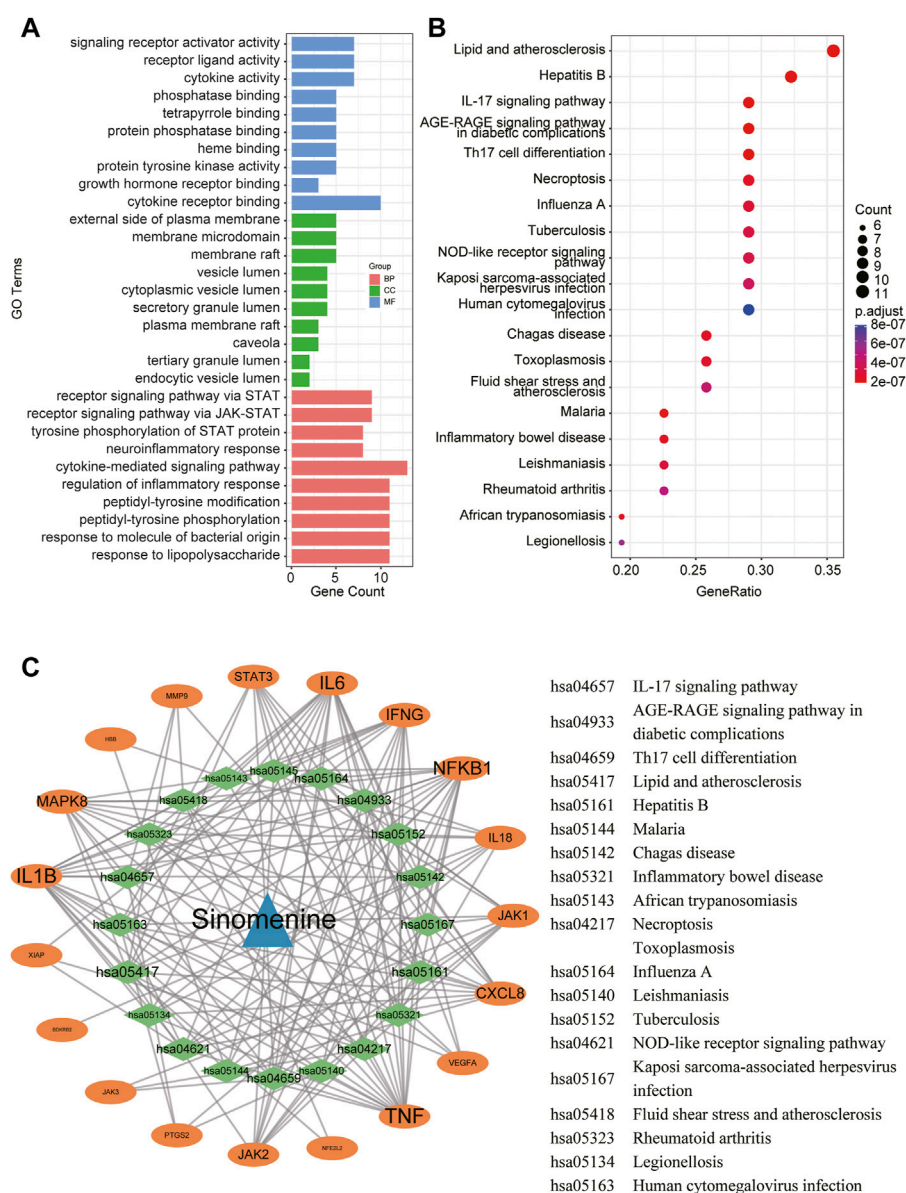


FIGURE 4

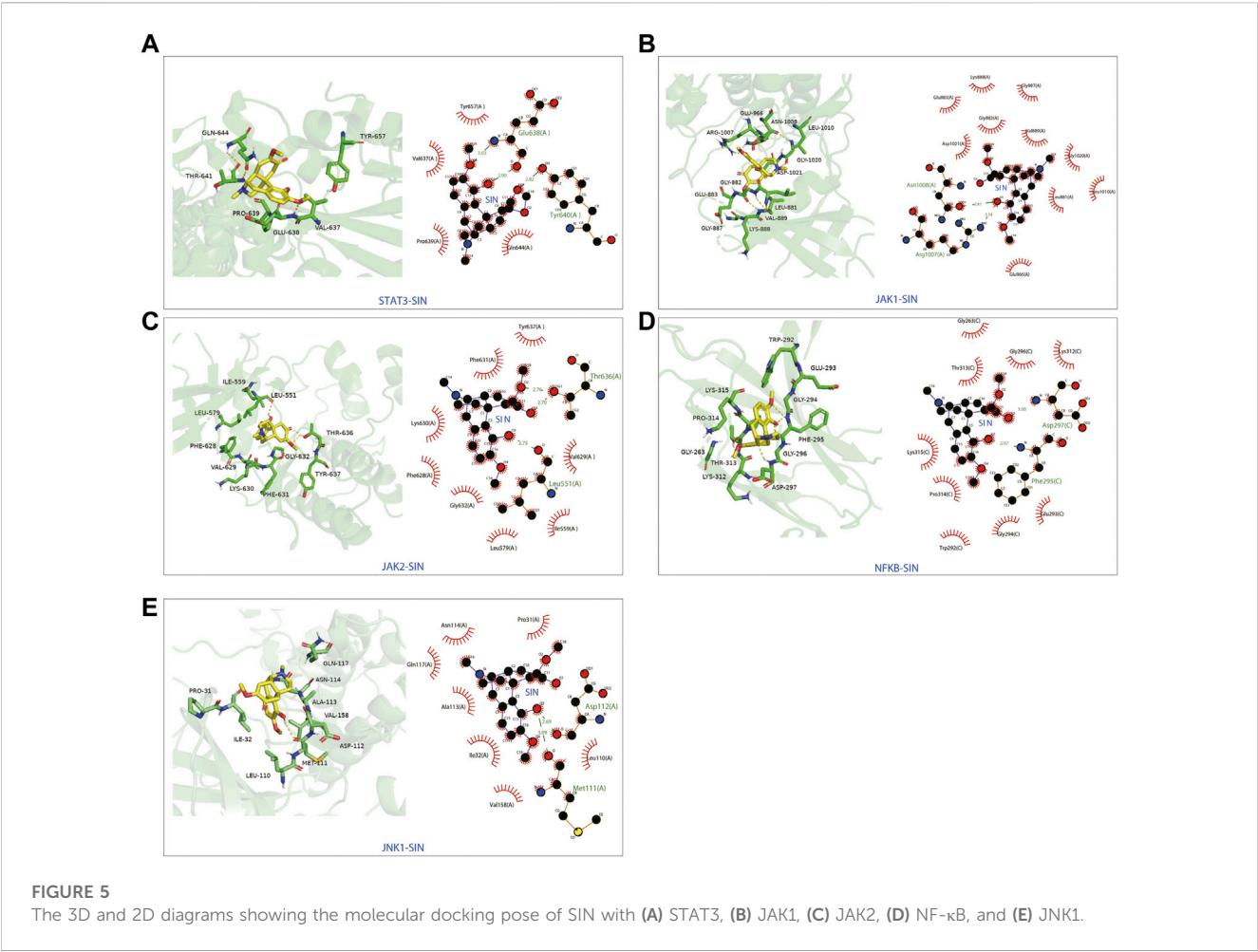
A network pharmacology approach was used to determine the fundamental mechanisms of SIN against sepsis. (A) The top 10 terms of GO analysis of the candidate targets of SIN against sepsis-induced MI. CC: cellular component, BP: biological process, MF: molecular function. (B) The top 20 pathways of KEGG pathway enrichment analysis that ranked by gene count. (C) The SIN-targets-pathway network showed detailed interactions between the hub targets and pathways. Orange cycles stand in for hub targets, while green circles represent the top 20 pathways that SIN uses to combat sepsis.

Therefore, the influence of SIN on the protein expression of these targets was further validated *in vivo*.

### 3.5 SIN regulated the inflammatory response via JNK/NF- $\kappa$ B pathway

The IL-17 pathway was found to be implicated in the effects of SIN on mitigating sepsis-induced cardiac dysfunction in this study via network pharmacology and KEGG enrichment analysis. The targets of SIN involved in the IL-17 signal pathway, including JNK1, NF- $\kappa$ B, IL-1 $\beta$ , *etc.*, were found and illustrated in Figure 4C.

Therefore, the effect of SIN on these targets were detected *in vivo*. As shown in Figure 6A, SIN demonstrated a significant dose-dependent effect in significantly reversing the increased inflammatory cytokines induced by CLP, including TNF- $\alpha$ , IL-1 $\beta$ , IL-6, IFN- $\gamma$ , and CXCL8. Based on molecular docking simulations, SIN was found to exhibit binding affinity with JNK1 and NF- $\kappa$ B, and may have potential impacts on the phosphorylation of these targets. The expression levels of p-JNK1 and p-NF- $\kappa$ B were shown to be significantly elevated by CLP, but reduced after treatment with SIN (Figure 6B). Taken together, these findings implied that the IL-17 signal pathway is expected to attenuate the effect of SIN on sepsis-induced MI.



**TABLE 1** The autodock score and hydrogen bonds of putative targets with sinomenine from molecular docking analysis.

Genes	PDB accession number	Autodock score (kcal/mol)	Hydrogen bonds
JAK1	4ei4	−6.23	Asn1008(OD1):SIN(O3)
			Arg1007(NH1):SIN(O3)
JAK2	7f7w	−5.54	Thr636(OG1):SIN(O2)
			Thr636(OG1):SIN(O1)
			Leu551(O):SIN(O3)
JNK1	3elj	−5.25	Asp112(O):SIN(O3)
			Met111(O):SIN(O3)
NF-κB	1ikn	−6.06	Asp297(N):SIN(O1)
			Phe295(N):SIN(O3)
STAT3	6nuq	−5.59	Glu638(N):SIN(O4)
			Tyr640(OH):SIN(O1)
			Glu638(O):SIN(O4)

Abbreviation: SIN, sinomenine; PDB, protein data bank.

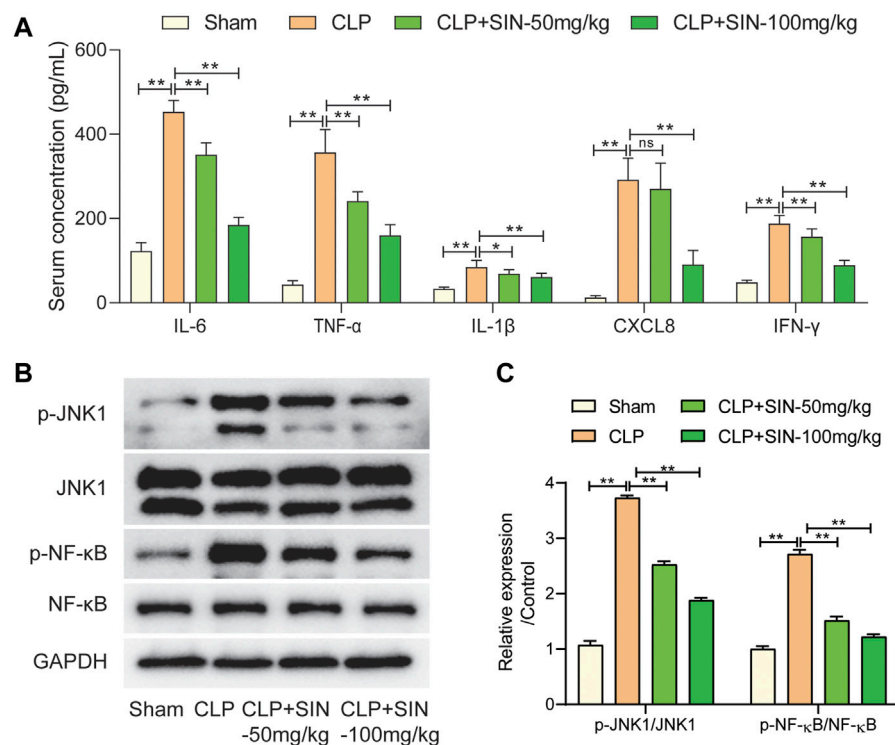


FIGURE 6

The regulation of SIN on the targets in the IL-17 signal pathway. (A) Comparison of serum concentration of inflammatory cytokines in each group. (B,C) The expression levels of JNK1/NF-κB and phosphorylated JNK1/NF-κB were examined.  $n = 6$ . \* $p < 0.05$ , \*\* $p < 0.01$ .

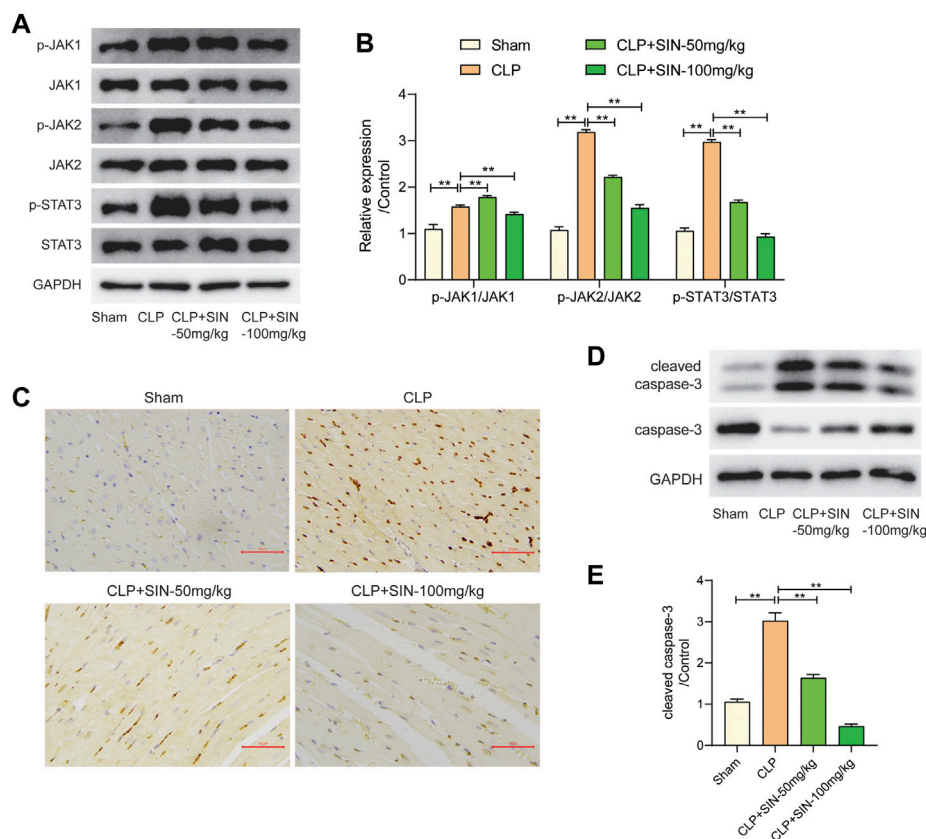
### 3.6 SIN decreased cardiomyocyte apoptosis by regulating JAK/STAT signal pathway

The JAK/STAT axis was associated with the progress of MI, and the biological processes involving JAK/STAT axis were found to be related to the protection of SIN from sepsis-induced MI. Here, we evaluated the impact of SIN on the phosphorylation levels of JAK1, JAK2, and STAT3, along with the apoptosis of myocardial tissues. The findings revealed a notable increase in the apoptotic level of cardiomyocytes and the expression of p-JAK1, p-JAK2 and p-STAT3 in the septic rats as compared with normal rats, whereas SIN treatment exhibited a dose-dependent effect in reducing cardiomyocyte apoptosis and phosphorylation of JAK1, JAK2 and STAT3 (Figures 7A,B). TUNEL assay performed on heart slices revealed an elevated number of TUNEL-positive nuclei in septic rats induced by CLP, whereas SIN demonstrated a dose-dependent effect in reducing the elevated number of TUNEL-positive nuclei in septic rats (Figure 7C). The ratio of cleaved-casp3/caspase-3 was significantly increased in the septic rats compared to the normal rats, but rats with sepsis that received SIN treatment showed a lower ratio of cleaved casp3/caspase-3 (Figures 7D,E). These data corroborated the apoptosis results detected by TUNEL assays. These finding suggested that SIN treatment might ameliorate activated cardiomyocyte apoptosis in rats with sepsis via the JAK/STAT signal pathway.

## 4 Discussion

Sepsis, a major contributor to infection-related death, poses a challenge for healthcare systems around the world. The key to treating sepsis in clinical settings revolves around the administration of antimicrobial medicines to combat underlying infection. In recent decades, natural compounds exhibiting antibacterial and anti-inflammatory properties have been increasingly utilized for the prevention of human diseases during the past few decades (McBride et al., 2020). Natural alkaloid SIN, having anti-inflammatory, immunoregulatory (Liu et al., 2020), anti-angiogenic (Feng et al., 2019) and other diverse pharmacological effects, is obtained from *Sinomenium acutum* Rehder. Previous studies have confirmed that SIN improved lung injury in sepsis by mediating gut homeostasis (Wang et al., 2020), and its hydrochloride salt could protect against polymicrobial sepsis via autophagy (Jiang et al., 2015). However, the potential therapeutic effects of SIN on sepsis-induced MI remain unclear. The findings of the present study suggested that SIN was favorable to improve the cardiac dysfunction of rats with sepsis, which was specifically manifested in the aspects of reducing the mortality rate, improving the cardiac functions, as well as ameliorating the MI of septic rats. Accumulating evidence has demonstrated the multiple and diverse mechanisms underlying the diverse functions of SIN in different diseases. SIN has a significant inhibitory effect on Glioblastoma multiforme (GBM) in advanced gliomas (Zheng



**FIGURE 7**

The effects of SIN on the JAK/STAT pathway and cardiomyocyte apoptosis. (A,B) The expression levels of JAK1, JAK2, STAT3, and phosphorylated JAK1, JAK2, and STAT3 were assessed via Western blot. (C) The apoptosis level of cardiomyocytes was detected by TUNEL staining. (D,E) The protein expression of caspase-3 and cleaved-caspase-3 were evaluated.  $n = 6$ . \* $p < 0.05$ , \*\* $p < 0.01$ .

et al., 2021). It prevents acute lung injury in sepsis by regulating intestinal microbiota and restoring intestinal barrier through aryl hydrocarbon receptor/NRF2-dependent pathway (Song W. et al., 2021), and also regulates the rationality of neuroimmune interaction to exert analgesic effects (Lai et al., 2022). Nonetheless, the specific mechanism of the action of SIN on sepsis-induced MI remains poorly understood.

In this study, a total of 33 putative targets were obtained by merging the SIN-targets and sepsis-related targets, and there were 3 genes with a higher degree: TNF- $\alpha$ , IL-6, and IL-1 $\beta$ . Many studies have confirmed the regulation of SIN on the gene expression or secretion of these 3 cytokines in sepsis (Deng et al., 2022; Pei et al., 2022). The results of the current study also confirmed that SIN could induce increased secretion of TNF- $\alpha$ , IL-6, and IL-1 $\beta$  in CLP-induced sepsis. TNF- $\alpha$  is a vital pro-inflammatory cell cytokine that can trigger inflammatory cascades and cause multiple clinical symptoms in patients with sepsis, such as hypotension, disseminated intravascular coagulation, and organ failure. Recent studies suggested that TNF- $\alpha$  (-238 G/A) polymorphism was associated with the progression of sepsis (Georgescu et al., 2020). It has been considered as promising therapeutic target for treating sepsis and MI (Jin et al., 2013), and anti-TNF- $\alpha$  immunotherapy has been developed for treating sepsis (Qu et al., 2018). IL-6 is another pro-inflammatory cytokine that is critical in immune and inflammatory

responses. High levels of IL-6 have been shown to be associated with an increased risk of severe sepsis and a higher mortality rate (Deng et al., 2022). Additionally, IL-6 (174G/C) polymorphism was proved to be associated with an increased susceptibility to sepsis (Hu et al., 2019). Uncoupling of IL-6 signaling and Microtubule-associated protein 1 light chain 3 (LC3)-associated phagocytosis was reported to cause immunoparalysis during sepsis (Akoumianaki et al., 2021). Genetic variants in IL-1 $\beta$  has been confirmed to be a risk factor for sepsis and MI (Varljen et al., 2020; Pan et al., 2021), and contribute to the clinical course of sepsis (Montoya-Ruiz et al., 2016). Therefore, it suggested that SIN might protect against sepsis-induced MI via targeting TNF- $\alpha$ , IL-6, and IL-1 $\beta$ .

Multiple pathways, including the IL-17 signal pathway, the NOD-like receptor signaling pathway, and the TNF signaling pathway, were identified to be prospective targets of SIN against sepsis. IL-17 is a pro-inflammatory cytokine that could activate Interleukin 16 (IL-16) production (Yao et al., 1995). It is crucial in the development of several malignancies, inflammatory and autoimmune diseases, and infectious diseases. It is of great pathophysiological significance in sepsis via IL-17-mediated response and signal transduction (Ge et al., 2020). In sepsis, recent investigations suggested that IL-17 may function as a biomarker and a therapeutic target (Bosmann and Ward, 2012; Ahmed Ali et al., 2018). In this study, there were nine targets of SIN

involved in the IL-17 signal pathway, including JNK1, NF- $\kappa$ B, CXCL8, IL-6, TNF- $\alpha$ , cyclooxygenase-2 (COX2), IL-1 $\beta$ , IFN- $\gamma$ , and Matrix metalloproteinase-9 (MMP9). According to the results of this study, SIN administration reduced the elevated secretion of inflammatory cytokines, including CXCL8, IL-6, TNF- $\alpha$ , COX2, IL-1 $\beta$ , and IFN- $\gamma$  in the CLP model, and reversed the increased phosphorylation levels of JNK1 and NF- $\kappa$ B induced by CLP. It has been reported that SIN can reduce the phosphorylation levels of JNK1 and NF- $\kappa$ B in macrophages, thus slowing down the inflammatory response caused by Lipopolysaccharides (LSP)-induced sepsis (Teng et al., 2012). A series of compounds were revealed to be effective in improving the LSP-induced sepsis via activation or deactivation of JNK1 and NF- $\kappa$ B (Hsu et al., 2013; Rocca et al., 2021). Meanwhile, JNK1 and NF- $\kappa$ B are closely related to cardiac pathologies (Xia et al., 2016; Singh et al., 2020). As a result, it can be inferred that SIN regulated the inflammatory cytokines via targeting JNK1 and NF- $\kappa$ B in the IL-17 signal pathway, which could eventually contribute to the its protective role in sepsis-induced MI. This finding may be helpful in determining new therapy directions for sepsis-induced MI.

GO analysis demonstrated that the candidate targets of SIN against sepsis were mainly associated with biological processes of cytokines, inflammation, and protein phosphorylation. Notably, our findings indicate a robust engagement of the Janus Kinase-Signal Transducers and Activators of Transcription (JAK-STAT) axis of the candidate targets of SIN against sepsis-induced MI. Western blot analysis showed a significant regulation of the phosphorylation of JAK2 and STAT3 after SIN exposure in CLP rats. JAKs-STATs signal pathways, known as the pivotal downstream signaling components of cytokine receptors, play a crucial role in mediating the biological effects of cytokines (Villarino et al., 2017). It contributes to organ damage and other dysfunctions in sepsis and offers novel therapeutic possibilities for sepsis (Cai et al., 2015; Clere-Jehl et al., 2020). Moreover, the JAK-STAT pathway is an integral part of myocardial response to various cardiac injuries and plays a prominent role in cardioprotective therapies (Barry et al., 2007). Cardiomyocyte apoptosis is robustly confirmed to be associated with the development of sepsis (Li et al., 2019), and regulated via the JAK-STAT pathway (Zhang et al., 2022). Therefore, we detected the cardiomyocyte apoptosis in each group in this study, and the findings demonstrated that the level of cardiomyocyte apoptosis in SIN groups was much lower than that in the CLP group. These findings suggested that SIN may prevent MI in sepsis by controlling cardiomyocyte apoptosis through the JAK-STAT pathway.

## 5 Conclusion

In conclusion, SIN improved the mortality rate and cardiac function of septic rats, and ameliorated sepsis-induced MI. Potential targets and pathways of SIN against sepsis were identified through network pharmacology analysis integrating molecular docking

simulation. The proposition that SIN protects against sepsis-induced MI via targeting multiple proteins and regulating cytokine secretion and cardiomyocyte apoptosis was finally experimentally validated.

## 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

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of Zhejiang Center of Laboratory Animals (Reference Number: ZJCLA-IACUC-20020101).

## Author contributions

LS and ZC contributed to the concept and design of the study. LS wrote the first draft of the manuscript. YN and ZH wrote sections of the manuscript. ZC, YN and ZH performed the experiments. LS, ZC and YN designed and/or performed the statistical analysis. 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.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2023.1138858/full#supplementary-material>

## References

- Ahmed Ali, M., Mikhael, E. S., Abdelkader, A., Mansour, L., El Essawy, R., El Sayed, R., et al. (2018). Interleukin-17 as a predictor of sepsis in polytrauma patients: A prospective cohort study. *Eur. J. Trauma Emerg. Surg. Off. Publ. Eur. Trauma Soc.* 44 (4), 621–626. doi:10.1007/s00068-017-0841-3
- Akoumianaki, T., Vaporidi, K., Diamantaki, E., Pène, F., Beau, R., Gresnigt, M. S., et al. (2021). Uncoupling of IL-6 signaling and LC3-associated phagocytosis drives immunoparalysis during sepsis. *Cell Host Microbe* 29 (8), 1277–1293.e6. doi:10.1016/j.chom.2021.06.002
- Barry, S. P., Townsend, P. A., Latchman, D. S., and Stephanou, A. (2007). Role of the JAK-STAT pathway in myocardial injury. *Trends Mol. Med.* 13 (2), 82–89. doi:10.1016/j.molmed.2006.12.002
- Bayer, J., Vaghela, R., Drechsler, S., Osuchowski, M. F., Erben, R. G., and Andrukhova, O. (2021). The bone is the major source of high circulating intact fibroblast growth factor-23 in acute murine polymicrobial sepsis induced by cecum ligation puncture. *PLoS One* 16 (5), e0251317. doi:10.1371/journal.pone.0251317
- Bosmann, M., and Ward, P. A. (2012). Therapeutic potential of targeting IL-17 and IL-23 in sepsis. *Clin. Transl. Med.* 1 (1), 4. doi:10.1186/2001-1326-1-4
- Cai, B., Cai, J. P., Luo, Y. L., Chen, C., and Zhang, S. (2015). The specific roles of JAK/STAT signaling pathway in sepsis. *Inflammation* 38 (4), 1599–1608. doi:10.1007/s10753-015-0135-z
- Clere-Jehl, R., Mariotte, A., Meziani, F., Bahram, S., Georgel, P., and Helms, J. (2020). JAK-STAT targeting offers novel therapeutic opportunities in sepsis. *Trends Mol. Med.* 26 (11), 987–1002. doi:10.1016/j.molmed.2020.06.007
- Corrales, M., Sierra, A., Doizi, S., and Traxer, O. (2022). Risk of sepsis in retrograde intrarenal surgery: A systematic review of the literature. *Eur. Urol. Open Sci.* 44, 84–91. doi:10.1016/j.euros.2022.08.008
- Deng, P., Tang, N., Li, L., Zou, G., Xu, Y., and Liu, Z. (2022). Diagnostic value of combined detection of IL-1 $\beta$ , IL-6, and TNF- $\alpha$  for sepsis-induced cardiomyopathy. *Med. Clin. Barc.* 158 (9), 413–417. doi:10.1016/j.medcli.2021.04.025
- Feng, Z. T., Yang, T., Hou, X. Q., Wu, H. Y., Feng, J. T., Ou, B. J., et al. (2019). Sinomenine mitigates collagen-induced arthritis mice by inhibiting angiogenesis. *Biomed. Pharmacother.* 113, 108759. doi:10.1016/j.biopha.2019.108759
- 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
- Georgescu, A. M., Banescu, C., Azamfirei, R., Hutanu, A., Moldovan, V., Badea, I., et al. (2020). Evaluation of TNF- $\alpha$  genetic polymorphisms as predictors for sepsis susceptibility and progression. *BMC Infect. Dis.* 20 (1), 221. doi:10.1186/s12879-020-4910-6
- Hsu, C. C., Lien, J. C., Chang, C. W., Chang, C. H., Kuo, S. C., and Huang, T. F. (2013). Yuwen02f1 suppresses LPS-induced endotoxemia and adjuvant-induced arthritis primarily through blockade of ROS formation, NFkB and MAPK activation. *Biochem. Pharmacol.* 85 (3), 385–395. doi:10.1016/j.bcp.2012.11.002
- Hu, P., Chen, Y., Pang, J., and Chen, X. (2019). Association between IL-6 polymorphisms and sepsis. *Innate Immun.* 25 (8), 465–472. doi:10.1177/1753425919872818
- Jiang, Y., Gao, M., Wang, W., Lang, Y., Tong, Z., Wang, K., et al. (2015). Sinomenine hydrochloride protects against polymicrobial sepsis via autophagy. *Int. J. Mol. Sci.* 16 (2), 2559–2573. doi:10.3390/ijms16022559
- Jiang, W., Fan, W., Gao, T., Li, T., Yin, Z., Guo, H., et al. (2020). Analgesic mechanism of sinomenine against chronic pain. *Pain Res. Manag.* 2020, 1876862. doi:10.1155/2020/1876862
- Jin, J., Chen, F., Wang, Q., Qiu, Y., Zhao, L., and Guo, Z. (2013). Inhibition of TNF- $\alpha$  by cyclophosphamide reduces myocardial injury after ischemia-reperfusion. *Ann. Thorac. Cardiovasc. Surg. official J. Assoc. Thorac. Cardiovasc. Surg. Asia* 19 (1), 24–29. doi:10.5761/atcs.0a.11.01877
- Lai, W. D., Wang, S., You, W. T., Chen, S. J., Wen, J. J., Yuan, C. R., et al. (2022). Sinomenine regulates immune cell subsets: Potential neuro-immune intervene for precise treatment of chronic pain. *Front. Cell Dev. Biol.* 10, 1041006. doi:10.3389/fcell.2022.1041006
- Li, N., Zhou, H., Wu, H., Wu, Q., Duan, M., Deng, W., et al. (2019). STING-IRF3 contributes to lipopolysaccharide-induced cardiac dysfunction, inflammation, apoptosis and pyroptosis by activating NLRP3. *Redox Biol.* 24, 102125. doi:10.1016/j.redox.2019.102125
- Li, X. M., Li, M. T., Jiang, N., Si, Y. C., Zhu, M. M., Wu, Q. Y., et al. (2021). Network pharmacology-based approach to investigate the molecular targets of sinomenine for treating breast cancer. *Cancer Manag. Res.* 13, 1189–1204. doi:10.2147/CMAR.S282684
- Lin, S. P., Wei, J. X., Hu, J. S., Bu, J. Y., Zhu, L. D., Li, Q., et al. (2021). Artemisinin improves neurocognitive deficits associated with sepsis by activating the AMPK axis in microglia. *Acta Pharmacol. Sin.* 42 (7), 1069–1079. doi:10.1038/s41401-021-00634-3
- Liu, W., Zhang, Y., Zhu, W., Ma, C., Ruan, J., Long, H., et al. (2018). Sinomenine inhibits the progression of rheumatoid arthritis by regulating the secretion of inflammatory cytokines and monocyte/macrophage subsets. *Front. Immunol.* 9, 2228. doi:10.3389/fimmu.2018.02228
- Liu, Y., Sun, Y., Zhou, Y., Tang, X., Wang, K., Ren, Y., et al. (2020). Sinomenine hydrochloride inhibits the progression of plasma cell mastitis by regulating IL-6/JAK2/STAT3 pathway. *Int. Immunopharmacol.* 81, 106025. doi:10.1016/j.intimp.2019.106025
- Liu, Y., Yang, W., Sun, X., Sang, M., and Jiao, R. (2021). Protective effect of sinomenine on lipopolysaccharide-induced injury of H9c2 rat cardiomyocytes (in Chinese). *Her. Med.* 40 (05), 611–616.
- McBride, M. A., Patil, T. K., Bohannon, J. K., Hernandez, A., Sherwood, E. R., and Patil, N. K. (2020). Immune checkpoints: Novel therapeutic targets to attenuate sepsis-induced immunosuppression. *Front. Immunol.* 11, 624272. doi:10.3389/fimmu.2020.624272
- McLaughlin, J., Chowdhury, N., Djurkovic, S., Shahab, O., Sayiner, M., Fang, Y., et al. (2020). Clinical outcomes and financial impacts of malnutrition in sepsis. *Nutr. Health* 26 (3), 175–178. doi:10.1177/0260106020930145
- Mishra, S. K., and Choudhury, S. (2018). Experimental protocol for cecal ligation and puncture model of polymicrobial sepsis and assessment of vascular functions in mice. *Methods Mol. Biol.* 1717, 161–187. doi:10.1007/978-1-4939-7526-6\_14
- Montoya-Ruiz, C., Jaimes, F. A., Rugeles, M. T., López, J., Bedoya, G., and Velilla, P. A. (2016). Variants in LTA, TNF, IL1B and IL10 genes associated with the clinical course of sepsis. *Immunol. Res.* 64 (5–6), 1168–1178. doi:10.1007/s12026-016-8860-4
- Pan, Q., Hui, D., and Hu, C. (2021). A variant of IL1B is associated with the risk and blood lipid levels of myocardial infarction in eastern Chinese individuals. *Immunol. Investig.* 51, 1162–1169. doi:10.1080/08820139.2021.1914081
- Pei, X., Wu, Y., Yu, H., Li, Y., Zhou, X., Lei, Y., et al. (2022). Protective role of lncRNA TTN-AS1 in sepsis-induced myocardial injury via miR-29a/E2F2 Axis. *Cardiovasc. drugs Ther.* 36 (3), 399–412. doi:10.1007/s10557-021-07244-5
- Qu, W., Han, C., Li, M., Zhang, J., and Jiang, Z. (2018). Anti-TNF- $\alpha$  antibody alleviates insulin resistance in rats with sepsis-induced stress hyperglycemia. *J. Endocrinol. investigation* 41 (4), 455–463. doi:10.1007/s40618-017-0742-7
- Rocca, C., De Bartolo, A., Grande, F., Rizzuti, B., Pasqua, T., Giordano, F., et al. (2021). Cateslytin abrogates lipopolysaccharide-induced cardiomyocyte injury by reducing inflammation and oxidative stress through toll like receptor 4 interaction. *Int. Immunopharmacol.* 94, 107487. doi:10.1016/j.intimp.2021.107487
- Salomao, R., Ferreira, B. L., Salomao, M. C., Santos, S. S., Azevedo, L. C. P., and Brunialti, M. K. C. (2019). Sepsis: Evolving concepts and challenges. *Braz. J. Med. Biol. Res.* 52 (4), e8595. doi:10.1590/1414-431X20198595
- Shiroorkar, P. N., Afzal, O., Kazmi, I., Al-Abbasi, F. A., Altamimi, A. S. A., Gubbiyappa, K. S., et al. (2020). Cardioprotective effect of tangeretin by inhibiting PTEN/AKT/mTOR Axis in experimental sepsis-induced myocardial dysfunction. *Molecules* 25 (23), 5622. doi:10.3390/molecules25235622
- Singh, K., Gupta, A., Sarkar, A., Gupta, I., Rana, S., Sarkar, S., et al. (2020). Arginyltransferase knockdown attenuates cardiac hypertrophy and fibrosis through TAK1-JNK1/2 pathway. *Sci. Rep.* 10 (1), 598. doi:10.1038/s41598-019-57379-7
- Song, L., Zhang, H., Hu, M., Liu, C., Zhao, Y., Zhang, S., et al. (2021). Sinomenine inhibits hypoxia induced breast cancer side population cells metastasis by PI3K/Akt/mTOR pathway. *Bioorg. Med. Chem.* 31, 115986. doi:10.1016/j.bmc.2020.115986
- Song, W., Yang, X., Wang, W., Wang, Z., Wu, J., and Huang, F. (2021b). Sinomenine ameliorates septic acute lung injury in mice by modulating gut homeostasis via aryl hydrocarbon receptor/Nrf2 pathway. *Eur. J. Pharmacol.* 912, 174581. doi:10.1016/j.ejphar.2021.174581
- Teng, P., Liu, H. L., Zhang, L., Feng, L. L., Huai, Y., Deng, Z. S., et al. (2012). Synthesis and biological evaluation of novel sinomenine derivatives as anti-inflammatory agents. *Eur. J. Med. Chem.* 50, 63–74. doi:10.1016/j.ejmech.2012.01.036
- Varljen, T., Sekulovic, G., Rakic, O., Maksimovic, N., Jekic, B., Novakovic, I., et al. (2020). Genetic variant rs16944 in IL1B gene is a risk factor for early-onset sepsis susceptibility and outcome in preterm infants. *Inflamm. Res.* 69 (2), 155–157. doi:10.1007/s00011-019-01301-4
- Villarino, A. V., Kanno, Y., and O'Shea, J. J. (2017). Mechanisms and consequences of Jak-STAT signaling in the immune system. *Nat. Immunol.* 18 (4), 374–384. doi:10.1038/ni.3691
- Wang, W., Yang, X., Chen, Q., Guo, M., Liu, S., Liu, J., et al. (2020). Sinomenine attenuates septic-associated lung injury through the Nrf2-Keap1 and autophagy. *J. Pharm. Pharmacol.* 72 (2), 259–270. doi:10.1111/jphp.13202
- Xia, J. B., Liu, G. H., Chen, Z. Y., Mao, C. Z., Zhou, D. C., Wu, H. Y., et al. (2016). Hypoxia/ischemia promotes CXCL10 expression in cardiac microvascular endothelial cells by NFkB activation. *Cytokine* 81, 63–70. doi:10.1016/j.cyto.2016.02.007



- Xie, S., Qi, X., Wu, Q., Wei, L., Zhang, M., Xing, Y., et al. (2021). Inhibition of 5-lipoxygenase is associated with downregulation of the leukotriene B4 receptor 1/Interleukin-12p35 pathway and ameliorates sepsis-induced myocardial injury. *Free Radic. Biol. Med.* 166, 348–357. doi:10.1016/j.freeradbiomed.2021.02.034
- Xu, P., Zhang, W. Q., Xie, J., Wen, Y. S., Zhang, G. X., and Lu, S. Q. (2020). Shenfu injection prevents sepsis-induced myocardial injury by inhibiting mitochondrial apoptosis. *J. Ethnopharmacol.* 261, 113068. doi:10.1016/j.jep.2020.113068
- Yao, Z., Fanslow, W. C., Seldin, M. F., Rousseau, A. M., Painter, S. L., Comeau, M. R., et al. (1995). Herpesvirus Saimiri encodes a new cytokine, IL-17, which binds to a novel cytokine receptor. *Immunity* 3 (6), 811–821. doi:10.1016/1074-7613(95)90070-5
- Zeng, M. Y., and Tong, Q. Y. (2020). Anti-inflammation effects of sinomenine on macrophages through suppressing activated TLR4/NF- $\kappa$ B signaling pathway. *Curr. Med. Sci.* 40 (1), 130–137. doi:10.1007/s11596-020-2156-6
- Zhang, Q., Wang, L., Wang, S., Cheng, H., Xu, L., Pei, G., et al. (2022). Signaling pathways and targeted therapy for myocardial infarction. *Signal Transduct. Target. Ther.* 7 (1), 78. doi:10.1038/s41392-022-00925-z
- Zheng, X., Li, W., Xu, H., Liu, J., Ren, L., Yang, Y., et al. (2021). Sinomenine ester derivative inhibits glioblastoma by inducing mitochondria-dependent apoptosis and autophagy by PI3K/AKT/mTOR and AMPK/mTOR pathway. *Acta Pharm. Sin. B* 11 (11), 3465–3480. doi:10.1016/j.apsb.2021.05.027
- Zhu, M., Wang, H., Chen, J., and Zhu, H. (2021). Sinomenine improve diabetic nephropathy by inhibiting fibrosis and regulating the JAK2/STAT3/SOCS1 pathway in streptozotocin-induced diabetic rats. *Life Sci.* 265, 118855. doi:10.1016/j.lfs.2020.118855



## OPEN ACCESS

## EDITED BY

Carlos F. Lagos,  
Universidad San Sebastian, Chile

## REVIEWED BY

Berna Dogan,  
Istanbul Technical University, Türkiye  
Anupam Nath Jha,  
Tezpur University, India

## \*CORRESPONDENCE

Julio Caballero,  
✉ jcaballero@utalca.cl

RECEIVED 03 May 2023

ACCEPTED 12 June 2023

PUBLISHED 23 June 2023

## CITATION

Castillo-Campos L, Velázquez-Libera JL and Caballero J (2023), Computational study of the binding orientation and affinity of noncovalent inhibitors of the papain-like protease (PLpro) from SARS-CoV-1 considering the protein flexibility by using molecular dynamics and cross-docking.

*Front. Mol. Biosci.* 10:1215499.

doi: 10.3389/fmolb.2023.1215499

## COPYRIGHT

© 2023 Castillo-Campos, Velázquez-Libera and Caballero. 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.

# Computational study of the binding orientation and affinity of noncovalent inhibitors of the papain-like protease (PLpro) from SARS-CoV-1 considering the protein flexibility by using molecular dynamics and cross-docking

Luis Castillo-Campos, José Luis Velázquez-Libera and Julio Caballero\*

Centro de Bioinformática, Simulación y Modelado (CBSM), Facultad de Ingeniería, Universidad de Talca, Talca, Chile

The papain-like protease (PLpro) from zoonotic coronaviruses (CoVs) has been identified as a target with an essential role in viral respiratory diseases caused by Severe Acute Respiratory Syndrome-associated coronaviruses (SARS-CoVs). The design of PLpro inhibitors has been proposed as an alternative to developing potential drugs against this disease. In this work, 67 naphthalene-derived compounds as noncovalent PLpro inhibitors were studied using molecular modeling methods. Structural characteristics of the bioactive conformations of these inhibitors and their interactions at the SARS-CoV-1 PLpro binding site were reported here in detail, taking into account the flexibility of the protein residues. Firstly, a molecular docking protocol was used to obtain the orientations of the inhibitors. After this, the orientations were compared, and the recurrent interactions between the PLpro residues and ligand chemical groups were described (with LigRMSD and interaction fingerprints methods). In addition, efforts were made to find correlations between docking energy values and experimentally determined binding affinities. For this, the PLpro was sampled by using Gaussian Accelerated Molecular Dynamics (GaMD), generating multiple conformations of the binding site. Diverse protein conformations were selected and a cross-docking experiment was performed, yielding models of the 67 naphthalene-derived compounds adopting different binding modes. Representative complexes for each ligand were selected to obtain the highest correlation between docking energies and activities. A good correlation ( $R^2 = 0.948$ ) was found when this flexible docking protocol was performed.

## KEYWORDS

papain-like protease, PLpro inhibitors, SARS-CoV, naphthalene-derived compounds, docking, molecular dynamics

## 1 Introduction

Zoonotic coronaviruses (CoVs) are important viral pathogens whose most recent species, the Severe Acute Respiratory Syndrome (SARS)-CoV-2, has been causing a worldwide emergency due to its rapid spread since the end of 2019. Previous CoV events caused by the SARS-CoV-1 (2002–2003) and the Middle East Respiratory Syndrome (MERS)-CoV (2012) were antecedents that showed the danger constituted by CoVs. After the SARS-CoV-1 appeared in Guangdong province in China in November 2002, affecting three continents and causing many deaths (Wang and Chang, 2004), researchers investigated the mechanisms of viral infection to discover options to provide treatment for patients infected with zoonotic CoVs. The results of these investigations made it possible to identify molecular targets currently being investigated to find specific drugs against CoVs. Research to modulate these targets has included the repurposing of already approved drugs (De Savi et al., 2020; Gordon et al., 2020; Indari et al., 2022; Khataniar et al., 2022) and the design of new specific drugs (Cannalire et al., 2022).

Infection with CoVs triggers the encoding of several protein targets with recognized functions relevant to the virus infection. The proteases 3CLpro and PLpro were identified as responsible for preprocessing translated multidomain polypeptides from the viral RNA genome (Hilgenfeld, 2014; Zhu et al., 2021). Since 2003, details of the structure and functions of 3CLpro have been reported; its structural and mechanistic aspects have been elucidated, offering multiple avenues as starting points for the design of antiviral compounds directed against CoVs (Ullrich and Nitsche, 2020). On the other hand, the less studied PLpro also plays critical biochemical events for coronavirus replication. It is vital in viral pathogenesis and is associated with processes of deubiquitination and deISGylation of host cell proteins (Báez-Santos et al., 2015). In association with viral protein processing, its enzymatic activity triggers the host antiviral immune response antagonism.

The architecture of PLpro consists of four domains: the palm domain, the thumb, the fingers, and an independent terminal domain similar to the ubiquitin domains. The binding site of PLpro is at the intersection between the palm and thumb domains (Ratia et al., 2006), formed by a catalytic triad composed of the residues Cys112-His273-Asp287 (in the SARS-CoV-1 PLpro) and subsites that can be specifically occupied by the substrate RLRGG (the C-terminus of ubiquitin). Closed and open conformations of the binding site are available because of structural changes in the six-residue BL2 loop, modulating substrate recognition (Chaudhuri et al., 2011).

Targeting PLpro has become an attractive strategy to stop the viral replication and infection caused by CoVs. In this sense, the design of PLpro inhibitors has been proposed (Calleja et al., 2022). In recent years, Ratia et al. synthesized a series of noncovalent naphthalene-derived compounds as SARS-CoV-1 PLpro inhibitors by high-throughput screening (Ratia et al., 2008; Ghosh et al., 2009; Ghosh et al., 2010; Báez-Santos et al., 2014a). They act as reversible competitive PLpro inhibitors by binding to the S<sub>3</sub>-S<sub>4</sub> subsites (Supplementary Figure S1). When bound, these compounds induce the reorientation of the Y269 side chain, generating the closure of the BL2 loop. Some of these compounds were co-crystallized with PLpro, allowing an initial source to generate more structural information explaining what structural aspects

contribute to the differences in the reported activities. With this in mind, we carried out computational modeling studies of the congeneric family of 67 naphthalene-derived compounds reported by Ratia et al. (2008); Ghosh et al. (2010), Báez-Santos et al. (2014a), Ghosh et al. (2009), providing relevant information about their binding modes and the causes of their differential activities. We assumed this information could be helpful for designing new potential PLpro inhibitors.

## 2 Materials and methods

### 2.1 Preparation of naphthalene-derived compounds

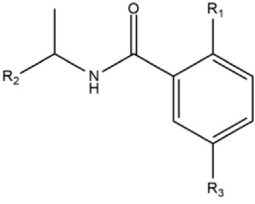
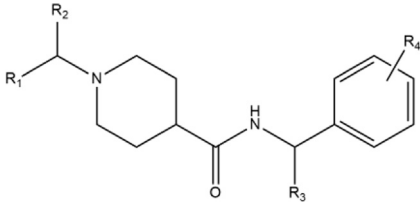
The 67 structures of naphthalene-derived compounds and their IC<sub>50</sub> values were collected from references of Ratia et al. (2008), Ghosh et al. (2010), Báez-Santos et al. (2014a), Ghosh et al. (2009). The chemical structures for each compound are in Table 1. Each compound has a name formed by the letters A, B, C, and D to differentiate the article of origin, followed by the compound identification in the article (compounds from references (Ratia et al., 2008; Ghosh et al., 2010; Báez-Santos et al., 2014a), and (Ghosh et al., 2009) are named A\_x, B\_x, C\_x, and D\_x, respectively). Table 1 represents a set of 24 inhibitors (compounds A\_x and D\_x) that contain a benzamide and a set of 43 compounds (compounds B\_x and C\_x) that contain a piperidine ring.

The structures were drawn in Maestro Molecular Editor (Maestro 12.8.117, Schrödinger LLC, New York, NY, USA, 2021) and processed using the Maestro's module LigPrep. The protonation states were estimated using Epik (Shelley et al., 2007) under a physiological pH value of 7. In the case of compounds containing two possible enantiomers or presented in racemic form, both were chosen for molecular docking experiments to explore interactions at the PLpro binding site.

### 2.2 Preparation of SARS-CoV-1 PLpro structures

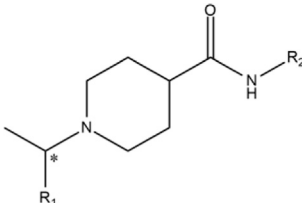
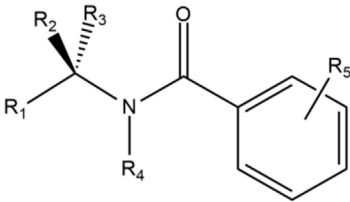
The three-dimensional (3D) crystallographic structures of the SARS-CoV-1 PLpro were obtained from the Protein Data Bank (PDB). We selected those structures co-crystallized with non-covalent inhibitors derived from naphthalene in the S<sub>3</sub> and S<sub>4</sub> sub-sites of the protease. Four PLpro-ligand structures were selected with the PDB IDs 3E9S (with GRL0617, resolution 2.50 Å) (Ratia et al., 2008), 3MJ5 (with B\_15g, resolution 2.63 Å) (Ghosh et al., 2010), 4OVZ (with C\_3j, resolution 2.50 Å), and 4OW0 (with C\_3k, resolution 2.10 Å) (Báez-Santos et al., 2014a). The Protein Preparation Wizard (Schrödinger LLC, New York, NY, USA, 2021) was used to improve PDB models. Missing atoms were assigned, and hydrogen atoms were added to have all the atoms represented and positioned explicitly. Crystallographic water molecules were removed, and native zinc ions were retained. Hydrogen bonding networks were optimized by reorienting hydroxyl groups, thiol groups, asparagine and glutamine amide groups, and histidine imidazole rings. Predictions of the protonation states of the ionizable groups were performed.

TABLE 1 Structures and activities of naphthalene-derived compounds as SARS-CoV-2 PLpro inhibitors.

						
Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	IC <sub>50</sub> (μM)
7724772 (S)	Me	2-naphthyl	H	—	—	>200
7724772 (R)	Me	2-naphthyl	H	—	—	8.7 ± 0.7
A_3 (R)	Me	2-naphthyl	H	—	—	14.5 ± 0.9
A_4 (R)	Me	2-naphthyl	H	—	—	>200
A_5 (R)	Me	1-naphthyl	H	—	—	2.3 ± 0.1
A_6 (R)	Me	1-naphthyl	NHAc	—	—	2.6 ± 0.1
A_7 (R)	Me	1-naphthyl	NO <sub>2</sub>	—	—	7.3 ± 0.9
GRL0617	Me	1-naphthyl	NH <sub>2</sub>	—	—	0.6 ± 0.1
						
65778771	1-naphthyl	H	H	2-OMe	—	59.2 ± 7.8
B_7a	1-naphthyl	H	H	4-OMe	—	116 ± 30
B_7b	1-naphthyl	H	H	3-OMe	—	30 ± 3
B_15a	1-naphthyl	Me	H	4-OMe	—	1.21 ± 0.04
B_15b	1-naphthyl	Me	H	3-OMe	—	0.34 ± 0.01
B_15c	1-naphthyl	Me	H	2-OMe	—	0.34 ± 0.01
B_15d	2-naphthyl	Me	H	3-OMe	—	13.2 ± 0.6
B_15e	2-naphthyl	Me	H	4-OMe	—	34.8 ± 4.0
B_15f	2-naphthyl	Me	H	3-OMe	—	5.8 ± 0.1
B_15g (R)	1-naphthyl	Me	H	3,4-O-CH <sub>2</sub> -O	—	0.32 ± 0.01
B_15h (S)	1-naphthyl	Me	H	3,4-O-CH <sub>2</sub> -O	—	0.56 ± 0.03
B_15i	1-naphthyl	H	H	3,4-O-CH <sub>2</sub> -O	—	~45
B_15j	2-naphthyl	H	H	3,4-O-CH <sub>2</sub> -O	—	~100
B_15k	1-naphthyl	Gem-dimethyl	H	3,4-O-CH <sub>2</sub> -O	—	>200
C_1a (R,S)	1-naphthyl	CH <sub>2</sub> Me	H	3,4-O-CH <sub>2</sub> -O	—	17.2 ± 0.03
C_1b (R,S)	1-naphthyl	CH <sub>2</sub> OH	H	3,4-O-CH <sub>2</sub> -O	—	32.0 ± 4.5
C_1c (R,S)	1-naphthyl	CH <sub>2</sub> OMe	H	3,4-O-CH <sub>2</sub> -O	—	>100
C_1d (R,S)	1-naphthyl	CH <sub>2</sub> Ph	H	3,4-O-CH <sub>2</sub> -O	—	>100
C_2a (R)	1-naphthyl	Me	H	H	—	2.2 ± 0.1
C_2b (R)	1-naphthyl	Me	(R)-Me	H	—	13.5 ± 1.2
C_2c (R)	1-naphthyl	Me	(S)-Me	H	—	12.7 ± 0.3

(Continued on following page)

TABLE 1 (Continued) Structures and activities of naphthalene-derived compounds as SARS-CoV-1 PLpro inhibitors.

Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	IC <sub>50</sub> (μM)
C_2d (R)	1-naphthyl	Me	(R)-CH <sub>2</sub> OMe	H	—	18.0 ± 1.9
C_2e (R)	1-naphthyl	Me	(S)-CH <sub>2</sub> OMe	H	—	1.9 ± 0.1
C_3a (R)	1-naphthyl	Me	H	4-Et	—	0.47 ± 0.01
C_3b (R)	1-naphthyl	Me	H	4-CO-NH-Me	—	0.60 ± 0.02
C_3c (R)	1-naphthyl	Me	H	3-CO-NH-Me	—	0.63 ± 0.01
C_3d (R)	1-naphthyl	Me	H	4-NH-CO-Me	—	5.7 ± 0.5
C_3e (R)	1-naphthyl	Me	H	3-NH-CO-Me	—	0.39 ± 0.01
C_3f (R)	1-naphthyl	Me	H	3-CH <sub>2</sub> -NH-CO-Me	—	20.4 ± 1.2
C_3g (R)	1-naphthyl	Me	H	3-Cl	—	27.2 ± 4.1
C_3h (R)	1-naphthyl	Me	H	4-Cl	—	0.58 ± 0.02
C_3i (R)	1-naphthyl	Me	H	3,4-diF	—	29.2 ± 2.1
C_3j (R)	1-naphthyl	Me	H	4-F	—	0.49 ± 0.01
C_3k (R)	1-naphthyl	Me	H	3-F	—	0.15 ± 0.01
						
C_4a (R,S)	8-quinolinyl	3-F-Ph-CH <sub>2</sub>	—	—	—	7.0 ± 0.7
C_4b (R,S)	5-quinolinyl	3-F-Ph-CH <sub>2</sub>	—	—	—	4.5 ± 0.2
C_4c (R,S)	5-isoquinolinyl	3-F-Ph-CH <sub>2</sub>	—	—	—	6.8 ± 0.3
C_4d (R,S)	1-isoquinolinyl	3-F-Ph-CH <sub>2</sub>	—	—	—	30.8 ± 2.6
C_5a (R)	1-naphthyl	3-pyridinyl-CH <sub>2</sub>	—	—	—	26.3 ± 2.3
C_5b (R)	1-naphthyl	4-pyridinyl-CH <sub>2</sub>	—	—	—	18.3 ± 0.9
C_5c (R)	1-naphthyl	2-methoxy-4-pyridinyl-CH <sub>2</sub>	—	—	—	0.35 ± 0.02
C_6a (R)	1-naphthyl	4-Cl-Ph-CH <sub>2</sub> CH <sub>2</sub>	—	—	—	1.6 ± 0.3
C_6b (R)	1-naphthyl	3-F-Ph-CH <sub>2</sub> CH <sub>2</sub>	—	—	—	1.9 ± 0.1
						
D_2	1-naphthyl	Me	H	H	2-Me and 5-CH <sub>2</sub> NH <sub>2</sub>	0.46 ± 0.03
D_5a	2-naphthyl	Me	H	H	3-Me	14.8 ± 5.0
D_5b	2-naphthyl	Me	H	H	4-Me	29.1 ± 3.8
D_5c	2-naphthyl	Me	H	H	2-OMe	90 ± 26
D_5d	2-naphthyl	Me	H	H	3-OMe	13.5 ± 6.8
D_5e	2-naphthyl	Me	H	H	4-OMe	149 ± 43
D_5f	2-naphthyl	Me	H	H	2,6-diMe	12.1 ± 0.7

(Continued on following page)



**TABLE 1 (Continued) Structures and activities of naphthalene-derived compounds as SARS-CoV-2 PLpro inhibitors.**

Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	IC <sub>50</sub> (μM)
D_9	2-naphthyl	Me	H	H	4-NH <sub>2</sub>	46.1 ± 13.0
D_21	1-naphthyl	Me	H	Me	2-Me	22.6 ± 6.9
D_23	1-naphthyl	Me	H	H	4-NH <sub>2</sub>	24.8 ± 1.0
D_29	1-naphthyl	Me	Me	H	2-Me and 5-NH <sub>2</sub>	11.1 ± 1.3
D_33	1-naphthyl	Me	H	H	2-Me and 5-CN	5.2 ± 0.5
D_40	1-naphthyl	Me	H	H	2-CH <sub>2</sub> OMe and 5-NH <sub>2</sub>	2.7 ± 0.1
D_32	1-naphthyl	Me	H	H	2-Me and 5-I	1.4 ± 0.3
D_47	1-naphthyl	Me	H	H	2-Me and 5-CH <sub>2</sub> NHBoc	4.8 ± 0.4
D_49	1-naphthyl	Me	H	H	2-Me and 5-CH <sub>2</sub> NHMe	1.3 ± 0.1

Finally, the structures were minimized using the OPLS force field (Harder et al., 2016).

Given the conformational diversity of the binding site of the naphthalene derivatives, we performed a previous analysis of the structures to know in more detail about the flexibility of the binding site of these compounds. For that, we aligned the structures coded 3MJ5, 4OVZ, and 4OW0 with the 3E9S structure and compared the orientation of the residues distributed at 5 Å from the ligand with root-mean-square deviation (RMSD) calculations (details in the [Supplementary Table S1](#)). This information was used as material for the development of subsequent analyses. In the comparison between structures, very little variation was observed in the conformations of the active site residues. Two orientations for Gln270 were observed: the first orientation at 3E9S and the second orientation at 3MJ5, 4OVZ, and 4OW0. The remaining residues did not show significant differences among them. Therefore, we selected 3E9S and 4OW0 (the latter having a better resolution than its analogues) to perform the docking calculations.

## 2.3 Docking calculations

Ligand-receptor docking calculations were performed using Glide from the Schrödinger suite to obtain binding modes (Friesner et al., 2004). The ligand array was docked inside the protein binding site using a 20 Å × 20 Å × 20 Å grid centered on residues corresponding to PLpro subsites S<sub>3</sub> and S<sub>4</sub>. Glide standard (SP) and extra precision (XP) modules were used. Glide SP is a more indulgent function and allows the identification of ligands with a reasonable tendency to bind. On the other hand, the extra precision module (XP) is a more strict function, which penalizes poses that violate physical-chemistry principles (Friesner et al., 2006). Using these modules together allowed access to good quality solutions. Glide SP was used to evaluate the ability of the protocol to find poses with similar interactions to those present in the crystallographic structures; meanwhile, the less indulgent XP function was used to obtain the final docking poses, which were used to start the analysis.

Default settings were used, where a flexible ligand was sampled in a rigid protein. Firstly, conformers were generated for each ligand. During this process, ring conformations were discarded if their

energies were higher than that of the lowest conformation by more than 2.5 kcal/mol. No more than 5000 poses per ligand were selected to pass to the grid refinement calculation. The rough-score cutoff (relative to the best rough score accumulated so far) for keeping poses for refinement was 100. Then, at most 400 poses (in SP) or 800 poses (in XP) per ligand were kept for energy minimization. During minimization, the distance-dependent dielectric constant setting was 2.0, and the maximum number of minimization steps (conjugate gradient minimization algorithm) was 100. The best five poses were considered for selecting the best pose.

The best pose for each ligand was chosen by employing two criteria. The first one corresponds to a score-based criterion, where the Emodel score was considered to find the best pose for a given ligand and the GlideScore to rank compounds based on their binding to the receptor. After this, an interaction-based criterion was considered, i.e., we selected poses that present interactions similar to that of the co-crystallized naphthalene-derived compounds.

## 2.4 LigRMSD

When docking congeneric compounds, we expect the binding mode to be conserved with respect to those of co-crystallized compounds in the PLpro structures selected for this study. Therefore, we compared the binding poses obtained by molecular docking calculations using the LigRMSD web server (Velázquez-Libera et al., 2020). LigRMSD allows selecting the maximum common substructure between the molecules being compared, establishing matching graphs between them, and calculating the RMSD between the equivalent atoms with respect to the reference. The match is defined using the values “%Ref” and “%Mol”. “%Ref” indicates the percentage of common graphs between a docked compound and a selected reference, related to the total number of atoms of the selected reference. On the other hand, “%Mol” is the percentage of common graphs between the docked compound and the selected reference, with respect to the total number of atoms of the docked compound. These values obtained from the LigRMSD server represent the maximum similarity between the compounds being compared, so

high values of “%Ref” and “%Mol” are associated with high similarity between the compared compounds.

Based on this, we compared the poses obtained using multiple references. The poses of the co-crystallized ligand GRL0617 and 6577871 were used as references for compounds docked inside the PDB with code 3E9S. In addition, the poses of the co-crystallized compound C\_3k and 7724772 were used as references for compounds docked inside the PDB with code 4OW0.

## 2.5 Interaction fingerprint (IFP)

Recurrent chemical interactions between the docked poses of ligands and residues in the SARS-CoV-1 PLpro binding site were captured by Interaction fingerprints (IFPs) (Deng et al., 2004). Maestro's Interaction Fingerprint panel was used to build them. This method describes the presence or absence of chemical interactions between ligands and binding residues using bits for the subsequent construction of an interaction matrix. Each bit describes if a specific type of interaction takes place between the ligand and a protein residue, considering hydrophobic (H), polar (P), and aromatic (Ar) interactions. It is also possible to detect whether a residue is acting as a hydrogen bond (HB) acceptor (A) or donor (D) and electrostatic interactions with charged groups (Ch). For this study, it was counted as an interaction when a PLpro residue is within a maximum cutoff distance of 4.0 Å between the heavy atoms with respect to the ligand atoms.

## 2.6 Gaussian accelerated molecular dynamics (GaMD) and correlation analysis

Molecular dynamics (MD) simulations were performed to obtain a diverse sampling of the SARS-CoV-1 PLpro binding site. They had to be carried out with ligands at the binding site to ensure that the site remained open, allowing for the inclusion of other ligands in the subsequent cross-docking calculations. When placing a ligand, induce-fit effects may occur due to a specific ligand. To mitigate the induced effects resulting from a single ligand, the PDB protein structures with codes 3E9S and 4OW0, complexed with the ligands GRL0617 and C\_3k, were used to generate four PLpro-ligand models (in the case of the structure with code 4OW0, only the first chain was used). Two of these models were the original structures 3E9S and 4OW0, containing the ligands GRL0617 and C\_3k, respectively. The other two models were the structures previously obtained by docking 3E9S with C\_3k and 4OW0 with GRL0617. This approach aimed to introduce greater variation in the starting structures.

Protein structures were prepared using the Protein Preparation Wizard (Schrödinger LLC, New York, NY, USA, 2021). From this, force field parameters and coordinate files were constructed using LEAP from Amber (Case et al., 2005). A regular truncated octahedral TIP3P water box with 12 Å between the solute and the edges of the box was used for the simulations. The system minimization was carried out for 10,000 steps. Two rounds of equilibration were then performed. The system was heated to 310K for 1 ns using an isothermal-isovolumetric (NVT)

assembly, followed by an isothermal-isobaric (NPT) equilibration for 80 ns.

To perform Gaussian accelerated molecular dynamics (GaMD) (Miao and McCammon, 2017), the pmemd.cuda implementation of Amber20 was used to generate four trajectories. We used the LiGaMD method (Miao et al., 2020), based on GaMD, which was necessary for more efficient sampling simulations of protein-ligand complexes' binding and unbinding process. First, a 60-ns MD simulation was performed. The first 10 ns correspond to a conventional preparatory MD, without statistical collection, followed by 50 ns of LiGaMD. Next, a production simulation was performed, which starts at 50 ns and extends up to 150 ns. The VMD (Humphrey et al., 1996) and CPPTRAJ (Roe and Cheatham, 2013) tools were used to analyze the trajectories.

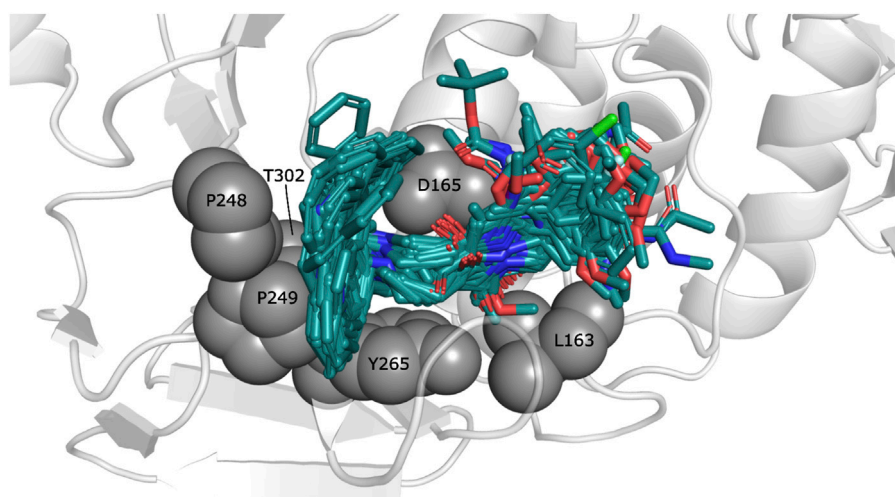
The trajectories generated for all systems were grouped using the K-means participle algorithm to obtain greater conformational diversity. An internal script using the scikit-learn library (Varoquaux et al., 2015) was used to perform the protocol. The different clusters were obtained considering six distance descriptors; (a) RMSD value of the Q270 residue; (b) distance between the more proximal carboxylate oxygen of the side chain of D165 and the nitrogen at the side chain of Q270; (c) distance between the hydroxyl group of Y269 and the nitrogen of the side chain of Q270; (d) distance between the nitrogen in the side chain of K158 and the oxygen at the side chain of Q270; (e) distance between the backbone oxygen of residue N268 and the nitrogen of C271, and (f) distance between the hydroxyl group of Y265 and the oxygen backbone of N268. Based on this, the possible clusters were represented by a dendrogram or “cluster tree,” where the root corresponds to the largest cluster containing all the sampled states, and each leaf refers to a single cluster.

The clustering process allowed us to find representative protein structures from the trajectories. The obtained protein structures were used as receptors of molecular cross-docking with each of the compounds under study, resulting in different poses for each ligand. The same docking settings described in Section 2.3 were employed for cross-docking. An *in-house* Python script (Muñoz-Gutiérrez et al., 2016) was used to select a representative complex for each ligand to best fit the correlations between the energy values calculated from the docking process and the logarithmic activities of the series of naphthalene-derived compounds. The result of the protocol corresponds to protein-ligand complexes showing the highest correlations.

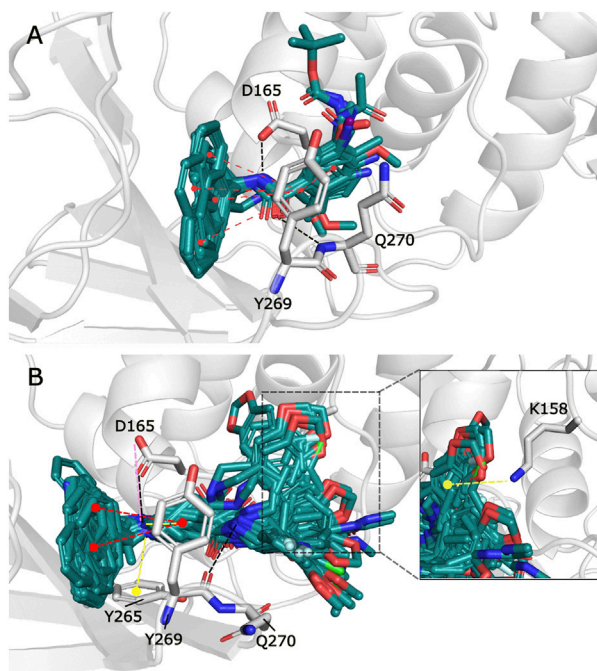
## 3 Results and discussion

### 3.1 Docking predictions

The ligands were docked to study the molecular basis of the interactions between the naphthalene-derived compounds and the SARS-CoV-1 PLpro (docking scoring energies are reported in the Supplementary Table S2). It can be seen that all ligands in the series adopt the same binding mode, placing the naphthylmethylamine group at the S<sub>4</sub> subsite of the enzyme (Figure 1). It has been previously verified that this subsite is specific for leucine and can accommodate large hydrophobic groups (Rut et al., 2020). Olsen

**FIGURE 1**

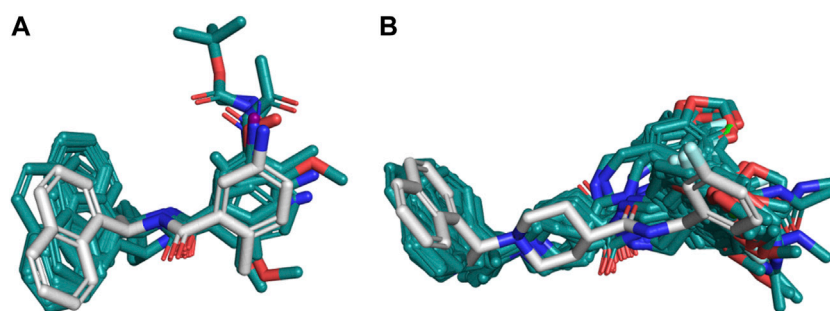
Docked structures within the SARS-CoV-1 PLpro binding site. Docked ligands are represented by sticks. Relevant residues at  $S_3$  and  $S_4$  subsites are represented by spheres.

**FIGURE 2**

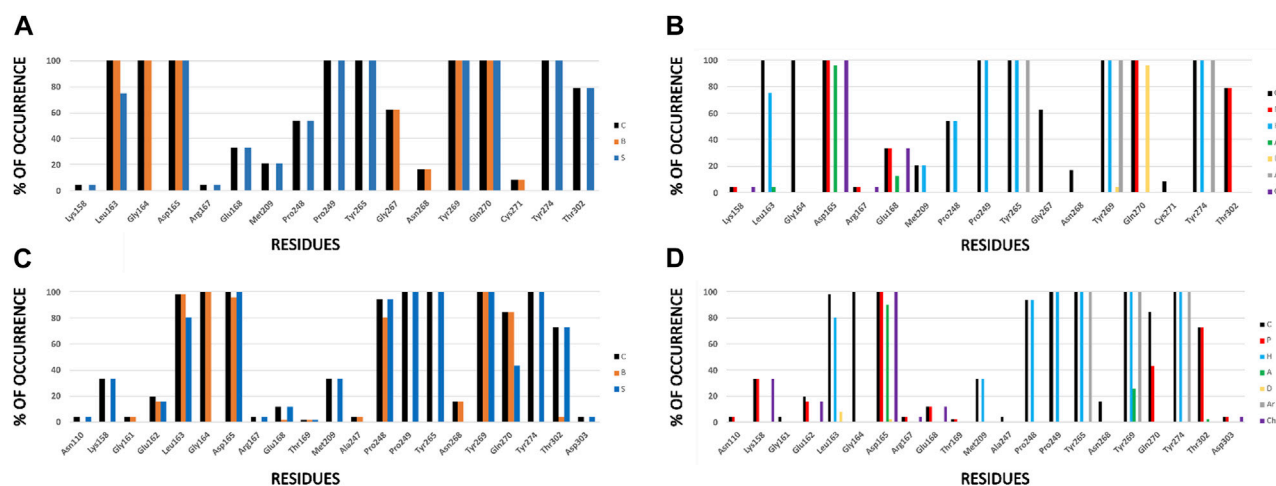
Docking poses for congeneric series of naphthalene-derived compounds at the SARS-CoV-1 PLpro binding site. (A) Compounds from series A, D, 7724772, and GRL0617 docked inside the structure with PDB code 3E9S. (B) Compounds from series B, C, and 6577871 docked inside the structure with PDB code 4OW0 (a rotation of a selection is at the right to observe interactions with Lys158). Ligands are represented by cyan sticks, while protein residues involved in interactions are represented with white sticks. Interactions are represented by dashed lines with the following coloring scheme: red color lines correspond to pi-pi stacking interactions, yellow color lines to pi-cation interactions, black color lines to HBs, and magenta color lines to salt bridges.

et al. observed that the 2-benzothiazolyl and (4-hydroxyphenyl) ethyl groups of the covalent inhibitors VIR250 and VIR251 occupy opposite sides of the broad  $S_4$  pocket of SARS-CoV-1 and SARS-CoV-2 PLpro (Rut et al., 2020; Patchett et al., 2021); chemical groups at  $S_4$  can be oriented closer to the Pro249 or closer to the Pro248. Our docking results show that the naphthylmethylamine group can occupy both sides of the  $S_4$  subsite.

The observed interactions are consistent with those reported for crystals having co-crystallized naphthalene-derived compounds (Ratia et al., 2008; Báez-Santos et al., 2014a). Compounds from series A, D, 7724772, and GRL0617 have HB interactions between the benzamide carbonyl of the inhibitors and the backbone NH of Gln270 (Figure 2A); the same interaction is absent in the poses obtained for compounds from series B, C, and 6577871 (Figure 2B). This occurs since the BL2 hinged loop exists in different conformations for each of the studied protein states. For the structure with PDB code 4OW0, the side chain of the Gln270 residue is moved away from the inhibitors, preventing HB formation between its backbone and compounds from series B, C, and 6577871 (Báez-Santos et al., 2014b). On the hand, the residue Tyr269 (also at the BL2 loop) does not have a considerable displacement between the structures with codes 3E9S and 4OW0 and is involved in pi-pi stacking interactions. This residue, and the residue Asp165 (forming HBs with donors of the ligands), are of great importance for stabilizing the naphthalene-derived compounds (Figure 2). Asp165 also forms a salt bridge with the protonated piperidine of compounds from series B, C, and 6577871. The aromatic group of the residue Tyr265 forms a pi-cation interaction with the same protonated piperidine groups (Figure 2B). It is also pertinent to point out that the residue Lys158 establishes pi-cation interactions with several aromatic substituents placed in its vicinity (Figure 2B to the right).

**FIGURE 3**

Structural similarity of the docking poses with respect to references 3E9S and 4OW0. **(A)** Compounds from series A and D compared to compound GRL0617 co-crystallized on 3E9S as reference. **(B)** Compounds from series B and C compared to compound C\_3k co-crystallized on 4OW0 as reference. For each of the cases, the reference is represented as white sticks, while the poses obtained by docking are shown in cyan.

**FIGURE 4**

IFPs that describe interactions between docked compounds and SARS-CoV-2 PLpro crystals. **(A, B)** Interactions of compounds from series A, D, 7724772, and GRL0617 with residues at the PLpro crystal with code 3E9S. **(C, D)** Interactions of compounds from series B, C, and 6577871 with residues at the PLpro crystal with code 4OW0. Interactions in the graphs at the left **(A, C)** are presented as percentage of occurrence of contacts [C], interactions with the backbone of the residue [B], and interactions with the side chain of the residue [S]. Interactions in the graphs at the right **(B, D)** are presented as percentage of occurrence of chemical interactions: contacts [C], polar [P], hydrophobic [H], HBs where the residue is an acceptor [A], HBs where the residue is a donor [D], aromatic [Ar], and electrostatic with charged groups [Ch].

The poses obtained from the molecular docking of the 67 naphthalene-derived inhibitors were compared with their similar inhibitors GRL0617 and C\_3k co-crystallized on the structures with codes 3E9S and 4OW0, respectively. This comparison was carried out using the LigRMSD server, which identifies common graphs between molecules and calculates the RMSD between the equivalent atoms in each graph (Velázquez-Libera et al., 2020). It is accepted in the literature that RMSD values less than 2 Å reflect a meaningful spatial relationship between the compared structures (Warren et al., 2006; Plewczynski et al., 2011; Sasmal et al., 2020). The results of this analysis are detailed in the Supplementary Table S3. The comparisons where the co-crystallized inhibitor GRL0617 in 3E9S was used as reference helped to characterize the orientations of the ligands from series A and D. When GRL0617 is used as reference, these compounds exhibited %

Ref values higher than 85%, likewise the %Mol values in most of the cases (except for compound D\_47 with %Mol = 70.97). Most RMSD values in the range of 0.25 Å to 1.5 Å were obtained, with only five compounds (A\_6, D\_21, D\_33, D\_40, and the redocked conformation of GRL0617) showing RMSD values between 2.30 and 2.51 Å. The naphthalene groups in these five compounds were positioned opposite to the same group in the reference, but their main scaffolds were oriented correctly. Therefore, docking poses of the complete set of ligands from series A and D were oriented similarly to the co-crystallized compound GRL0617. On the other hand, compound C\_3k, co-crystallized in 4OW0, was used as a reference to characterize the orientations of the ligands from series B and C. When C\_3k is used as reference, these compounds exhibited %Ref values higher than 84%, likewise the %Mol values in most of the cases (except for

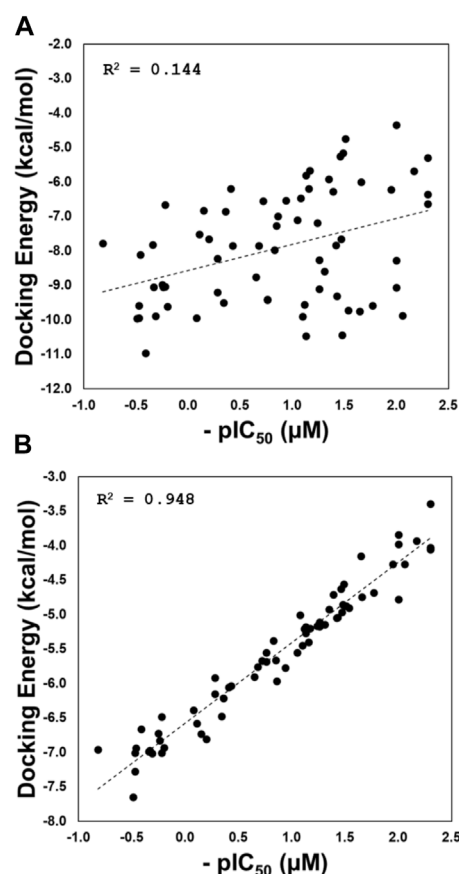


compound C\_1d with %Mol = 75.68). As an exception, compounds C\_6a and C\_6b have %Ref and %Mol values of 75.86 and 73.33 respectively; these values also indicate that there is similarity with C\_3k. Compounds from series B had RMSD values  $< 2 \text{ \AA}$  with only the compound B\_15j showing RMSD = 4.04  $\text{\AA}$ . The majority of compounds from series C had RMSD values  $< 2 \text{ \AA}$ ; however, six of the non-optically active compounds (C\_2d, C\_3i, C\_5b, C\_5c, C\_6a, and C\_6b) had higher values, and the optically active compounds C\_1a (S), C\_1b (R), C\_1c (R), C\_1d (R and S), C\_4a (R and S), C\_4c (R), and C\_4d (R) have also RMSD values  $> 2 \text{ \AA}$ . In some cases, the aromatic groups from the benzylamine (or similar substituents in C\_5a-c and C\_6a-b) of the studied compounds from series B and C, rotated in the opposite direction with respect to the reference, inducing  $\pi$ -cation interactions with Lys158 (Figure 2B to the right). This greater flexibility led to higher RMSD values; however, a visual analysis shows that a similar orientations with respect to the reference compound were obtained, which was reflected in the coincidence between the scaffolds of the compared compounds (Figure 3).

For a better understanding of the interactions between the docked ligands and PLpro, an IFP was performed. This analysis allows annotating the recurrent chemical interactions observed between the compounds of the congeneric series and the protease binding site. The graphs of the types of chemical interactions occurring per residue are reported. The IFPs for the 24 compounds from series A, D, 7724772, and GRL0617 docked in the PLpro crystal with code 3E9S are in Figures 4A, B, and the IFPs for the 43 compounds from series B, C, and 6577871 docked in the PLpro crystal with code 4OW0 are in Figures 4C, D.

For both protein crystal structures, the residues implicated in the formation of interactions at the protein-ligand interface are similar (Figure 4). Hydrophobic contributions and aromatic contacts with residues Tyr265, Tyr269, and Tyr274 occur in 100% of the docked structures. These residues form an aromatic box that contribute to attraction and stabilization of the naphthalene-derived inhibitors; specifically, Tyr269 is essential for closing the BL2 loop to adopt the closed conformation of the binding site (Báez-Santos et al., 2014b). IFPs show that Tyr269 was also identified as an HB donor with  $\sim 5\%$  of compounds from series A and D, and as an HB acceptor with  $\sim 25\%$  of compounds from series B and C. These roles can be present when including substituents with specific polar groups (Figure 2).

The residues Pro248 and Pro249 favored the occurrence of hydrophobic contacts at the protein-ligand interface. Hydrophobic contacts of Pro249 had 100% of occurrence, while Pro248 also had high hydrophobic contributions, with  $\sim 55\%$  and  $\sim 90\%$  of occurrence in the structures 3E9S and 4OW0, respectively. Several residues were also identified that contributed to form electrostatic interactions at the SARS-CoV-1 PLpro binding site. Asp165 has polar interactions with the docked poses with 100% of occurrence. This residue acts as HB acceptor with more than 90% of occurrence in 3E9S and 4OW0, respectively. It reflects that this residue forms HBs with benzamide NH group of compounds from series A and D, and also forms HBs (and salt bridges) with the protonated piperidine of compounds from series B and C (Figure 2). The residue Gln270 from the BL2 loop had 100% of occurrence of



**FIGURE 5**  
Regression plots of the docking scoring energies versus experimental activities ( $pIC_{50}$ ) for the docking experiments performed in structures with codes 3E9S and 4OW0 (A), and for the cross-docking protocol (B).

polar contacts and is an HB donor in  $\sim 95\%$  of the docked compounds in 3E9S. It had  $\sim 40\%$  of occurrence of polar contacts when forming complexes between ligands and the structure with code 4OW0.

Other noteworthy IFPs are detailed as followed. Gly164 had contacts with 100% of occurrence in 3E9S and 4OW0. Leu163, its backbone, had contacts with all the ligands, and its side chain had hydrophobic interactions with  $\sim 75\%$  and  $\sim 80\%$  of occurrence in 3E9S and 4OW0, respectively. Lys158 had polar and charged contributions in  $\sim 5\%$  of the structures docked in 3E9S, and the same contributions in  $\sim 30\%$  of the structures docked in 4OW0. Glu168 had polar and charged contacts with  $\sim 30\%$  of occurrence and acted as HB acceptor with  $\sim 10\%$  of occurrence in 3E9S; in contrast, it had polar and charged contacts with  $\sim 10\%$  of occurrence in 4OW0. Finally, Thr302 had polar contributions in  $\sim 80\%$  of the structures docked in 3E9S, and the same contributions in  $\sim 70\%$  of the structures docked in 4OW0.

The analysis presented with the IFPs shows two variants of how two sets of non-covalent inhibitors bind to the  $S_3$ - $S_5$  subsites. It is possible to observe some interactions that seem essential and others that appear occasionally. The IFPs show how substituents of the studied sets are distributed at  $S_4$ . The naphthalene group can be



**TABLE 2** List of structures used as receptors for cross-docking experiments and molecules involved in the structure-activity relationship model with the highest  $R^2$ .

Model	Conformation	Ligands	Model	Conformation	Ligands
3E9S	c2	B_7b, B_15j, B_15k, C_2c, C_3d, C_4d	4OW0	c13	B_7a, B_15i, C_1c, C_1d, D_21, D_23
	c3	B_15d, C_1b, D_32		c15	7724772(R), B_15b, C_3h, C_3j, C_4a, C_4b, D_5a, D_5f
	c4	C_1a, C_2a, C_3g, C_3i, C_5b, D_2		c16	A_7, B_15h, C_4c, D_5b, D_29, D_33, D_47
	c6	7724772(S), A_4, C_2b, D_5c		c17	A_6, C_3b, C_3k, C_5c, C_6a, D_40, GRL0617
	c8	6577871, A_3, B_15e, C_2d, C_3f, D_9		c19	A_5, B_15c, B_15g, C_2e, C_3a, C_3c, C_6b, D_5d, D_49
	c9	B_15a, B_15f, C_3e, C_5a, D_5e			

oriented closer to Pro249 or in the opposite direction, closer to Pro248 (similar to the structures of complexes between PLpro with the covalent inhibitors VIR250 and VIR251) (Rut et al., 2020; Patchett et al., 2021).

### 3.2 Binding site flexibility and correlation results

In order to increase the conformational sampling of the SARS-CoV-1 PLpro binding site in the presence of naphthalene-derived inhibitors, four GaMD simulations were performed following the protocol described in the Materials and Methods section. Two of them were carried out on the solvated PDB structure with code 3E9S in complex with GRL0617 and C\_3K, while the others two simulations were developed on the solvated structure with code 4OW0 coupled to the same ligands. Stability of the GaMD trajectories using the RMSD of the positions for the backbone PLpro atoms as a function of simulation time was evaluated; RMSD was reasonably stable during the production simulation for all the systems (Supplementary Figure S2).

From the GaMD simulations, six distance descriptors (Materials and Methods section) were considered to perform a partition clustering process by means of the K-means algorithm. This clustering algorithm assigns all MD conformations into one large grouping. The largest cluster was divided into two subclusters iteratively until each conformation forms a single cluster (Abramyan et al., 2016). The value of  $k$  in the algorithm was defined using the “elbow method” as well as a dendrogram or cluster tree plot, thus confirming that the data set contains five clusters (Shi et al., 2021). This process, applied to four GaMD simulations, resulted in twenty representative and structurally diverse PLpro conformations, named c0-c19 in this manuscript (c0-c9 and c10-c19 were derived from GaMD simulations of the models constructed from structures with codes 3E9S and 4OW0, respectively). These structures were used to perform the cross-docking methodology (67 compounds were docked in twenty PLpro structures with diverse conformations of the binding site). It is important to remark that significative variations were identified in the binding sites for c0-c19, mainly in the BL2 loop (Supplementary Figure S3).

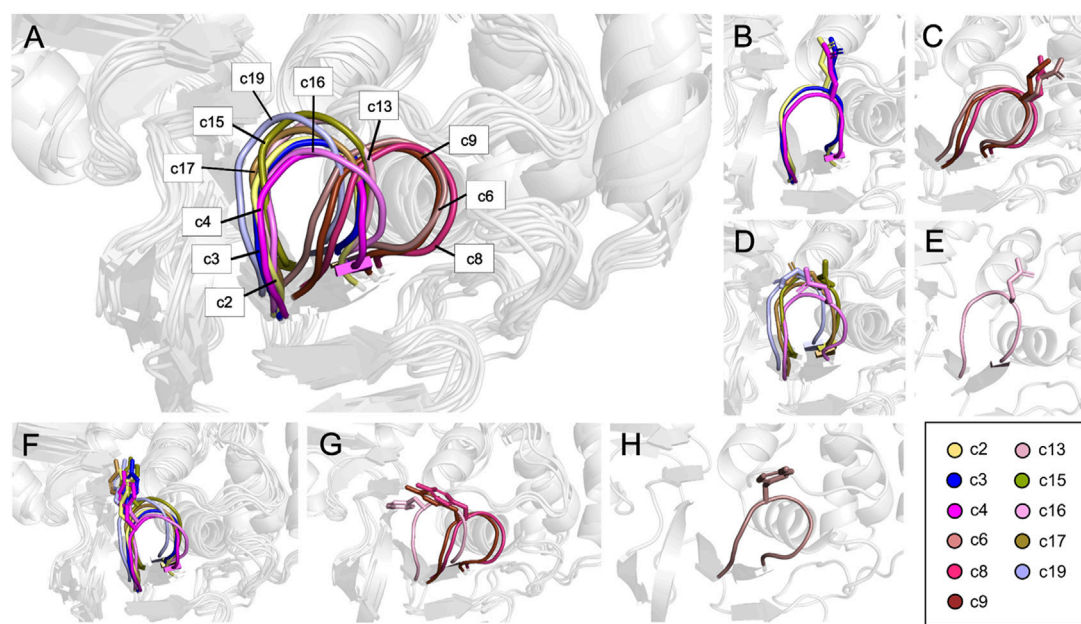
The cross-docking yielded twenty different poses for each ligand. The orientations of these poses were verified with LigRMSD (Velázquez-Libera et al., 2020) to ensure the presence of reasonable solutions. Representative PLpro-inhibitor complexes

for each ligand were selected after application of the *in-house* Python script (Muñoz-Gutierrez et al., 2016) that optimize correlations between the calculated and experimental activities. This script yielded the set of PLpro-inhibitor complexes that produce the best correlation between the docking scoring energies and experimental PLpro inhibitory activities (scoring energies for the representative complexes are reported in the Supplementary Table S2).

The results for correlations are depicted in Figure 5. The correlation considering the docking experiments performed in these structures with codes 3E9S and 4OW0 is poor ( $R^2 = 0.144$ ; Figure 5A). This result is expected. It is well-known in literature that current docking scoring functions such as GlideScore have demonstrated satisfactory performance in docking and screening power tests; however, these functions may not be as effective when it comes to evaluating scoring power, which reflects the ability to establish a strong linear correlation between predicted and experimental binding affinities (Ferrara et al., 2004; Plewczynski et al., 2011; Su et al., 2019). To address this issue, one approach is to incorporate a flexible receptor binding site (Baumgartner and Evans, 2018). Our script employs various conformational states obtained through GaMD simulations, which allows for flexibility in the binding site. As demonstrated in Figure 5B, our method has significantly improved the correlation between predicted and experimental binding affinities, achieving an  $R^2$  value of 0.948.

The high correlation reflects a successful explanation of the structure-activity relationship through the proposed protocol. Eleven of the twenty PLpro conformations were selected by the model, these conformations are listed in Table 2. This table also shows the list of compounds docked in each PLpro conformation to obtain the structure-activity relationship model with the highest  $R^2$  value.

The GaMD and clustering process was performed to obtain different conformations of the SARS-CoV-1 PLpro binding site, and this was achieved mainly due to large changes in the BL2 loop (Supplementary Figure S3). Different versions of the binding site were obtained, which in turn differ from the binding sites in the PDB structures coded 4OW0 and 3E9S. There are some differences in the BL2 loop when comparing the 4OW0 and 3E9S structures. The residues Tyr269 and Gln270 adopt different conformations between these structures, representing a more opened (4OW0) and more closed (3E9S) state of the BL2 loop (Supplementary Figure S4). The MD and clustering protocol produced other binding site variation options, increasing flexibility, and creating new structural conformations that were a starting point for the cross-docking



**FIGURE 6**

Residues conforming the BL2 loop in different clusters. **(A)** Representation of the BL2 loop for the 11 structures that maximize the structure-activity correlation. **(B)** Conformation I for Q270. **(C)** Conformation II for Q270. **(D)** Conformation III for Q270. **(E)** Conformation IV for Q270. **(F)** Conformation I for Y269. **(G)** Conformation II for Y269. **(H)** Conformation III for Y269.

calculations. The structural conformations c0-c19 have differences in the SARS-CoV-1 PLpro binding site. The analysis of the conformational variations observed for the residues that constitute this site is reported in the [Supplementary Table S4](#). From this table, it can be seen that most of the structural units being compared have RMSD values greater than 2.0Å, reflecting displacements between the parts being compared. In some cases, these variations are related to specific fluctuations that do not reflect the fluctuations of the macromolecule or the portion being compared as a whole. Consequently, a root mean square fluctuation (RMSF) analysis was performed considering the residues that are part of the binding site ([Supplementary Figure S5](#)). RMSF shows that the BL2 loop residues Tyr269 and Gln270 are the most mobile residues within the binding site. Therefore, RMSD analyses were performed on these residues ([Supplementary Table S5](#)). It is observed that most of the structures presented RMSD values higher than 2.0Å reflecting the conformational diversity of the BL2 loop between the conformations c0-c19. The high RMSD values in this part of the binding site reflect the possibility of great flexibility that justify the use of our GaMD and clustering protocol, instead of the rigid structures coming from PDB.

From the twenty conformations c0-c19, eleven participated in the model that maximizes the structure-activity correlation, when six and five were derived from the 3E9S and 4OW0 structures, respectively. RMSD analyses for these eleven conformations were performed considering the residues with the highest fluctuations (Tyr269 and Gln270) in the binding site (using values in [Supplementary Table S5](#)), and high RMSD values for most cases were observed. The [Figure 6](#) shows a visual inspection of the residues Tyr269 and Gln270 in the eleven conformations that are in the model that maximizes the structure-activity correlation. Gln270 presents four different orientations named I, II, III, and IV ([Figures 6B–E](#)), while

Tyr269 can be grouped in three different orientations named I, II, and III (represented in the [Figures 6F–H](#)). The remaining residues at the binding site do not present considerable changes. Three structures (c2, c3, and c4) adopted conformation I for Tyr269 and I for Gln270, including 15 inhibitors. Four structures (c15, c16, c17, and c19) adopted conformation I for Tyr269 and III for Gln270, including 31 inhibitors. Two structures (c8 and c9) adopted conformation II for Tyr269 and II for Gln270, including 11 inhibitors. The combination of conformations II of Gln270 and III of Tyr269 was present in c6 that contains 4 inhibitors, and the combination of conformations IV of Gln270 and II of Tyr269 was present in c13 that contains 6 inhibitors. In all cases, the ligand poses included in the model with the highest  $R^2$  value had the expected interactions with the residues corresponding to the BL2 loop. A visual analysis shows that the complexes in this model share interaction profiles similar to each other and concordant with the crystallographic structures. [Table 2](#) shows that the most active compounds (c\_3k, B\_15g, B\_15c, and B\_15b) were selected in the PLpro conformations c15, c17, and c19, which adopt conformation I for Gln270 and conformation I for Tyr269, as previously mentioned ([Figures 6B, F](#)). Consequently, these receptor conformations are proposed as the most suitable for a potential exploration of new potent compounds.

Compound interactions with PLpro binding site residues for protein-ligand complexes in the highest correlation model were verified using IFPs. Previously, the most important residues were shown in an IFP analysis made on the complexes obtained by docking. It was expected that such important residues should be maintained in the complexes obtained by cross-docking. The [Supplementary Figure S6](#) shows that the recurrent chemical interactions between the compounds and the PLpro binding site were kept in the protein-ligand complexes present in the model with the highest correlation. The most important interactions

with the residues Leu163, Gly164, Asp165, Pro248, Pro249, Tyr265, Tyr269, Gln270, Tyr274, and Thr302 previously identified, were also present in the IFPs in [Supplementary Figure S6](#). Interestingly, both series of compounds have remarkably increased polar interactions with the side chain of Arg167 (30% of occurrence). On the other hand, compounds from series B, C, and 6577871 have remarkably increased polar interactions with the side chain of the residue Glu162 (with more than 40% of occurrence).

The high conformational variation of the two residues composing the BL2 loop implies changes in the volume and shape of the binding site, which has an influence on the specific interactions of the studied compounds. The conformational diversity in the receptor binding site contributes to the ligands adopting conformations that maximized the correlation between docking scoring and  $pIC_{50}$  values.

Our results suggest that it is very relevant to consider the flexibility of the PLpro binding site for the study of its inhibitors. The flexibility of proteins poses a significant challenge when it comes to ligand docking, as the binding site can exist in various conformations ([Caballero, 2021](#)). Docking protocols in the literature widely employ combinations of docking and MD simulations ([Munoz et al., 2012](#); [Sharma et al., 2016](#); [Śledź and Caflisch, 2018](#)). These methods have shown that incorporating multiple protein conformations enhances the results. For instance, Strecker and Meyer conducted a recent study in which they compared docking using several crystal structures and structures obtained from MD simulations ([Strecker and Meyer, 2018](#)). They assessed the impact of structure selection and discovered that binding site shapes not observed in any crystal structure in the PDB were accessible through 500-ns MD simulations. They demonstrated that these structures significantly contributed to accurate binding pose predictions, improved ability to distinguish active compounds (screening utility), and enhanced scoring accuracy. Our results are in agreement with what was shown in this study.

Before 2019, there were few studies on SARS-CoV-1 PLpro inhibitors using computational methods; however, some recent studies have focused on the study of SARS-CoV-2 PLpro; in some of these works, the flexibility of the PLpro binding site was studied in some way ([Ferreira et al., 2022](#); [Santos et al., 2022](#); [Singh et al., 2022](#)). Among the recent studies, we would like to highlight the work of Garland et al. ([Garland et al., 2023](#)). The authors virtually examined the ZINC20 database ([Irwin et al., 2020](#)) using a docking method and filtering with a pharmacophore to identify possible noncovalent PLpro modulators. Using this methodology, the authors discovered the compound VPC-300195 ( $IC_{50} = 15 \mu M$ ). The authors found a limited diversity of active compounds, which they attributed to the rigidity of the PLpro active site in crystal structures. In part, this report proposes that the inclusion of flexibility in the binding site is necessary for future designs.

## 4 Conclusion

A set of 67 naphthalene-derived compounds as noncovalent PLpro inhibitors were studied using a flexible molecular docking protocol. In summary, the following four steps were carried out: i) the structures of the protein-ligand complexes were obtained with a rigid docking, ii)

multiple conformations of the PLpro binding site were obtained by using GaMD, iii) a cross-docking was performed between the 67 compounds and selected PLpro conformations, and iv) protein-ligand complexes that represent the highest correlation between docking energies and experimental activities were selected. As a result, a set of complexes was identified where the ligands interact with a flexible binding site of PLpro. The proposed methodology proved successful, and a correlation value of  $R^2 = 0.948$  was obtained in the aforementioned last step. Considering the flexibility of the protein by using various PDB structures and the GaMD sampling of the receptor was fundamental to achieving the proposed objective. When using a rigid docking, it is ignored that ligands can be bound with significant protein conformational changes, therefore taking into account flexibility of the binding site results in a more rational approach. Overall, the strategy employed in this article serves as a good approach to studying PLpro ligands with computational tools, and the method reflects a possible conformational selection approach. Performing a detailed structural study of the inhibitory role of naphthalene derivatives acting against the SARS-CoV-1 PLpro allows us to contribute positively to the research field aimed at the design and computational evaluation of more potent candidates against this protease.

## 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

The work was completed by cooperation of all authors. JC was responsible for the study of concept and design of the project. LCC and JLV performed the docking, LigRMSD, IFPs, GaMD calculations, and correlation analysis. LCC and JC drafted and revised the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This research was funded by FONDECYT Regular grant number 1210138 (JC).

## 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

- Abramyan, T. M., Snyder, J. A., Thypambil, A. A., Stuart, S. J., and Latour, R. A. (2016). Cluster analysis of molecular simulation trajectories for systems where both conformation and orientation of the sampled states are important. *J. Comput. Chem.* 37, 1973–1982. doi:10.1002/jcc.24416
- Báez-Santos, Y. M., Barraza, S. J., Wilson, M. W., Agius, M. P., Mielech, A. M., Davis, N. M., et al. (2014a). X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. *J. Med. Chem.* 57, 2393–2412. doi:10.1021/jm401712t
- Báez-Santos, Y. M., Mielech, A. M., Deng, X., Baker, S., and Mesecar, A. D. (2014b). Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. *J. Virol.* 88, 12511–12527. doi:10.1128/JVI.01294-14
- Báez-Santos, Y. M., St John, S. E., and Mesecar, A. D. (2015). The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antivir. Res.* 115, 21–38. doi:10.1016/j.antiviral.2014.12.015
- Baumgartner, M. P., and Evans, D. A. (2018). Lessons learned in induced fit docking and metadynamics in the drug design data resource grand challenge 2. *J. Comput. Aided Mol. Des.* 32, 45–58. doi:10.1007/s10822-017-0081-y
- Caballero, J. (2021). The latest automated docking technologies for novel drug discovery. *Expert Opin. Drug Discov.* 16, 625–645. doi:10.1080/17460441.2021.1858793
- Calleja, D. J., Lessene, G., and Komander, D. (2022). Inhibitors of SARS-CoV-2 PLpro. *Front. Chem.* 10, 876212. doi:10.3389/fchem.2022.876212
- Cannalire, R., Cerchia, C., Beccari, A. R., Di Leva, F. S., and Summa, V. (2022). Targeting SARS-CoV-2 proteases and polymerase for COVID-19 treatment: State of the art and future opportunities. *J. Med. Chem.* 65, 2716–2746. doi:10.1021/acs.jmedchem.0c01140
- 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
- Chaudhuri, R., Tang, S., Zhao, G., Lu, H., Case, D. A., and Johnson, M. E. (2011). Comparison of SARS and NL63 papain-like protease binding sites and binding site dynamics: Inhibitor design implications. *J. Mol. Biol.* 414, 272–288. doi:10.1016/j.jmb.2011.09.030
- De Savi, C., Hughes, D. L., and Kvaerno, L. (2020). Quest for a COVID-19 cure by repurposing small-molecule drugs: Mechanism of action, clinical development, synthesis at scale, and outlook for supply. *Org. Process Res. Dev.* 24, 940–976. doi:10.1021/acs.oprd.0c00233
- Deng, Z., Chuaqui, C., and Singh, J. (2004). Structural interaction fingerprint (SIFt): A novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* 47, 337–344. doi:10.1021/jm030331x
- Ferrara, P., Gohlke, H., Price, D. J., Klebe, G., and Brooks, C. L. (2004). Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* 47, 3032–3047. doi:10.1021/jm030489h
- Ferreira, G. M., Pillaiyar, T., Hirata, M. H., Poso, A., and Kronenberger, T. (2022). Inhibitor induced conformational changes in SARS-CoV-2 papain-like protease. *Sci. Rep.* 12, 11585. doi:10.1038/s41598-022-15181-y
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749. doi:10.1021/jm0306430
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for Protein-Ligand complexes. *J. Med. Chem.* 49, 6177–6196. doi:10.1021/jm051256o
- Garland, O., Ton, A.-T., Moradi, S., Smith, J. R., Kovacic, S., Ng, K., et al. (2023). Large-scale virtual screening for the discovery of SARS-CoV-2 papain-like protease (PLpro) non-covalent inhibitors. *J. Chem. Inf. Model* 63, 2158–2169. doi:10.1021/acs.jcim.2c01641
- Ghosh, A. K., Takayama, J., Aubin, Y., Ratia, K., Chaudhuri, R., Baez, Y., et al. (2009). Structure-based design, synthesis, and biological evaluation of a series of novel and reversible inhibitors for the severe acute respiratory syndrome-coronavirus papain-like protease. *J. Med. Chem.* 52, 5228–5240. doi:10.1021/jm900611t
- Ghosh, A. K., Takayama, J., Rao, K. V., Ratia, K., Chaudhuri, R., Mulhearn, D. C., et al. (2010). Severe acute respiratory syndrome coronavirus papain-like novel protease inhibitors: Design, synthesis, protein-ligand X-ray structure and biological evaluation. *J. Med. Chem.* 53, 4968–4979. doi:10.1021/jm1004489
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi:10.1038/s41586-020-2286-9
- Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., et al. (2016). OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* 12, 281–296. doi:10.1021/acs.jctc.5b00864
- Hilgenfeld, R. (2014). From SARS to MERS: Crystallographic studies on coronavirus proteases enable antiviral drug design. *FEBS J.* 281, 4085–4096. doi:10.1111/febs.12936
- Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Indari, O., Kumar Singh, A., Tiwari, D., Chandra Jha, H., and Nath Jha, A. (2022). Deciphering antiviral efficacy of malaria box compounds against malaria exacerbating viral pathogens- Epstein Barr virus and SARS-CoV-2, an *in silico* study. *Med. Drug Discov.* 16, 100146. doi:10.1016/j.medidd.2022.100146
- Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., et al. (2020). ZINC20—a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* 60, 6065–6073. doi:10.1021/acs.jcim.0c00675
- Khataniar, A., Pathak, U., Rajkhowa, S., and Jha, A. N. (2022). A comprehensive review of drug repurposing strategies against known drug targets of COVID-19. *COVID* 2, 148–167. doi:10.3390/covid2020011
- Miao, Y., Bhattarai, A., and Wang, J. (2020). Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. *J. Chem. Theory Comput.* 16, 5526–5547. doi:10.1021/acs.jctc.0c00395
- Miao, Y., and McCammon, J. A. (2017). Gaussian accelerated molecular dynamics: Theory, implementation, and applications. *Annu. Rep. Comput. Chem.* 13, 231–278. doi:10.1016/bs.arcc.2017.06.005
- Munoz, C., Adasme, F., Alzate-Morales, J. H., Vergara-Jaque, A., Kniess, T., and Caballero, J. (2012). Study of differences in the VEGFR2 inhibitory activities between semaxanib and SU5205 using 3D-QSAR, docking, and molecular dynamics simulations. *J. Mol. Graph Model* 32, 39–48. doi:10.1016/j.jmgm.2011.10.005
- Muñoz-Gutiérrez, C., Adasme-Carreño, F., Fuentes, E., Palomo, I., and Caballero, J. (2016). Computational study of the binding orientation and affinity of PPAR $\gamma$  agonists: Inclusion of ligand-induced fit by cross-docking. *RSC Adv.* 6, 64756–64768. doi:10.1039/C6RA12084A
- Patchett, S., Lv, Z., Rut, W., Békés, M., Drag, M., Olsen, S. K., et al. (2021). A molecular sensor determines the ubiquitin substrate specificity of SARS-CoV-2 papain-like protease. *Cell. Rep.* 36, 109754. doi:10.1016/j.celrep.2021.109754
- Plewczynski, D., Łaźniewski, M., Augustyniak, R., and Ginalski, K. (2011). Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* 32, 742–755. doi:10.1002/jcc.21643
- Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S., et al. (2008). A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16119–16124. doi:10.1073/pnas.0805240105
- Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C., et al. (2006). Severe acute respiratory syndrome coronavirus papain-like protease: Structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5717–5722. doi:10.1073/pnas.0510851103
- Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi:10.1021/ct400341p
- Rut, W., Lv, Z., Zmudzinski, M., Patchett, S., Nayak, D., Snipas, S. J., et al. (2020). Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti-COVID-19 drug design. *Sci. Adv.* 6, eabd4596. doi:10.1126/sciadv.abd4596
- Santos, L. H., Kronenberger, T., Almeida, R. G., Silva, E. B., Rocha, R. E. O., Oliveira, J. C., et al. (2022). Structure-based identification of naphthoquinones and derivatives as novel inhibitors of main protease mpro and papain-like protease

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1215499/full#supplementary-material>



PLpro of SARS-CoV-2. *J. Chem. Inf. Model* 62, 6553–6573. doi:10.1021/acs.jcim.2c00693

Sasmal, S., El Khoury, L., and Mobley, D. L. (2020). D3R grand challenge 4: Ligand similarity and MM-GBSA-based pose prediction and affinity ranking for BACE-1 inhibitors. *J. Comput. Aided Mol. Des.* 34, 163–177. doi:10.1007/s10822-019-00249-1

Sharma, R. K., Espinoza-Moraga, M., Poblete, H., Douglas, R. G., Sturrock, E. D., Caballero, J., et al. (2016). The dynamic nonprime binding of sampatrilat to the C-domain of angiotensin-converting enzyme. *J. Chem. Inf. Model* 56, 2486–2494. doi:10.1021/acs.jcim.6b00524

Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. (2007). Epik: A software program for pK(a) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* 21, 681–691. doi:10.1007/s10822-007-9133-z

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., and Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP J. Wirel. Commun. Netw.* 2021, 31. doi:10.1186/s13638-021-01910-w

Singh, E., Jha, R. K., Khan, R. J., Kumar, A., Jain, M., Muthukumaran, J., et al. (2022). A computational essential dynamics approach to investigate structural influences of ligand binding on Papain like protease from SARS-CoV-2. *Comput. Biol. Chem.* 99, 107721. doi:10.1016/j.compbiolchem.2022.107721

Sledz, P., and Caflisch, A. (2018). Protein structure-based drug design: From docking to molecular dynamics. *Curr. Opin. Struct. Biol.* 48, 93–102. doi:10.1016/j.sbi.2017.10.010

Strecker, C., and Meyer, B. (2018). Plasticity of the binding site of renin: Optimized selection of protein structures for ensemble docking. *J. Chem. Inf. Model* 58, 1121–1131. doi:10.1021/acs.jcim.8b00010

Su, M., Yang, Q., Du, Y., Feng, G., Liu, Z., Li, Y., et al. (2019). Comparative assessment of scoring functions: The CASF-2016 update. *J. Chem. Inf. Model* 59, 895–913. doi:10.1021/acs.jcim.8b00545

Ullrich, S., and Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorg Med. Chem. Lett.* 30, 127377. doi:10.1016/j.bmcl.2020.127377

Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., and Mueller, A. (2015). Scikit-learn: Machine learning without learning the machinery. *Getmob. Mob. Comp. Comm.* 19, 29–33. doi:10.1145/2786984.2786995

Velázquez-Libera, J. L., Durán-Verdugo, F., Valdés-Jiménez, A., Núñez-Vivanco, G., and Caballero, J. (2020). LigRMSD: A web server for automatic structure matching and RMSD calculations among identical and similar compounds in protein-ligand docking. *Bioinformatics* 36, 2912–2914. doi:10.1093/bioinformatics/btaa018

Wang, J.-T., and Chang, S.-C. (2004). Severe acute respiratory syndrome. *Curr. Opin. Infect. Dis.* 17, 143–148. doi:10.1097/00001432-200404000-00013

Warren, G. L., Andrews, C. W., Capelli, A.-M., Clarke, B., LaLonde, J., Lambert, M. H., et al. (2006). A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 49, 5912–5931. doi:10.1021/jm050362n

Zhu, W., Shyr, Z., Lo, D. C., and Zheng, W. (2021). Viral proteases as targets for coronavirus disease 2019 drug development. *J. Pharmacol. Exp. Ther.* 378, 166–172. doi:10.1124/jpet.121.000688





## OPEN ACCESS

## EDITED BY

Gihyun Lee,  
Kyung Hee University, Republic of Korea

## REVIEWED BY

Won-Yung Lee,  
Gachon University, Republic of Korea  
Sang-Min Park,  
Chungnam National University, Republic  
of Korea

## \*CORRESPONDENCE

Sanghun Lee,  
✉ ezhani@kiom.re.kr

RECEIVED 20 March 2023

ACCEPTED 01 August 2023

PUBLISHED 10 August 2023

## CITATION

Tran MN, Baek S-J, Jun HJ and Lee S  
(2023), Identifying target organ location  
of Radix Achyranthis Bidentatae: a  
bioinformatics approach on active  
compounds and genes.  
*Front. Pharmacol.* 14:1187896.  
doi: 10.3389/fphar.2023.1187896

## COPYRIGHT

© 2023 Tran, Baek, Jun and Lee. 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.

# Identifying target organ location of Radix Achyranthis Bidentatae: a bioinformatics approach on active compounds and genes

Minh Nhat Tran<sup>1,2,3</sup>, Su-Jin Baek<sup>1</sup>, Hyeong Joon Jun<sup>1</sup> and Sanghun Lee<sup>1,2\*</sup>

<sup>1</sup>Korean Medicine Data Division, Korea Institute of Oriental Medicine, Daejeon, Republic of Korea, <sup>2</sup>Korean Convergence Medical Science, University of Science and Technology, Daejeon, Republic of Korea, <sup>3</sup>Faculty of Traditional Medicine, Hue University of Medicine and Pharmacy, Hue University, Hue, Vietnam

**Background:** Herbal medicines traditionally target organs for treatment based on medicinal properties, and this theory is widely used for prescriptions. However, the scientific evidence explaining how herbs act on specific organs by biological methods has been still limited. This study used bioinformatic tools to identify the target organ locations of Radix Achyranthis Bidentatae (RAB), a blood-activating herb that nourishes the liver and kidney, strengthens bones, and directs prescription to the lower body.

**Methods:** RAB's active compounds and targets were collected and predicted using databases such as TCMSP, HIT2.0, and BATMAN-TCM. Next, the RAB's target list was analyzed based on two approaches to obtain target organ locations. DAVID and Gene ORGANizer enrichment-based approaches were used to enrich an entire gene list, and the BioGPS and HPA gene expression-based approaches were used to analyze the expression of core genes.

**Results:** RAB's targets were found to be involved in whole blood, blood components, and lymphatic organs across all four tools. Each tool indicated a particular aspect of RAB's target organ locations: DAVID-enriched genes showed a predominance in blood, liver, and kidneys; Gene ORGANizer showed the effect on low body parts as well as bones and joints; BioGPS and HPA showed high gene expression in bone marrow, lymphoid tissue, and smooth muscle.

**Conclusion:** Our bioinformatics-based target organ location prediction can serve as a modern interpretation tool for the target organ location theory of traditional medicine. Future studies should predict therapeutic target organ locations in complex prescriptions rather than single herbs and conduct experiments to verify predictions.

## KEYWORDS

bioinformatics, Radix Achyranthis Bidentatae, target organ location, enrichment analysis, gene expression analysis

## 1 Introduction

The rapid growth of biological data, the development of algorithms, and the increase in computer power have made bioinformatics an important contributor to a deeper understanding of existing drugs and to the development of new drugs, in both traditional and modern medicine. Furthermore, herbal medicine-related 'omics' data and

methods for analyzing molecular mechanisms and biological pathways of herbal medicine have been used to provide innovative ideas (Gu and Chen, 2014). Several bioinformatics-based methods such as network pharmacology, herbal genomics, molecular dynamics simulation, and molecular docking provide knowledge and insight into herbal medicine from different perspectives. The mechanisms underlying herbs and prescriptions have been revealed and further detailed through “multi-target–multi-pathway” paradigms of network pharmacology (Liang et al., 2014; Zhang et al., 2019; Tran and Lee, 2022). These methods are also useful for understanding specific concepts of traditional medicine at the molecular scale such as the synergism of herbal pairs (Wang et al., 2017), or the Qi and Blood-tonifying effects of a group of herbs (Sun et al., 2016; Tran et al., 2022). The common feature of these studies is that the molecular mechanisms were determined using databases to generate large gene sets of interest analyses. Although these types of lists are now frequently produced by biological research, it is still a challenging task to comprehend how gene sets affect an organism’s biology at the tissue and organ level (Brookes and Robinson, 2015).

To overcome this obstacle, multiple tools have been established, allowing researchers to directly or indirectly identify the target organ or tissue locations connected to the genes of interest. For example, Database for Annotation, Visualization and Integrated Discovery (DAVID) is a popular tool that allows lists of genes to be enriched for shared biological pathways, disease associations, as well as tissue expressions (Sherman et al., 2022). Another tool based on gene list enrichment is the Gene ORGANizer, which, unlike DAVID, considers gene–phenotype associations to directly link genes to the human body parts affected by those genes (Gokhman et al., 2017). BRITE (Kanehisa et al., 2016) and Organ System Heterogeneity DB (Mannil et al., 2014) also provide direct linkages between genes and body parts; however, they consider only a few organs and tissues and were not developed for gene list analysis (Gokhman et al., 2017). A further approach for indirectly associating genes to organs is based on expression, which uses mRNA levels to identify the tissues and cell types in which a gene is active, rather than analyzing the enrichment in a list of genes. For instance, the Human Protein Atlas (HPA) is a database that contains localization and expression data for all essential human organs or tissues (Uhlén et al., 2015), and BioGPS provides abundance gene expression data corresponding to tissues or cells based on microarray analyses (Wu et al., 2016). Gene expression is influenced by physiological factors that differ depending on tissue type, and developmental stage. The particular gene expression patterns in organs and tissues provide critical insights regarding gene function (Su et al., 2004; Pan et al., 2013). Therefore, it is critical to assess the tissue mRNA expression patterns of diverse genes at the organ level to investigate the therapeutic effects of herbal target proteins on organs. However, to date, only few studies have used BioGPS to analyze target organ/tissue location for herbs and prescriptions such as *Rhodiola rosea* L. (Zhang X. et al., 2020), *Acori Tatarinowii* Rhizoma (Song et al., 2018), and Sanhe Decoction (Zhang et al., 2016). Furthermore, the use of BioGPS and the interpretation of its results in these studies remain insufficient. Additionally, other tools that have not been integrated to link herbs to anatomical body parts present an unexplored aspect of this line of research.

Target organ locations, including herbal channel tropism (HCT), is a foundational theory that has influenced traditional treatment for thousands of years (Liu et al., 2013). As per this theory, the therapeutic actions of an herb have selective effects on particular physiological organs or channels (World Health Organization, 2007). Numerous recent studies have demonstrated the value of using systems biology to assess the scientific significance of herbal medicine (Buriani et al., 2012). *Radix Achyranthis Bidentatae* (RAB), also known as Niuxi, is a blood-activating medicinal herb obtained from the dried roots of *Achyranthes bidentata* Bl. Traditionally, RAB is considered to enter and supplement the liver and kidneys, invigorate blood circulation, and reinforce tendons and bones; directing effects of the prescription to the lower part of the body have also been reported (Chinese Pharmacopoeia Commission, 2015). Research on RAB in modern medicine mostly focused on its pharmacological effects on the bone metabolism, nervous system, and immune system; joint-protection properties; and antioxidation and antitumor effects (He et al., 2017). Using a bioinformatics-based approach, the mechanisms underlying the effects of RAB in the treatment of diseases such as rheumatic arthritis (Fu et al., 2021), osteoarthritis (Zhang et al., 2020), bone trauma (Wu et al., 2021), and breast cancer (Ju et al., 2021) have been elucidated. Nevertheless, the understanding of pathology still differs between traditional and modern medicines, as the terms used for diseases are not identical between the two fields. Specifically, traditional medicine often divides disease models into patterns related to body components or organs, such as Liver and Kidney yin deficiency or Qi and Blood deficiency (Lam et al., 2019). Therefore, principles behind herb-based treatments of diseases are to apply the HCT theory to rebalance body constituents and organ systems. Understanding how multi-organ systems respond to medicinal herbs at a biological system level may assist in the development of improved diagnosis and treatment strategies for complex diseases. However, the application of bioinformatic tools to the association of herbs with organs has not been sufficiently investigated.

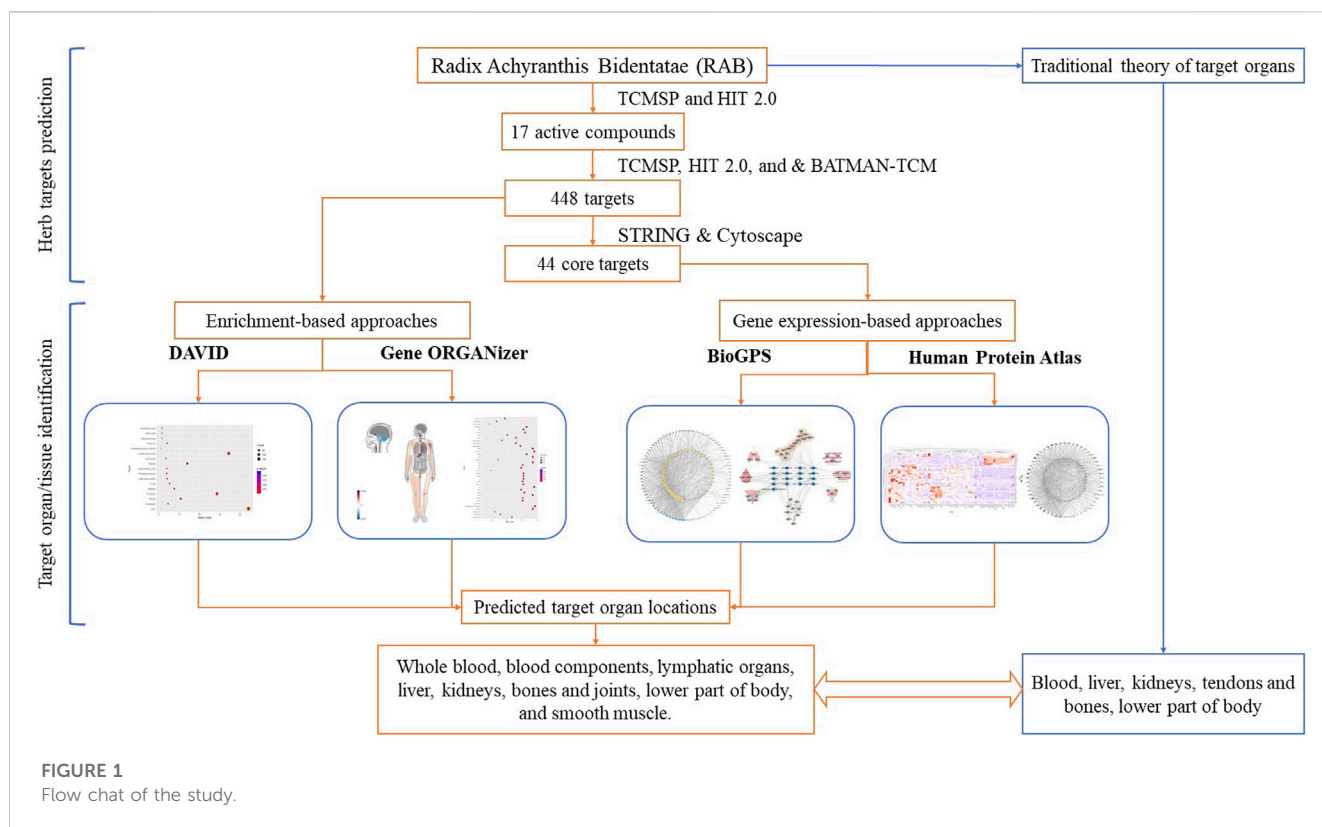
Therefore, in this study, we used different bioinformatic tools to analyze the target organ/tissue location of herbs, considering RAB as an example. Using gene expression and enrichment-based tools to analyze the targets of RAB, we provided new insight into the HCT mechanisms at the organ/tissue level. The flowchart of this study is shown in Figure 1.

## 2 Methods

### 2.1 Predicting the RAB targets

The RAB targets were predicted in two steps: first, the active compounds were screened, and second, different databases were utilized to predict the targets of the active compounds.

The compounds associated with RAB were retrieved from the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP) version 2.3 (Ru et al., 2014) and the Herbal-Ingredient-Target Platform (HIT) version 2.0 (Yan et al., 2022) using “Niuxi” or “Niu Xi” as keywords. Subsequently, the synonym names, Chemical Abstracts Service (CAS), PubChem



compound IDs (CID), and International Chemical Identifier (InChIKey) numbers of the compounds were introduced into the PubChem database to obtain the compound structures and ensure that the compounds are recognizable for further steps (Kim et al., 2021). The duplicated compounds between databases or those with no information on the PubChem database were removed from further analysis. Next, to evaluate absorption, distribution, metabolism, and excretion, the oral bioavailability (OB) and drug-likeness (DL) of compounds were obtained from the TCMSP. The active compounds with an OB  $\geq 30\%$  and DL  $\geq 0.18$  were selected (Xu et al., 2012).

Targets of RAB active compounds were obtained from the TCMSP version 2.3 (Ru et al., 2014), HIT version 2.0 (Yan et al., 2022), and Bioinformatics Analysis Tool for Molecular mechanism of TCM (BATMAN-TCM) databases (Liu et al., 2016). From TCMSP, compound-target linkages were acquired via two different approaches: 1) experimentally verified compound-target pairings were obtained from the HIT version 1.0 database; 2) For compounds lacking verified targets, the potential targets were predicted using the SysDT model (Ru et al., 2014). The HIT version 2.0 contains completely updated data compared to HIT version 1.0 by calibrating literature data from 2000 to 2010 and adding experimental data from 2010 to 2020; thus, yielding nearly twice as much data compared to that yielded by the previous version and additional features of target confidence indicators (Yan et al., 2022). The names, CIDs, and CASs of compounds were introduced in the HIT 2.0 database to obtain targets. The targets with levels of A, B, and C were selected for this study. BATMAN-TCM, a similarity-based approach that ranks probable compound-target linkages based on their similarity to the known drug-target interactions,

was used to predict potential targets of compounds (Liu et al., 2016). The CIDs for each compound were inputted into the BATMAN-TCM, and the predicted potential targets (including known targets) with a Score\_cutoff = 30 were selected. Targets were collected from the three databases. The official gene symbol of the “*Homo sapiens*” genes was obtained from the UniProt database (The UniProt Consortium, 2020), duplicate targets were deleted, and the remaining targets were used for enrichment-based analysis to identify target organ or tissue location.

Core targets were identified by importing all RAB targets into the Search Tool for the Retrieval of Interacting Genes (STRING) database to generate a protein-protein interaction (PPI) network (Szklarczyk et al., 2021). Further, targets were imported into Cytoscape software for topology analysis (Shannon et al., 2003). The PPI network was set up in the STRING database with a high confident interaction score  $>0.9$ , *homo sapiens* as species, and FDR stringency = 5%. The disconnected nodes were hidden in the network. Based on the topological analysis in Cytoscape software, targets with degree  $>2$  times the average were selected as core targets. These targets were used for gene expression-based analysis to identify target organ or tissue location.

## 2.2 Linking genes to the organs using enrichment-based analysis approaches

### 2.2.1 Tissue expression assessment using the DAVID tool

DAVID is a well-known bioinformatics tool that includes a web server as well as a platform for enrichment analyses and functional

annotation of gene lists. From the version released in 2006, a gene tissue expression annotation category was added in DAVID and continued to be updated (Huang et al., 2007b). Fisher's exact test is used in DAVID to calculate the gene enrichment within the annotation categories with EASE score ( $p$ -value). Importantly, the individual EASE scores were adjusted by multiple testing correction (adjusted  $p$ -value), such as Bonferroni, Benjamini, and false discovery rate (FDR) tests (Huang et al., 2007a).

In this study, the official names of RAB targets were imported into DAVID (version 2021) to analyze tissue expression annotations, with *Homo sapiens* as species and a  $p$ -value <0.05 (after Benjamini correction). Uniprot keyword annotations (UP\_TISSUE) were chosen for the Tissue expression category. The significant tissues were plotted as a bubble chart using the R package "ggplot2."

### 2.2.2 Gene ORGANizer tool

The phenotype-based tool Gene ORGANizer directly connects genes to the bodily parts they influence. It is based on a comprehensive curated database linking more than 7,000 genes to 150 anatomical parts utilizing more than 150,000 gene-organ associations based on DisGeNET and Human Phenotype Ontology databases. These data were converted into relationships between genes and anatomical regions where the phenotype was observed. Overall, the Gene ORGANizer connected 146 different body parts to target genes. The hypergeometric distribution was used in Gene ORGANizer to calculate the significance level of enrichment or depletion, and the mid-range correction was utilized to obtain  $p$ -values (Gokhman et al., 2017).

In this study, the list of RAB targets was imported into the Gene ORGANizer to analyze gene-organ associations, with a  $p$ -value <0.05, confident in curation level (inferred from data on humans), typical in frequency of phenotypes (appear in >50% of sick individuals), and FDR in multiple testing corrections. The significant organs and body parts were plotted as a bubble chart using the R package "ggplot2." The significantly enriched or depleted body parts were visualized as a heat map based on their enrichment or depletion level obtained from the Gene ORGANizer web platform.

## 2.3 Linking genes to the organs using gene expression-based approaches

### 2.3.1 BioGPS tool

BioGPS is a database for accessing and managing genomic annotation tools. It offers gene expression data from tissues or cells, based on microarray analyses (Wu et al., 2009). BioGPS provides a 'Gene expression/activity chart' plugin with a dataset collection function that allows pre-loading of approximately 8,000 datasets from EBI's ArrayExpress and NCBI's GEO repositories. These datasets are derived from nine microarray platforms containing data on humans, mice, and rats (Wu et al., 2016).

We constructed the gene-organ localization network using the GeneAtlas U133A gcrrna dataset. First, 84 organ tissue samples were used to determine the mRNA expression patterns of each RAB core targets. Second, the overall expression average value across all organs

was calculated. Third, genes were linked in the relevant organs where the mRNA expression level was higher than the average. Finally, gene-organ networks were constructed using Cytoscape 3.4.0.

### 2.3.2 HPA tool

The HPA is a crucial tool for identifying single gene expression patterns in tissue, blood, brain, and cell lines. It contains spatial data regarding the human proteome based on integrated omics methods. This database was generated based on 44 samples from the major tissues and organs of the human body, which were examined using 24,028 antibodies and 16,975 protein-encoding genes, together with RNA-sequencing data for 32 tissues (Uhlén et al., 2015).

In this study, we first downloaded RNA consensus tissue data for RAB core targets. Based on transcriptomics data from HPA (Uhlén et al., 2015) and the Genotype-Tissue Expression (Lonsdale et al., 2013), consensus transcript expression levels per gene were summarized in 54 tissues. The highest transcripts per kilobase million (TPM) value for each gene across the two data sources was used to compute the consensus normalized expression (nTPM) value. The gene expression was then converted to a Z-score for comparing the expression of each gene in different tissues. The R package "ComplexHeatmap" was used to generate a heat map of the core gene expression levels. Lastly, a gene-tissue location network was established by linking the gene to the relevant organs or tissues where it was overexpressed, using Cytoscape. Gene expressions with Z-score >0 were considered to represent overexpression.

## 3 Results

### 3.1 Predicting targets of Radix Achyranthis Bidentatae

By retrieving data from the TCMSP and HIT 2.0 databases, 176 and 13 related compounds were obtained for RAB, respectively. A total of 17 RAB active compounds were screened using OB and DL filtering criteria (Table 1). Beta-sitosterol,  $\beta$ -hydroxyecdysone, and spinasterol were active compounds present in both databases.

By entering the information of 17 active compounds into the databases, a total of 448 RAB targets were identified (Supplementary Table S1). Specifically, the TCMSP, HIT2.0, and BATMAN-TCM databases identified 207, 312, and 78 targets, respectively. After removing duplicates, 448 targets were selected. These targets were used for enrichment analyses to identify target organ locations.

Further, the 448 RAB targets were introduced into the STRING database to generate a PPI network (Figure 2A). Topological analysis of this network was performed using Cytoscape, and we identified 44 core targets, which had a degree centrality two-fold higher than the mean value (Figure 2B). Next, these targets were used for gene expression analyses to identify the target organ location.

### 3.2 Linking genes to the organs using enrichment approaches

Using the DAVID tool, various tissues were enriched from the RAB target list as shown in Figure 3. The top five most enriched tissues included liver, fibroblast, blood, placenta, and platelets.



**TABLE 1** List of *Radix Achyranthis Bidentatae* active compounds.

Molecule name	Pubchem ID	OB (%)	DL	TCMSP ID	HIT2.0 ID
Poriferasta-7,22E-dien-3beta-Ol	5283663	42.98	0.76	MOL001006	
Spinoside A	5281325	41.75	0.4	MOL012537	
B-ecdysterone	27545171	44.23	0.82	MOL012542	C0653
Berberine	2353	36.86	0.78	MOL001454	
Coptisine	72322	30.67	0.86	MOL001458	
Wogonin	5281703	30.68	0.23	MOL000173	
Delta 7-stigmastenol	12315385	37.42	0.75	MOL002643	
Baicalein	5281605	33.52	0.21	MOL002714	
Baicalin	64982	40.12	0.75	MOL002776	
Epiberberine	160876	43.09	0.78	MOL002897	
Beta-sitosterol	222284	36.91	0.75	MOL000358	C1178
Inophyllum E	455251	38.81	0.85	MOL003847	
Kaempferol	5280863	41.88	0.24	MOL000422	
Spinasterol	5281331	42.98	0.76	MOL004355	C0750
Stigmasterol	5280794	43.83	0.76	MOL000449	
Palmatine	19009	64.6	0.65	MOL000785	
Quercetin	5280343	46.43	0.28	MOL000098	

Interestingly, the liver ( $p = 3.1\text{e-}12$ ) was the most enriched organ; along with the kidneys ( $p = 1.4\text{e-}3$ ), which are the two channel tropisms of RAB in traditional medicine. In addition, blood ( $p = 1.5\text{e-}7$ ) and blood-related components, including platelets ( $p = 5.2\text{e-}5$ ), peripheral blood ( $p = 4.6\text{e-}4$ ), plasma ( $p = 1.0\text{e-}2$ ), and endothelial cells ( $p = 4.9\text{e-}2$ ), were also significantly enriched.

Using the Gene ORGANizer, we identified several organs and body parts that exhibited gene enrichment or depletion as shown in Figure 4. In the bubble chart of Figure 4A, the terms were divided into three categories, including organs, bones, and body parts. Cranial nerves, cerebellum, lymph nodes, intestine, and peripheral nervous system were the most significantly affected organs. The results also showed that the RAB targets affect several bones, including scapula, tibia, fibula, and femur. Interestingly, depleted organs or body parts (blue color) were mostly located within the head, specifically cranial nerves (FDR =  $2.20\text{e-}05$ ), cerebellum (FDR =  $5.78\text{e-}05$ ), outer ear (FDR = 0.01), and nose (FDR = 0.03); in contrast, the enriched organs and body parts (red color) were mainly distributed in the lower body and limbs, including hip, wrist, elbow (FDR = 0.01), knee, thigh, shoulder (FDR = 0.01), and ankle (FDR = 0.02) (Figures 4A,B). This suggests that RAB targets tend to increase their phenotype expression in lower parts of the body rather than the head.

### 3.3 Linking genes to the organs using gene expression approaches

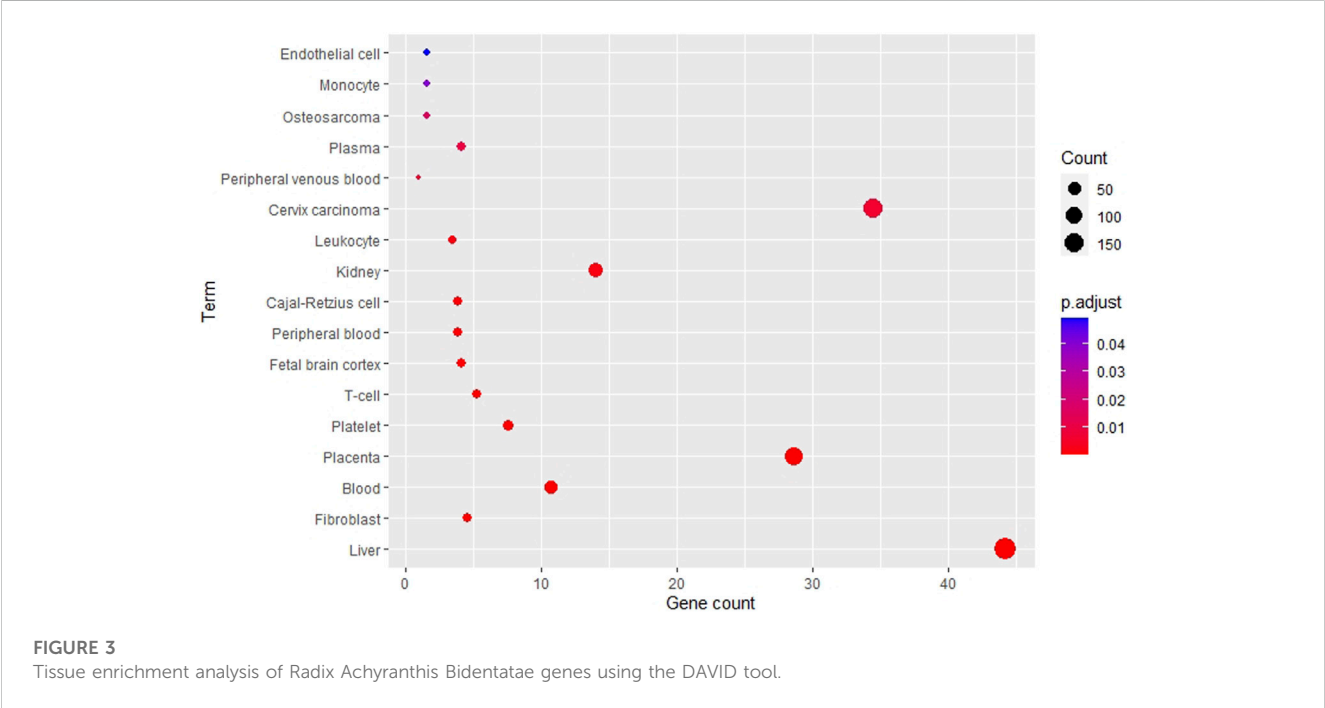
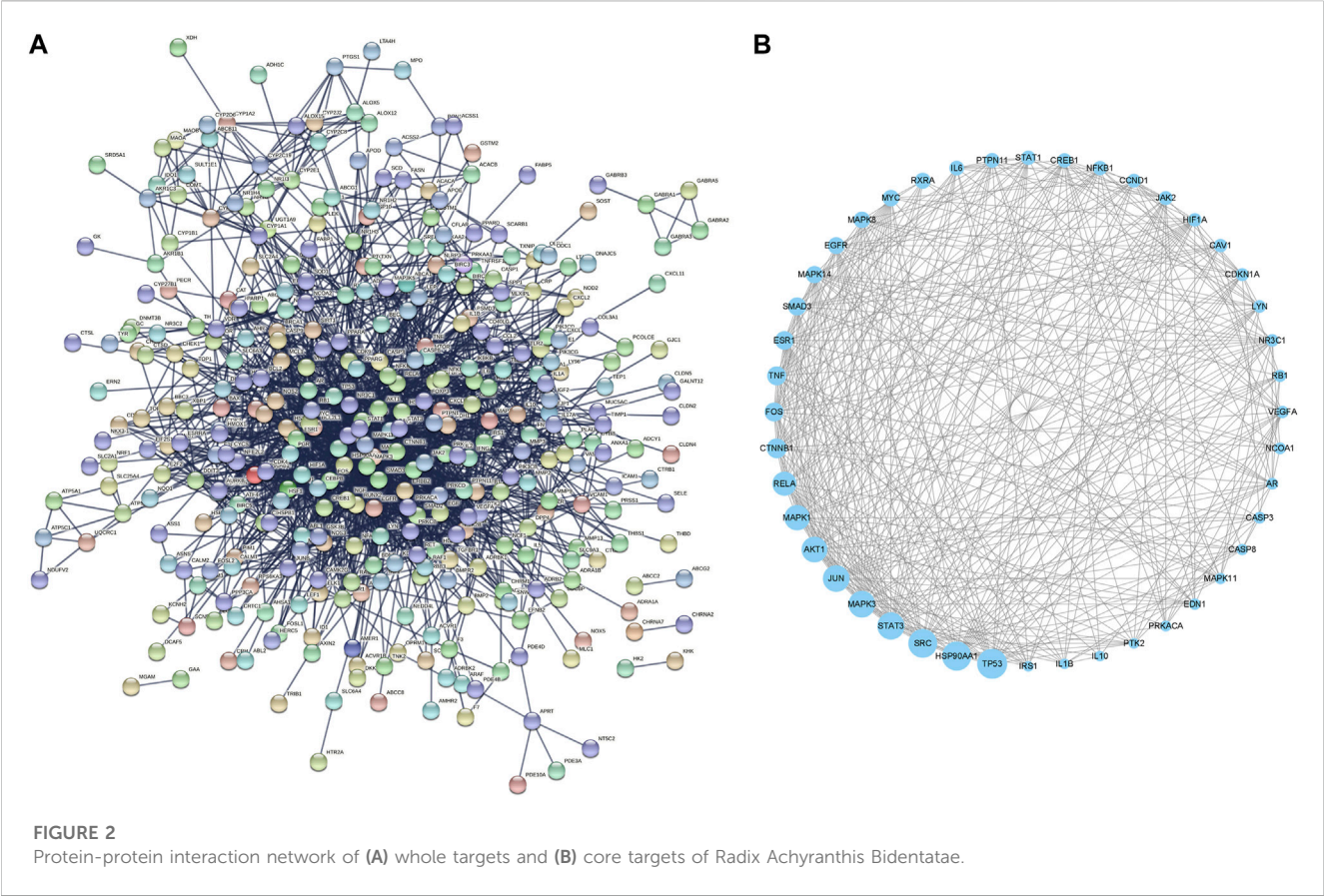
Specific tissue mRNA expression patterns provide vital insights into gene function. Therefore, it is critical to know the tissue mRNA

expression patterns of numerous genes at the organ level to investigate the HCT and traditional effects of RAB. In this study, using gene expression data from the BioGPS and the HPA databases for 44 core targets spread throughout several different tissues and organs, we created target organ location networks to better understand the effects of RAB at the organ- and tissue-level.

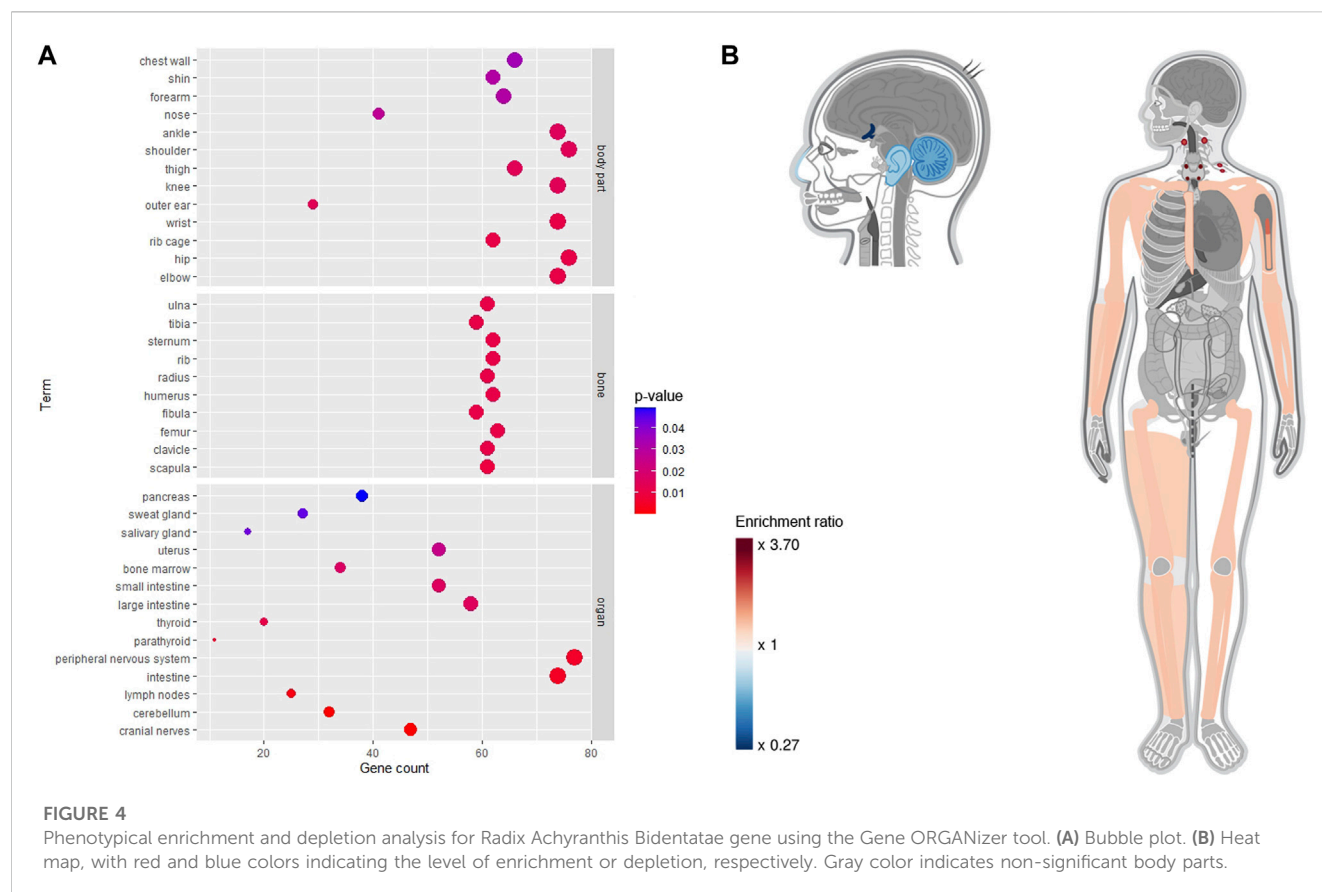
The expression patterns of 44 core targets in 84 normal tissues obtained from the BioGPS database are shown in Figure 5A. These core targets were mostly found in human tissues; however, their mRNA expression levels varied. Networks of RAB targets-organ locations are shown in Figure 5B, where the linkages between genes and organs are represented using the mRNA expressions in each organ. In particular, of 84 organ tissues, CD33<sup>+</sup> myeloid cells showed the highest number of RAB target overexpression, with 33 of the 44 core targets being overexpressed. This was followed by 32 targets being overexpressed in CD56<sup>+</sup> natural killer cells, 31 in CD34<sup>+</sup> hematopoietic stem cells, 30 in B lymphoblasts, 29 in whole blood, 28 in smooth muscle, and 25 in CD14<sup>+</sup> monocytes and BDCA4<sup>+</sup> dendritic cells. These results clearly indicate that most targets were overexpressed simultaneously in numerous tissues.

Next, we focused and regrouped on top 15 target organs in the network, including bone marrow (grouped by CD33<sup>+</sup>, CD34<sup>+</sup>, and B lymphoblasts), whole blood (grouped by CD14<sup>+</sup>, CD56<sup>+</sup>, and whole blood), lymphoid tissues (BDCA4<sup>+</sup>, Raji, CD19<sup>+</sup>, and CD4<sup>+</sup>), prostate, smooth muscle, cardiac myocytes, thyroid, and lung. The regrouping was based on cell location (such as lymphoblast, CD14<sup>+</sup>, or CD4<sup>+</sup>) and organ of origin (bone marrow, blood, or lymphoid tissues) during hematopoiesis from stem cells (Orkin and Zon, 2008). RAB targets were linked to the tissues where their mRNA expression levels were the highest. We specifically analyzed





the expression profiles of various tissues and discovered that the 14 genes (*TNF*, *MAPK14*, *CTNNB1*, *HSP90AA1*, *CREB1*, *NFKB1*, *NR3C1*, *TP53*, *PTPN11*, *STAT1*, *MYC*, *FOS*, *JAK2*, and *JUN*) showed higher expression levels in the bone marrow than in other tissues, 9 genes (*SRC*, *MAPK8*, *ESR1*, *LYN*, *EGFR*, *MAPK11*, *RELA*, *CASP8*, and *CASP3*) were lymphoid tissue-specific, 4 were associated with



whole blood (*NCOA1*, *RXRA*, *MAPK1*, and *PRKACA*), smooth muscle (*IL1B*, *CDKN1A*, *HIF1A*, and *IL6*), and the prostate (*VEGFA*, *AR*, *AKT1*, and *CCND1*), and 3 were associated with lungs (*EDN1*, *MAPK3*, and *STAT3*), cardiac myocytes (*RB1*, *IL10*, and *SMAD3*), and thyroid (*CAV1*, *PTK2*, and *IRS1*). Based on their expression patterns, 17 active compounds and 44 core targets were organized into an organ location network (Figure 5C).

As shown in Figure 6A, 44 core targets had their expression profiles mapped onto 54 normal tissues from the HPA database. Core targets had the highest total expression in organs such as bone marrow (4,806.8 nTPM), adipose tissue (4,725.4 nTPM), fallopian tube (4,271.6 nTPM), skin (3,940.5 nTPM), and cerebral cortex (3,936.3 nTPM).

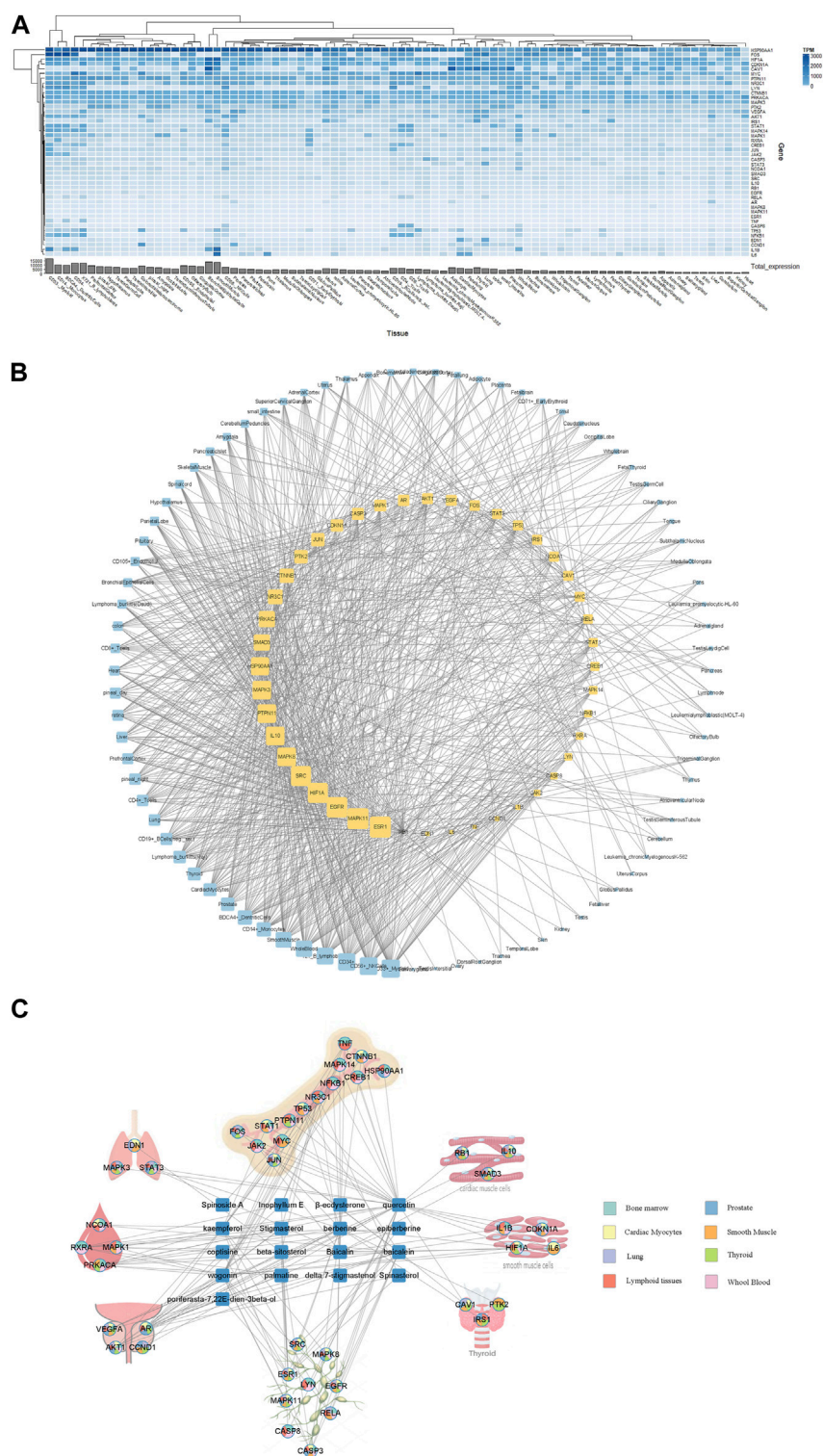
Expression profiles have been converted to Z-scores (Figure 6B), and a network of tissue distribution-potential genes was constructed based on the Z-scores. The network is shown in Figure 6C, including 98 nodes and 860 edges. Each edge represented overexpression (Z-score >0) of the potential gene in a specific tissue or organ. The network analysis results showed that 44 targets were distributed among the adipose tissue (degree = 35), urinary bladder (degree = 28), breast (degree = 28), esophagus (degree = 27), lung and spleen (degree = 26), as well as bone marrow, liver, and smooth muscle (degree = 23). In particular, *STAT3* (degree = 30) and *NFKB1* (degree = 29) were the genes most closely related to organs or tissues.

Ultimately, to obtain a general overview of RAB-related organs following the use of four analytical tools, a comparison of the most affected organs or tissues is described in Table 2.

## 4 Discussion

The target organ location represents the foundation of the traditional theory and describes an herb's selective effects on specific parts of the body. In traditional medicine, RAB is a blood-activating herb that regulates blood circulation, strengthens bones as well as muscles, and nourishes the liver and kidneys. In prescriptions, it can direct other herbs to have therapeutic effects on the lower part of the body (Chinese Pharmacopoeia Commission, 2015). Classical pharmacological and emerging bioinformatics-based studies have shown that RAB has pharmacological as well as therapeutic effects on various diseases; however, its HCT and effects on anatomical organs have not been adequately investigated. This study revealed the link between RAB and target organ locations by using four different bioinformatic tools. Collectively, RAB targets were associated with whole blood, blood components, and lymphatic organs. Using the DAVID tool, we showed that RAB targets were highly enriched in whole blood, liver, and kidneys. Furthermore, the Gene ORGANizer analysis showed the effects of RAB on lower body parts as well as bones and joints. Alternatively, BioGPS and HPA indicated high RAB-associated gene expression levels in bone marrow, lymphoid tissue, and smooth muscle. Our study revealed the similarities between the target organ locations predicted through bioinformatics and those indicated by traditional medicine.

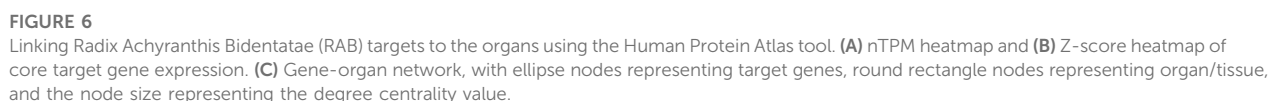
In this study, we proposed two different approaches to investigate RAB-associated genes: enrichment- and expression-based analyses. Whereas enrichment analyses assess the entire



**FIGURE 5** Linking *Radix Achyranthis Bidentatae* (RAB) targets to organs using the BioGPS tool. **(A)** Core target gene expression. **(B)** Gene-organ network. Yellow and blue nodes represent genes and organs, respectively. Edges represent the gene displayed above average mRNA expressions in each organ, whereas the node size represents the degree centrality value. **(C)** Compound-gene-organ network of most significant organs. The node pie chart illustrates the organs in which each target has a high expression level.

gene list and usually result in target organ locations (e.g., heart), expression analyses show the expression of individual genes and typically concentrate on particular tissues or cell types (e.g., cardiomyocytes) instead of entire organs or systems. Nevertheless, our results still showed a general trend for RAB-associated genes to be associated with whole blood, blood





improvements in hemorheological properties (He et al., 2017). A water extract of RAB, for example, was shown to decrease the erythrocyte aggregation index, hematocrit, and whole blood viscosity in wild type rats *in vivo* and significantly prolonged plasma recalcification time, kaolin partial thromboplastin time,

TABLE 2 Top 15 target organ/tissue location of the four tools.

No.	Enrichment-based (sorted by <i>p</i> -value)		Expression-based (sorted by network degree)	
	DAVID	Gene ORGANizer	BioGPS	Human Protein Atlas
1	Liver	cranial nerves	CD33 <sup>+</sup> myeloid cells	adipose tissue
2	Fibroblast	cerebellum	CD56 <sup>+</sup> NK cells	breast
3	Blood	lymph nodes	CD34 <sup>+</sup>	urinary bladder
4	Placenta	intestine	721 B lymphoblasts	esophagus
5	Platelet	peripheral nervous system	Whole blood	spleen
6	T-cell	parathyroid	Smooth muscle	lung
7	Peripheral blood	thyroid	BDCA4 <sup>+</sup> dendritic cells	bone marrow
8	Fetal brain cortex	large intestine	CD14 <sup>+</sup> monocytes	liver
9	Cajal-Retzius cell	small intestine	Prostate	smooth muscle
10	Kidney	bone marrow	Cardiac Myocytes	cervix
11	Leukocyte	uterus	Thyroid	gallbladder
12	Peripheral venous blood	salivary gland	Lymphoma burkitts (Raji)	ovary
13	Cervix carcinoma	sweat gland	CD19 <sup>+</sup> B Cells (neg._sel.)	endometrium
14	Plasma	pancreas	Lung	skeletal muscle
15	Osteosarcoma	N/A	CD4 <sup>+</sup> T cells	tongue

and prothrombin time *in vitro* (Li et al., 1990). RAB has been demonstrated to affect the blood rheological property index of an acute blood stasis rat model by significantly lowering the thrombus sizes, the hematocrit, the platelet adhesion rate, the platelet adhesion value, whole blood viscosity, shear relative viscosity, and fibrinogen content (Si et al., 2007). RAB has also been suggested as a potential regulator that affects a variety of immune cells, including natural killer cells, T-lymphocytes, and macrophages (He et al., 2017). In addition, considering whole blood, blood components, and lymphatic organs, each approach explains a distinct aspect of RAB target organ locations.

Enrichment is a common approach for addressing trends within a list of genes, where gene sets are formed based on shared biological or functional properties identified by reference knowledge of the biological domain (Mathur et al., 2018). Our DAVID-based enrichment analysis showed that RAB-associated genes are involved in several blood components such as platelet, peripheral blood, plasma, and endothelial cell (Figure 3). Plasma, with albumin as the predominant protein, is a primary moderator of fluid flow across body compartments and a primary regulator of oncotic pressure in blood vessels (Panali et al., 2012). The mechanoreceptors of endothelial cells allow them to detect shear stress induced by blood flow across their surface, allowing the blood vessel to adjust its wall thickness and diameter to adapt the blood circulation. Endothelial cells also promote quick responses to brain impulses for blood vessel dilatation by releasing the Nitric oxide, which causes smooth muscle in the vessel wall to relax (Alberts et al., 2002). Alternatively, RAB-associated genes were also highly enriched in the liver and kidney organs, which is in line with the HCT theory and the proclaimed nourishing effect on the liver and kidney. Although the concept of viscera in traditional and modern

medicine is not identical, experimental studies clarifying this discrepancy are lacking. Therefore, this gap in knowledge needs to be addressed, and further experiments should be suggested. For the Gene ORGANizer, the list of genes is enriched or depleted against the genome background. Interestingly, for the RAB-associated genes, the head and neck areas were depleted, and the lower body was enriched, with a focus on the bones (Figure 4B); thus, indicating that RAB tends to affect the lower part of the body rather than the upper part. RAB has been previously shown to have an effect on muscles and bones. It can enhance bone strength, and inhibit bone loss by adjusting phosphorus excretion and urinary calcium (He et al., 2010; Zhang et al., 2012), as well as create an environment that is favorable for ossification by increasing blood flow during bone reconstruction (Jiang et al., 2014).

The gene expression-based approach, unlike the enrichment approach that can process the entire gene list, shows the expression of individual genes and places them in a general context. The mRNA expression patterns in tissues provide valuable information for deciphering gene function. In previous studies on target organ locations of herbs, genes with high expression in certain tissues were investigated; however, the selection of cutoff points for high gene expression and subsequent construction of the gene-organ network also varied widely between studies (Wang et al., 2015; Zhang et al., 2016; Zhang et al., 2020; Wang et al., 2020). In this study, we normalized gene expression by mean and Z-score to select genes with high expression. The higher mRNA expression in a tissue compared with the average expression of 84 tissues for each gene was used for analysis using the BioGPS tool (Wang et al., 2015). Furthermore, we selected a cutoff Z-score >0 for the HPA tool. Both HPA and BioGPS showed high expression level for RAB-associated genes in lymphatic organs such as bone marrow and



lymphoid tissues. In addition, smooth muscle was also a tissue type with notable RAB-associated gene expression (Figures 5, 6). The primary site of new blood cell production (hemopoiesis) is bone marrow (Lucas, 2021) and it is similar to the role of invigorating herbs in traditional medicine. RAB has a considerable impact on uterine smooth muscle depending on the species and the different physiological conditions (Guo et al., 1997; Yijun et al., 2002). This ties in the role of RAB in the treatment of menstrual pain caused by blood stasis (Ran et al., 2021).

While enrichment and expression-based approaches were shown to be useful for identifying target organ locations in this study, both of them are associated with a number of limitations. First, various tools have different tissue resolutions. For instance, some datasets study the brain as a whole, whereas others study various parts separately (e.g., thalamus, cerebellum, midbrain). Consequently, information on the impact of entire organs is incomplete or skewed. Second, the datasets used have a massive bias towards certain organs (e.g., skin, blood, and brain), whereas other tissues are less explored or unavailable. Third, samples for gene expression analyses are typically collected postmortem, at specific developmental stages, and from selected organ parts. As a result, the collected data rarely accounts for all temporal and structural variations within organs. Fourth, the study was only conducted on a single herb; hence, the results may not reflect the accuracy and objectivity of these approaches. Future studies need to predict target organ locations for multiple herbs in the same herbal group or prescription.

Moreover, the use of bioinformatic tools and databases is always associated with several inherent limitations. For example, prediction of herbal targets by TCMSP and BATMAN-TCM can lead to false positives. Such false positives can introduce incorrect connections in biological networks, like protein-protein interactions; thus, distorting the overall understanding of the biological pathways and leading to inaccurate interpretations. To overcome this limitation, we used two approaches: 1) We combined the TCMSP and BATMAN-TCM databases with HIT 2.0, which includes experimentally verified compound-target pairings (Yan et al., 2022); TCMSP itself also incorporates experimentally validated targets from the HIT 1.0 database (Ru et al., 2014). 2) We used a higher cutoff score ( $\geq 30$ ) than the recommended threshold ( $\geq 20$ ) for BATMAN-TCM (Liu et al., 2016). Although the PPI network is essential for understanding cell physiology in both normal and diseased states, it has certain drawbacks as the interaction sets obtained from current literature are inadequate (Menche et al., 2015) and biased toward more research protein (Rolland et al., 2014). Future research must use objective methods to examine PPI networks. For example, Guney et al. suggested network-based drug-disease proximity to provide an unbiased assessment of the therapeutic efficacy of a pharmacological protein (Guney et al., 2016).

In conclusion, we used enrichment- and expression-based approaches to detect the target organ locations of RAB using different bioinformatic tools. We discovered that RAB targets were commonly involved in whole blood, blood components, bone marrow, and lymphoid tissue. Furthermore, each tool explained a particular aspect of RAB's effects on target organ locations: Enriched genes identified using the DAVID tool were observed in whole blood and its components, as well as the liver and kidneys. The Gene ORGANizer showed that the effects of RAB focus on the lower body

and impacts bones as well as joints. Lastly, BioGPS and HPA showed high RAB-associated gene expression levels in bone marrow and lymphoid tissue and significant expression levels in smooth muscle. Our study revealed the partial similarities between the RAB target organ locations predicted through bioinformatics and the traditional effects of RAB. Thus, to the best of our knowledge, our study presents the first example for further studies attempting to link genes to target anatomical organs and the HCT theory. In the future, this approach can be extended to predict the target organ location of complex prescriptions rather than individual herbs. Additionally, experimental studies that verify the predicted target organ locations are also necessary.

## 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

Conceptualization: SL and MNT. Methodology and formal analysis: MNT. Writing—original draft: MNT. Writing—review and editing: S-JB, HJJ, and SL. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by the Collection of Clinical Big Data and Construction of Service Platform for Developing Korean Medicine Doctor with Artificial Intelligence research project (grant number KSN1922110).

## 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/fphar.2023.1187896/full#supplementary-material>

## References

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). "Blood vessels and endothelial cells," in *Molecular biology of the cell* (New York: Garland Science).
- Brookes, A. J., and Robinson, P. N. (2015). Human genotype–phenotype databases: Aims, challenges and opportunities. *Nat. Rev. Genet.* 16 (12), 702–715. doi:10.1038/nrg3932
- Buriani, A., Garcia-Bermejo, M. L., Bosisio, E., Xu, Q., Li, H., Dong, X., et al. (2012). Omic techniques in systems biology approaches to traditional Chinese medicine research: Present and future. *J. Ethnopharmacol.* 140 (3), 535–544. doi:10.1016/j.jep.2012.01.055
- Chinese Pharmacopoeia Commission (2015). *Chinese pharmacopoeia*. Beijing, China: China Medical Science Press.
- Fanali, G., di Masi, A., Trezza, V., Marino, M., Fasano, M., and Ascenzi, P. (2012). Human serum albumin: From bench to bedside. *Mol. Aspects Med.* 33 (3), 209–290. doi:10.1016/j.mam.2011.12.002
- Fu, J., Wu, H., Wu, H., Deng, R., and Sun, M. (2021). Deciphering the metabolic profile and pharmacological mechanisms of *Achyranthes bidentata* blume saponins using ultra-performance liquid chromatography quadrupole time-of-flight mass spectrometry coupled with network pharmacology-based investigation. *J. Ethnopharmacol.* 274, 114067. doi:10.1016/j.jep.2021.114067
- Gokhman, D., Kelman, G., Amartely, A., Gershon, G., Tsur, S., and Carmel, L. (2017). Gene ORGANizer: Linking genes to the organs they affect. *Nucleic Acids Res.* 45 (W1), W138–W145. doi:10.1093/nar/gkx302
- Gu, P., and Chen, H. (2014). Modern bioinformatics meets traditional Chinese medicine. *Briefings Bioinforma.* 15 (6), 984–1003. doi:10.1093/bib/bbt063
- Guney, E., Menche, J., Vidal, M., and Barabasi, A. L. (2016). Network-based *in silico* drug efficacy screening. *Nat. Commun.* 7 (1), 10331. doi:10.1038/ncomms10331
- Guo, S., Che, X., and Fan, X. (1997). Effects of *Achyranthes bidentata* saponin a on animal uteri. *Touranl Xian Med. Univ.* 18, 216–218.
- He, C. C., Hui, R. R., Tezuka, Y., Kadota, S., and Li, J. X. (2010). Osteoprotective effect of extract from *Achyranthes bidentata* in ovariectomized rats. *J. Ethnopharmacol.* 127 (2), 229–234. doi:10.1016/j.jep.2009.11.016
- He, X., Wang, X., Fang, J., Chang, Y., Ning, N., Guo, H., et al. (2017). The genus *Achyranthes*: A review on traditional uses, phytochemistry, and pharmacological activities. *J. Ethnopharmacol.* 203, 260–278. doi:10.1016/j.jep.2017.03.035
- Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., et al. (2007a). The DAVID gene functional classification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 8 (9), R183. doi:10.1186/gb-2007-8-9-r183
- Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., et al. (2007b). DAVID bioinformatics resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* 35 (2), W169–W175. doi:10.1093/nar/gkm415
- Jiang, Y., Zhang, Y., Chen, W., Liu, C., Li, X., Sun, D., et al. (2014). *Achyranthes bidentata* extract exerts osteoprotective effects on steroid-induced osteonecrosis of the femoral head in rats by regulating RANKL/RANK/OPG signaling. *J. Transl. Med.* 12 (1), 334. doi:10.1186/s12967-014-0334-7
- Ju, Y., Liang, H., Du, K., Guo, Z., and Meng, D. (2021). Isolation of triterpenoids and phytosterols from *Achyranthes bidentata* Bl. to treat breast cancer based on network pharmacology. *Nat. Prod. Res.* 35 (24), 5939–5942. doi:10.1080/14786419.2020.1805603
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44 (D1), D457–D462. doi:10.1093/nar/gkv1070
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* 49 (D1), D1388–D1395. doi:10.1093/nar/gkaa971
- Lam, W. C., Lyu, A., and Bian, Z. (2019). ICD-11: Impact on traditional Chinese medicine and world healthcare systems. *Pharm. Med.* 33 (5), 373–377. doi:10.1007/s40290-019-00295-y
- Li, X., Li, W., and Chen, G. (1990). Experimental study on blood activating effect of *Achyranthes bidentata*. *J. Tradit. Chin. Med. Res.* 3, 27–29.
- Liang, X., Li, H., and Li, S. (2014). A novel network pharmacology approach to analyse traditional herbal formulae: The liu-wei-di-huang pill as a case study. *Mol. Biosyst.* 10 (5), 1014–1022. doi:10.1039/C3MB70507B
- Liu, P., Liu, S., Chen, G., and Wang, P. (2013). Understanding channel tropism in traditional Chinese medicine in the context of systems biology. *Front. Med.* 7 (3), 277–279. doi:10.1007/s11684-013-0273-3
- Liu, Z., Guo, F., Wang, Y., Li, C., Zhang, X., Li, H., et al. (2016). BATMAN-TCM: A bioinformatics analysis tool for molecular mechanism of traditional Chinese medicine. *Sci. Rep.* 6 (1), 21146. doi:10.1038/srep21146
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45 (6), 580–585. doi:10.1038/ng.2653
- Lucas, D. (2021). Structural organization of the bone marrow and its role in hematopoiesis. *Curr. Opin. Hematol.* 28 (1), 36–42. doi:10.1097/moh.0000000000000621
- Mannil, D., Vogt, I., Prinz, J., and Campillos, M. (2014). Organ system heterogeneity DB: A database for the visualization of phenotypes at the organ system level. *Nucleic Acids Res.* 43 (D1), D900–D906. doi:10.1093/nar/gku948
- Mathur, R., Rotroff, D., Ma, J., Shojaie, A., and Motsinger-Reif, A. (2018). Gene set analysis methods: A systematic comparison. *BioData Min.* 11 (1), 8. doi:10.1186/s13040-018-0166-8
- Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., et al. (2015). Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 347 (6224), 1257601. doi:10.1126/science.1257601
- Orkin, S. H., and Zon, L. I. (2008). Hematopoiesis: An evolving paradigm for stem cell biology. *Cell* 132 (4), 631–644. doi:10.1016/j.cell.2008.01.025
- Pan, J. B., Hu, S. C., Shi, D., Cai, M. C., Li, Y. B., Zou, Q., et al. (2013). PaGenBase: A pattern gene database for the global and dynamic understanding of gene function. *PLOS ONE* 8 (12), e80747. doi:10.1371/journal.pone.0080747
- Ran, D., Hong, W., Yan, W., and Mengdie, W. (2021). Properties and molecular mechanisms underlying geniposide-mediated therapeutic effects in chronic inflammatory diseases. *J. Ethnopharmacol.* 273, 113958. doi:10.1016/j.jep.2021.113958
- Rolland, T., Taşan, M., Charleaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., et al. (2014). A proteome-scale map of the human interactome network. *Cell* 159 (5), 1212–1226. doi:10.1016/j.cell.2014.10.050
- 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. cheminformatics* 6, 13. doi:10.1186/1758-2946-6-13
- 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 (11), 2498–2504. doi:10.1101/gr.1239303
- Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., et al. (2022). David: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 50 (W1), W216–W221. doi:10.1093/nar/gkac194
- Si, L., Huang, S., Li, T., and Guan, S. (2007). Effect of *Achyranthes bidentata* saponins on blood rheology index of acute blood stasis model rats. *Clin. J. Tradit. Chin. Med.* 19, 356–358.
- Song, Z., Yin, F., Xiang, B., Lan, B., and Cheng, S. (2018). Systems pharmacological approach to investigate the mechanism of *Acori Tatarinowii rhizoma* for alzheimer's disease. *Evidence-Based Complementary Altern. Med.* 2018, 5194016. doi:10.1155/2018/5194016
- Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci.* 101 (16), 6062–6067. doi:10.1073/pnas.0400782101
- Sun, J., Zhang, L., He, Y., Zhang, K., Wu, L., Fan, Y., et al. (2016). To unveil the molecular mechanisms of Qi and blood through systems biology-based investigation into Si-Jun-Zi-Tang and Si-Wu-tang formulae. *Sci. Rep.* 6 (1), 34328. doi:10.1038/srep34328
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids Res.* 49 (D1), D605–D612. doi:10.1093/nar/gkaa1074
- The UniProt Consortium (2020). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49 (D1), D480–D489. doi:10.1093/nar/gkaa1100
- Tran, M. N., Kim, S., Nguyen, Q. H. N., and Lee, S. (2022). Molecular mechanisms underlying qi-invigorating effects in traditional medicine: Network pharmacology-based study on the unique functions of qi-invigorating herb group. *Plants* 11 (19), 2470. doi:10.3390/plants11192470
- Tran, M. N., and Lee, S. (2022). The molecular mechanisms of *Panax ginseng* in treating type 2 diabetes mellitus: Network pharmacology analysis and molecular docking validation. *Evidence-Based Complementary Altern. Med.* 2022, 3082109. doi:10.1155/2022/3082109
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347 (6220), 1260419. doi:10.1126/science.1260419
- Wang, Y., Guo, G., Yang, B. R., Xin, Q. Q., Liao, Q. W., Lee, S. M. Y., et al. (2017). Synergistic effects of Chuanxiong-Chishao herb-pair on promoting angiogenesis at network pharmacological and pharmacodynamic levels. *Chin. J. Integr. Med.* 23 (9), 654–662. doi:10.1007/s11655-017-2408-x
- Wang, Y., Ru, Y., Zhuo, G., Sheng, M., Wang, S., Ma, J., et al. (2020). Investigation of the potential mechanism governing the effect of the *shen zhu san* on COVID-19 by network pharmacology. *Evidence-Based Complementary Altern. Med.* 2020, 8468303. doi:10.1155/2020/8468303
- Wang, Y., Zheng, C., Huang, C., Li, Y., Chen, X., Wu, Z., et al. (2015). Systems pharmacology dissecting holistic medicine for treatment of complex diseases: An

example using cardiocerebrovascular diseases treated by TCM. *Evidence-Based Complementary Altern. Med.* 2015, 980190. doi:10.1155/2015/980190

World Health Organization (2007). *WHO international standard terminologies on traditional medicine in the Western Pacific Region*. Manila: WHO Regional Office for the Western Pacific.

Wu, C., Jin, X., Tsueng, G., Afrasiabi, C., and Su, A. I. (2016). BioGPS: Building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res.* 44 (D1), D313–D316. doi:10.1093/nar/gkv1104

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., et al. (2009). BioGPS: An extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* 10 (11), R130. doi:10.1186/gb-2009-10-11-r130

Wu, L., Hao, Y., Dai, C., Zhang, Z., Ijaz, M., Ibrahim, S. M., et al. (2021). Network pharmacological study of *Achyranthis bidentatae* Radix effect on bone trauma. *BioMed Res. Int.* 2021, 5692039. doi:10.1155/2021/5692039

Xu, X., Zhang, W., Huang, C., Li, Y., Yu, H., Wang, Y., et al. (2012). A novel chemometric method for the prediction of human oral bioavailability. *Int. J. Mol. Sci.* 13 (6), 6964–6982. doi:10.3390/ijms13066964

Yan, D., Zheng, G., Wang, C., Chen, Z., Mao, T., Gao, J., et al. (2022). Hit 2.0: An enhanced platform for herbal ingredients' targets. *Nucleic Acids Res.* 50 (D1), D1238–D1243. doi:10.1093/nar/gkab1011

Yijun, Y., Xiaomin, Q., Tianzhen, Z., and Jingdong, X. (2002). Study of exciting mechanism about *achyranthes bidentata* blume on the spike activity of the uterine smooth muscle in virgin rats I. *J. Lanzhou Med. Coll.* 28 (1), 15–18.

Zhang, L., Shi, X., Huang, Z., Mao, J., Mei, W., Ding, L., et al. (2020a). Network pharmacology approach to uncover the mechanism governing the effect of *Radix Achyranthis bidentatae* on osteoarthritis. *BMC Complementary Med. Ther.* 20 (1), 121. doi:10.1186/s12906-020-02909-4

Zhang, R., Hu, S. J., Li, C., Zhang, F., Gan, H. Q., and Mei, Q. B. (2012). *Achyranthes bidentata* root extract prevent OVX-induced osteoporosis in rats. *J. Ethnopharmacol.* 139 (1), 12–18. doi:10.1016/j.jep.2011.05.034

Zhang, R., Zhu, X., Bai, H., and Ning, K. (2019). Network pharmacology databases for traditional Chinese medicine: Review and assessment. *Front. Pharmacol.* 10, 123. doi:10.3389/fphar.2019.00123

Zhang, W., Tao, Q., Guo, Z., Fu, Y., Chen, X., Shar, P. A., et al. (2016). Systems pharmacology dissection of the integrated treatment for cardiovascular and gastrointestinal disorders by traditional Chinese medicine. *Sci. Rep.* 6 (1), 32400. doi:10.1038/srep32400

Zhang, X., Zhu, J., Yan, J., Xiao, Y., Yang, R., Huang, R., et al. (2020b). Systems pharmacology unravels the synergic target space and therapeutic potential of *Rhodiola rosea* L. For non-small cell lung cancer. *Phytomedicine* 79, 153326. doi:10.1016/j.phymed.2020.153326



## OPEN ACCESS

## EDITED BY

David Ramirez,  
University of Concepcion, Chile

## REVIEWED BY

Yeray Rodriguez Nuñez,  
Andres Bello University, Chile  
Vinicius Gonçalves Maltarollo,  
Federal University of Minas Gerais, Brazil  
Jans Alzate-Morales,  
Universidad de Talca, Chile

## \*CORRESPONDENCE

Carmen Gil,  
✉ carmen.gil@csic.es

## †PRESENT ADDRESS

Tiziana Ginex,  
Pharmacelera, Barcelona, Spain

RECEIVED 22 June 2023

ACCEPTED 04 August 2023

PUBLISHED 11 August 2023

## CITATION

Ginex T, Madruga E, Martinez A and Gil C  
(2023), MBC and ECBL libraries:  
outstanding tools for drug discovery.  
*Front. Pharmacol.* 14:1244317.  
doi: 10.3389/fphar.2023.1244317

## COPYRIGHT

© 2023 Ginex, Madruga, Martinez and Gil.  
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.

# MBC and ECBL libraries: outstanding tools for drug discovery

Tiziana Ginex<sup>1†</sup>, Enrique Madruga<sup>1,2</sup>, Ana Martinez<sup>1,2</sup> and  
Carmen Gil<sup>1\*</sup>

<sup>1</sup>Centro de Investigaciones Biológicas "Margarita Salas" (CIB-CSIC), Madrid, Spain, <sup>2</sup>Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain

Chemical libraries have become of utmost importance to boost drug discovery processes. It is widely accepted that the quality of a chemical library depends, among others, on its availability and chemical diversity which help in rising the chances of finding good hits. In this regard, our group has developed a source for useful chemicals named Medicinal and Biological Chemistry (MBC) library. It originates from more than 30 years of experience in drug design and discovery of our research group and has successfully provided effective hits for neurological, neurodegenerative and infectious diseases. Moreover, in the last years, the European research infrastructure for chemical biology EU-OPENSOURCE has generated the European Chemical Biology library (ECBL) to be used as a source of hits for drug discovery. Here we present and discuss the updated version of the MBC library (MBC v.2022), enriched with new scaffolds and containing more than 2,500 compounds together with ECBL that collects about 100,000 small molecules. To properly address the improved potentialities of the new version of our MBC library in drug discovery, up to 44 among physicochemical and pharmaceutical properties have been calculated and compared with those of other well-known publicly available libraries. For comparison, we have used ZINC20, DrugBank, ChEMBL library, ECBL and NuBBE along with an approved drug library. Final results allowed to confirm the competitive chemical space covered by MBC v.2022 and ECBL together with suitable drug-like properties. In all, we can affirm that these two libraries represent an interesting source of new hits for drug discovery.

## KEYWORDS

drug discovery, chemical libraries, virtual screening, molecular diversity, chemical space

## 1 Introduction

Virtual high throughput screening (vHTS) represents a gold standard in modern drug discovery workflows especially for Pharma and Biotech companies (Subramaniam et al., 2008; Tanrikulu et al., 2013). Integration and complementation of *in silico* tools to classical HTS has boosted the capability of rapidly exploring a wider chemical space for the effective identification of new hits with indirect beneficial effects also on further steps of drug discovery as hit-to-lead optimization (Bajorath, 2002). The *in silico* techniques generally applied in this context are based on a common principle that is the accurate and effective assessment of the chemical complementarity between the protein target of interest and small molecules. In the case of ligand-based techniques as the mainstream QSAR-based (Neves et al., 2018) or pharmacophore-based (Kim et al., 2010) virtual screening, preliminary and



well-curated experimental data for representative chemical scaffolds are needed. These data in conjunction with a proper selection of relevant atomic or molecular descriptors for the compiled list of active compounds are then used to guide the search for new compounds. In the case of structure-based approaches as docking-based virtual screening, each compound in the chemical library is screened for its binding affinity toward a given target by using properly tuned scoring functions (Neves et al., 2021). Best scored molecular candidates from both approaches can be used for preliminary proof-of-concept. In case active compounds are found, they can go to further structural optimization with the ultimate goal of maximizing both biological effect and pharmacokinetic properties.

One strategy that can be envisaged to rise the chances to find effective compounds is represented by the use of big non-enumerated libraries. Nowadays, we are assisting to the growing of huge chemical libraries with an average number of compounds from  $10^{10}$  to  $10^{20}$  (Nicolaou et al., 2016). Despite they enable access to an impressive large chemical space, a common bottleneck is still represented here by the *in silico* tools since a complete structure-based screening of such huge libraries would require unaffordable computational costs and time. Besides, such huge chemical libraries have a number of compounds with properties far from being optimal to be considered as hits. To overcome these limitations, machine learning models from the implementation of Bayesian optimization algorithms for docking-based virtual screening would significantly reduce the computing time making the screening of large chemical library possible (Graff et al., 2021).

Another more feasible possibility could be represented by the use of focused chemical libraries. These kinds of libraries are generally small, drug-like collections and come from a focused enumeration of compounds acting on specific targets as kinases (Kéri et al., 2005), protein-protein interactions (PPI) (Sperandio et al., 2010), G-coupled receptors (Jimonet and Jäger, 2004) among others. The use of such small libraries indeed would allow to shorten the computing time. Best candidates from preliminary screening can be then used to setup *ad hoc* optimization strategies aimed at improving activity toward a specific protein target of interest (Balakin et al., 2006; Mayr and Bojanic, 2009). The advantage of using quality-focused libraries resides in the fact that properties for compounds are already partially optimized. Moreover, a clear linkage between structure and biological activity is also guaranteed.

Our laboratory has developed an in-house quality-focused chemical library named Medicinal and Biological Chemistry (MBC) library, that condensates more than 30 years of medicinal chemistry research in our group. It contains compounds with a standard chemical purity of at least 95% by HPLC and is available both electronically and physically upon request. Since its first publication in 2017 (Sebastián-Pérez et al., 2017), the MBC library has grown significantly reaching a total of 2,577 curated compounds with annotated data about activity and purity. Compounds of the first version of the MBC library (MBC v.2016) have been designed mainly as potential drugs for neurological and neurodegenerative diseases but can be also used as a useful reservoir for the treatment of other diseases. The actual version (MBC v.2022) has been enriched by novel chemical series that have been developed for different targets as those responsible for neglected or infectious diseases, among others. The most representative chemical families of the new MBC v.2022 library are reported in Figure 1.

To validate the quality of the updated MBC library, up to 44 physicochemical and pharmaceutical properties have been calculated for all the compounds with particular attention to drug-likeness properties. The original version of the library (MBC v.2016) has been compared with the new one (MBC v.2022) to quantify the level of improvement of the new version. To exclude structural redundancy, special attention has been deserved to the analysis of the structure (i.e., Tanimoto similarity) and substructure (i.e., Bemis-Murcko algorithm) variability. Finally, to provide a wider perspective, the MBC v.2022 library has been also compared with others well-known chemical libraries such as ZINC20 (Irwin et al., 2020), DrugBank (Wishart et al., 2018), ChEMBL (Mendez et al., 2019), NuBBE (Saldívar-González et al., 2019) and the Approved drug library from Selleck Chemicals together with the European chemical biology library (ECBL) (Horvath et al., 2014). This last library was assembled by the European research infrastructure consortium (ERIC) for chemical biology named EU-OPENSOURCE (EU-OS) (Frank, 2014; Brennecke et al., 2019). This ERIC integrates high-capacity screening platforms throughout Europe with the ECBL (Horvath et al., 2014) and medicinal chemistry expert laboratories making available new hit discoveries for a selected target and the *hit-to-lead* optimization.

## 2 Materials and methods

### 2.1 Database's collection

The MBC v.2016 (1,096 compounds) (Sebastián-Pérez et al., 2017), MBC v.2022 (2,577 compounds), ECBL (101,021 compounds; <https://www.eu-openscreen.eu/services/database.html>), ZINC20 (Irwin et al., 2020) (10,723,360 compounds; <https://files.docking.org/zinc20-ML/>), DrugBank v.5.0 (Wishart et al., 2018) (10,981 compounds; <https://go.drugbank.com/releases/latest>), ChEMBL v.31 (Mendez et al., 2019) (1,908,325 compounds; <https://chembl.gitbook.io/chembl-interface-documentation/downloads>), NuBBE (Saldívar-González et al., 2019) (2,223 compounds; <https://nubbe.iq.unesp.br/portal/nubbe-search.html>) and Approved drug library (3,104 compounds; <https://www.selleckchem.com/screening/fda-approved-drug-library.html>) databases downloaded from their websites in September 2022 were considered for comparison. For comparative purposes, focused subsets of the freely available databases ZINC20, DrugBank, ChEMBL were considered. Briefly, the in-stock drug-like subset was used in case of the ZINC20 database. Regarding the ChEMBL database, only small molecules was selected, discarding other entries (as antibodies or enzymes) out of the scope of this study. Finally, for the DrugBank library, biotechnology products were ignored. All the material and data produced for this study along with the python scripts used to reproduce all the graphics are available at <https://doi.org/10.5281/zenodo.8212104>.

### 2.2 Database's preparation

For all the databases, the 3D structures were generated with the LigPrep module of the Schrödinger suite (Schrödinger, 2022; LigPrep, Schrödinger, LLC, New York, NY, 2022) in accordance with our previous study (Sebastián-Pérez et al., 2017). In brief,



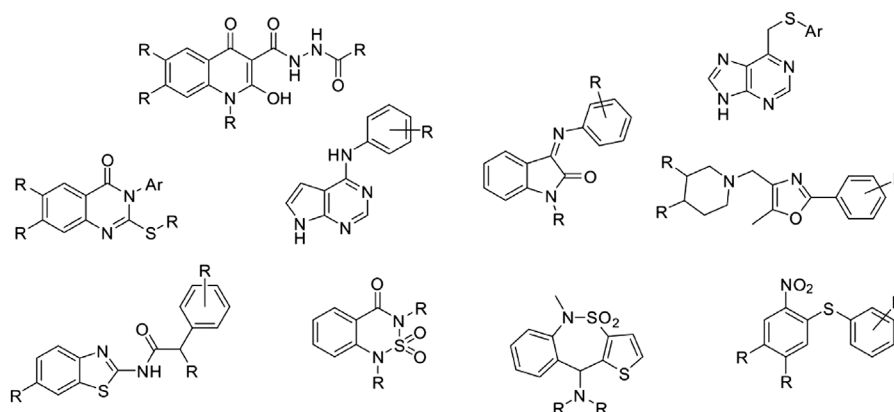


FIGURE 1

Chemical scaffolds for the first 10 most populated clusters of the MBC v.2022 library.

molecules were protonated according to the protonation state at physiological pH. All counterions were removed and no tautomers were generated. Finally, stereochemistry was retained according to the original entries. In-house python scripts along with the pandas-1.4.4, matplotlib-3.6.0 and seaborn-0.12.0 modules of python3 were used to produce all the graphics and statistics reported in this study.

## 2.3 Properties' calculation

Pharmaceutically relevant principal descriptors for all the compounds (see Table 1) were calculated using the QikProp v.6.8 (Schrödinger Release, 2022: QikProp, Schrödinger, LLC, New York, NY, 2022). As stated in the manual, QikProp is unable to calculate properties for not neutralizable quaternary ammonium compounds, so we were forced to exclude these compounds from all the analyzed databases. For this reason, 7.39%, 3.14%, 0.20%, 4.13%, 0.61%, 2.97%, 4.62%, and 0.14% of the prepared compounds respectively from MBC v.2016, MBC v.2022, ECBL, DrugBank, ZINC, ChEMBL, approved drug library and NuBBE were excluded. The probability of a false readout in a screening assay was determined by HitDexter3.0 (Stork et al., 2020; Stork et al., 2021). Similarly, 3.61% and 0.15% of MBC v.2022 and ECBL were not able to be processed. With regard to Veber and Ghose filters, both were calculated with RDKit (Landrum, 2016; Bento et al., 2020). The corresponding measurements and thresholds can be found elsewhere (Ghose et al., 1999; Veber et al., 2002).

## 2.4 Structure similarity analysis

A wide chemical space as a result of a large chemical diversity in chemical libraries is of utmost importance in rising the chances of finding effective and thus promising hits in drug discovery (Gerry and Schreiber, 2018). In this scenario, the Tanimoto coefficient has been routinely used to evaluate chemical similarity or variability (Bajusz et al., 2015). The Tanimoto coefficient ( $T_c$ ) between two points,  $a$  and  $b$ , with  $k$  dimensions is calculated according to Eq. 1

$$T_c = \frac{\sum_{j=1}^k a_j \times b_j}{\left( \sum_{j=1}^k a_j^2 + \sum_{j=1}^k b_j^2 - \sum_{j=1}^k a_j \times b_j \right)} \quad (1)$$

The pairwise comparison of fingerprints—one for the query and one for the target structure - allows to obtain the global similarity between two molecules ( $T_c$ ) which can vary between 0.0 (no similarity) and 1.0 (maximum similarity or identity).

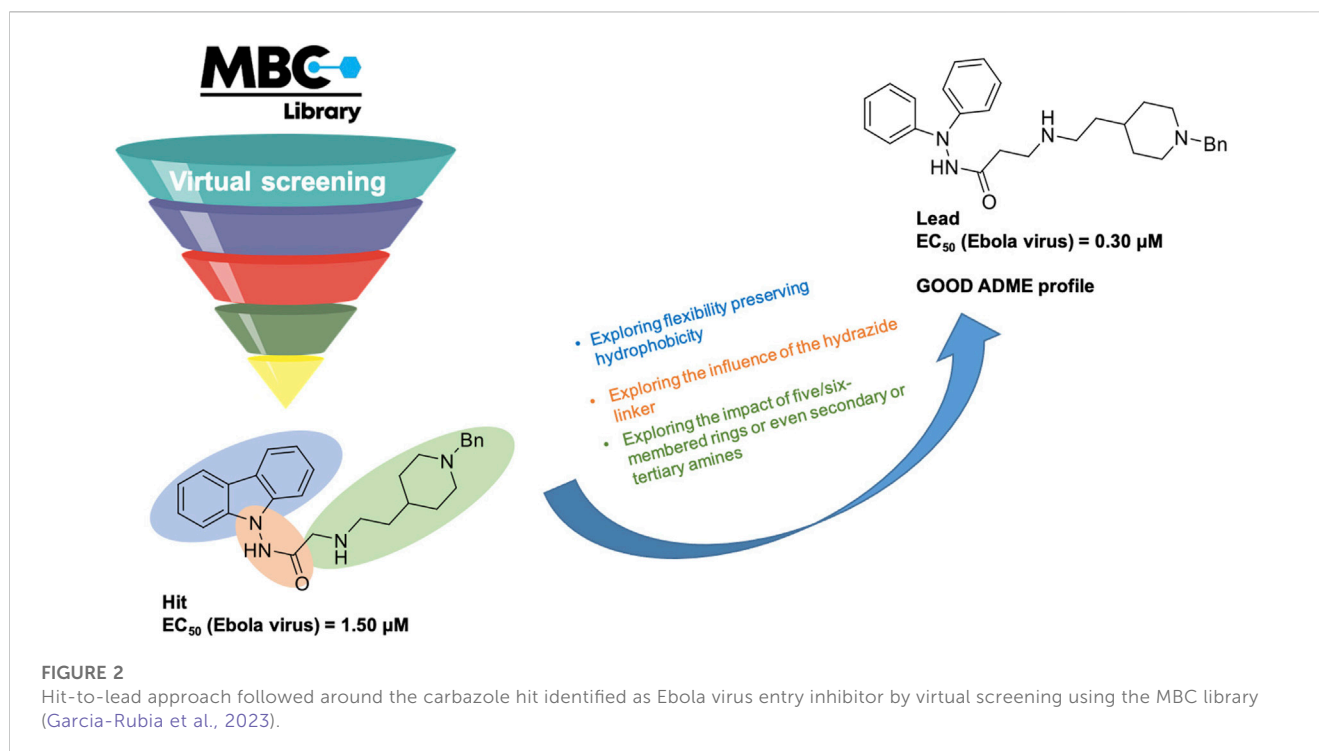
Tanimoto similarity matrixes for the MBC v.2016, MBC v.2022 and ECBL libraries were generated with RDKit (Landrum, 2016; Bento et al., 2020). Accordingly, the SMILES codes for each molecule of the previously cited datasets were first converted in RDKit molecules and molecular fingerprints were thus calculated. Comparison of the so generated RDKit fingerprints allowed to generate a NxM matrix whose dimensions depends on the length of the analyzed database. Accordingly,  $1,096 \times 1,096$ ,  $2,577 \times 2,577$  and  $101,021 \times 101,021$  Tanimoto matrixes were generated respectively for the MBC v.2016, MBC v.2022 and ECBL libraries and plotted (Figure 4). SMILES codes for compounds bearing a quaternary ammonium - for which QikProp was unable to calculate properties - were retained for this analysis.

## 2.5 Substructure similarity analysis

A quite common scaffold representation is the Murcko framework proposed by Bemis and Murcko (Hu et al., 2016). Given a query molecule, the method employs a systematic dissection into four parts: ring systems, linkers, side chains, and the Murcko framework that is the union of ring systems and linkers in a molecule. The information obtained by this analysis can be used for different purpose as, for instance, database enumeration. In this work, the Bemis-Murcko scaffolds have been calculated for each input RDKit molecule by using the ChemAxon Bemis-Murcko node of the KNIME platform (<https://www.knime.com/knime-analytics-platform>). The resulted scaffolds were finally clustered according to their canonical SMILES codes.

**TABLE 1** Quantitative distributions for the most relevant pharmacokinetic properties of the MBC and ECBL libraries calculated with QikProp.

Property	Intervals	MBC v.2016	MBC v.2022	ECBL
Lipinski's rule of 5	0 violations	85.1%	85.3%	98.7%
	1 violation	13.4%	12.0%	1.0%
	2 violations	1.5%	2.3%	0.2%
	3 violations	0.0%	0.4%	0.1%
	4 violations	0.0%	0.0%	<0.1%
Jorgensen's rule of 3	0 violations	76.0%	76.1%	92.4%
	1 violation	23.5%	23.1%	7.3%
	2 violations	0.5%	0.8%	0.2%
	3 violations	0.0%	0.0%	<0.1%
Veber filter	Meet the criteria	96.4%	95.1%	95.4%
Ghose filter	Meet the criteria	85.8%	85.8%	94.9%
MW (Da)	0–200	9.6%	5.9%	0.2%
	201–300	36.5%	34.4%	21.0%
	301–400	39.2%	42.0%	64.0%
	401–500	12.9%	14.4%	14.1%
	>500	1.8%	3.3%	0.7%
Nr. of rotatable bonds	0–5	82.1%	79.5%	66.9%
	6–10	7.1%	17.1%	27.0%
	>10	0.8%	3.4%	6.1%
donorHB (HBD)	≤5	99.9%	99.8%	99.9%
	>5	0.1%	0.2%	0.1%
accptHB (HBA)	≤10	99.4%	98.0%	94.0%
	>10	0.6%	2.0%	6.0%
QPlogP <sub>o/w</sub>	≤5	91.0%	87.5%	99.5%
	>5	9.0%	12.5%	0.5%
QPlogS	–12.0/–7.0	11.6%	6.7%	0.4%
	–6.9/–3.0	73.3%	82.9%	67.0%
	–2.9/2.0	15.1%	10.4%	32.6%
QPlogBB	–9.0/–5.0	0.0%	0.0%	<0.1%
	–4.0/–1.0	61.2%	62.0%	20.2%
	–0.9/2.0	38.8%	38.0%	79.8%
Human oral absorption in GI	0%–50%	1.9%	2.1%	0.7%
	51%–75%	9.6%	8.5%	12.5%
	76%–100%	88.5%	89.4%	86.8%
Probability of highly promiscuous activities in target-based assays	0.00–0.50	-	93.2%	99.3%
	0.51–0.75	-	0.7%	0.1%
	0.76–1.00	-	2.6%	0.4%



### 3 Results and discussion

Successful screening projects begin with the selection of appropriate chemical libraries in terms of size, quality, and chemical diversity. Unlike ultra-large libraries, which are computationally expensive to use, quality-focused chemical libraries represent a useful source of chemical entities enriched with active chemotypes that may be designed to efficiently combine chemical diversity along with a significant reduction of the computational resources eventually required for a screening campaign. Moreover, these libraries may be built according to drug likeness properties (e.g., ADME/Tox properties), offering promising starting-points that can accelerate hits finding and *hit-to-lead* protocols (Gong et al., 2017).

#### 3.1 The Medicinal and Biological Chemistry (MBC) library from 2016 to 2022

The MBC library has originated from more than 30 years of experience in drug discovery of our research group. It has been conceived as a collection of focused sets of chemical probes with common therapeutic profiles mostly in the field of neurodegenerative and infectious diseases as such as Alzheimer's and Parkinson's diseases, amyotrophic lateral sclerosis (ALS), schistosomiasis, and leishmaniasis, among others. It represents a source of fully accessible, ready-to-use compounds with proved efficacy. The library has been growing from 1,096 compounds in 2016 to 2,577 compounds in 2022 with a significant exploitation in the field of infectious diseases. The utility of the MBC library to initiate drug discovery programs is reflected mainly in the neurodegenerative and anti-infective fields. Particularly, successful families of CK1 inhibitors with a benzothiazole core (Salado et al.,

2014; Martínez-González et al., 2020) and CDC7 inhibitors with a 6-mercaptapurine scaffold (Rojas-Prats et al., 2021) useful for ALS were developed till the *in vivo* proof of concept after initial hit identification using the MBC library as reported. Very recently new mitophagy modulators having chemically diverse scaffolds were also discovered (Maestro et al., 2023). In the anti-infective field it is remarkable the discovery of *N'*-phenylacetohydrazide derivatives as potent Ebola virus entry inhibitors (García-Rubia et al., 2023) starting with a carbazole hit identified from the MBC library (Lasala et al., 2021) (Figure 2). The increased potentialities of the new version of our MBC library have been addressed here and compared with the previous version (Sebastián-Pérez et al., 2017).

#### 3.2 The European Chemical Biology Library (ECBL)

The selected EU-OPENSREEN (EU-OS) compound collection is centrally stored and managed at the EU-OS laboratory facility on the Research-Campus Berlin-Buch (Germany). All compound structures and primary screening data will be published in the open-access European Chemical Biology Database, where they are made available to a wide scientific audience. The European research infrastructure EU-OS collaboratively develops novel molecular tool compounds and early therapeutic candidate molecules together with external users from various disciplines of the life sciences. Access to the EU-OS resources is open to researchers from academia and industry from countries inside and outside of the European Union. The current version (v.2022) of the ECBL integrates 101,021 available, ready-to-use compounds with unbiased chemical diversity, designed by five renowned academic computational chemistry groups. To maximize the coverage of chemical space, criteria followed by these groups in

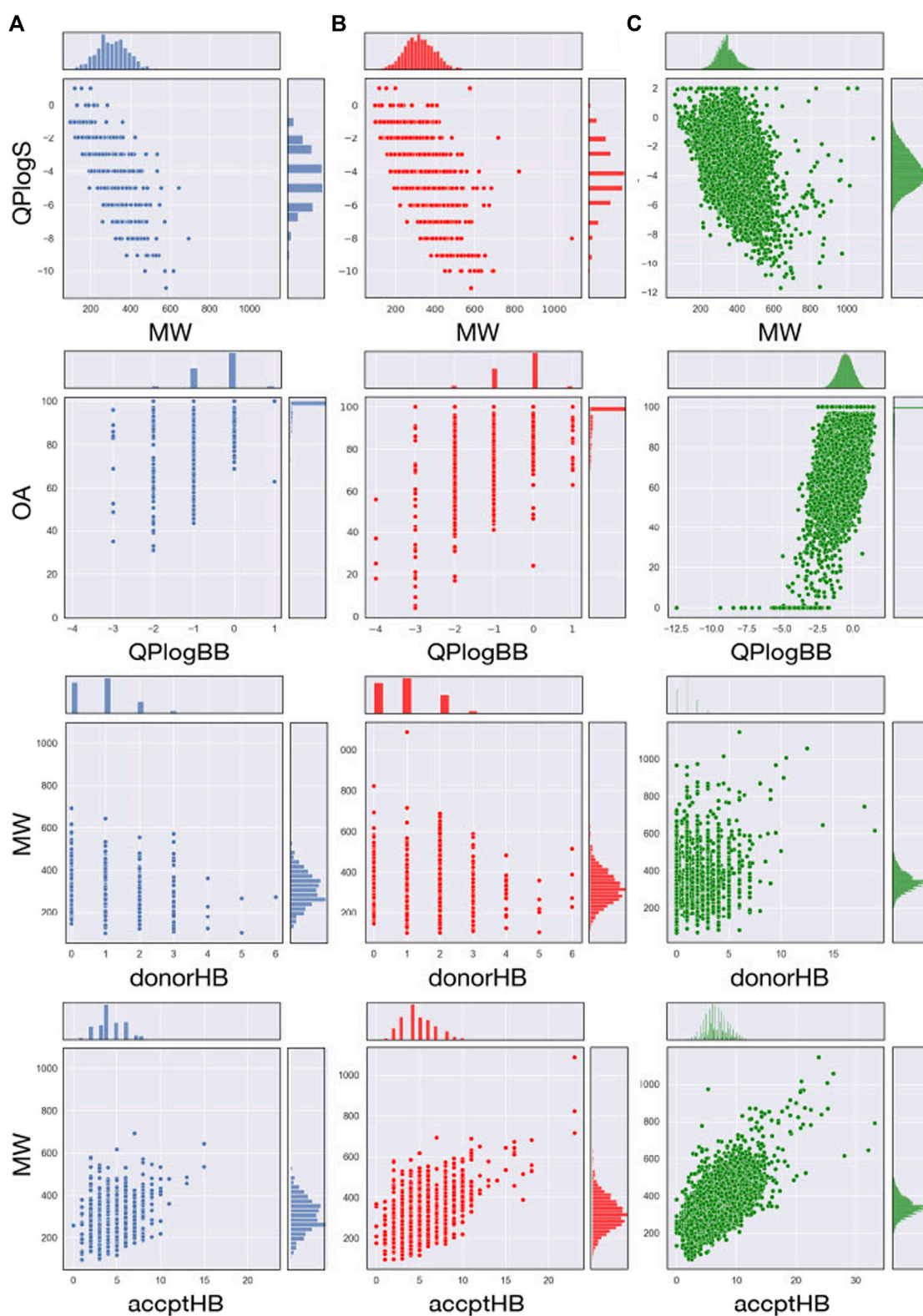


FIGURE 3

Joint scatter and distribution plots of the main principal descriptors and properties calculated with QikProp for the MBC v.2016 (A; in blue), MBC v.2022 (B; in red) and for ECBL (C; in green) libraries.

the molecules selection were completely different but chemical stability, drug-likeness criteria, and practical availability were pursued in all the cases (Horvath et al., 2014). Recently, this library have started to be used providing valuable hits to fight against COVID-19 (Kuzikov et al., 2021; Schuller et al., 2021).

### 3.3 Comparative analysis

The most relevant physicochemical and pharmaceutical properties for each of the compound of the MBC library have been predicted with the QikProp module of Schrödinger. Quantitative distributions for the most relevant pharmacokinetic properties of the two versions of the MBCs library are reported in Table 1. For a comparative description, relative dispersion and distribution for some salient pharmacokinetics (PK) properties as molecular weight (MW), predicted octanol/water logP (QP logP), predicted logS (QP logS), hydrogen-bond donor (HBD), hydrogen-bond acceptor (HBA), predicted blood-brain barrier (BBB) permeability (QP logBB) and human oral absorption (OA) are showed in Figure 3.

A traditional method to evaluate drug likeness is represented by the Lipinski's Rule of Five (Lipinski et al., 2001; Lipinski, 2004). In line with the previous version of the library, the 97.3% (85.3% with 0 violations and 12.0% with 1 violation) of the compounds in the MBC v.2022 have less than 2 violations of the Lipinski's Rule of Five. Less than 3% have more than 2 violations. The Jorgensen's Rule of Three (Lionta et al., 2014) is another widely followed rule for lead like properties and states that the aqueous solubility measured as logS should be greater than -5.7, the apparent Caco-2 cell permeability should be faster than 22 nm/s, and the number of primary metabolites should be less than 7. In both versions of the MBC library, the majority of the compounds (76.0%) have no violations with the 23% showing only 1 violation.

Molecular flexibility, number of hydrogen-bonding donor/acceptor groups and molecular weight are critical parameters for drug likeness. In the MBC library, most of compounds (90.8% of the MBC v.2022 and 88.6 of the MBC v.2016 library) have a MW between the recommended interval of 200–500 Da (see Table 1; Figure 3). More than 90% of the MBC compounds have up to 10 rotatable bonds. More than 98% have less than 5 hydrogen-bond donor and less than 10 hydrogen-bond acceptor groups (see Figure 3).

Lipophilicity (Ginex et al., 2019) and thus solubility (Bergström and Larsson, 2018) have a great impact on the pharmacokinetic profile of a potential drug. Most of the compounds of our MBC library have suitable lipophilicity and solubility predicted values (more than 80% with QP logP<sub>o/w</sub> below 5 and QP logS between -6.5 and 0.5; see Table 1; Figure 3). This is also reflected in a good BBB predicted permeability with about 60% of compounds with QPlogBB values between -1 and -4 (see Table 1). Here, a close look at the distributions for the QPlogBB values reported in Figure 3 allows to see that most of the compounds specifically fall within 0 and -1. Finally, more than 95% of the compounds have a predicted oral absorption (hOA) rate in the gastrointestinal (GI) tract higher than 50% (see Table 1; Figure 3). Moreover, potentially promiscuous compounds should be carefully treated and analyzed in order to avoid false-positive results. There is

a wide range of strategies to afford this, from classical substructure detection [e.g., Pan Assay Interference Compounds (PAINS) filter (Baell and Holloway, 2010)] to more refined machine learning methodologies (Blaschke et al., 2019). In this sense, the probability of triggering a positive result in a target-based screening, understood as a false positive due to the chemical promiscuity of the molecule, was calculated here using HitDexter 3.0 server (<https://nerdd.univie.ac.at/hitdexter3>), a machine learning approach that shows how the vast majority of the MBC library (93.2%) avoid this alert.

The analysis of the most relevant physicochemical and pharmacokinetic properties for ECBL has been reported in Table 1 and Figure 3C. In brief, the 98.7% and the 92.4% of the compounds have no violations of respectively the Lipinski's Rule of Five and Jorgensen Rule of Three which is globally indicative of the high pharmaceutical relevance of the dataset. As demonstrated by the data in Table 1 and plots in Figure 3C, this dataset guarantees an excellent coverage of the drug-like chemical space with MW lower than 600 Da, a number of rotatable bonds lower than 10, less than 5 hydrogen-bond donor and less than 10 hydrogen-bond acceptor atoms. QPlogBB values fall in the interval of 2 and -4 (see Table 1), with the majority of the compounds within 1 and -2.5 (see Figure 3). The good characterization, data curation and immediate availability of the compounds of the ECBL make it also a good reservoir of potential hits. Finally, the HitDexter program shows how the 99.3% of this library present low probability of trigger a false-positive readout in target-based assays.

### 3.4 Tanimoto similarity

Beside physicochemical and PK properties, a wide chemical variability or diversity is also a pivotal feature since it could influence the success rate of a screening protocol (López-Vallejo et al., 2012). The use of small-sized libraries with low chemical variability and high structural redundancy could in fact reduce the possibility to find useful hits.

In this regard, the Tanimoto metric has been widely used to evaluate molecule similarity thus it represents a valid way to measure the qualitative chemical variability of a compound's library (Sankara Rao et al., 2011; Bajusz et al., 2015; Xia and Yan, 2017). The chemical variability of our MBC library has been subjected to Tanimoto-based fingerprint similarity analysis (see Figure 4). With the exception of a small cluster of *structurally-related*, relatively similar compounds with values among 0.5 and 0.7 (white/pink square in the similarity matrix for MBC v.2016 and v.2022), a clear predominance of fingerprint values lower than 0.5 is generally observed thus confirming the suitable chemical diversity of our library. Regarding the ECBL, no similarity clusters were found as shown by the low values of the Tanimoto coefficients. This once again highlights the valuable chemical diversity of this library.

### 3.5 Scaffold clustering

Bemis and Murcko outlined a popular method for deriving scaffolds from molecules by removing side chain atoms (Bemis and Murcko, 1996). Widely speaking, the Bemis-Murcko framework



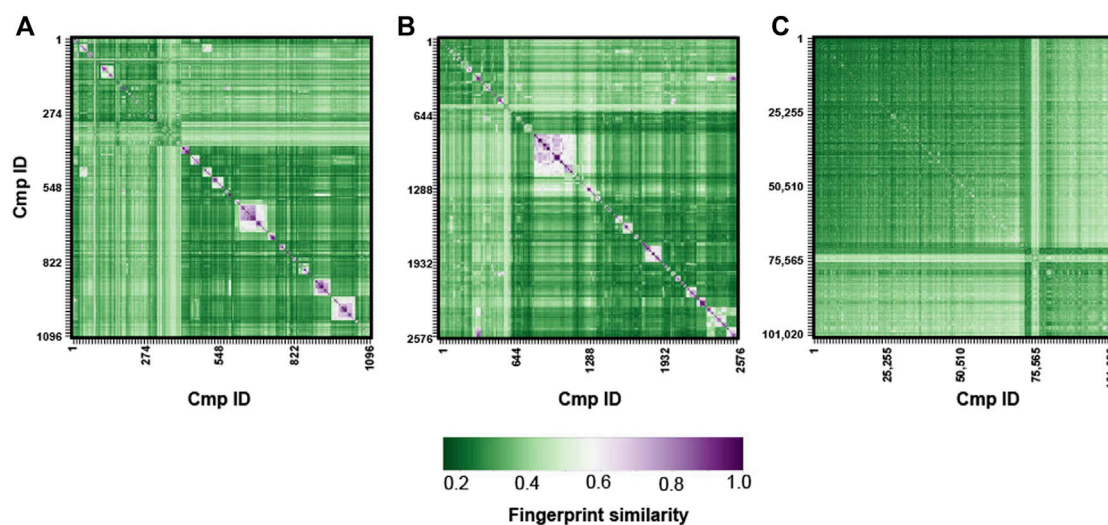


FIGURE 4

Tanimoto-based fingerprint similarity analysis for the MBC v.2016 (A), MBC v.2022 (B) and ECBL (C) libraries.

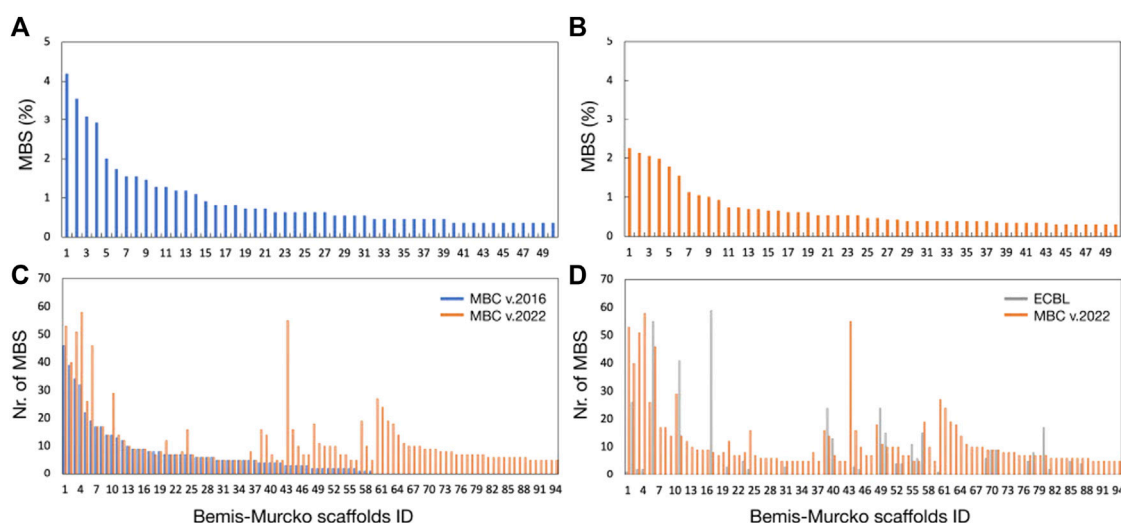


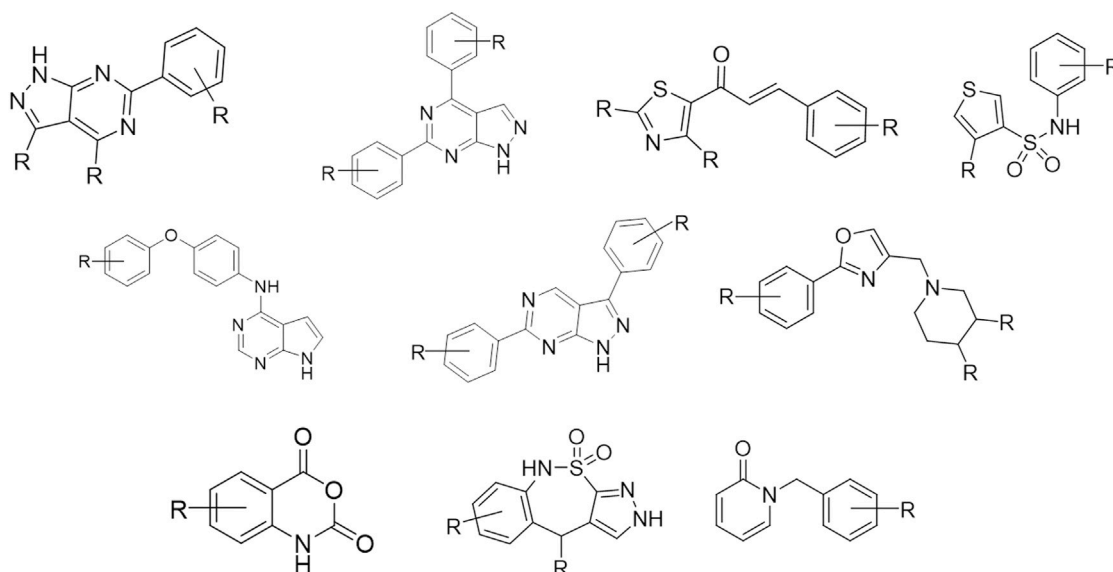
FIGURE 5

Bemis-Murcko scaffold distribution (%) for the MBC v.2016 (A) and the MBC v.2022 (B). Analysis of the common scaffolds between MBC v.2022 and MBC v.2016 or ECBL is reported respectively in (C, D). For clarity, only the scaffolds with a population >5 compounds have been shown.

algorithm represents an effective indicator of the chemical diversity of the chemical libraries. The ChemAxon extended version of the Bemis-Murcko framework algorithm implemented in KNIME (Hu et al., 2016) has been used here to perform scaffold clustering to check the chemical diversity of the MBC v.2022 and the ECBL. More details about the procedure can be found in the *Materials & Methods* section. For comparative purposes, the old version of the MBC library (v.2016) has been also included in the analysis (see Figure 5).

A total of 465 and 1123 Bemis-Murcko scaffolds have been found respectively for the MBC v.2016 and v.2022 libraries thus confirming the enrichment in chemical variability of the new version. The most representative new scaffolds from MBC

v.2022 with respect to the v.2016 are depicted in Figure 6. As observed in Figures 5A,B, a high level of chemical diversity generally characterizes the MBC v.2022 database with only 2 Bemis-Murcko scaffolds having a population above the 2% of the structures present in the MBC v.2022 database, showing that the vast majority of the compounds are distributed over different chemotypes (Figure 5B). Library expansion could be due to two possible factors as 1) the enrichment of already present scaffolds by means of further enumeration or 2) the introduction of totally new chemical entities. In case of the MBC library, analysis of the common scaffolds (see Figure 5C) between the two versions allowed to see that the library expansion generally came from the introduction of new chemical species with a limited enumeration of



**FIGURE 6**  
Most representative new scaffolds from the MBC v.2022 with respect to the MBC v.2016.

the scaffolds already present in the previous version (see Bemis-Murcko scaffolds IDs 1-59 in Figure 5C). Finally, the uniqueness of the chemical scaffolds collected in the MBC v.2022 library has been evaluated with respect to the ECBL. On a total of 94 scaffolds with population higher than 5 compounds, there are 62 unique structures in the MBC v.2022 library (see Figure 5D) and only 32 are shared with the ECBL.

### 3.6 Comparison with other libraries

The physicochemical and drug-like properties of the MBC library have been also compared with those of some publicly available chemical databases as ZINC, DrugBank, ChEMBL, NuBBE and approved drug library from Selleck Chemicals (henceforth referred to as Approved drug library). ZINC (Irwin et al., 2020) is a freely available database of commercially available compounds developed by the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF). One of the most recent versions, ZINC20 (Irwin et al., 2020), contains over 10 millions of drug-like compounds. The DrugBank (Wishart et al., 2018) is a freely available database that includes drug information, drug targets, 3D structure and metabolic pathways. The database contains about 11,000 small compounds. ChEMBL (Mendez et al., 2019) is an open-source database developed by the European Bioinformatics Institute (EMBL-EBI) in Cambridge (United Kingdom). It also contains structures from the U.S. Food and Drug Administration (FDA). Information about approved products (from the FDA Orange Book), including dosage information and administration routes, is also included in the database. Currently, the database contains about 1.9 million drug-like compounds. The NuBBE database (Saldivar-González et al., 2019) is a natural product library created in 2013 that aims to collect the chemical structural diversity of Brazilian natural

biodiversity, resulting in an extraordinary curated source of 2,223 natural compounds. Finally, the approved drug library used in this work is a collection of compounds downloaded from Selleck Chemicals that are ready to be used for HTS. The 3,104 compounds in this library are approved by different regulatory agencies such as the FDA or the European Medicines Agency (EMA), among others.

Density distributions relative to the molecular weight, SASA, QP  $\log P_{o/w}$ , QP  $\log S$ , donorHB (HBD) and acceptHB (HBA) properties are reported in Figure 7. As shown, most of the compounds fall within the range of Lipinski's rule of Five (that is, less than 500 Da) for MW with the MBC and ECBL having a slightly better fit among the analyzed databases. The solvent accessible surface area (SASA) for ECBL, ChEMBL and ZINC ranges between 400 and 900 Å<sup>2</sup>. A slightly shifter profile can be seen for MBC, DrugBank and the Approved drugs library, with SASA values from 200 to 800 Å<sup>2</sup>. Regarding NuBBE, the distribution seems to be an intermediate case compared to the previous ones, covering wider values from 200 to 1000 Å<sup>2</sup> with a maximum population density close to that of the ECBL or ChEMBL. In the case of hydrogen-bond donor (donorHB) and acceptor (acceptHB) properties, all the libraries apart from DrugBank, Approved drug library and ChEMBL for HBD, the vast majority agree with Lipinski's rule of five and are in the range of 0.0–6.0 for HBD and 2.0–20.0 for HBA. This is likely due to the presence of small peptidomimetics and complex sugars in the previously cited libraries. Regarding lipophilicity (QP  $\log P_{o/w}$ ), similar distributions in the range of –2.5 to 7.5 have been registered for all the libraries except for DrugBank and Approved drug library that have also some compounds with QP  $\log P_{o/w}$  values below –2.5. Finally, compounds from MBC, ECBL, and ZINC are in the optimal range of solubility (–6.5 < QP  $\log S$  < 0.5). NuBBE and ChEMBL show a similar distribution with a small set of compounds with QP  $\log S$  values lower than –6.5. However, the remaining compounds show an appropriate solubility profile. DrugBank and Approved drug library slightly deviate from the ideal range having a small number of compounds with QP  $\log S$  values than 0.5.

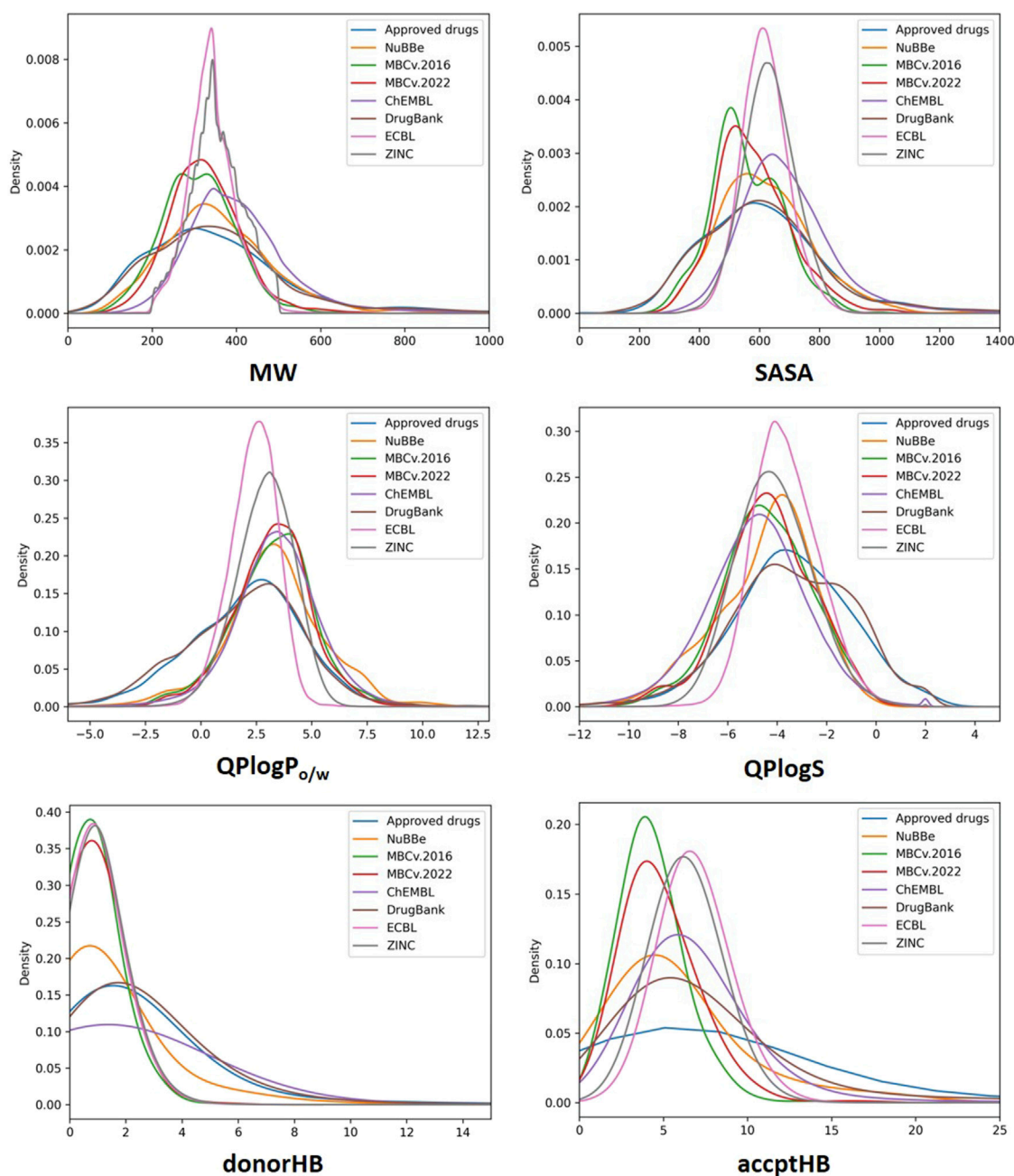


FIGURE 7

Probability density distribution of MW, SASA, QP  $\log P_{o/w}$ , QP  $\log S$ , donorHB and accptHB properties for all the analyzed libraries.

## 4 Conclusion

The screening of quality-focused libraries could represent a way to efficiently provide a useful source of probes that help characterize the protein targets emerging from research studies. In this direction, efficient synthetic routes, availability, good characterization, suitable physicochemical and pharmacokinetic properties can really make the difference since they can contribute to rise the success rate and shorten the drug discovery process.

We here presented an updated version of our *in-house* MBC library which is a unique collection of small molecules with enriched

drug-like properties and chemical diversity. From the first publication in 2016, the library has been constantly enriched with new compounds becoming 2.3 times bigger than the previous version with over 2,500 *ready-to-use* chemical compounds. To test its potential impact on drug discovery, the quality and variability of the chemical structures collected in the new version of the MBC library has been analyzed by using the QikProp module of Schrödinger and RDKit. As official partners and active collaborators of the EU-OPENSOURCE ERIC, we also presented and discussed the potentialities of the open-access European Chemical Biology Library (ECBL) that collect data from several

countries in EU. Finally, a wider comparison with other well-known publicly available libraries has been provided and discussed. Results of this analysis remark the high quality in terms of structural diversity and drug-like properties of the MBC and ECBL, making them suitable reservoirs of hits for drug discovery.

## Data availability statement

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found in the article/Supplementary Material.

## Author contributions

Conceptualization, CG and AM; methodology, TG; formal analysis, TG and EM; investigation, TG, EM, AM, and CG; resources, CG; writing—original draft preparation, TG and EM; writing—review and editing, AM and CG; funding acquisition, AM and CG. All authors contributed to the article and approved the submitted version.

## Funding

This research was funded by MCIN/AEI/10.13039/501100011033 (Grant No PID 2019-105600RB-I00 to AM and PID 2021-122223OB-I00 to CG) and ERDF “A way of making

Europe” by the European Union, Education Spanish Ministry (Grant FPU20/03743 to EM) and the European Commission Horizon 2020 Framework Programme (Project EU-OPENSREEN-DRIVE—GA 823893).

## 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/fphar.2023.1244317/full#supplementary-material>

## References

- Baell, J. B., and Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 53 (7), 2719–2740. doi:10.1021/jm901137j
- Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov.* 1 (11), 882–894. doi:10.1038/nrd941
- Bajusz, D., Rácz, A., and Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform* 7, 20. doi:10.1186/s13321-015-0069-3
- Balakin, K. V., Tkachenko, S. E., Kiselyov, A. S., and Savchuk, N. P. (2006). Focused chemistry from annotated libraries. *Drug Discov. Today Technol.* 3 (4), 397–403. doi:10.1016/j.ddtec.2006.12.006
- Bemis, G. W., and Murcko, M. A. (1996). The properties of known drugs. 1. molecular frameworks. *J. Med. Chem.* 39 (15), 2887–2893. doi:10.1021/jm9602928
- Bento, A. P., Hersey, A., Félix, E., Landrum, G., Gaulton, A., Atkinson, F., et al. (2020). An open source chemical structure curation pipeline using RDKit. *J. Cheminform* 12 (1), 51. doi:10.1186/s13321-020-00456-1
- Bergström, C. A. S., and Larsson, P. (2018). Computational prediction of drug solubility in water-based systems: qualitative and quantitative approaches used in the current drug discovery and development setting. *Int. J. Pharm.* 540 (1–2), 185–193. doi:10.1016/j.ijpharm.2018.01.044
- Blaschke, T., Miljković, F., and Bajorath, J. (2019). Prediction of different classes of promiscuous and nonpromiscuous compounds using machine learning and nearest neighbor analysis. *ACS Omega* 4 (4), 6883–6890. doi:10.1021/acsomega.9b00492
- Brennecke, P., Rasina, D., Aubi, O., Herzog, K., Landskron, J., Cautain, B., et al. (2019). EU-OPENSREEN: a novel collaborative approach to facilitate chemical biology. *SLAS Discov.* 24 (3), 398–413. doi:10.1177/2472555218816276
- Frank, R. (2014). EU-OPENSREEN—a European infrastructure of open screening platforms for chemical biology. *ACS Chem. Biol.* 9 (4), 853–854. doi:10.1021/cb500189k
- García-Rubia, A., Lasala, F., Ginex, T., Morales-Tenorio, M., Olal, C., Heung, M., et al. (2023). N'-Phenylacetohydrazide derivatives as potent Ebola virus entry inhibitors with an improved pharmacokinetic profile. *J. Med. Chem.* 66 (8), 5465–5483. doi:10.1021/acs.jmedchem.2c01785
- Gerry, C. J., and Schreiber, S. L. (2018). Chemical probes and drug leads from advances in synthetic planning and methodology. *Nat. Rev. Drug Discov.* 17 (5), 333–352. doi:10.1038/nrd.2018.53
- Ghose, A. K., Viswanadhan, V. N., and Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. a qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* 1 (1), 55–68. doi:10.1021/cc9800071
- Ginex, T., Vazquez, J., Gilbert, E., Herrero, E., and Luque, F. J. (2019). Lipophilicity in drug design: an overview of lipophilicity descriptors in 3D-QSAR studies. *Future Med. Chem.* 11 (10), 1177–1193. doi:10.4155/fmc-2018-0435
- Gong, Z., Hu, G., Li, Q., Liu, Z., Wang, F., Zhang, X., et al. (2017). Compound libraries: recent advances and their applications in drug discovery. *Curr. Drug Discov. Technol.* 14 (4), 216–228. doi:10.2174/1570163814666170425155154
- Graff, D. E., Shakhnovich, E. I., and Coley, C. W. (2021). Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem. Sci.* 12 (22), 7866–7881. doi:10.1039/d0sc06805e
- Horvath, D., Lisurek, M., Rupp, B., Kühne, R., Specker, E., von Kries, J., et al. (2014). Design of a general-purpose European compound screening library for EU-OPENSREEN. *ChemMedChem* 9 (10), 2309–2326. doi:10.1002/cmdc.201402126
- Hu, Y., Stumpfe, D., and Bajorath, J. (2016). Computational exploration of molecular scaffolds in medicinal chemistry. *J. Med. Chem.* 59 (9), 4062–4076. doi:10.1021/acs.jmedchem.5b01746
- Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., et al. (2020). ZINC20-A free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* 60 (12), 6065–6073. doi:10.1021/acs.jcim.0c00675
- Jimonet, P., and Jäger, R. (2004). Strategies for designing GPCR-focused libraries and screening sets. *Curr. Opin. Drug Discov. Devel* 7 (3), 325–333.
- Kéri, G., Székelyhidi, Z., Bánhegyi, P., Varga, Z., Hegymegi-Barakonyi, B., Szántai-Kis, C., et al. (2005). Drug discovery in the kinase inhibitory field using the Nested Chemical Library technology. *Assay. Drug Dev. Technol.* 3 (5), 543–551. doi:10.1089/adt.2005.3.543
- Kim, K. H., Kim, N. D., and Seong, B. L. (2010). Pharmacophore-based virtual screening: a review of recent applications. *Expert Opin. Drug Discov.* 5 (3), 205–222. doi:10.1517/17460441003592072



- Kuzikov, M., Costanzi, E., Reinshagen, J., Esposito, F., Vangeel, L., Wolf, M., et al. (2021). Identification of inhibitors of SARS-CoV-2 3CL-Pro enzymatic activity using a small molecule *in vitro* repurposing screen. *ACS Pharmacol. Transl. Sci.* 4 (3), 1096–1110. doi:10.1021/acspstsci.0c00216
- Landrum, G. (2016). *RDKit: open-source cheminformatics software*. Available at: <https://www.rdkit.org/>
- Lasala, F., García-Rubia, A., Requena, C., Galindo, I., Cuesta-Geijo, M. A., García-Dorival, I., et al. (2021). Identification of potential inhibitors of protein-protein interaction useful to fight against Ebola and other highly pathogenic viruses. *Antivir. Res.* 186, 105011. doi:10.1016/j.antiviral.2021.105011
- Lionta, E., Spyrou, G., Vassilatis, D. K., and Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr. Top. Med. Chem.* 14 (16), 1923–1938. doi:10.2174/1568026614666140929124445
- Lipinski, C. A. (2004). Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* 1 (4), 337–341. doi:10.1016/j.ddtec.2004.11.007
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46 (1–3), 3–26. doi:10.1016/s0169-409x(00)00129-0
- López-Vallejo, F., Giulianotti, M. A., Houghten, R. A., and Medina-Franco, J. L. (2012). Expanding the medicinally relevant chemical space with compound libraries. *Drug Discov. Today* 17 (13–14), 718–726. doi:10.1016/j.drudis.2012.04.001
- Maestro, I., Madrugá, E., Boya, P., and Martínez, A. (2023). Identification of a new structural family of SGK1 inhibitors as potential neuroprotective agents. *J. Enzyme Inhib. Med. Chem.* 38 (1), 2153841. doi:10.1080/14756366.2022.2153841
- Martínez-González, L., Rodríguez-Cueto, C., Cabezano, D., Bartolomé, F., Andrés-Benito, P., Ferrer, I., et al. (2020). Motor neuron preservation and decrease of *in vivo* TDP-43 phosphorylation by protein CK-1δ kinase inhibitor treatment. *Sci. Rep.* 10 (1), 4449. doi:10.1038/s41598-020-61265-y
- Mayr, L. M., and Bojanic, D. (2009). Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* 9 (5), 580–588. doi:10.1016/j.coph.2009.08.004
- 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 (D1), D930–D940. doi:10.1093/nar/gky1075
- Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., and Andrade, C. H. (2018). QSAR-based virtual screening: advances and applications in drug discovery. *Front. Pharmacol.* 9, 1275. doi:10.3389/fphar.2018.01275
- Neves, B. J., Mottin, M., Moreira-Filho, J. T., Sousa, B. K. d. P., Mendonça, S. S., and Andrade, C. H. (2021). "Chapter 4 - best practices for docking-based virtual screening," in *Molecular docking for computer-aided drug design*. Editor M. S. Coumar (China: Academic Press), 75–98.
- Nicolaou, C. A., Watson, I. A., Hu, H., and Wang, J. (2016). The proximal Lilly collection: mapping, exploring and exploiting feasible chemical space. *J. Chem. Inf. Model.* 56 (7), 1253–1266. doi:10.1021/acs.jcim.6b00173
- Rojas-Prats, E., Martínez-González, L., Gonzalo-Consuegra, C., Liachko, N. F., Perez, C., Ramirez, D., et al. (2021). Targeting nuclear protein tdp-43 by cell division cycle kinase 7 inhibitors: a new therapeutic approach for amyotrophic lateral sclerosis. *Eur. J. Med. Chem.* 210, 112968. doi:10.1016/j.ejmech.2020.112968
- Salado, I. G., Redondo, M., Bello, M. L., Perez, C., Liachko, N. F., Kraemer, B. C., et al. (2014). Protein kinase CK-1 inhibitors as new potential drugs for amyotrophic lateral sclerosis. *J. Med. Chem.* 57 (6), 2755–2772. doi:10.1021/jm500065f
- Saldívar-González, F. I., Valli, M., Andricopulo, A. D., da Silva Bolzani, V., and Medina-Franco, J. L. (2019). Chemical space and diversity of the NuBBE database: a chemoinformatic characterization. *J. Chem. Inf. Model.* 59 (1), 74–85. doi:10.1021/acs.jcim.8b00619
- Sankara Rao, A., Durga Bhavani, S., Sobha Rani, T., Raju S. B., and Narahari Sastry, G. (2011). "Study of diversity and similarity of large chemical databases using Tanimoto measure", in: *Computer Networks and Intelligent Computing*, eds. K. R. Venugopal and L. M. Patnaik: Germany, Springer Berlin Heidelberg), 40–50.
- Schrödinger Release (2022). *Schrödinger Release 2022-2*. New York, NY: QikProp, Schrödinger, LLC.
- Schrödinger (2022). *Schrödinger Release 2022-2*. New York, NY: LigProp, Schrödinger, LLC.
- Schuller, M., Correy, G. J., Gahbauer, S., Fearon, D., Wu, T., Díaz, R. E., et al. (2021). Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* 7 (16), eabf8711. doi:10.1126/sciadv.abf8711
- Sebastián-Pérez, V., Roca, C., Awale, M., Reymond, J. L., Martínez, A., Gil, C., et al. (2017). Medicinal and Biological Chemistry (MBC) library: an efficient source of new hits. *J. Chem. Inf. Model.* 57 (9), 2143–2151. doi:10.1021/acs.jcim.7b00401
- Sperandio, O., Reynès, C. H., Camproux, A. C., and Villoutreix, B. O. (2010). Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov. Today* 15 (5–6), 220–229. doi:10.1016/j.drudis.2009.11.007
- Stork, C., Embruch, G., Šicho, M., de Bruyn Kops, C., Chen, Y., Svozil, D., et al. (2020). Nerdd: a web portal providing access to *in silico* tools for drug discovery. *Bioinformatics* 36 (4), 1291–1292. doi:10.1093/bioinformatics/btz695
- Stork, C., Mathai, N., and Kirchmair, J. (2021). Computational prediction of frequent hitters in target-based and cell-based assays. *Artif. Intell. Life Sci.* 1, 100007. doi:10.1016/j.aillsci.2021.100007
- Subramaniam, S., Mehrotra, M., and Gupta, D. (2008). Virtual high throughput screening (vHTS)-a perspective. *Bioinformation* 3 (1), 14–17. doi:10.6026/97320630003014
- Tanrikulu, Y., Krüger, B., and Proschak, E. (2013). The holistic integration of virtual screening in drug discovery. *Drug Discov. Today* 18 (7–8), 358–364. doi:10.1016/j.drudis.2013.01.007
- Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45 (12), 2615–2623. doi:10.1021/jm020017n
- 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 (D1), D1074–D1082. doi:10.1093/nar/gkx1037
- Xia, Z., and Yan, A. (2017). Computational models for the classification of mPGES-1 inhibitors with fingerprint descriptors. *Mol. Divers* 21 (3), 661–675. doi:10.1007/s11030-017-9743-x





## OPEN ACCESS

## EDITED BY

Xinhua Qu,  
Shanghai Jiao Tong University, China

## REVIEWED BY

Jun Wan,  
The University of Queensland, Australia  
Venkatesh Katari,  
University of Toledo, United States

## \*CORRESPONDENCE

Zhengxiao Ouyang,  
✉ ouyangzhengxiao@csu.edu.cn  
Wenzhao Li,  
✉ liwenzhao@csu.edu.cn

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 18 June 2023

ACCEPTED 07 September 2023

PUBLISHED 02 October 2023

## CITATION

Li T, Li W, Guo X, Tan T, Xiang C and Ouyang Z (2023), Unraveling the potential mechanisms of the anti-osteoporotic effects of the *Achyranthes bidentata*–*Dipsacus asper* herb pair: a network pharmacology and experimental study. *Front. Pharmacol.* 14:1242194. doi: 10.3389/fphar.2023.1242194

## COPYRIGHT

© 2023 Li, Li, Guo, Tan, Xiang and Ouyang. 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.

# Unraveling the potential mechanisms of the anti-osteoporotic effects of the *Achyranthes bidentata*–*Dipsacus asper* herb pair: a network pharmacology and experimental study

Tao Li<sup>1†</sup>, Wenzhao Li<sup>1\*</sup>, Xiaoning Guo<sup>1</sup>, Tingting Tan<sup>2</sup>,  
Cheng Xiang<sup>1†</sup> and Zhengxiao Ouyang<sup>1\*</sup>

<sup>1</sup>Department of Orthopedics, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China, <sup>2</sup>Department of Immunology, School of Basic Medical Science, Central South University, Changsha, Hunan, China

**Background:** Osteoporosis is a prevalent bone metabolism disease characterized by a reduction in bone density, leading to several complications that significantly affect patients' quality of life. The *Achyranthes bidentata*–*Dipsacus asper* (AB–DA) herb pair is commonly used in Traditional Chinese Medicine (TCM) to treat osteoporosis. This study aimed to investigate the therapeutic compounds and potential mechanisms of AB–DA using network pharmacology, molecular docking, molecular dynamics simulation, and experimental verification.

**Methods:** Identified compounds of AB–DA were collected from the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP), Traditional Chinese Medicine Information Database (TCM-ID), TCM@Taiwan Database, BATMAN-TCM, and relevant literature. The main bioactive ingredients were screened based on the criteria of "OB (oral bioavailability) ≥ 30, DL (drug-likeness) ≥ 0.18." Potential targets were predicted using the PharmMapper and SwissTargetPrediction websites, while disease (osteoporosis)-related targets were obtained from the GeneCards, DisGeNET, and OMIM databases. The PPI network and KEGG/GO enrichment analysis were utilized for core targets and pathway screening in the STRING and Metascape databases, respectively. A drug–compound–target–pathway–disease network was constructed using Cytoscape software to display core regulatory mechanisms. Molecular docking and dynamics simulation techniques explored the binding reliability and stability between core compounds and targets. *In vitro* and *in vivo* validation experiments were utilized to explore the anti-osteoporosis efficiency and mechanism of sitogluside.

**Results:** A total of 31 compounds with 83 potential targets for AB–DA against osteoporosis were obtained. The PPI analysis revealed several hub targets, including AKT1, CASP3, EGFR, IGF1, MAPK1, MAPK8, and MAPK14. GO/KEGG analysis indicated that the MAPK cascade (ERK/JNK/p38) is the main pathway involved in treating osteoporosis. The D–C–T–P–T network demonstrated

therapeutic compounds that mainly consisted of iridoids, steroids, and flavonoids, such as sitogluside, loganic acid, and  $\beta$ -ecdysterone. Molecular docking and dynamics simulation analyses confirmed strong binding affinity and stability between core compounds and targets. Additionally, the validation experiments showed preliminary evidence of antiosteoporosis effects.

**Conclusion:** This study identified iridoids, steroids, and flavonoids as the main therapeutic compounds of AB–DA in treating osteoporosis. The underlying mechanisms may involve targeting core MAPK cascade (ERK/JNK/p38) targets, such as MAPK1, MAPK8, and MAPK14. *In vivo* experiments preliminarily validated the anti-osteoporosis effect of sitogluside. Further in-depth experimental studies are required to validate the therapeutic value of AB–DA for treating osteoporosis in clinical practice.

#### KEYWORDS

osteoporosis, *Achyranthes bidentata*, *Dipsacus asper*, network pharmacology, MAPK

## 1 Introduction

Osteoporosis is a prevalent bone metabolism disease that affects over 200 million people worldwide (McDonald et al., 2021; Grewe et al., 2022). It is characterized by a reduction in bone density, which

greatly increases the risk of fractures (NIH Consensus Development Panel on Osteoporosis Prevention et al., 2001). Osteoporotic fractures and associated complications can have a significant and lasting impact on patients' quality of life, sometimes even threatening their lives, and place a considerable cost on society

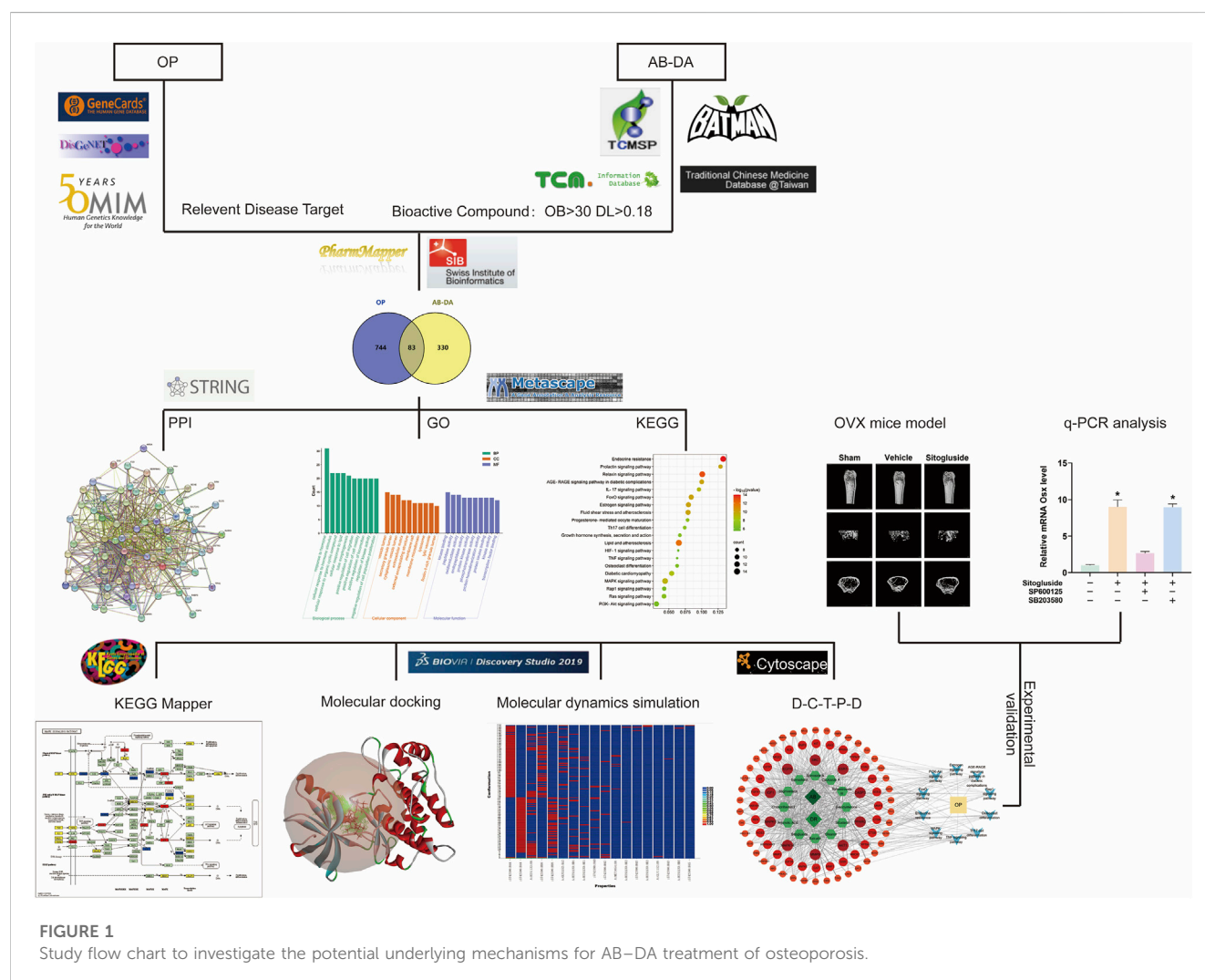
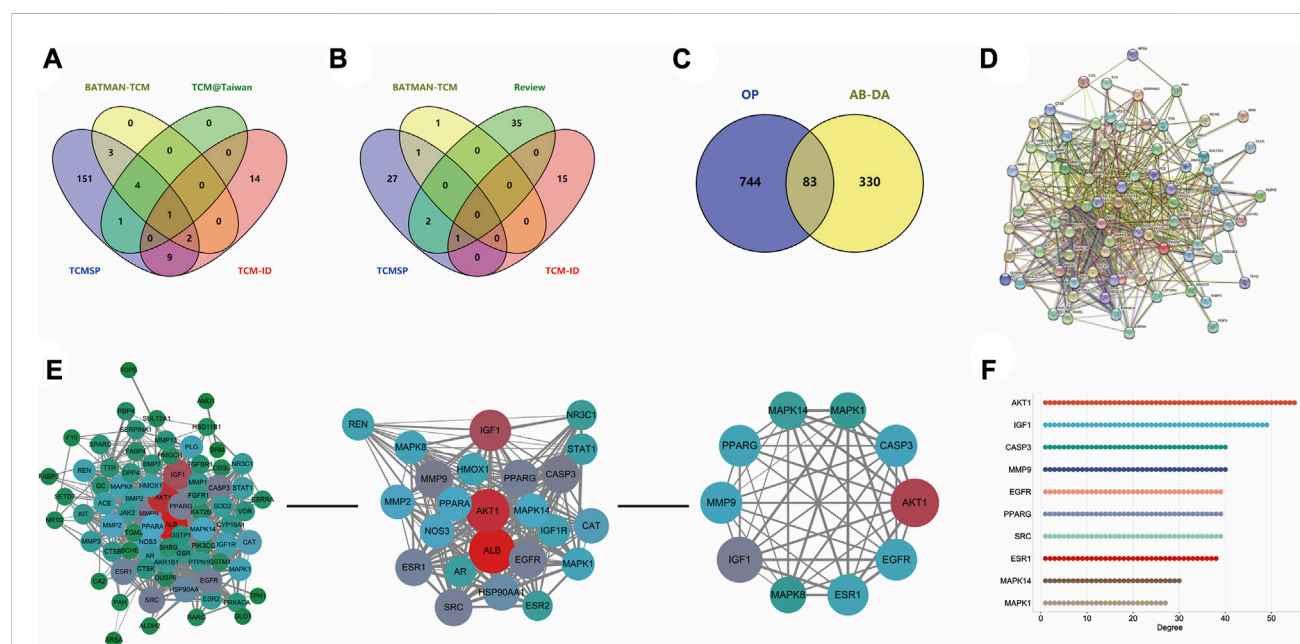
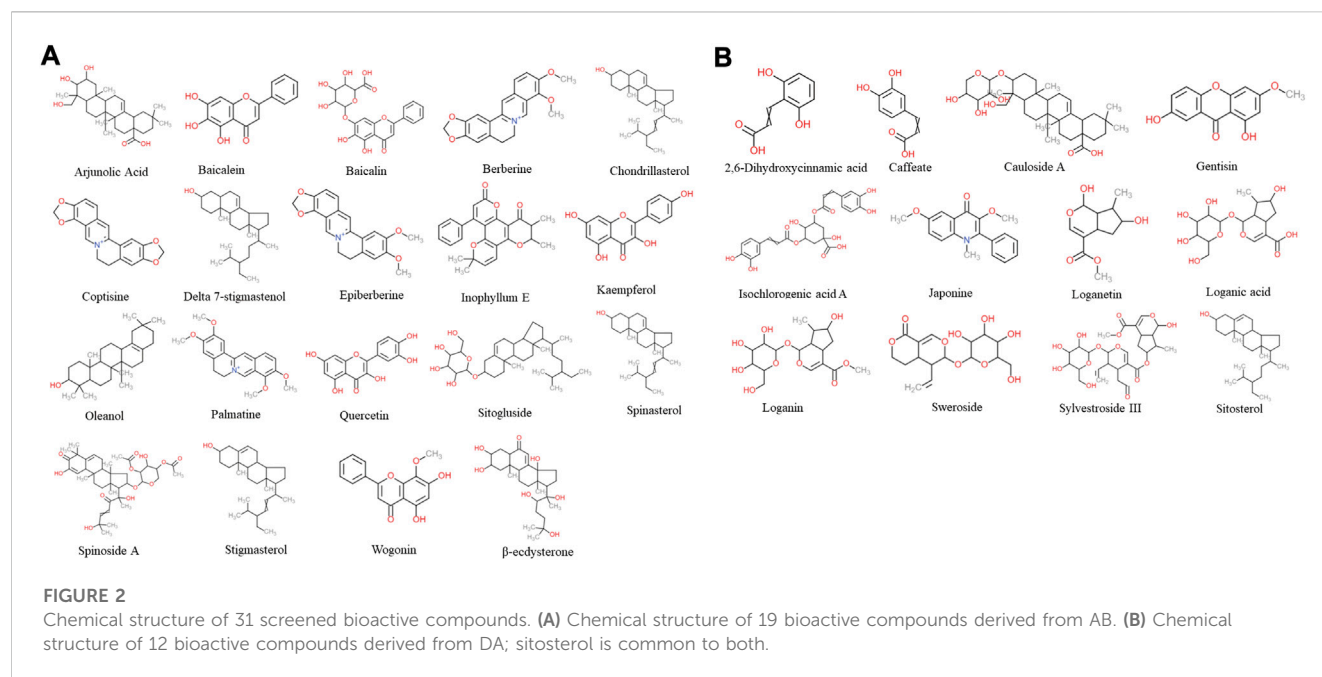


FIGURE 1

Study flow chart to investigate the potential underlying mechanisms for AB–DA treatment of osteoporosis.

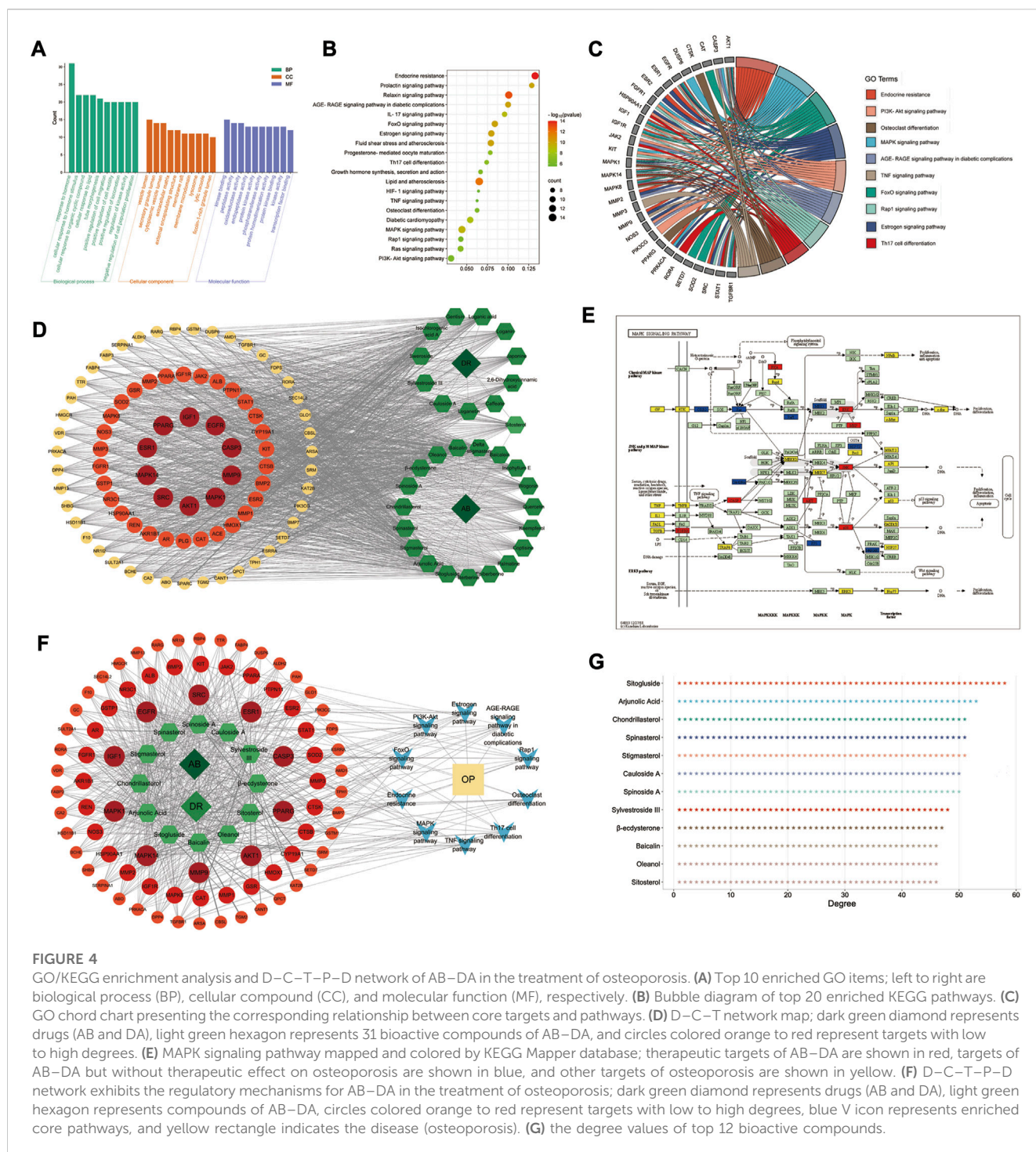


and individuals (Compston et al., 2019; Liang et al., 2022). With the aging of the global population, osteoporosis has become a pressing health problem (Jiang et al., 2020).

Anti-osteoporosis drugs such as estrogen, raloxifene, bisphosphonates, calcitonin, and parathyroid hormone (PTH) are commonly used in clinical practice (Chen et al., 2022). These

medications, however, have side effects and severe responses that restrict their long-term usage. For example, bisphosphonates may cause jaw osteonecrosis and renal impairment (Li et al., 2021). There is, therefore, an urgent need to identify potential anti-osteoporosis drugs that are both more effective and safer.





Traditional Chinese Medicine (TCM) has a rich history and has been widely used in Asia to treat various diseases, including osteoporosis (Huang et al., 2022). TCM is cost-effective and has fewer side effects than chemosynthetic drugs, making it more suitable for long-term use (Mukwaya et al., 2014). The application of TCM in modern society provides a new pathway for complementary and alternative medicine (CAM) treatment (Chu et al., 2022). Over the years, many TCM treatments and prescriptions have been used to treat various orthopedic diseases,

especially osteoporosis and fractures, with great success (Suvarna et al., 2018; Peng et al., 2022). TCM's prescription to treat osteoporosis can also play a comprehensive role in regulating body function and relieving pain (Feng et al., 2022). A Chinese herb pair, generally composed of two kinds of herbal medicine, is the essence of TCM prescriptions. Compared with all herbs in prescriptions, studying and elucidating the complex pharmacological mechanism of herb pairs is simpler and more beneficial (Liu et al., 2020).

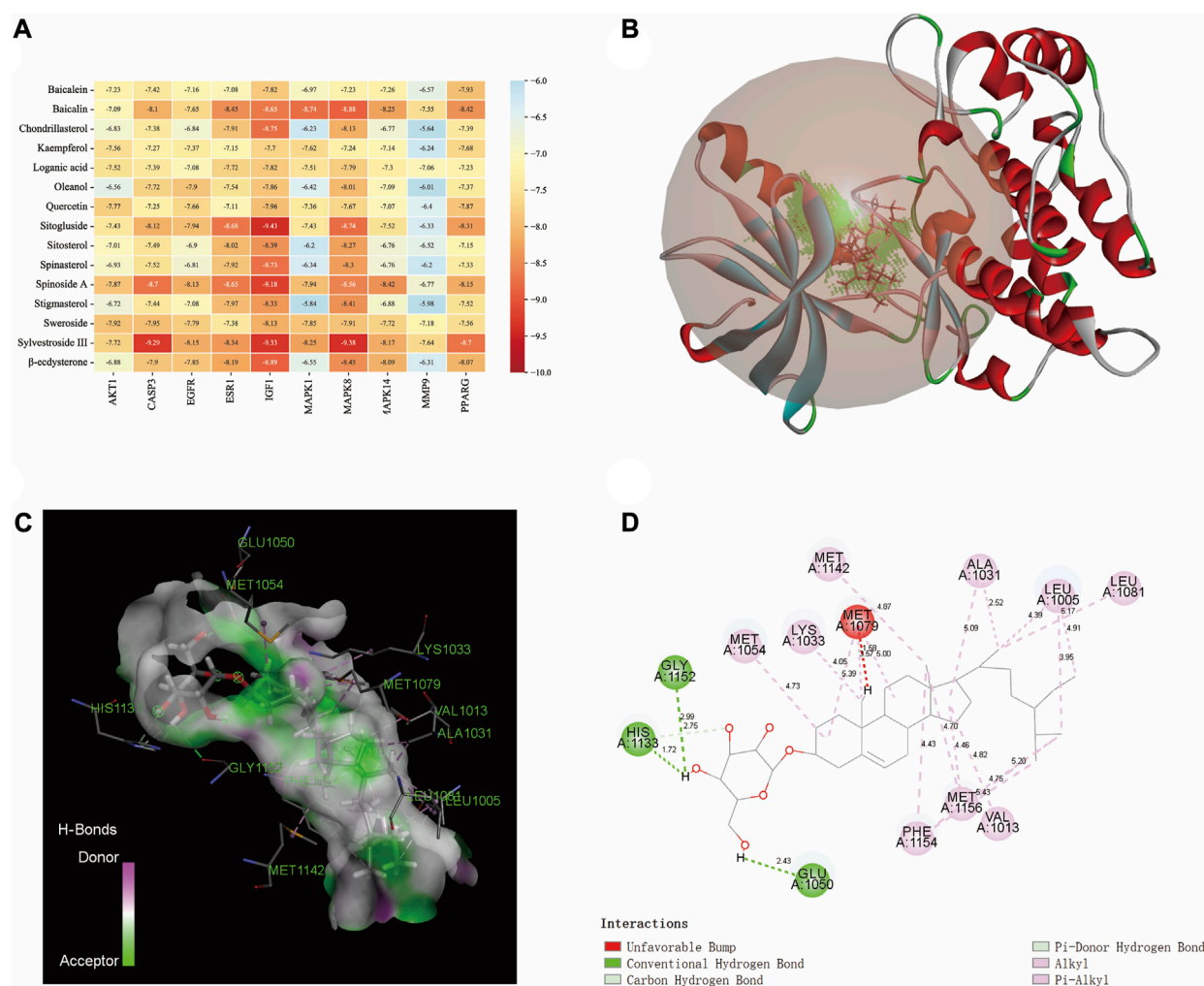


FIGURE 5

Molecular docking between bioactive compounds and core targets. (A) Binding affinity heatmap of compound ligand–protein receptor complexes, showing stronger binding affinity. (B) Binding details of the sitogluside–IGF1 complex (3D). (C) Binding details of the sitogluside–IGF1 complex (spatial structure). (D) Binding details of intermolecular force types of the sitogluside–IGF1 complex (2D).

*Achyranthes bidentata* (AB), also known as Niu Xi, is included in the Chinese, Japanese, and Korean pharmacopoeia (He et al., 2017). Additionally, its dried roots are regularly used in TCM for osteoporosis (Yan et al., 2019). A number of biological activities, including anti-osteoporosis (He et al., 2010; Zhang et al., 2012; Jiang et al., 2014; Suh et al., 2014; Zhang M. et al., 2018; Zhang S. et al., 2018), anti-tumor (Jin et al., 2007), and anti-oxidant (Huang et al., 2015), have been demonstrated by contemporary pharmacological research on AB extracts. *Dipsacus asper* (DA), also known in Chinese as Xu Duan —meaning “to rebuild fractures and unite bones”—is discussed in Shennong’s Classic of Material Medicine, which is the earliest source (Tao et al., 2020). DA can be used to treat muscle pain and bone repair, golden sores, and collapses (Wu et al., 2022). According to modern pharmacological investigations, numerous disorders, including osteoporosis and osteoarthritis, have been successfully treated with DA (Liu et al., 2019; Yu et al., 2019; Zhang et al., 2019). Jiegudan capsules, which contain AB and DA, are a common traditional Chinese medicine prescription to treat osteoporosis. Although many compounds

have been isolated from AB and DA, the potential pharmacological mechanisms of AB–DA herb pairs and their interactions with osteoporosis-related targets and pathways remain unclear and need further exploration.

In recent years, the use of network pharmacology has become increasingly popular for exploring the interaction network of TCM therapy (Shuai et al., 2020). Molecular docking, a virtual screening technology that simulates the behavior of small-molecule ligands at the binding sites of receptor proteins, has also gained popularity for developing novel drugs (Pagadala et al., 2017). This research aims to elucidate the potential mechanism in TCM of the AB–DA herb pair for treating osteoporosis, bioinformatics prediction by network pharmacology, molecular docking, and molecular dynamics simulation, and verify these via alkaline phosphatase (ALP) activity, osteoblast mineralization assays, Western blot, q-PCR, and an ovariectomy (OVX) osteoporosis mouse model. A flow chart outlining the study’s approach is presented in Figure 1.



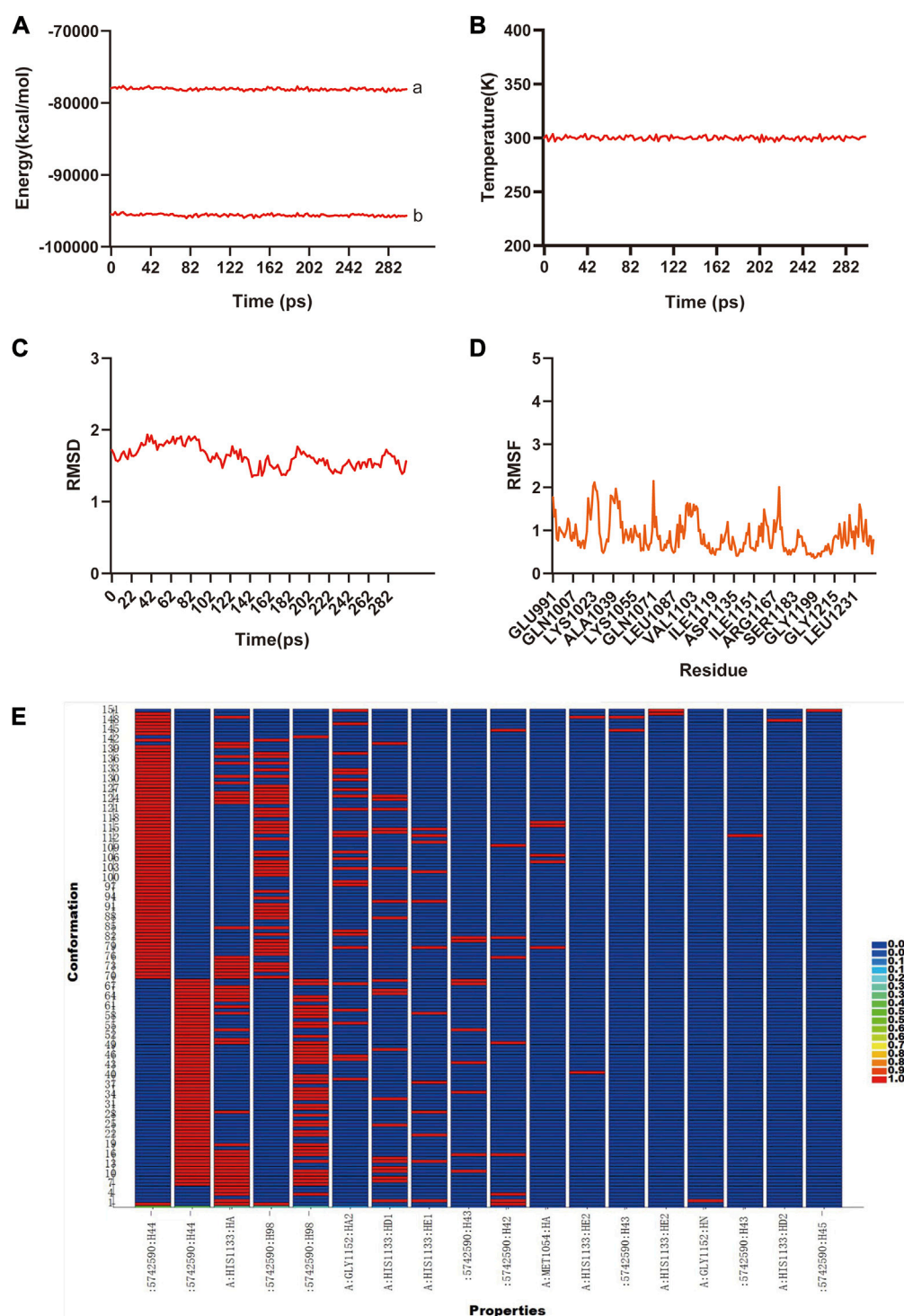


FIGURE 6

The molecular dynamic simulations to calculate the binding stability of sitogluside-IGF1 complex. (A), The total energy (a) and potential energy (b) curves of the whole ensemble in 300 ps simulation, showed the stability. (B), The temperature alteration of whole ensemble is controllable in 300 ps simulation. (C), The RMSD curve present the conformational alternation of receptor made by ligand. (D), The RMSF curve showed the conformational alternation of amino acid residues. (E), The hydrogen bond heat map of sitogluside-IGF1 complex.

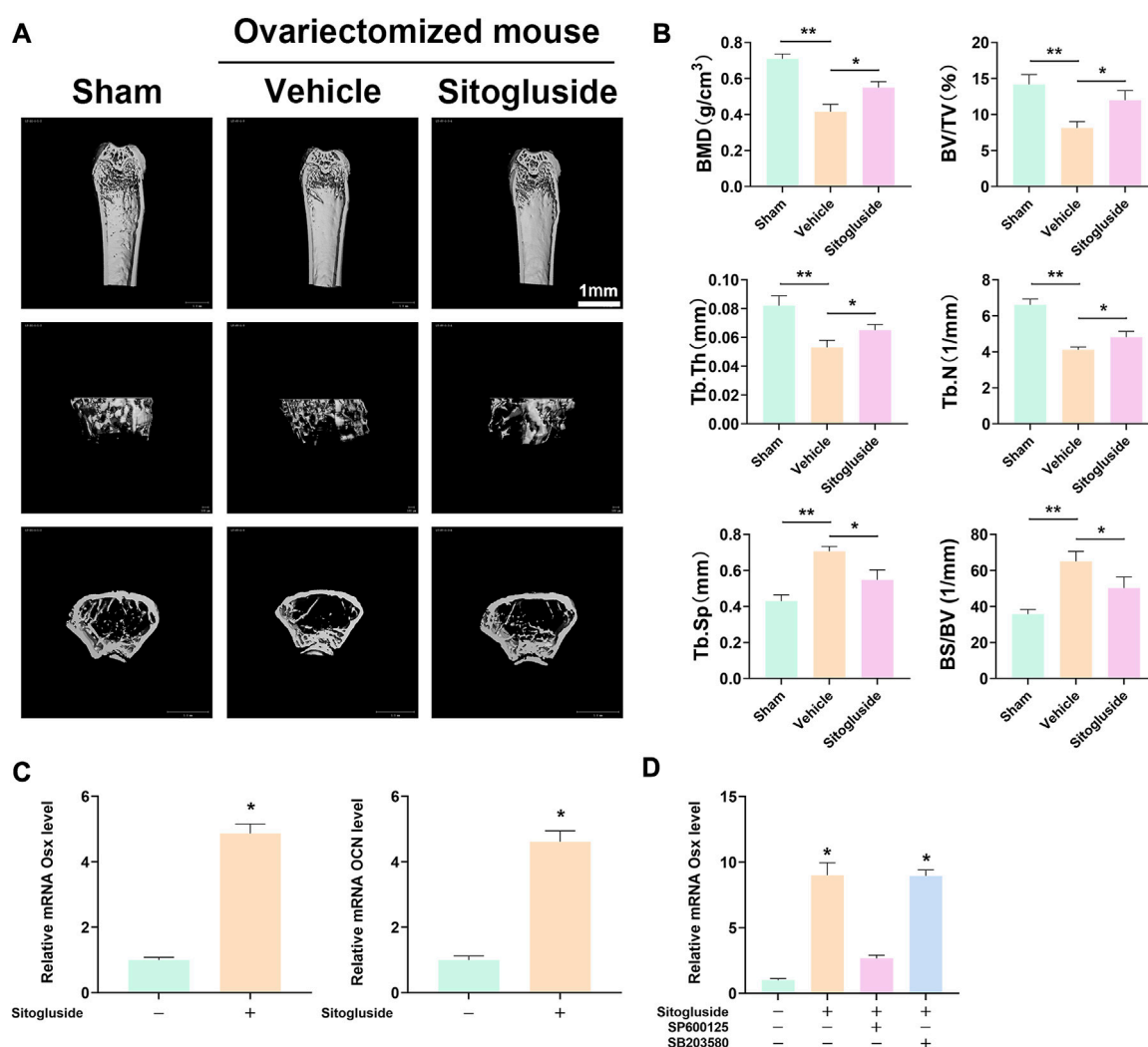


FIGURE 7

*In vivo* OVX model and q-PCR experiments to validate the potential anti-OP effect of sitogluside. (A)  $\mu$ CT scanning of mice tibia showing the bone loss alleviated by sitogluside's efficiency. (B) Statistical results of osteoporosis phenotype parameters of OVX mice, including BMD, Tb. Th, BV/TV, and Tb. N. (C) q-PCR results of osteogenic biomarker, which is shown to promote the osteogenic effect of sitogluside. (D) q-PCR results of OSX, which, treated with sitogluside or combined with the specific inhibitor of JNK (SP600125) and p38 (SB203580), indicates that sitogluside could target the JNK pathway to promote osteogenic genes in the treatment of OP.

## 2 Materials and methods

### 2.1 Screening of bioactive compounds of AB–DA

The TCM@Taiwan (<http://tcm.cmu.edu.tw/zh-tw/>) Database (Chen, 2011), Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP <http://lsp.nwu.edu.cn/tcmsp.php>) (Ru et al., 2014), BATMAN-TCM (L et al., 2016) (a bioinformatics analytical tool for the molecular mechanisms of TCM: <http://bionet.ncpsb.org.cn/>), Traditional Chinese Medicine Information Database (Kang et al., 2013) (TCM-ID <http://bidd.group>), and relevant literature were utilized to acquire all identified AB–DA compounds. ADME (absorption, distribution, metabolism, and excretion) properties were applied to screen bioactive ingredients, and the screening

criteria were set as “oral bioavailability (OB)  $\geq$  30, drug-likeness (DL)  $\geq$  0.18” (Gao et al., 2022) to screen the compounds from the TCMSP database. Similarly, compounds from different sources were screened in the SwissADME (Daina et al., 2017) database using their pharmacokinetic properties (<http://www.swissadme.ch>).

### 2.2 Relevant targets of AB–DA compound and osteoporosis

To predict potential targets based on their spatial configuration, the compounds generated in the previous step were imported into the SwissTargetPrediction (Daina et al., 2019) (<http://www.swisstargetprediction.ch/>) and PharmMapper databases (Liu et al., 2010) (<http://www.lilab-ecust.cn/pharmmapper/>). The UniProt ID of the target was converted into a standardized gene name using the

UniProt database (Holzhüter and Geertsma, 2022) (<https://www.uniprot.org>). The keyword “osteoporosis” was searched in the GeneCards (Barshir et al., 2021) (<https://www.genecards.org/>), DisGeNET (Piñero et al., 2020) (<https://www.disgenet.org>), and Online Mendelian Inheritance in Man (Li et al., 2012) (OMIM, <https://omim.org>) databases to obtain relevant targets. Then, the overlapping targets identified by Venn diagram were considered as targets of AB–DA for the treatment of osteoporosis after merging and removing duplicates.

## 2.3 PPI network for core target selection

To identify potential hub genes, a protein–protein interaction (PPI) network was established in the STRING (Szklarczyk et al., 2021) (<http://string-db.org>, Version 11.5) database, with a focus on the anti-osteoporosis efficacy of AB–DA in *Homo sapiens* and an interaction score threshold of 0.4. Topological analysis was performed, and the core targets of AB–DA for treating osteoporosis were accurately selected by using the Cytoscape plug-ins MCODE (molecular complex detection) and CytoHubba (Ye et al., 2022).

## 2.4 GO and KEGG pathway enrichment analyses for the core pathways

Analysis of Gene Ontology (GO) functions, including cellular component (CC), molecular function (MF), and biological process (BP), and Kyoto Encyclopedia of Genome and Genome (KEGG) pathway enrichment analysis, was utilized to clarify the key anti-osteoporosis mechanism of the AB–DA herb pair. When entering the targets into the Metascape database (<http://www.metascape.org/>), the cut-off *p*-value, minimum overlap value, and concentration value were set to 0.01, 3, and 1.5, respectively (Zhou et al., 2019). False-positive rate (FPR) analysis was eliminated using the Benjamini–Hochberg method with a *q*-value of 0.05 or lower (Zou et al., 2016). The enriched findings were displayed as bar and bubble plots on the bioinformatics website using the R package (<http://www.bioinformatics.com.cn/>). Comprehensive information on the most significantly enriched pathway was then extracted and colored (Kanehisa and Sato, 2020). Finally, a herb–compound–target–pathway–disease network was created using Cytoscape software (v.3.9.1, <https://cytoscape.org/>) to present the complicated network of the AB–DA herb pair in the treatment of osteoporosis.

## 2.5 Molecular docking to validate binding affinity

The SwissDock platform (Grosdidier et al., 2011) (<http://www.swissdock.ch/>) is an online molecular docking (MD) tool to determine the binding affinity from each binding site between small molecule ligands and receptor proteins. The X-ray diffraction of the protein crystal structure of key targets were downloaded from the Protein Data Bank (PDB) database ([www.rcsb.org](http://www.rcsb.org)) (Nakamura et al., 2022). The binding sites were ranked based on their binding affinity scores, with the site having the

smallest score considered the best binding site. Discovery Studio 2019 software (<https://www.3ds.com>) was used to visualize the binding details (Sultana et al., 2022).

## 2.6 Molecular dynamics simulation to validate binding stability

To investigate the stability of the complexes between small-molecule ligands and proteins, molecular dynamics simulations (MDS) were performed using the Standard Dynamics Cascade subunit of the Discovery Studio 2019 software. The ligand–protein complex with the lowest binding affinity score according to molecular docking analysis was selected (Hu et al., 2022). In this simulation system, water molecules are used to fill the solvent chamber, and Cl and Na<sup>+</sup> ions are used to maintain an electrically neutral state. The simulation time was set as 300 ps, and the heating, balancing, and manufacturing phases were carried out after the system was balanced by an NPT ensemble, which fixed the pressure, temperature, and particle number. The analysis of the locus was performed using root mean square fluctuation (RMSF), root mean square deviation (RMSD), and hydrogen bond properties to produce the results.

## 2.7 Reagent and cell culture

Sitogluside, identified as one of the most promising bioactive compounds in the AB–DA herb pair, was further investigated for its anti-osteoporotic effects and associated mechanisms. Sitogluside was procured from the Dalian Meilunbio company and solubilized in dimethyl sulfoxide (DMSO). Human fetal osteoblast (hFOB) cells were obtained from the American Type Culture Collection (ATCC) and cultured in six-well plates using Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Gibco, United States), 1% penicillin–streptomycin (P/S), and 0.3 mg/mL Geneticin (G418). The JNK-specific inhibitor (SP600125) and p38-specific inhibitor (SB203580) were acquired from Absin (Shanghai). The cells were maintained in a humidified sterile atmosphere at 34 °C. They were subsequently treated with sitogluside when they reached a confluence of 60%–80% per well.

## 2.8 Alkaline phosphatase activity, Alizarin red staining mineralization, and osteoclast differentiation assays

For alkaline phosphatase (ALP) activity analysis, the BCIP/NBT Alkaline Phosphatase color development kit (Beyotime Institute of Biotechnology, Shanghai, China) was used according to the manufacturer's procedure. Briefly, osteoblast precursor cells were seeded at  $3 \times 10^4$  cells/well in 24-well plates and grown for 14 days in osteogenic media (DMEM +10% FBS +1% P/S +100 M ascorbic acid +2 mM 2-glycerophosphate +10 nM dexamethasone). The stained culture plates were photographed using a microscope (Leica image analysis system, Q500MC) and quantified using ImageJ software (National Institutes of Health, Bethesda, MD, United States). In

addition, a 2% ARS reagent (Beyotime Institute of Biotechnology) was used to detect matrix mineralization, with the same protocol as ALP assay except the dye. In addition, we also investigated the effect of sitogluside on the osteoclast to fully illustrate the anti-OP effects. The RAW264.7 (osteoclast precursor) cells were treated with a different concentration of sitogluside and supplemented with 50 ng/mL receptor activator of NF- $\kappa$ B ligand (RANKL) as an osteoclast formation stimulator. Cells were then fixed in paraformaldehyde and stained with tartrate-resistant acid phosphatase (TRAP) after the intervention.

## 2.9 Ovariectomized mouse model

This study received ethical approval from the Animal Care Committee of the Second Xiangya Hospital of Central South University. A total of 30 10-week-old female C57BL/6 mice were procured from SLAC Laboratory Animal Co. Ltd. (SLACCAS, Shanghai, China). They were acclimatized in specific pathogen-free (SPF) cages for 1 week, during which measures were taken to minimize animal suffering through the use of anesthesia and sterile techniques during the surgical procedures. Bilateral ovariectomy (OVX) or sham surgery (retroperitoneal incision without ovariectomy) was performed based on group assignment, as described in the following paragraph.

Following the surgical procedures, the mice were randomly assigned to one of three groups: sham group (non-OVX mice,  $n = 10$ ), vehicle group (OVX mice,  $n = 10$ ), and sitogluside group (OVX mice intraperitoneally injected with sitogluside at a dose of 10 mg/kg/day,  $n = 10$ ).

After 12 weeks, all mice were euthanized by cervical dislocation, and the right femurs were dissected and fixed in 4% paraformaldehyde (PFA) for 48 h. High-resolution micro-computed tomography ( $\mu$ CT40, Scanco, Zurich, Switzerland) was employed for bone analysis at the following parameters: scanning voltage = 80 kV, electric current = 80  $\mu$ A, and resolution = 10  $\mu$ m. The relevant trabecular bone volume fractions (BV/TV), trabecular number (Tb. N), trabecular thickness (Tb. Th), and trabecular separation (Tb. Sp) were subsequently calculated to assess the protective efficacy of sitogluside in OVX mice. In addition, to investigate the biosafety of sitogluside in the OVX model, the main organs were also obtained and detected using hematoxylin-eosin (H&E) staining. The heart, liver, spleen, lung, and kidney were hence fixed with formalin and embedded in paraffin cut to a 4  $\mu$ m section. They were dewaxed in xylene, rehydrated with concentration gradient ethanol, and then stained with H&E for histological examinations and morphometric analysis. The serum of the mice was also collected to examine the biomarkers of alanine aminotransferase (ALT), creatine kinase (CK), and blood urea nitrogen (BUN).

## 2.10 Quantitative real-time PCR analysis

Human osteoblast (hFOB) cells were treated with 40  $\mu$ M sitogluside or combined with the specific inhibitor of JNK and p38, depending on the groups, with ascorbic acid added in the osteogenic media. The cells were then harvested using the RNeasy

Mini kit (QIAGEN, CA, United States) to extract total RNA following the manufacturer's protocol. Subsequently, cDNA synthesis was performed using the reverse transcriptase kit (Takara Biotechnology, Japan). Real-time PCR analysis was performed using the SYBR Premix Ex Taq kit (Takara Biotechnology, Japan). The PCR parameters were set as follows: 40 cycles (denaturation at 95 °C for 10 s and amplification at 60 °C for 30 s). The resultant data were recorded as cycle threshold (Ct) values, and the  $2^{-\Delta\Delta CT}$  method was employed for further analysis of RNA expression. In addition, to determine the modulated effect of sitogluside with JNK and osteogenic genes, the knockdown and activation of JNK expression were applied. ShRNA (shGnai3) was thus used to downregulate the expression level of JNK, and ASM was used as an activator to increase the p-JNK level (Meng et al., 2021). The alterations of osteogenic biomarkers were detected via q-PCR.

## 2.11 Western blot analysis

The hFOB cells were harvested by trypsin and then lysed in RIPA for 30 min on ice. Cell lysates were centrifuged at 12,000 g for 15 min at 4 °C; the supernatant was collected, and protein content was quantified via the BSA protein assay kit following the manufacturer's instruction. Proteins were separated by electrophoresis on 10%–12% SDS-PAGE at 100 V for 1.5 h and transferred onto a 0.45  $\mu$ m polyvinylidene difluoride (PVDF) membrane at 250 mA for 1 h. The PVDF membrane was blocked with 5% non-fat milk in TBST buffer for 1 h at room temperature and incubated with primary antibody at 4 °C overnight. They were then incubated with secondary antibody for 1 h at room temperature and detected using the Chemiluminescence Kit.

## 2.12 Statistical analysis

Statistical analyses were conducted using GraphPad Prism 8.0.2 software (San Diego, United States). The data are presented as means  $\pm$  standard deviation (SD). Data comparisons were performed using one-way analysis of variance (ANOVA), and statistical significance was determined by a  $p$ -value of less than 0.05.

# 3 Results

## 3.1 Relevant targets of AB–DA compound and osteoporosis

According to TCM databases, 185 components of AB and 82 compounds of DA were obtained in this work (Figures 3A, B). After filtration by screening criteria, 19 potential bioactive compounds of AB were obtained: arjunolic acid, baicalein, baicalin, berberine, chondrillasterol, coptisine, delta-7-stigmastenol, epiberberine, inophyllum E, kaempferol, oleanol, palmatine, quercetin, sitogluside, spinasterol, spinoside A, stigmasterol, wogonin, and  $\beta$ -ecdysterone. The 12 compounds from DA were 2,6-dihydroxycinnamic acid, caffeate, cauloid A, gentisin, isochlorogenic acid A, japonine, loganetin, loganic acid,



loganin, sweroside, and sylvestroside III. Sitosterol is a common compound of both AB and DA (details shown in [Supplementary Table S1](#); structures shown in [Figure 2](#)).

After merging and duplicating results, 413 potential targets associated with 31 bioactive compounds were obtained, and 827 relevant targets of osteoporosis were acquired. Some 83 overlapped genes on Venn between AB–DA ingredients and osteoporosis were regarded as potential therapeutic targets ([Figure 3C](#); details shown in [Supplementary Table S2](#)).

## 3.2 PPI network of AB–DA against osteoporosis

A total of 83 targets of AB–DA against osteoporosis were imported to the STRING database; after deleting the disconnect targets, a PPI network with 78 nodes and 702 edges was constructed. Cytoscape software was utilized for further visualization, and plug-ins of MCODE and CytoHubba based on the topological parameters were applied to screen the hub genes, including AKT1, IGF1, CASP3, MMP9, EGFR, PPARG, ESR1, MAPK1, MAPK8, and MAPK14 ([Figures 3D–F](#)).

## 3.3 GO and KEGG pathway enrichment analyses of AB–DA against osteoporosis

GO and KEGG pathway enrichment analyses were performed in the Metascape database, and 83 targets with 875 GO items were enriched. It contained 772 biological processes (BP), 29 cellular components (CC), and 74 molecular functions (MF) items ( $p < 0.01$ , adjusted  $q < 0.05$ ) (details shown in [Figure 4A](#)).

A total of 135 KEGG pathways were significantly enriched, which mainly included the MAPK signaling pathway, osteoclast differentiation, PI3K-Akt signaling pathway, and endocrine resistance ([Figures 4B, C](#)). We further mapped and colored the regulation details of the MAPK cascade in the KEGG mapper database; in that map, red objects represent targets of AB–DA against osteoporosis, blue objects show the targets of AB–DA without the therapeutic effects of osteoporosis, and the other untargeted targets of osteoporosis are colored yellow ([Figure 4E](#)).

## 3.4 Drug–compound–target–pathway–disease network analysis

A drug–compound–target network was constructed using Cytoscape to illustrate core compounds and targets ([Figure 4D](#)); the core bioactive compounds included sitogluside, arjunolic acid, chondrillasterol, stigmasterol, spinasterol, spinoside A, cauloside A, sylvestroside III,  $\beta$ -ecdysterone, sitosterol, oleanol, and baicalin (ranked by degree value) ([Figure 4G](#)). A drug–compound–target–pathway–disease (D–C–T–P–D) network was then constructed to exhibit the complex molecular mechanisms of AB–DA anti-osteoporosis with multi-compound, multi-target, and multi-pathway characteristics. The dark green diamond represents the drugs (AB and DA), the light green hexagon represents the compounds of AB–DA, the circles colored orange

to red represent targets with low to high degrees, the blue V icon represents the core enriched pathways, and the yellow rectangle indicates the disease (osteoporosis) ([Figure 4F](#)).

## 3.5 Molecular docking

The binding affinity between the core compounds and core targets are shown by heatmap ([Figure 5A](#)). According to relevant theories of molecular docking, the results of binding affinity  $< -5.0$  kcal/mol suggest that there is a good spontaneous binding activity between molecule ligands and protein receptors, and results  $< -7.0$  kcal/mol are stronger. Our research results showed that all the compounds had good binding activity with core targets, with binding affinities ranging from  $-5.64$  kcal/mol to  $-9.43$  kcal/mol. Sitogluside has the highest binding activity with IGF1. As shown in [Figures 5B–D](#), the molecular interaction forces between IGF1 and sitogluside include  $\pi$ -donor hydrogen bond,  $\pi$ -alkyl bond, conventional hydrogen bond, carbon–hydrogen bond, and alkyl bond. The distances between the sitogluside atoms and amino acid residues of IGF1 range from  $1.72$  Å (number 1133, histidine residue) to  $5.43$  Å (number 1154, phenylalanine residue).

## 3.6 Molecular dynamics simulation

The previous analysis of molecular docking showed the strong binding affinity between AB–DA compounds and core targets, and molecular dynamics simulation was utilized to identify the stability of the ligand–protein complex after docking by conformation alternation with potential energy under Newton's law of motion. After the 300 ps simulation, the energy and temperature alternate tendency of the ensemble, hydrogen bond, RMSD, and RMSF changes of ligand–receptor interaction were calculated for stability analysis. RMSD was used to analyze the conformational alternation of receptors made by the ligand, and the results showed that curves and fluctuations only occurred at the beginning of the 80 ps simulation, and then tended to be stable ([Figure 6A](#)). The RMSF curve was utilized to monitor the conformational alternation of amino acid residues, and results show that the whole process is stable with only some small random fluctuations—which also reflects the whole ensemble's stability to some extent ([Figure 6B](#)). The same tendency is also observed in the hydrogen bond heatmap ([Figure 6C](#)). The energies and temperature alternation of the whole ensemble were also stable and controlled. Therefore, the overall results exhibited good stability between the small-molecule ligand–protein receptor complex.

## 3.7 Sitogluside promotes mineralization and ALP activity in osteoblast without any effect on osteoclast

To further investigate the impact of sitogluside on alkaline phosphatase (ALP) activity and mineralization in osteoblasts, we conducted ALP activity and ARS experiments. As illustrated in [Supplementary Figure S3D](#), the sitogluside group exhibited



increased ALP activity and mineralization compared to the control group. However, when shGnai3 intervention was introduced, the stimulatory effect of sitogluside on ALP activity and mineralization in osteoblasts was partially diminished. Subsequently, with the addition of the p-JNK activator ASM, the promotion of ALP activity and mineralization was restored. Additionally, our results indicated that sitogluside might not have a significant effect on osteoclast formation (Supplementary Figure S2).

### 3.8 Sitogluside administration protects against OVX-induced bone loss

The OVX model was employed to further investigate the anti-osteoporotic effects of sitogluside *in vivo*. As depicted in Figure 7A, the  $\mu$ CT scan results demonstrated a significant loss of bone in the vehicle group (OVX mice) compared to the sham group, confirming the successful establishment of the OVX model. Additionally, the sitogluside-treated group exhibited a mitigating effect on bone loss compared to the vehicle group. Specifically, the sitogluside-treated OVX mice demonstrated increased bone mineral density (BMD), trabecular thickness (Tb. Th), bone volume fraction (BV/TV), and trabecular number (Tb. N), while exhibiting a decreased bone-surface-to-bone-volume ratio (BS/BV) and trabecular separation (Tb. Sp). These findings collectively indicate the anti-osteoporotic efficacy of sitogluside in the OVX mouse model (Figure 7B). Furthermore, histological examination of major organs in the drug-treated group, including the heart, liver, spleen, lungs, and kidneys, was conducted using (H&E) staining. Additionally, mouse serum was analyzed to assess cardiac, hepatic, and renal functions. As depicted in Supplementary Figure S1, the experimental results indicate that sitogluside did not exert significant toxic effects on mouse organs.

### 3.9 Effects of sitogluside on osteoblast-related genes

To gain further insights into the underlying mechanism of sitogluside's anti-osteoporotic effects, we examined the expression of osteogenic-related genes using q-PCR and Western blot. As shown in Figure 7, treatment with sitogluside for 48 h resulted in a significant increase in the expression of osteogenic markers, such as Osterix (OSX) and osteocalcin (OCN). These findings suggest that sitogluside has the potential to promote osteogenic activity. Additionally, we investigated the core target and pathway associated with sitogluside in the treatment of osteoporosis via blocking the JNK and p38 cascade (Figure 7D). Our results revealed that targeting JNK could reduce the osteogenic efficacy of sitogluside while blocking the p38 pathway without significant expression changes of OSX. To further investigate whether sitogluside exerts its osteogenic effects through the JNK pathway, we employed shRNA to silence the JNK pathway. As shown in Supplementary Figure S3C, compared to the sitogluside group, silencing the JNK pathway with shRNA (shGnai3) resulted in a partial reduction in the expression levels of osteogenic markers, including Runx2, OSX, OCN, and ALP. However, when treated with the p-JNK activator anisomycin (ASM), these osteogenic markers increase. Moreover, as

shown in Supplementary Figures S3A, B, the phosphorylation level of p-JNK increases with sitogluside intervention. These further suggest that sitogluside may promote osteoblast differentiation through the JNK signaling pathway. We also investigated the effect of sitogluside on the Smad 1/5/8 protein in the BMP signaling pathway, which also showed increasing expression (Supplementary Figures S3A, B).

## 4 Discussion

Osteoporosis is regarded as a silent disease without clinical symptoms before complications are apparent (Anthamatten and Parish, 2019). Current conventional therapy mainly focuses on symptom prevention with long-term supplements of calcium and on intervention to regulate bone metabolism. The curative effects depend on the individual response and have mild or severe side effects (Ensrud and Crandall, 2017). AB and DA have been two of the most important herbs for bone diseases in TCM therapy for more than 2,000 years (He et al., 2017; Tao et al., 2020), and play an important role in many classic anti-osteoporosis drugs. Recently, researchers have isolated more than 100 ingredients and verified numerous bioactivities (He et al., 2017; Tao et al., 2020). Our research obtained 31 bioactive compounds of AB–DA based on the screening criteria and applied the main therapeutic compounds to anti-osteoporosis. Compounds of AB–DA such as baicalin, kaempferol, oleanol, quercetin, and sitosterol are widespread in herbal medicine, and many studies have shown their anti-osteoporosis activity. It has been reported that kaempferol could ameliorate the inhibitory effects of osteogenesis by activating JNK and p38 pathways in the glucocorticoid-induced and ovariectomy-induced osteoporosis model (Wong et al., 2019). Vakili et al. (2021) also constructed an OVX osteoporosis model treated with quercetin; results showed that it might modulate cell autophagy and apoptosis to alleviate osteoporosis; the potential mechanisms mainly involve Wnt, NF- $\kappa$ B, and MAPK cascades. Moreover, asperosaponin VI is the quality indicator of DA and could induce the differentiation of osteoblastic cells by increased expression of BMP2, promote osteogenesis and angiogenesis via regulating the OPG/RANKL signaling pathway, and inhibit the differentiation of osteoclast (Chen et al., 2022).  $\beta$ -Ecdysterone is an iconic ingredient of AB and could upregulate the activity of alkaline phosphatase in mesenchymal stem cells by modulating the expression of estrogen receptors (Gao et al., 2008) and could also inhibit apoptosis and autophagy induced by the dexamethasone of osteoblast cells *in vivo* and *in vitro* (Tang et al., 2018). This body of evidence suggests that the numerous bioactive compounds retrieved from AB–DA have certain anti-OP effects and partly confirm our findings; however, more accurate and in-depth study is still necessary to explore the AB–DA herb pair.

After screening the core therapeutic ingredients, a PPI network was constructed to determine core targets; our results showed that AKT1, CASP3, EGFR, ESR1, IGF1, MAPK1, and MAPK14 are important for AB–DA treatment of osteoporosis. It is widely reported that RAC- $\alpha$  serine/threonine-protein kinase (AKT1)

regulates a series of biological processes, including cell proliferation, growth, metabolism, and angiogenesis (Heron-Milhavet et al., 2011). A targeted knockdown AKT1 mouse model showed that AKT1 deficiency would induce osteoclast-osteogenesis disorder and diminish the RANKL (NF- $\kappa$ B ligand) and MCSF (macrophage colony-stimulating factor) receptors on multinucleated osteoclasts. This is evidence for AKT1 as an intermediary to regulate osteoblast and osteoclast differentiation (Mukherjee and Rotwein, 2012). Wang et al. (2022) also showed that modulating the expression of AKT1 could relieve osteoporosis. Epidermal growth factor receptor (EGFR) could activate the downstream of MEK-ERK, PI3K-AKT, and NF- $\kappa$ B signaling, transferring extracellular cues into cellular response (Runkle et al., 2016). It was reported that promoting the phosphorylation of EGFR and ERK1/2 could alleviate the apoptosis induced by H<sub>2</sub>O<sub>2</sub> of MC3T3-E1 (osteoblast cell) (Yang et al., 2019), exhibiting the potential targeted therapeutic value of EGFR on osteoporosis. Insulin-like growth factor I (IGF1) is the most abundant growth promotor in the bone matrix and also regulates glycogen synthesis in osteoblasts; it plays an important role in bone homeostasis maintenance and osteoblast differentiation by mediating the mTOR (mammalian target of rapamycin) signaling pathway (Xian et al., 2012). The results of molecular docking and molecular dynamics simulation also showed a stronger binding affinity between AB-DA compounds and IGF1. All the aforementioned evidence shows that the hub genes regulated by AB-DA compounds are important and meaningful in the treatment of osteoporosis.

The GO/KEGG pathway enrichment analysis of 83 potential therapeutic targets showed that AB-DA anti-osteoporosis mainly involved the MAPK signaling pathway, osteoclast differentiation, and the PI3K-Akt signaling pathway. As shown in Figure 4E, the colored target details of the MAPK pathway—ERK (MAPK1), JNK(MAPK8), and p38 (MAPK14)—are three classical cascades of the MAPK pathway, all of which were involved in the potential mechanism of AB-DA treating osteoporosis. ERK cascade mediated cell growth and differentiation via cytoskeletal rearrangement, and upregulating the phosphorylation of ERK (p-ERK) might promote osteoblast differentiation (Jing et al., 2019). The JNK and p38 cascades would be activated by extracellular stimulation, such as pro-inflammatory and physical stress. Lee et al. (2019) showed that downregulated osteoclast-related gene expression was associated with JNK cascade inhibition and that suppressing the p38 cascade would also relieve osteoporosis (Wang et al., 2018). Liu et al. (2022) indicated that vitexin could act against osteoporosis by promoting osteogenesis and angiogenesis in an ovariectomized rat model; the potential underlying mechanism might upregulate the PI3K-AKT cascade. Regulated PI3K-AKT signaling could also mediate the biological function of osteoclast (Jiang et al., 2022). Thus, our research mined the herb databases and screened therapeutic compounds of AB-DA with appropriate pharmacokinetic properties. In summary, the 31 therapeutic compounds have different targets and regulate different signaling pathways with a synergistic effect against osteoporosis, showing the complex molecular mechanisms with “multi-compound,” “multi-target,” and “multi-pathway.”

Furthermore, *in vitro* and *in vivo* experimental validation were both employed to explore the anti-osteoporotic effects of sitogluside

(also known as daucosterol) and its underlying molecular mechanism. Some research has indicated the regulation of both the osteoblast and osteoclast formation of AB-DA (He et al., 2017; Tao et al., 2020). In this present research, our findings show the potential effects of sitogluside with osteoblast differentiation and mineralization, but the effects on osteoclast were not significant; this was also verified by the collaborative therapeutic effects of AB-DA compounds. Moreover, q-PCR and Western blot analyses showed that sitogluside might upregulate the JNK cascade to promote osteogenics, such as Runx2, Osx, and OCN, and knockdown or block with the inhibitor could partly reduce its efficacy. Previous research has also reported that daucosterol could increase the p-JNK expression to exert an anti-prostate cancer effect (Gao et al., 2019). Huang et al. reported that the JNK kinase pathway with downstream OSX belonged to the non-canonical Smad-independent BMP signaling pathway to promote osteogenics. Hence, our study illustrated that a potential mechanism for sitogluside in the treatment of osteoporosis was to promote the JNK pathway and non-canonical BMP signaling to regulate downstream osteogenic genes. In addition, the OVX mouse model also showed anti-osteoporosis efficacy *in vivo* without observable toxicity. Our results therefore highlight the potential therapeutic value of sitogluside in the treatment of osteoporosis. However, further systematic and in-depth research is required to explore other bioactive compounds of the AB-DA combination.

There were some limitations to our research. First, it is based on the bioinformatic analysis of network pharmacology, molecular docking, and molecular dynamics simulation. The results illustrated the core therapeutic ingredients, core targets, and core signaling pathways of AB-DA in the treatment of osteoporosis, although we have preliminarily verified the anti-osteoporotic effect of sitogluside. Our findings thus need more wet experiments *in vitro* and *in vivo* for corroboration. The second limitation is that all the bioactive compounds were filtered by the ADME properties; the content of specific ingredients were not considered, and the toxicological information was ignored. Therefore, the next task for us is to comprehensively verify our network pharmacological findings with experiments and evaluate the biosafety of these ingredients for optimally utilizing AB-DA treatment of osteoporosis during clinical practice (Zou et al., 2021).

## 5 Conclusion

This study represents the first comprehensive investigation into the bioactive compounds of the AB-DA herb pair. Our findings illustrate that the primary therapeutic compounds responsible for treating osteoporosis are iridoids, steroids, and flavonoids. Additionally, we propose that the underlying mechanisms of action may involve targeting key core targets, including MAPK1, MAPK8, and MAPK14, to modulate the MAPK cascade (ERK/JNK/p38). Furthermore, *in vitro* and *in vivo* experiments have provided preliminary validation of the anti-osteoporotic effect of the most potent bioactive compound, sitogluside. Nevertheless, further in-depth experiments are needed to fully harness the therapeutic potential of AB-DA for treating osteoporosis in clinical practice.

## 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](#).

## Ethics statement

The animal study was approved by the Animal Care Committee of the Second Xiangya Hospital of Central South University. The study was conducted in accordance with the local legislation and institutional requirements.

## Author contributions

ZO designed this study; CX wrote the manuscript; TL collected the data and performed statistical analysis; WL and TT revised the manuscript; XG created all tables, figures, and prepared the [Supplementary Material](#). All authors contributed to the article and approved the submitted version.

## Funding

This research was funded by the National Natural Science Foundation of China (No. 82371600), the Natural Science Foundation of Hunan Province (No. 2021JJ40810 and 2023JJ40833), the Natural Science Foundation of Hunan Province of Outstanding Young Scholars (Grant No. 2021JJ20086), the Scientific Research Projects of the Health Commission of Hunan Province (project number B202304077119), the Natural Science Foundation of Hunan Province (grant number 2021JJ30930), and the Fundamental Research Funds for the Central Universities of Central South University (grant number 2023ZZTS0558).

## References

- Anthamatten, A., and Parish, A. (2019). Clinical update on osteoporosis. *J. Midwifery Womens Health* 64, 265–275. doi:10.1111/jmwh.12954
- Barshir, R., Fishilevich, S., Iny-Stein, T., Zelig, O., Mazor, Y., Guan-Golan, Y., et al. (2021). GeneCaRNA: A comprehensive gene-centric database of human non-coding RNAs in the GeneCards suite. *J. Mol. Biol.* 433, 166913. doi:10.1016/j.jmb.2021.166913
- Chen, C. Y. (2011). TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS One* 6, e15939. doi:10.1371/journal.pone.0015939
- Chen, F., Liang, Q., Mao, L., Yin, Y., Zhang, L., Li, C., et al. (2022). Synergy effects of Asperosaponin VI and bioactive factor BMP-2 on osteogenesis and anti-osteoclastogenesis. *Bioact. Mater* 10, 335–344. doi:10.1016/j.bioactmat.2021.09.001
- Chu, H., Moon, S., Park, J., Bak, S., Ko, Y., and Youn, B. Y. (2022). The use of artificial intelligence in complementary and alternative medicine: A systematic scoping review. *Front. Pharmacol.* 13, 826044. doi:10.3389/fphar.2022.826044
- Compston, J. E., McClung, M. R., and Leslie, W. D. (2019). Osteoporosis. *Lancet* 393, 364–376. doi:10.1016/S0140-6736(18)32112-3
- 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
- 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
- Ensrud, K. E., and Crandall, C. J. (2017). Osteoporosis. *Ann. Intern. Med.* 167, I17–I32. doi:10.7326/aitc201708010
- Feng, S., Wang, T., Fan, L., An, X., Ding, X., Wang, M., et al. (2022). Exploring the potential therapeutic effect of Eucommia ulmoides-Dipsaci Radix herbal pair on osteoporosis based on network pharmacology and molecular docking technology. *RSC Adv.* 12, 2181–2195. doi:10.1039/d1ra05799e
- Gao, L., Cai, G., and Shi, X. (2008). Beta-ecdysterone induces osteogenic differentiation in mouse mesenchymal stem cells and relieves osteoporosis. *Biol. Pharm. Bull.* 31, 2245–2249. doi:10.1248/bpb.31.2245
- Gao, P., Huang, X., Liao, T., Li, G., Yu, X., You, Y., et al. (2019). Daucosterol induces autophagic-dependent apoptosis in prostate cancer via JNK activation. *Biosci. Trends* 13, 160–167. doi:10.5582/bst.2018.01293
- Gao, Y., Ji, W., Lu, M., Wang, Z., Jia, X., Wang, D., et al. (2022). Systemic pharmacological verification of Guizhi Fuling decoction in treating endometriosis-associated pain. *J. Ethnopharmacol.* 297, 115540. doi:10.1016/j.jep.2022.115540
- Grewé, J. M., Knapstein, P. R., Donat, A., Jiang, S., Smit, D. J., Xie, W., et al. (2022). The role of sphingosine-1-phosphate in bone remodeling and osteoporosis. *Bone Res.* 10, 34. doi:10.1038/s41413-022-00205-0

## Acknowledgments

The authors sincerely appreciate the TCMSP, TCM@Taiwan, BATMAN-TCM, TCM-ID, PharmMapper, Swiss Institute of Bioinformatics, UniProt, GeneCards, DisGeNET, OMIM, STRING, and Metascape databases and Cytoscape and Discovery Studio 2019 software for data collection, statistical analysis, and figure plotting.

## 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/fphar.2023.1242194/full#supplementary-material>

### SUPPLEMENTARY TABLE S1

Detailed information of AB–DA herb pair bioactive compounds.

### SUPPLEMENTARY TABLE S2

Protein name and Entrez ID of all targets.

- Grosdidier, A., Zoete, V., and Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic acids Res.* 39, W270–W277. doi:10.1093/nar/gkr366
- He, C. C., Hui, R. R., Tezuka, Y., Kadota, S., and Li, J. X. (2010). Osteoprotective effect of extract from *Achyranthes bidentata* in ovariectomized rats. *J. Ethnopharmacol.* 127, 229–234. doi:10.1016/j.jep.2009.11.016
- He, X., Wang, X., Fang, J., Chang, Y., Ning, N., Guo, H., et al. (2017). The genus *Achyranthes*: A review on traditional uses, phytochemistry, and pharmacological activities. *J. Ethnopharmacol.* 203, 260–278. doi:10.1016/j.jep.2017.03.035
- Heron-Milhavet, L., Khouya, N., Fernandez, A., and Lamb, N. J. (2011). Akt1 and Akt2: Differentiating the action. *Histol. Histopathol.* 26, 651–662. doi:10.14670/HH-26.651
- Holzthüter, K., and Geertsma, E. R. (2022). Uniport, not proton-symport, in a non-mammalian SLC23 transporter. *J. Mol. Biol.* 434, 167393. doi:10.1016/j.jmb.2021.167393
- Hu, Y., Wu, Y., Jiang, C., Wang, Z., Shen, C., Zhu, Z., et al. (2022). Investigative on the molecular mechanism of licorice flavonoids anti-melanoma by network pharmacology, 3D/2D-QSAR, molecular docking, and molecular dynamics simulation. *Front. Chem.* 10, 843970. doi:10.3389/fchem.2022.843970
- Huang, A. Y., Xiong, Z., Liu, K., Chang, Y., Shu, L., Gao, G., et al. (2022). Identification of kaempferol as an OSX upregulator by network pharmacology-based analysis of qianggu Capsule for osteoporosis. *Front. Pharmacol.* 13, 1011561. doi:10.3389/fphar.2022.1011561
- Huang, Y., Li, N., Wan, J. B., Zhang, D., and Yan, C. (2015). Structural characterization and antioxidant activity of a novel heteropolysaccharide from the submerged fermentation mycelia of *Ganoderma capense*. *Carbohydr. Polym.* 134, 752–760. doi:10.1016/j.carbpol.2015.08.067
- Jiang, M., Liu, R., Liu, L., Kot, A., Liu, X., Xiao, W., et al. (2020). Identification of osteogenic progenitor cell-targeted peptides that augment bone formation. *Nat. Commun.* 11, 4278. doi:10.1038/s41467-020-17417-9
- Jiang, T., Gu, H., and Wei, J. (2022). Echinacoside inhibits osteoclast function by down-regulating PI3K/akt/C-fos to alleviate osteolysis caused by periprosthetic joint infection. *Front. Pharmacol.* 13, 930053. doi:10.3389/fphar.2022.930053
- Jiang, Y., Zhang, Y., Chen, W., Liu, C., Li, X., Sun, D., et al. (2014). *Achyranthes bidentata* extract exerts osteoprotective effects on steroid-induced osteonecrosis of the femoral head in rats by regulating RANKL/RANK/OPG signaling. *J. Transl. Med.* 12, 334. doi:10.1186/s12967-014-0334-7
- Jin, L. Q., Zheng, Z. J., Peng, Y., Li, W. X., Chen, X. M., and Lu, J. X. (2007). Opposite effects on tumor growth depending on dose of *Achyranthes bidentata* polysaccharides in C57BL/6 mice. *Int. Immunopharmacol.* 7, 568–577. doi:10.1016/j.intimp.2006.12.009
- Jing, Z., Wang, C., Yang, Q., Wei, X., Jin, Y., Meng, Q., et al. (2019). Luteolin attenuates glucocorticoid-induced osteoporosis by regulating ERK/Lrp-5/GSK-3 $\beta$  signaling pathway *in vivo* and *in vitro*. *J. Cell Physiol.* 234, 4472–4490. doi:10.1002/jcp.27252
- Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci. a Publ. Protein Soc.* 29, 28–35. doi:10.1002/pro.3711
- Kang, H., Tang, K., Liu, Q., Sun, Y., Huang, Q., Zhu, R., et al. (2013). HIM-herbal ingredients *in-vivo* metabolism database. *J. cheminformatics* 5, 28. doi:10.1186/1758-2946-5-28
- Liu, Z., Guo, F., Wang, Y., Li, C., Zhang, X., Li, H., et al. (2016). BATMAN-TCM: A bioinformatics analysis tool for molecular mechanism of traditional Chinese medicine. *Sci. Rep.* 6, 21146. doi:10.1038/srep21146
- Lee, K. M., Lee, C. Y., Zhang, G., Lyu, A., and Yue, K. K. M. (2019). Methylglyoxal activates osteoclasts through JNK pathway leading to osteoporosis. *Chem. Biol. Interact.* 308, 147–154. doi:10.1016/j.cbi.2019.05.026
- Li, T., Jiang, G., Hu, X., Yang, D., Tan, T., Gao, Z., et al. (2021). Punicalin attenuates breast cancer-associated osteolysis by inhibiting the NF- $\kappa$ B signaling pathway of osteoclasts. *Front. Pharmacol.* 12, 789552. doi:10.3389/fphar.2021.789552
- Li, Z., Ying, B., Liu, X., Zhang, X., and Yu, H. (2012). An examination of the OMIM database for associating mutation to a consensus reference sequence. *Protein & Cell* 3, 198–203. doi:10.1007/s13238-012-2037-2
- Liang, B., Burley, G., Lin, S., and Shi, Y. C. (2022). Osteoporosis pathogenesis and treatment: Existing and emerging avenues. *Cell Mol. Biol. Lett.* 27, 72. doi:10.1186/s11658-022-00371-3
- Liu, K., Liu, Y., Xu, Y., Nandakumar, K. S., Tan, H., He, C., et al. (2019). Asperosaponin VI protects against bone destructions in collagen induced arthritis by inhibiting osteoclastogenesis. *Phytomedicine Int. J. phytotherapy Phytopharm.* 63, 153006. doi:10.1016/j.phymed.2019.153006
- Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J., et al. (2010). PharmMapper server: A web server for potential drug target identification using pharmacophore mapping approach. *Nucleic acids Res.* 38, W609–W614. doi:10.1093/nar/gkq300
- Liu, Y., Xue, Q., Li, A., Li, K., and Qin, X. (2020). Mechanisms exploration of herbal pair of Huangqi-DanShen on cerebral ischemia based on metabolomics and network pharmacology. *J. Ethnopharmacol.* 253, 112688. doi:10.1016/j.jep.2020.112688
- Liu, Y., Zhu, S., Liu, J., Chen, Y., Zhong, S., Lian, D., et al. (2022). Vitexin regulates angiogenesis and osteogenesis in ovariectomy-induced osteoporosis of rats via the VDR/PI3K/AKT/eNOS signaling pathway. *J. Agric. Food Chem.* 71, 546–556. doi:10.1021/acs.jafc.2c07005
- McDonald, M. M., Khoo, W. H., Ng, P. Y., Xiao, Y., Zamerli, J., Thatcher, P., et al. (2021). Osteoclasts recycle via osteomorphs during RANKL-stimulated bone resorption. *Cell* 184, 1940–1947.e13. doi:10.1016/j.cell.2021.03.010
- Meng, L., Yuan, L., Ni, J., Fang, M., Guo, S., Cai, H., et al. (2021). Mir24-2-5p suppresses the osteogenic differentiation with Gna3 inhibition presenting a direct target via inactivating JNK-p38 MAPK signaling axis. *Int. J. Biol. Sci.* 17, 4238–4253. doi:10.7150/ijbs.60536
- Mukherjee, A., and Rotwein, P. (2012). Selective signaling by Akt1 controls osteoblast differentiation and osteoblast-mediated osteoclast development. *Mol. Cell Biol.* 32, 490–500. doi:10.1128/MCB.06361-11
- Mukwaya, E., Xu, F., Wong, M. S., and Zhang, Y. (2014). Chinese herbal medicine for bone health. *Pharm. Biol.* 52, 1223–1228. doi:10.3109/13880209.2014.884606
- Nakamura, R., Ogawa, S., Takahashi, Y., and Fujishiro, T. (2022). Cycloserine enantiomers inhibit PLP-dependent cysteine desulfurase SufS via distinct mechanisms. *FEBS J.* 289, 5947–5970. doi:10.1111/febs.16455
- NIH Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy (2001). Osteoporosis prevention, diagnosis, and therapy. *Jama* 285, 785–795. doi:10.1001/jama.285.6.785
- Pagadala, N. S., Syed, K., and Tuszyński, J. (2017). Software for molecular docking: A review. *Biophys. Rev.* 9, 91–102. doi:10.1007/s12551-016-0247-1
- Peng, Z., Xu, R., and You, Q. (2022). Role of traditional Chinese medicine in bone regeneration and osteoporosis. *Front. Bioeng. Biotechnol.* 10, 911326. doi:10.3389/fbioe.2022.911326
- Piñero, J., Ramírez-Anguita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., et al. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic acids Res.* 48, D845–d855. doi:10.1093/nar/gkz1021
- 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. cheminformatics* 6, 13. doi:10.1186/1758-2946-6-13
- Runkle, K. B., Kharbanda, A., Stypulkowski, E., Cao, X. J., Wang, W., Garcia, B. A., et al. (2016). Inhibition of DHHC20-mediated EGFR palmitoylation creates a dependence on EGFR signaling. *Mol. Cell* 62, 385–396. doi:10.1016/j.molcel.2016.04.003
- Shuai, Y., Jiang, Z., Yuan, Q., Tu, S., and Zeng, F. (2020). Deciphering the underlying mechanism of eucommiae cortex against osteoporotic fracture by network pharmacology. *eCAM* 2020, 7049812. doi:10.1155/2020/7049812
- Suh, K. S., Lee, Y. S., and Choi, E. M. (2014). The protective effects of *Achyranthes bidentata* root extract on the antitumor A induced damage of osteoblastic MC3T3-E1 cells. *Cytotechnology* 66, 925–935. doi:10.1007/s10616-013-9645-4
- Sultana, N., Chung, H. J., Emon, N. U., Alam, S., Taki, M. T. I., Rudra, S., et al. (2022). Biological functions of *Dillenia pentagyna* roxb. Against pain, inflammation, fever, diarrhea, and thrombosis: Evidenced from *in vitro*, *in vivo*, and molecular docking study. *Front. Nutr.* 9, 911274. doi:10.3389/fnut.2022.911274
- Suvarna, V., Sarkar, M., Chaubey, P., Khan, T., Sherje, A., Patel, K., et al. (2018). Bone health and natural products- an insight. *Front. Pharmacol.* 9, 981. doi:10.3389/fphar.2018.00981
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids Res.* 49, D605–d612. doi:10.1093/nar/gkaa1074
- Tang, Y. H., Yue, Z. S., Xin, D. W., Zeng, L. R., Xiong, Z. F., Hu, Z. Q., et al. (2018).  $\beta$ -Ecdysterone promotes autophagy and inhibits apoptosis in osteoporotic rats. *Mol. Med. Rep.* 17, 1591–1598. doi:10.3892/mmr.2017.8053
- Tao, Y., Chen, L., and Yan, J. (2020). Traditional uses, processing methods, phytochemistry, pharmacology and quality control of *dipsacus asper* wall. Ex C.B. Clarke: A review. *J. Ethnopharmacol.* 258, 112912. doi:10.1016/j.jep.2020.112912
- Vakili, S., Zal, F., Mostafavi-Pour, Z., Savardashtaki, A., and Koohpeyma, F. (2021). Quercetin and vitamin E alleviate ovariectomy-induced osteoporosis by modulating autophagy and apoptosis in rat bone cells. *J. Cell Physiol.* 236, 3495–3509. doi:10.1002/jcp.30087
- Wang, W. D., Kang, W. B., Zhou, X. Q., Yao, G. F., and Wang, X. J. (2018). Mitochondrial protein OPA mediates osteoporosis induced by radiation through the P38 signaling pathway. *Eur. Rev. Med. Pharmacol. Sci.* 22, 8091–8097. doi:10.26355/eurrev\_201812\_16499



- Wang, Y., Liu, L., Qu, Z., Wang, D., Huang, W., Kong, L., et al. (2022). Tanshinone ameliorates glucocorticoid-induced bone loss via activation of AKT1 signaling pathway. *Front. Cell Dev. Biol.* 10, 878433. doi:10.3389/fcell.2022.878433
- Wong, S. K., Chin, K. Y., and Ima-Nirwana, S. (2019). The osteoprotective effects of kaempferol: The evidence from *in vivo* and *in vitro* studies. *Drug Des. Devel Ther.* 13, 3497–3514. doi:10.2147/DDDT.S227738
- Wu, H., Lv, Y., Wei, F., Li, C., Ge, W., and Du, W. (2022). Comparative analysis of anti-osteoporosis efficacy in *Radix Dipsaci* before and after processing with salt based on spectrum-effect relationship. *J. Pharm. Biomed. analysis* 221, 115078. doi:10.1016/j.jpba.2022.115078
- Xian, L., Wu, X., Pang, L., Lou, M., Rosen, C. J., Qiu, T., et al. (2012). Matrix IGF-1 maintains bone mass by activation of mTOR in mesenchymal stem cells. *Nat. Med.* 18, 1095–1101. doi:10.1038/nm.2793
- Yan, C., Zhang, S., Wang, C., and Zhang, Q. (2019). A fructooligosaccharide from *Achyranthes bidentata* inhibits osteoporosis by stimulating bone formation. *Carbohydr. Polym.* 210, 110–118. doi:10.1016/j.carbpol.2019.01.026
- Yang, F., Lin, Z. W., Huang, T. Y., Chen, T. T., Cui, J., Li, M. Y., et al. (2019). Ligustilide, a major bioactive component of *Angelica sinensis*, promotes bone formation via the GPR30/EGFR pathway. *Sci. Rep.* 9, 6991. doi:10.1038/s41598-019-43518-7
- Ye, X. W., Wang, H. L., Cheng, S. Q., Xia, L. J., Xu, X. F., and Li, X. R. (2022). Network pharmacology-based strategy to investigate the pharmacologic mechanisms of coptidis rhizoma for the treatment of alzheimer's disease. *Front. aging Neurosci.* 14, 890046. doi:10.3389/fnagi.2022.890046
- Yu, J. H., Yu, Z. P., Wang, Y. Y., Bao, J., Zhu, K. K., Yuan, T., et al. (2019). Triterpenoids and triterpenoid saponins from *Dipsacus asper* and their cytotoxic and antibacterial activities. *Phytochemistry* 162, 241–249. doi:10.1016/j.phytochem.2019.03.028
- Zhang, M., Wang, Y., Zhang, Q., Wang, C., Zhang, D., Wan, J. B., et al. (2018a). UPLC/Q-TOF-MS-based metabolomics study of the anti-osteoporosis effects of *Achyranthes bidentata* polysaccharides in ovariectomized rats. *Int. J. Biol. Macromol.* 112, 433–441. doi:10.1016/j.ijbiomac.2018.01.204
- Zhang, R., Hu, S. J., Li, C., Zhang, F., Gan, H. Q., and Mei, Q. B. (2012). *Achyranthes bidentata* root extract prevent OVX-induced osteoporosis in rats. *J. Ethnopharmacol.* 139, 12–18. doi:10.1016/j.jep.2011.05.034
- Zhang, S., Zhang, Q., Zhang, D., Wang, C., and Yan, C. (2018b). Anti-osteoporosis activity of a novel *Achyranthes bidentata* polysaccharide via stimulating bone formation. *Carbohydr. Polym.* 184, 288–298. doi:10.1016/j.carbpol.2017.12.070
- Zhang, W., Xue, K., Gao, Y., Huai, Y., Wang, W., Miao, Z., et al. (2019). Systems pharmacology dissection of action mechanisms of *Dipsaci Radix* for osteoporosis. *Life Sci.* 235, 116820. doi:10.1016/j.lfs.2019.116820
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10, 1523. doi:10.1038/s41467-019-09234-6
- Zou, M. L., Chen, Z. H., Teng, Y. Y., Liu, S. Y., Jia, Y., Zhang, K. W., et al. (2021). The Smad dependent TGF- $\beta$  and BMP signaling pathway in bone remodeling and therapies. *Front. Mol. Biosci.* 8, 593310. doi:10.3389/fmolb.2021.593310
- Zou, X., Holmes, E., Nicholson, J. K., and Loo, R. L. (2016). Automatic spectroscopic data categorization by clustering analysis (ASCLAN): A data-driven approach for distinguishing discriminatory metabolites for phenotypic subclasses. *Anal. Chem.* 88, 5670–5679. doi:10.1021/acs.analchem.5b04020





## OPEN ACCESS

## EDITED BY

Concepción González-Bello,  
Universidade de Santiago de Compostela,  
Spain

## REVIEWED BY

Elena Cichero,  
University of Genoa, Italy  
Pranav Shah,  
National Center for Advancing  
Translational Sciences (NIH),  
United States

## \*CORRESPONDENCE

Fernanda I. Saldívar-González,  
✉ fer.saldivarg@gmail.com  
José L. Medina-Franco,  
✉ medinajl@unam.mx

RECEIVED 12 August 2023

ACCEPTED 16 October 2023

PUBLISHED 02 November 2023

## CITATION

Saldívar-González FI,  
Navarrete-Vázquez G and  
Medina-Franco JL (2023), Design of a  
multi-target focused library for  
antidiabetic targets using a  
comprehensive set of chemical  
transformation rules.  
*Front. Pharmacol.* 14:1276444.  
doi: 10.3389/fphar.2023.1276444

## COPYRIGHT

© 2023 Saldívar-González, Navarrete-Vázquez and Medina-Franco. 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.

# Design of a multi-target focused library for antidiabetic targets using a comprehensive set of chemical transformation rules

Fernanda I. Saldívar-González<sup>1\*</sup>, Gabriel Navarrete-Vázquez<sup>2</sup> and José L. Medina-Franco<sup>1\*</sup>

<sup>1</sup>Department of Pharmacy, DIFACQUIM Research Group, School of Chemistry, Universidad Nacional Autónoma de México, Mexico City, Mexico, <sup>2</sup>Facultad de Farmacia, Universidad Autónoma Del Estado de Morelos, Cuernavaca, Morelos, Mexico

Virtual small molecule libraries are valuable resources for identifying bioactive compounds in virtual screening campaigns and improving the quality of libraries in terms of physicochemical properties, complexity, and structural diversity. In this context, the computational-aided design of libraries focused against antidiabetic targets can provide novel alternatives for treating type II diabetes mellitus (T2DM). In this work, we integrated the information generated to date on compounds with antidiabetic activity, advances in computational methods, and knowledge of chemical transformations available in the literature to design multi-target compound libraries focused on T2DM. We evaluated the novelty and diversity of the newly generated library by comparing it with antidiabetic compounds approved for clinical use, natural products, and multi-target compounds tested *in vivo* in experimental antidiabetic models. The designed libraries are freely available and are a valuable starting point for drug design, chemical synthesis, and biological evaluation or further computational filtering. Also, the compendium of 280 transformation rules identified in a medicinal chemistry context is made available in the linear notation SMIRKS for use in other chemical library enumeration or hit optimization approaches.

## KEYWORDS

computer-aided drug design, diabetes, target-focused libraries, library enumeration, ligandbased design, multi-target design, transformation rules, reaction informatics

**Abbreviations:** ADME, absorption, distribution, metabolism, and excretion; AR, aldose reductase; DM, diabetes mellitus; FFA1, Free fatty acid receptor 1; GPR40, G-protein-coupled receptor 40; MetS, metabolic syndrome; MOE, Molecular Operating Environment; MW, molecular weight; PCA, principal component analysis; PMI, principal moments of inertia; PPARs, peroxisome proliferator-activated receptors; PTP1B, protein tyrosine phosphatase 1B; QED, quantitative estimate of drug-likeness; RB, rotatable bonds; T2DM, Type II diabetes mellitus; TMAP, tree map; TPSA, topological polar surface area; t-SNE, t-distributed stochastic neighbor embedding; SMILES, Simplified Molecular Input Line Entry Specification, SMIRKS: line notation for specifying reaction transforms; SUR, Sulfonylurea receptor.

## 1 Introduction

Type 2 diabetes mellitus (T2DM) is a metabolic disorder characterized by hyperglycemia caused by defects in insulin secretion and/or action due to a complex network of pathological conditions (Galicia-García et al., 2020). Currently, T2DM is one of the diseases with the highest socio-health impact and prevalence worldwide. Although pharmacotherapeutic options include different mechanisms of action, they are limited by side effects and lack of blood glucose control in the diabetic population (Shah et al., 2021). Another problem is that patients with this disease are prone to polypharmacy, which increases the risk of adverse effects and makes it difficult for patients to adhere to their treatment and receive proper follow-up from healthcare professionals (Dobrică et al., 2019). For this reason, new biological targets have been explored in multi-target approaches (Makhoba et al., 2020). Similarly, virtual libraries focused on single therapeutic targets have been developed using various computational approaches and their application in multi-target approaches is emerging.

Recent advances in computational methods and the incorporation of synthetic expert knowledge have inspired research groups to develop *de novo* and “make-on-demand” chemical libraries (Walters, 2019). Several companies use the so-called “novel molecular matter” in early-phase drug discovery (Korn et al., 2023). Specifically in the design of antidiabetic compounds, Otava released the chemical structures of ten libraries focused on DM-related targets designed under ligand- and structure-based

approaches or combinations of both (OTAVACHemicals, Ltd. - synthetic organic compounds for research and drug discovery, n.d.). ChemDiv developed methods for screening diverse and highly specialized focused compounds. Recently, ChemDiv released an “Annotated space library” with more than 18,000 chemical compounds covering 38 validated targets (including targets for T2DM) across 900 drugs launched in the last 10 years. Academic groups have also generated virtual libraries focused on T2DM. For example, Chen et al. used a generative method to design compounds targeted for GPR40 (Chen et al., 2021). To our knowledge, no *in silico* multi-target libraries have been disclosed for T2DM. However, there is published information on active compounds and pharmacophore models that can guide the design of multi-target compounds for T2DM (Artasensi et al., 2020; Lillich et al., 2020; Tassopoulou et al., 2022).

Figure 1 shows the structures of pharmacophores and multi-target compounds studied *in vivo* models for T2DM and metabolic syndrome (MetS). For example, dual peroxisome proliferator-activated receptor (PPAR)  $\alpha/\gamma$  agonists can improve insulin sensitivity and reduce triglyceride and blood glucose levels without the PPAR $\gamma$ -related weight gain since the latter effect is balanced by PPAR $\alpha$  agonistic activity (Balakumar et al., 2019). Attempts towards developing dual agonists for PPAR  $\alpha/\gamma$  yielded promising molecules that have reached clinical trials. However, many of these compounds have failed due to significant side effects. Compounds such as 1 (MHY908) and 2 (LT175) continue to be explored, as they have been shown to have

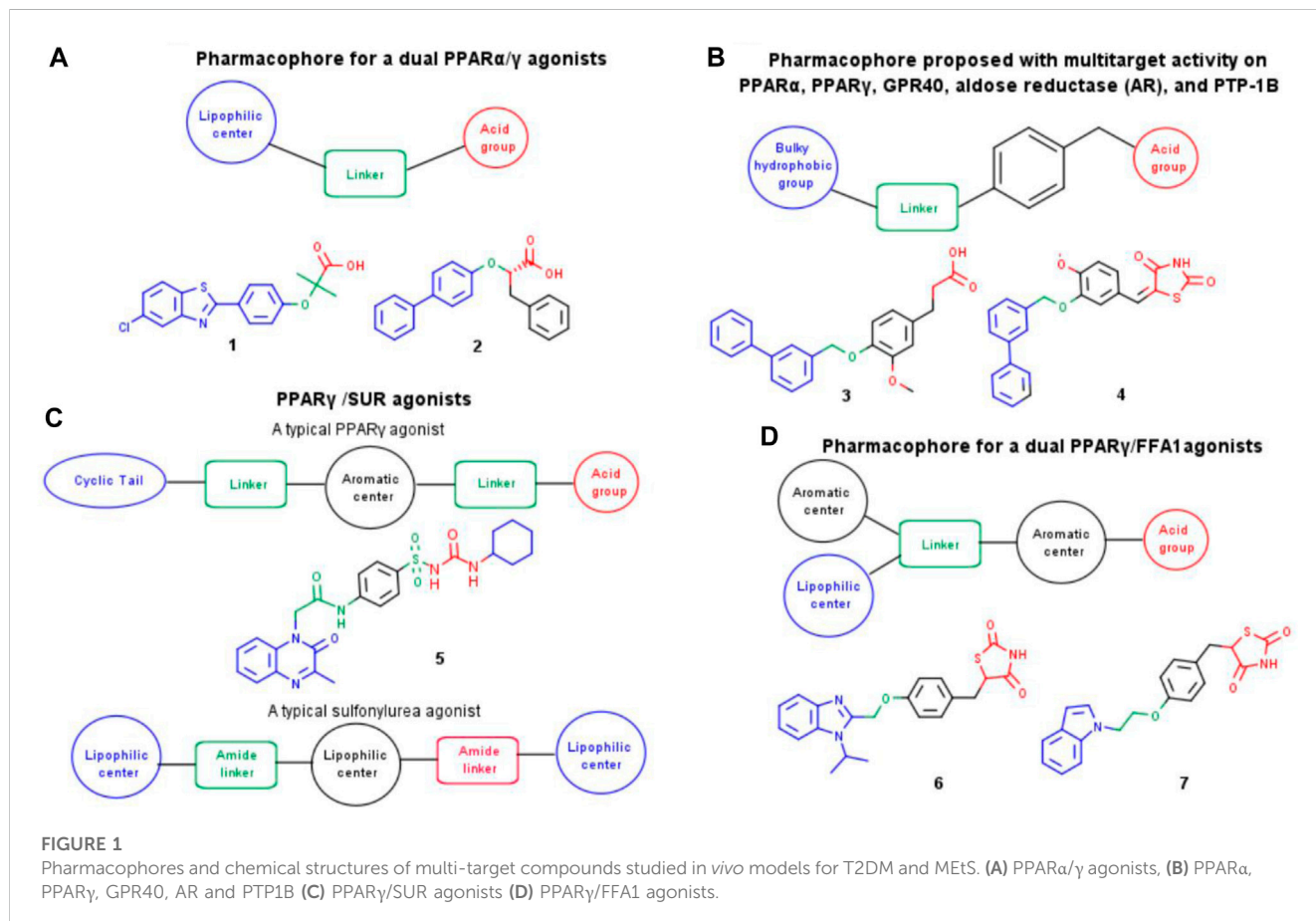


TABLE 1 Multi-target ligands studied *in vivo* models for type II diabetes mellitus.

Targets	Lead compounds	Implications in T2DM and MetS/Outcome	Reference
GLUT4, PPAR- $\alpha$ , PPAR- $\gamma$ , adiponectin	1	Antihyperglycemic, antidiabetic, and antidyslipidemic effects	Estrada-Soto et al. (2022)
PPAR $\alpha$ , PPAR $\gamma$ , GPR40*, AR, PTP1B	2	Antihyperglycemic and antidiabetic effects. Attractive to prevent or delay the development of diabetic complications	Colín-Lozano et al. (2018), Domínguez-Mendoza et al. (2021)
PPAR $\gamma$ , GLUT-4	3	Antihyperglycemic and antidiabetic effects	Gutiérrez-Hernández et al. (2019)
PPAR $\alpha$ , PPAR $\gamma$ , FATP-1, GLUT-4, PTP1B	4	Antihyperglycemic and antidiabetic effects	Herrera-Rueda et al. (2018)
sEH,PPAR $\gamma$	2	Antidiabetic, cardioprotective, renoprotective, hypotensive effects	Blöcher et al. (2016), Hye Khan et al. (2018)
DPP-4,GPR119	2	Antidiabetic, glucose homeostasis effects	Huan et al. (2017), Fang et al. (2020)
PPAR- $\alpha$ , $\gamma$	21	Antidiabetic and antidyslipidemic effects (13 PPAR $\alpha$ / $\gamma$ dual agonist compounds have reached clinical trials or the market)	Balakumar et al., 2007 (2019), Ammazalorso et al. (2019)
PPAR- $\alpha$ ,d	3	Antidiabetic, antidyslipidemic and anti-fatty liver effects	Hanf et al. (2014), Ren et al. (2020), Liu et al. (2021)
PPAR-d, $\gamma$	1	Antihyperglycemic and anti-fatty liver effects	Li et al. (2021)
PPAR- $\alpha$ ,d, $\gamma$	6	Antidiabetic and antidyslipidemic effects. Therapeutic potential for nonalcoholic steatohepatitis patients	Mahindroo et al. (2005), He et al. (2012), Boubia et al. (2018)
PPAR $\gamma$ , AT1	6	Antidiabetic and antihypertensive effects	Benson et al. (2004), Casimiro-Garcia et al. (2011), 2013; Lamotte et al. (2014), Choung et al. (2018)
PPAR $\gamma$ , GK	7	Antihyperglycemic, antidiabetic, improves insulin resistance and sensitize muscle cells to insulin response	Song et al. (2011), Li et al. (2014), Lei et al. (2015)
PPAR $\gamma$ , SUR	10	Improve insulin sensitivity and stimulate insulin secretion at the same time	Ibrahim et al., 2017a, 2017b
FFA1*, PPARd	5	Antidiabetic and anti-fatty liver effects	Li et al., 2019a, 2019b (2020), Hu et al. (2020), Zhou et al. (2022)
FFA1, PPAR $\gamma$	4	Antidiabetic and antihyperlipidemic effects	Darwish et al. (2018), Hidalgo-Figueroa et al. (2021)
FFA1, PPAR $\gamma$ , PPARd	1	Antidiabetic and antihyperglycemic effects	Li et al. (2018)
PPAR $\gamma$ , PTP1B	4	Antihyperglycemic and antiobesity effects	Bhattarai et al., 2009 (2010), Otake et al., 2012 (2015)
PPAR $\alpha$ / $\gamma$ /PTP1B	2	Antidiabetic, antidyslipidemic and antiobesity effects	Otake et al. (2011)
PARP-1 - AR	2	Nephroprotective effect and antioxidant potential	Chadha and Silakari (2017), Kumar et al. (2022)
SGLT1-SGLT2	5	Antihyperglycemic and antiobesity effects	Lapuerta et al. (2015), Kuo et al. (2018), Xu et al., 2018 (2020), He et al. (2019)

AT1, angiotensin II type 1 receptor; AR, aldose reductase; DPP-4, dipeptidyl peptidase-4; FATP-1, fatty acid transport protein 1; FFA1, free fatty acid receptor 1; GK,glucokinase; GLUT4, glucose transporter type 4; GPR119, G protein-coupled receptor 119; GPR-40, G-protein-coupled receptor 40; PARP-1, poly (ADP-ribose)polymerase-1; PPARs, peroxisome proliferator-activated receptors; PTP1B, protein tyrosine phosphatase 1B; SGLTs; ; sEH, soluble epoxide hydrolase; SUR, sulfonylurea receptor.

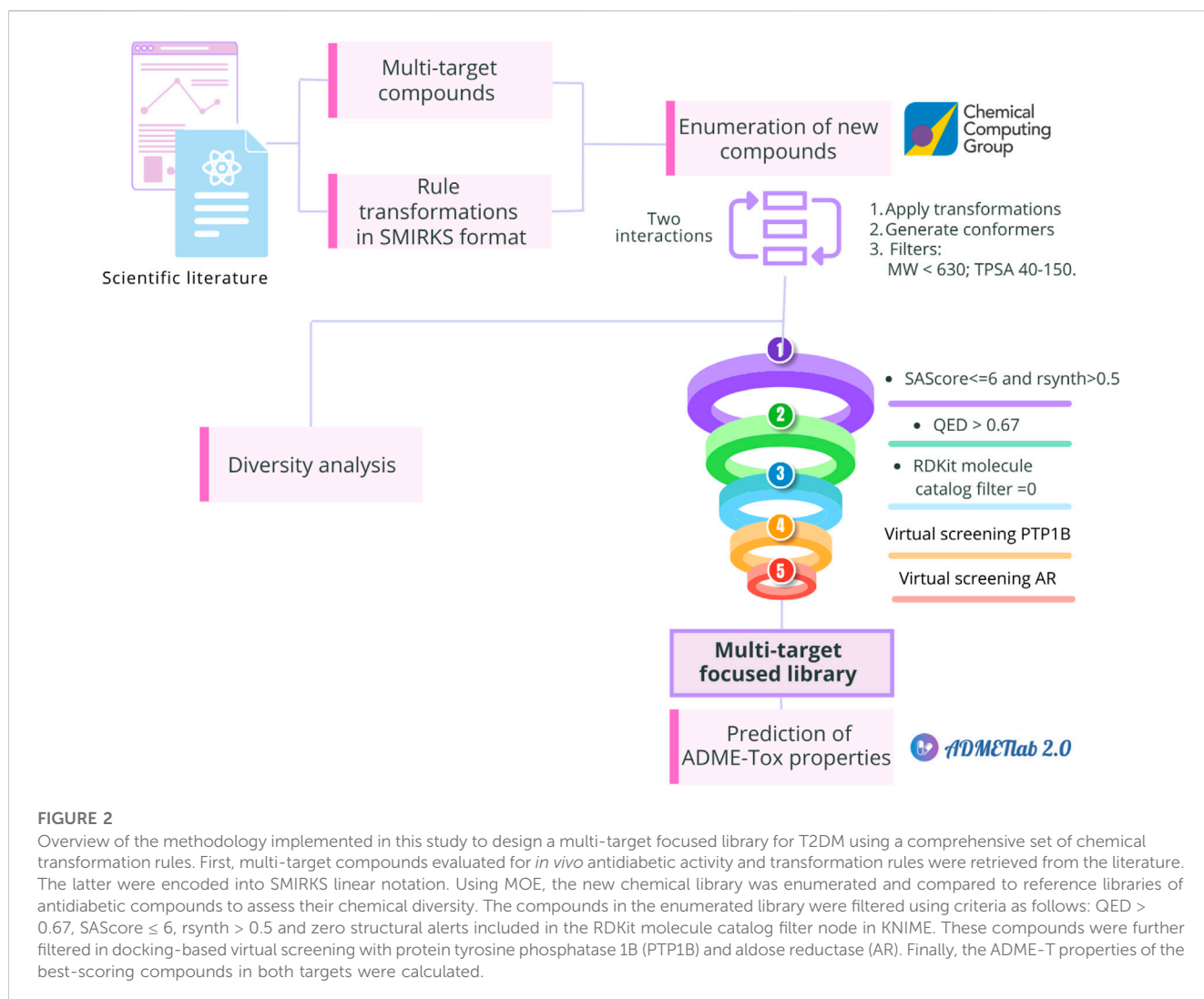
\*GPR40 is also known as FFA1.

beneficial effects on blood glucose and insulin resistance in animal models of T2DM (Gilardi et al., 2014; Park et al., 2016).

The development of multi-target compounds for GPR40, PTP1B, AR, PPAR $\alpha$ , and PPAR $\gamma$  may provide additional therapeutic benefits in preventing or delaying the development of diabetic complications (Domínguez-Mendoza et al., 2021). Navarrete's group designed compounds 3 and 4 (Figure 1) that showed robust *in vivo* antihyperglycemic activity (Domínguez-Mendoza et al., 2021). Dual PPAR $\gamma$ /SUR agonists, such as compound 5 (Ibrahim et al., 2017a), can improve insulin sensitivity and stimulate insulin secretion simultaneously, making them an attractive therapeutic option for patients with T2DM who have insulin resistance and decreased insulin secretion. Another

combination that has been explored is that of PPAR $\gamma$  and FFA1, also known as GPR40. Dual PPAR $\gamma$ /FFA1 agonists (compounds 6 and 7, Darwish et al., 2018) can improve insulin sensitivity, stimulate insulin secretion, and lower blood glucose levels.

Table 1 shows multi-target compounds for T2DM based on pharmacophores and virtual screening approaches that have been reported and tested *in vivo* and can guide the design of new chemical libraries. The table includes information on the combination of targets, the number of lead compounds that hit all targets, and the implications or outcome in T2DM and MetS. The table indicated that there are ninety-one multi-target compounds with *in vivo* antidiabetic activity. These have been tested in a total of twenty target combinations. The most studied biological target in multi-target

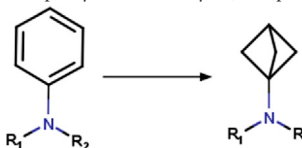
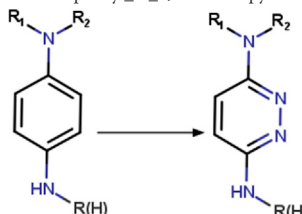
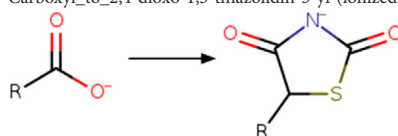
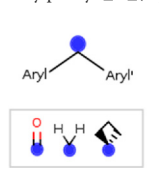
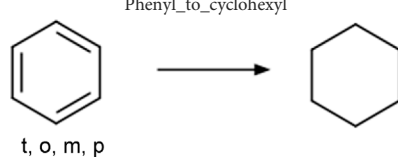
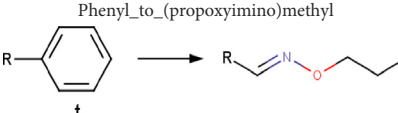


approaches is PPAR $\gamma$ . The most successful target combination is PPAR $\alpha/\gamma$  as several compounds directed to these two targets are in clinical studies such as ragaglitazar, imiglitazar, muraglitazar, tesaglitazar, naveglitazar, aleglitazar, saroglitazar, netoglitazone and lobeglitazone. In particular, PPARs have demonstrated clinical efficacy in metabolic diseases such as T2DM and those related to lipid metabolism, which has led to testing and optimization of different compounds such as fibrates and TZDs to simultaneously modulate multiple targets, with a synergistic effect on T2DM and MetS (Ammazzalorso et al., 2019). Other compounds under investigation or in any of the experimental phases by the FDA include elafibranor for PPAR $\alpha/\delta$  (Schattenberg et al., 2021), telmisartan and fimasartan for PPAR $\gamma$ /AT1 (Seo et al., 2022), lanifibranor and sodelglitazar for PPAR $\alpha/\delta/\gamma$  (Kamata et al., 2023) and licoglitazone and sotaglitazone for SGLT1-2, the latter already approved by the FDA in 2023 to reduce the risk of cardiovascular death and heart failure in adults with heart failure, T2DM, chronic kidney disease, and other cardiovascular risk factors (Packer, 2023).

An attractive approach to exploring and expanding the chemical space around the first hit compounds of single and multi-target compounds is the computational generation of chemical libraries that can be used in virtual screening campaigns (Walters, 2019).

Through chemical library enumeration, it is possible to find new bioactive compounds and generate therapeutic options for emerging and challenging molecular targets and complex diseases. It is also possible to control features such as library size, complexity, physicochemical properties, and structural diversity (Ruddigkeit et al., 2012). The goal is to help design high-quality analog series overcoming issues such as low potency, off-target activities, metabolic stability, or poor physicochemical properties for oral administration. One approach that can aid rigorous exploration of the chemistry around first hit compounds is using approaches based on transformation rules from empirical observation and systematic identification using chemoinformatics methods. A recent application of transformation rules was presented with DrugSpaceX, a database with more than 100 million compounds transformed from approved drug molecules (Yang et al., 2021). Although transformation rules are useful for generating *in silico* libraries, the list of rules currently available in the public domain is limited, in many cases due to the difficulty in collecting, curating, and annotating such information (Rarey et al., 2022). To address this problem, we compiled, organized, and made freely available an extensive list of transformation rules for generating compound libraries, as part of this work.

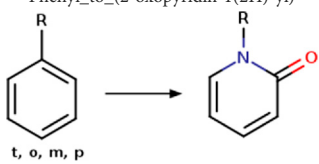
TABLE 2 Examples of transformation rules retrieved from the literature.

Group	Transformation	Transformation type	SMIRKS	Note	Reference
Aminophenyl	Aminophenyl_to_aminobicyclo[1.1.1]pentyl 	Ring substitution	[*:3]-[#7:2](-[*:1])-[c:4]1[c; D2][c; D2][c; D2][c; D2][c; D2]1>>[*:3]-[#7:2](-[*:1])[C:4]12[#6]-[#6](-[#6]1)-[#6]2	-Metabolic stability -Isosteric replacement	Sodano et al. (2020)
1,4-diaminophenyl	1,4-diaminophenyl_to_3,6-diaminopyridazinyl 	Ring modification	[#7H2:1]-[c; D3:2]1[c; D2:6][c; D2:5][c; D3:4][c; D2:3][c; D2]1>>[#7H2:1]-[c; D3:2]1[c; D2:6][c; D2:5][c; D3:4]nn1	-Metabolic stability	Zhang et al. (2020)
Carboxyl	Carboxyl_to_2,4-dioxo-1,3-thiazolidin-5-yl (ionized) 	Ring addition	[#8; D1][#6; A; !R:2]([[*:1]) = O>>[*:1]-[#6:2]-1-[#16]-[#6](=O)-[#7]-[#6]-1 = O [s:0:1]	-Similar acidic pKa -Increased sterics	Hidalgo-Figueroa et al. (2013), Domínguez-Mendoza et al. (2021)
Benzoylphenyl	benzoylphenyl_to_(3-phenyloxetan-3-yl)phenyl 	Linker modification	[#6; a:1][C; \$([#6:2] = O),\$([#6; A; @:2]1[#6]-[#6]-[#6]1),\$([#6; H2])][#6; a:3] >>[#6; a:3][#6; A@:2]1([[#6; a:1])[#6]-[#8:4]-[#6]1	-Isosteric replacement	Dubois et al. (2021)
			*Also apply to 1,1-diphenylmethyl and 1,1'-cyclobutane-1,1'-diyl diphenyl	-Metabolic stability -Reduce phototoxicity in benzofenones	
Phenyl	Phenyl_to_cyclohexyl  t, o, m, p	Ring substitution	[c; x2:2]1[c; x2:3][c; x2:4][c; x2:5][c; x2:6][c; x2:1]1>>[#6:5]-1-[#6:6]-[#6:1]-[#6:2]-[#6:3]-[#6:4]-1	-Bioisosteric replacement -Increase lipophilicity -Improved aqueous solubility -Enhanced oral bioavailability -Reproducible PK profiles	Press et al., 2012; Press et al., 2015; Huang et al., 2019; Subbaiah and Meanwell, 2021
Phenyl	Phenyl_to_(propoxyimino)methyl  t	Functional group change	[*:3]-[c; x2D3:1]1[c; x2D2][c; x2D2][c; x2D2][c; x2D2]1>>[#6]-[#6]-[#6]-[#8][\#7] = [#6:1][*:3]  rb:1:2.2:2.3:2.4:2.5:2.6:2;s:1:3	-Bioisostere replacement -Modulation of selectivity	Piemontese et al. (2015), Subbaiah and Meanwell (2021)

(Continued on following page)



TABLE 2 (Continued) Examples of transformation rules retrieved from the literature.

Group	Transformation	Transformation type	SMIRKS	Note	Reference
Phenyl	<p>Phenyl_to_(2-oxopyridin-1(2H)-yl)</p> 	Ring modification	<chem>[c; D2:2]1[c; x2:3][c; x2:4][c; x2:5][c; x2:6][c; D3:1]1&gt;&gt;O = [#6; x2:2]-1-[#7:1]-[#6:6] = [#6:5]-[#6:4] = [#6:3]-1</chem>	<ul style="list-style-type: none"> <li>-Improved potency</li> <li>-Reduce off-target</li> <li>-Metabolic stability</li> <li>-Enhanced solubility</li> <li>-Reduced lipophilicity</li> </ul>	Subbaiah and Meanwell (2021)

This study aimed to design a multi-target focused library for T2DM using a comprehensive set of chemical transformation rules herein collected, curated, and annotated. As shown in Figure 2, we evaluated the novelty and diversity of the focused library by comparing it with antidiabetic compounds approved for clinical use, natural products, and multi-target compounds reported with *in vivo* activity. To focus the library on attractive and synthetically feasible compounds, computational filters based on medicinal chemistry criteria were applied. Finally, virtual screening using molecular docking for PTP1B and AR was performed at Molecular Operating Environment (MOE) version 2022.02 (Chemical Computing Group (CCG), 2023) and the ADME-Tox properties of compounds with potential multi-target activity were calculated using ADMELab 2.0 (Xiong et al., 2021). Here, we selected PTP1B and AR considering the reference compounds used for enumeration and the currently available hypotheses and information from molecular dynamics models (Domínguez-Mendoza et al., 2021). In particular, this combination could be attractive to modulate insulin signaling, reducing insulin resistance and preventing or delaying diabetic complications such as nephropathies, neuropathies, and cardiomyopathies (Maccari and Ottanà, 2015). The significance of this work is twofold: making available a focused multi-targeted library for T2DM with full details of the methodology used to construct the compounds and making publicly available a general and extensive list of transformation rules to explore the chemical space of targets of therapeutic relevance.

## 2 Methods

### 2.1 Encoding chemical transformations in linear notation (SMIRKS format)

Transformation rules associated with hit optimization were retrieved from the literature. These transformation rules included modifications associated with molecular, physicochemical, pharmacological, ADME, safety, and toxicity parameters. Other rules are associated with structural diversity and bioisosteric changes. Most of the transformation rules found in the literature come from the addition, substitution or removal of a functional group. There are also rules that include cyclization and ring substitution by other rings or intramolecular hydrogen bonding

groups. In total, 280 transformation rules were collected and converted into SMIRKS notation (Daylight theory: SMIRKS - A reaction transform language, 2023) using MarvinSketch (Chemaxon, 2023). The transformation rules were considered in a protonation state of 7.4. Table 2 summarizes examples of the rules identified and their SMIRKS. The Supplementary material includes the complete list of transformation rules and the literature reference.

### 2.2 Enumeration of a new multi-target focused library

Compounds 3 and 4 (Figure 1) were selected as reference compounds to enumerate the new library focused on multi-target compounds. These compounds have robust antihyperglycemic activity *in vivo* and molecular dynamics studies with PTP1B and AR provide relevant structure ligand-protein interaction information for structure-based optimization studies (Domínguez-Mendoza et al., 2021). The best predicted scoring conformations of these compounds docked with a crystallographic structure of the PTP1B protein obtained from the Protein Data Bank (Berman et al., 2000) (PDB ID: 4Y14 (Krishnan et al., 2015)) and that maintained protein-ligand interactions reported in literature were used as the basis for the enumeration of the new library. To compare the effect on the number of compounds and the molecular diversity generated, we used the 175 transformation rules integrated into the MedChem module of MOE version 2022.02 and added 273 transformation rules that we constructed as part of this study. Two iterations were used. Compounds that had a molecular weight (MW) < 630 and topological surface area (TPSA) between 40 and 150 Å were kept. The threshold values were established based on the minimum and maximum values of the multi-target compounds designed for T2DM.

### 2.3 Evaluation of the chemical diversity of the multi-target focused library

The compounds generated in Section 2.2 were compared in terms of physicochemical properties with antidiabetic compounds approved for clinical use retrieved from DrugBank (Wishart et al.,

**TABLE 3** Reference compound datasets to assess the novelty and properties of the newly designed libraries.

Database	Size	Note	Reference
Approved antidiabetic drugs	42	36 approved drugs and 6 under investigation. Obtained from DrugBank. Only compounds for a single target.	Wishart et al. (2018)
DiaNat -DB	329	Antidiabetic compounds from medicinal plants	Madariaga-Mazón et al. (2021)
Multi-target compounds	91	From literature. This set includes 16 multi-target compounds under investigation or in experimental phase by FDA.	This study. See Section 2.1
Multi-target compMedChem	84,778	Compounds generated in MOE using 455 transformations rules	This study. See Section 2.3

2018); antidiabetic compounds from medicinal plants obtained from DiaNat-DB (Madariaga-Mazón et al., 2021), and multi-target compounds for T2DM studied *in vivo* assays. Compounds in SMILES format can be found in the Supplementary material. Prior to analysis, each molecule was prepared using the open-source cheminformatics toolkits RDKit (Landrum, n.d.) and MolVS (MolVS: Molecule Validation and Standardization—MolVS 0.1.1 documentation, n.d.). Compounds were standardized and those containing multiple components were split, keeping the largest component. Compounds with valence errors or chemical elements other than H, B, C, N, O, F, Si, P, S, Cl, Se, Br, and I were removed. The remaining compounds were neutralized and reionized to subsequently generate a canonical tautomer. Finally, duplicated compounds were deleted and canonical simplified molecular-input line-entry system (SMILES) (ignoring stereochemistry information) were generated as molecular representation (Weininger, 1988). Table 3 summarizes the number and source of compounds used in the comparison.

### 2.3.1 Property rules and synthetic accessibility

To profile the generated molecules based on common property rules and synthetic accessibility, we calculated seven descriptors for each molecular entity with the RDKit library: molecular weight (MW), octanol–water partition coefficient (logP), number of hydrogen bond acceptors (HBA), number of hydrogen bond donors (HBD), topological polar surface area (TPSA), the number of rotatable bonds (RotB) and the distribution of quantitative estimate of drug-likeness (QED) (Bickerton et al., 2012).

Among the several computational scores available to estimate the synthetic accessibility of generated compounds, and based on previous comparisons of scores. (Skoraczynski et al., 2023), in this work, we calculated rsynth (a ligand-based score implemented in MOE, version 2022.02) and SAScore (Ertl and Schuffenhauer, 2009) as a structure-based approach.

### 2.3.2 Chemical multiverse: visual representation and analysis

To have a comprehensive analysis of the chemical space, we used three well-established visualization methods and different types of descriptors. Noteworthy, analyzing the chemical space with distinct and complementary descriptors is crucial because each one provides a different and complementary perspective of the chemical space (aka, a “chemical multiverse” as discussed elsewhere (Medina-Franco et al., 2022)). In this study, we used principal component

analysis (PCA) and t-distributed stochastic neighbor embedding t-SNE based on six molecular properties of pharmacological interest, namely; HBD, HBA, logP, MW, RB, TPSA. As a third method to characterize the chemical space, we used was the Tree MAP (TMAP) algorithm (Probst and Reymond, 2020) based on ECFP4 as a structural fingerprint. Additionally, the molecular shape of the compounds in the newly generated library was evaluated using the principal moments of inertia (PMI) graph, which was carried out by calculating the lowest energy conformation of each compound using MMFF94x as a force field in MOE, version 2022.22. Once the lowest energy conformer was calculated, values of normalized PMI ratios, npr1 (I1/I3) and npr2 (I1/I3), were determined in MOE. Then, npr1 and npr2 were plotted on a triangular graph with the vertices (0,1), (0.5,0.5), and (1,1) representing a perfect rod, disc, and sphere, respectively.

## 2.4 Filtering of compounds with relevance in pharmaceutical chemistry

Various filters (e.g., calculated descriptors) can be used to improve the selection of enumerated compounds, including the removal of compounds with undesirable pharmaceutical properties and molecules very difficult to synthesize. In this work, we chose to filter by SAScore, rsynth, QED, and the filters included in the RDKit molecule catalog filter node in KNIME. This node removes compounds with Pan-assay interference compounds (PAINS) (Baell and Holloway, 2010), unwanted functionality due to potential toxicity or unfavorable pharmacokinetic properties, and problematic functional groups. The compounds in the enumerated library were filtered using criteria as follows: QED > 0.67, SAScore ≤ 6, rsynth > 0.5 and zero structural alerts. It should be noted that other cutoff values could be used for other applications.

## 2.5 Structure-based filtering

Although enumeration of compounds can be performed using the protein structure to obtain a score value in MOE, this can be computationally expensive considering the number of compounds that can potentially be generated. In addition, although the search algorithm can be selected, neither an algorithm to perform the rescoring nor a specific number of conformations to be obtained can be selected. For this reason, we followed a structure-based filtering to select the most promising compound subset as a multi-target library.

In this study, compounds that showed a docking score better than the reference compounds for PTP1B (PDB ID: 4Y14 (Krishnan et al., 2015)), were further filtered considering docking with another therapeutic target relevant to T2DM: AR (PDB ID: 4XZH (Ruiz et al., 2015)). Of note, other relevant targets in T2DM can be considered during the design of the reference compounds. Here, we selected PTP1B and AR considering the current available hypothesis and model information based on molecular dynamics (Domínguez-Mendoza et al., 2021). Also, we want to point out that several other docking programs could be used (including post consensus scoring analysis). However, testing different docking programs and exploring multiple consensus analysis schemes is out of the scope of this study that is focused on proposing a general approach to design multi-target focused libraries.

The protein preparation of PDB ID: 4Y14 and 4XZH was made with default settings of the QuickPrep module of MOE v. 2022.02. This module carries out the following processes: calibration of the structure by protonation, addition of all the lacking hydrogen atoms, elimination of water molecules 4.5 Å farther from the protein, addition of missing amino acids residues, neutralization of the endpoints adjoining empty residues and energy minimization. We used AMBER14:EHT forcefield (ff14SB (Maier et al., 2015)) as parameter for the energy minimization stage for the protein. The ligands were also prepared in MOE, we used MFF94x as forcefield. For docking, the receptor was considered rigid and the ligands flexible. We used the default settings for placement (method: triangle matcher, score function: London dG) and refinement (method: rigid receptor, score function: GBVI/WSA dG) (Vilar et al., 2008).

## 2.6 Prediction of ADME-Tox properties

The ADME-tox properties of the compounds that had successfully passed the filters described in Section 2.6 and Section 2.7 were evaluated using ADMETlab 2.0 platform (Xiong et al., 2021). This platform has been compared with other free online ADMET tools and has significant advantages (Dulsat et al., 2023). Based on these comparisons we used this tool because ADMETlab is a complete platform in terms of the large number of relevant parameters that can be predicted including elimination parameters such as clearance and half-life ( $t_{1/2}$ ). The latter two descriptors are particularly relevant in chronic diseases such as T2DM, where ideally compounds with a long half-life are sought to reduce the dose. Finally, ADMETlab 2.0 allows evaluating up to 500 compounds at the same time.

## 3 Results and discussion

### 3.1 Encoding chemical transformations in SMIRKS format

Based on the structures exploited in the multi-target pharmacophores for DMT2, 280 transformation rules were collected from the literature, of which 113 were bioisosteric modifications of the phenyl group (Subbaiah and Meanwell, 2021), 25 for the amide bond (Kumari et al., 2020), 36 for the carboxyl group (Bredael et al., 2022), 57 for the phosphate group, and 49 for other moieties including ester, alcohol, alkyl,

aminophenyl, and nitro to name a few examples. Compared to the preloaded MOE transformations (version 2022.22), these included only 10 for the phenyl group, 7 for the amide bond and 4 for the carboxyl group. The remaining transformations in MOE mainly concern cyclization and substitution of rings by other rings or intramolecular hydrogen bonding groups. It is important to mention that some examples of the rule transformations present in Table 2 have been applied in the optimization of antidiabetic compounds. For example, Huang et al. reported the effect of bioisosteric replacement of a phenyl ring in the biphenyl moiety with cyclohexyl motif in a GPR40 agonist (Huang et al., 2019). Another bioisosteric replacement of the phenyl ring in an antidiabetic compound was reported by (Piemontese et al., 2015). In that work, the authors replaced a terminal phenyl ring in a dual PPAR  $\alpha/\gamma$  agonist with the n-propyl oxime moiety. For its part, Dr. Navarrete-Vázquez's group has proposed various series of thiazolidine-2,4-dione and barbituric acid derivatives with robust antihyperglycemic activity *in vivo* as acid bioisosteres (Hidalgo-Figueroa et al., 2013; Domínguez-Mendoza et al., 2021).

### 3.2 Enumeration of new compounds

Using the transformation rules preloaded in MOE and performing two iterations, 6,838 molecules were obtained from compound 3 and 1834 from compound 4. After adding the transformation rules collected from the literature and keeping two iterations, the number of compounds increased to 52,185 (from compound 3) and to 32,593 (from compound 4), after curation of the enumerated library. This large increase in the number of compounds was expected, as it followed an iterative process, and the transformations considered in this work are focused on moieties that contain the reference compounds.

### 3.3 Evaluation of the chemical diversity of the multi-target focused library

#### 3.3.1 Property rules and synthetic accessibility

Figure 3 shows the distributions of each descriptor for the compounds of the generated library and the reference libraries using rain cloud plots (Allen et al., 2019). These plots allow visualization of the probability density and typical boxplot statistics such as median, mean, and confidence intervals. According to these plots, DiaNat-DB (329 antidiabetic compounds from medicinal plants) has a wider distribution in terms of properties of pharmaceutical importance. Since MW and TPSA values were used as filters to generate the new compounds, the distributions of these values for generated compounds are skewed and resemble the distribution of multi-target compounds and approved drugs. Comparing the plots of the other descriptors, we can see that the transformations used increased the range of properties such as LogP and HBD, HBA and chirality (Supplementary Table S1 in the Supplementary Material).

Figure 4 shows the distributions of the calculated descriptors to quantify synthetic accessibility and drug-likeness through QED. As can be seen in Figure 4, most of the generated compounds have a

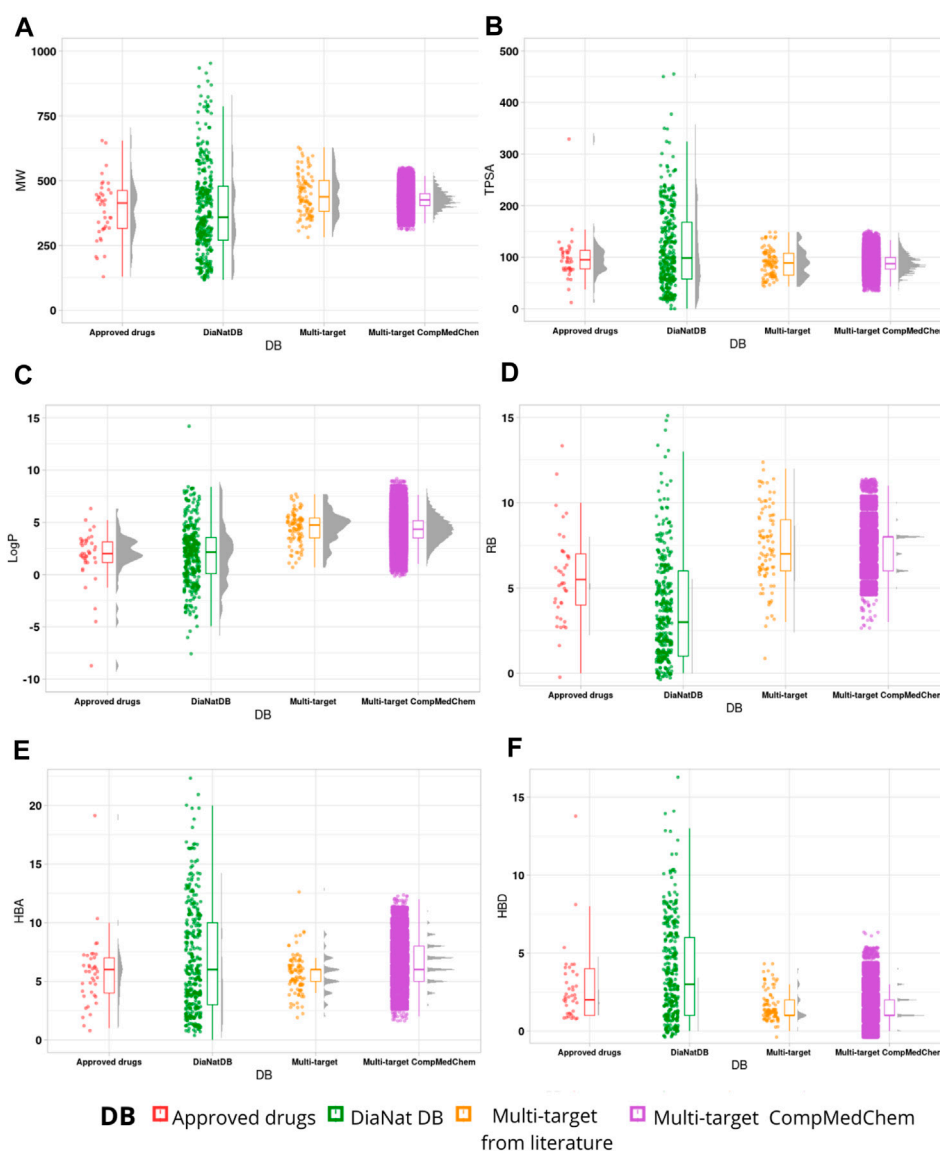


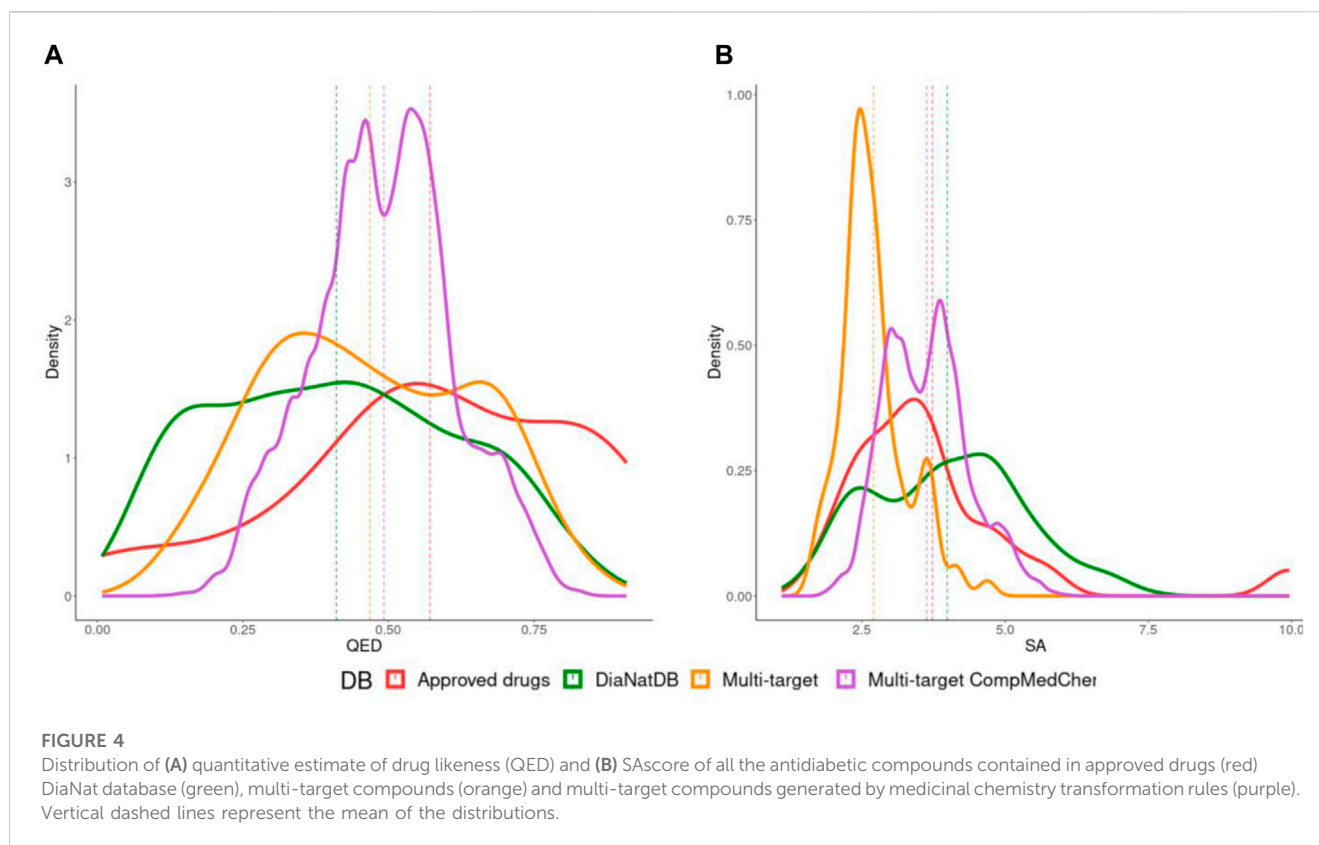
FIGURE 3

Rain cloud plots of the six physicochemical properties of pharmaceutical relevance: (A) molecular weight (MW), (B) topological polar surface area (TPSA), (C) partition coefficient octanol/water (log P), (D) number of rotatable bonds (RB), (E) hydrogen bond acceptors (HBA) and (F) hydrogen bond donors (HBD).

SAScore value of less than 6, indicating that they are, in principle, synthesizable (Ertl and Schuffenhauer, 2009). Regarding the QED value, which is a measure of drug-likeness based on the concept of desirability, a value greater than or equal to 0.67 represents attractive compounds, and the lower this value, the compounds are considered unattractive (QED 0.49~0.67) and too complex (QED < 0.49). These reference values were established based on the calculated physicochemical properties of marketed oral drugs and published human data (Bickerton et al., 2012). When we compare the QED distribution of all antidiabetic compounds, we find that the compounds generated by transformation rules have a higher mean value (0.49) than the multi-target compounds obtained from the literature (0.46). The summary statistics of these plots can be found in the Supplementary material.

### 3.3.2 Chemical multiverse of generated compounds

The concept of chemical multiverse (e.g., for the same compound datasets, different chemical spaces, each based on a different set of descriptors) was used to compare comprehensively the visual representation of the chemical space of the generated compounds with collections of reference compounds (Figure 5). The PCA of six physicochemical properties: MW, HBD, HBA, logP, TPSA and RB shows DiaNatDB is the most diverse database in terms of physicochemical properties (Figure 5A). Using a non-linear method such as t-SNE and the same descriptors, we obtain a different visualization that allows us to visualize clusters or groups of data points and their relative proximities (5b). In terms of molecular fingerprints (ECFP4)



(5c) and molecular shapes (5d), the library generated with transformation rules exhibited the largest structural and shape diversity, which is even larger than that of the DiaNat-DB database.

### 3.4 Filtering of compounds with relevance in pharmaceutical chemistry

To focus the library on attractive and synthetically feasible compounds, computational filters were applied as indicated in Section 2.6 of the Methods section. Table 4 summarizes the criteria used and the number of compounds remaining after applying the filters. As can be seen, the filter that considerably reduced the number of compounds was the QED value. The filtered compounds are characterized by not having chiral centers. If this is an important feature, you will need to consider it when applying the filters. Filtered compounds are listed in the Supplementary material.

### 3.5 Structure-based filtering

The compounds selected from the filtering described in Section 3.6 were subjected to a docking-based virtual screening with PTP1B and AR. The docking protocols are described in Section 2.6. The docking scores calculated with MOE for compounds 3 (PTP1B score:  $-7.04$  kcal/mol and AR score =  $-8.84$  kcal/mol) and 4 (PTP1B score:  $-7.91$  kcal/mol and AR score =  $-8.98$  kcal/mol) were used as cut-offs to select potential multi-target compounds. It

was also checked whether the docking of these compounds reproduced the interactions previously reported in molecular dynamics studies (Domínguez-Mendoza et al., 2021). Supplementary Table S2 shows the score values and interaction plots of compounds 3 and 4 with PTP1B and AR. The first virtual screening with PTP1B yielded 1,655 compounds with a lower score value than the reference compounds: 1,543 from compound 3 and 112 from compound 4. The virtual screening hit compounds were docked with AR yielding 816 hit compounds: 792 were from compound 3 and 24 from compound 4. Figure 6 shows examples of designed compounds that have the potential to be used in multi-target approaches. The figure shows the parent structure (3 and 4), the transformation rules and the compounds designed with their corresponding calculated docking scores with PTP1B and AR. We also include the 3D docking models for the proposed compounds, and their overlap with the reference compounds. The docking results for PTP1B and AR of the 816 compounds can be found in the Supplementary material.

### 3.6 Prediction of ADME-Tox properties

Table 5 shows the average and standard deviation for different ADME-Tox properties of approved DMT2 drugs, multi-target compounds with reported *in vivo* activity and the 816 compounds that showed the potential to interact with PTP1B and AR. The values described to make an empirical decision are taken directly from the ADMElab 2.0 documentation



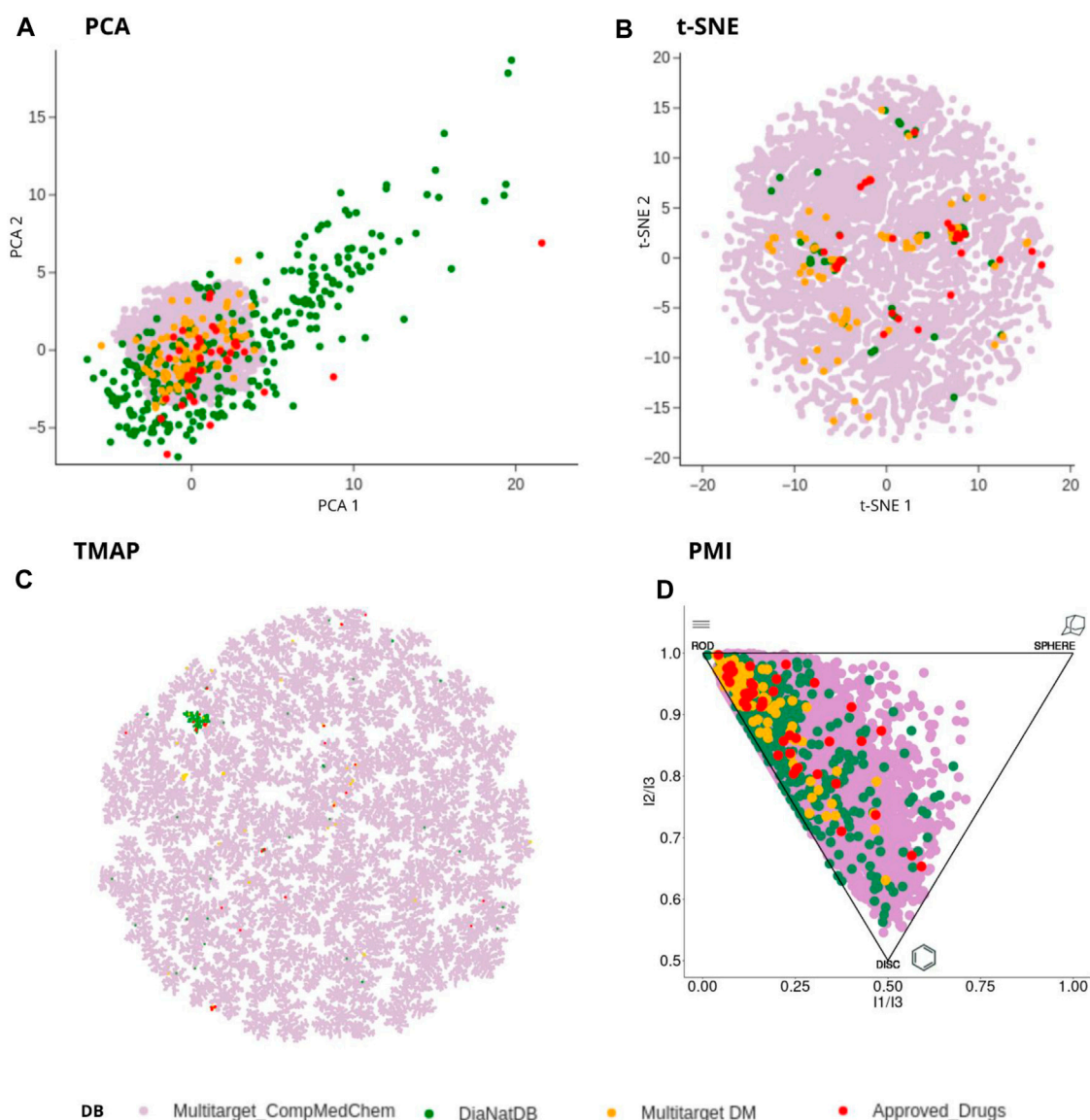


FIGURE 5

Visual representation of the chemical multiverse of antidiabetic compounds contained in approved drugs (red) DiaNat-DB (green), multi-target compounds (orange) and multi-target compounds generated by medicinal chemistry transformation rules (purple). (A) PCA of six physicochemical properties: MW, HBD, HBA, logP, TPSA and RB (B) t-SNE of six physicochemical properties (*vide supra*), (C) TMAP based on molecular ECFP4 fingerprint. (D) PMI space. Each corner on the triangular PMI plot indicates compounds with certain shape characteristics. The top left corner of the PMI represents compounds with rodlike shape, the top right corner represents compounds with spherical shape, and the bottom corner represents compounds with disc-like shape.

(<https://admetmesh.scbdd.com/explanatthat/overlap/index>). As can be seen, the subset of compounds selected from the multi-target library generated in this work exhibit optimal adsorption and distribution properties. It is also noteworthy that in terms of metabolism, ADMETlab 2.0 predicts that the compounds generated in this work have a lower probability of being inhibitors of various CYPs compared to approved drugs and multi-target compounds that have a moderate to high probability of being inhibitors of CYP2C9, CYP2C19, and CYP3A4. Knowing this information is important because the inhibition of some CYPs is associated with the risk of hypoglycemia. Knowing this information

is important because the inhibition of some CYPs is associated with the risk of hypoglycemia. For example, CYP2C9 metabolizes nateglinide, repaglinide, rosiglitazone, and most sulfonylureas, such as glibenclamide, glimepiride and glipizide (Holstein et al., 2012). Other examples are pioglitazone and repaginate metabolized with CYP3A4. ADMETlab 2.0 also predicts the probability of being substrates of various CYPs. This data and other properties calculated with ADMETlab 2.0 for each compound can be found in the Supplementary material. It is important to mention that the prediction of inhibitory activity in ADMETlab 2.0 is based on a dataset containing information on inhibitory activity of compounds

**TABLE 4** Filters applied and the number of compounds remaining for virtual screening.

Filters	Compound 3	Compound 4
Initial (ComMedChem rules)	72,349	33,661
Curated <sup>a</sup>	52,185	32,593
SAScore ≤ 6	52,185	32,552
Rsynth > 0.5	43,625	29,166
QED > 0.67	2,276	3,226
RDKit Molecule Catalog Filter = 0	1,543	451

<sup>a</sup>For diversity studies, the set of curated structures was used.

obtained by high-throughput quantitative screening with an *in vitro* bioluminescence assay (Veith et al., 2009). However, in the description of these data, it is mentioned that the inhibitory

activity in the assay may be due to compounds acting as substrates and that some weakly binding substrates may be classified as “inactive,” so that the predictions may need further confirmation.

In terms of excretion, clearance (CL) and half-life ( $T_{1/2}$ ) are important pharmacokinetic parameters that allow defining a drug's dosing frequency. In the case of ADMETlab 2.0, the half-life is not measured in units of time. The output value is the probability of falling into category 1 ( $T_{1/2} \leq 3$ ). That is, the greater the probability of falling into category 1, the more likely the substance is to be classified as “poor” because its  $T_{1/2}$  would be lower ( $T_{1/2} \leq 3$ ). For antidiabetic drugs, the average CL is 4.4716 mL/min/kg (poor) and the  $T_{1/2}$  is 0.4080 (medium). The discrepancy in predictions could be due to the difference in models or datasets. In the case of the generated multi-target compounds, they may not be optimal for reducing the frequency of administration. Finally, although the probability of compounds being hERG blockers is reduced, the

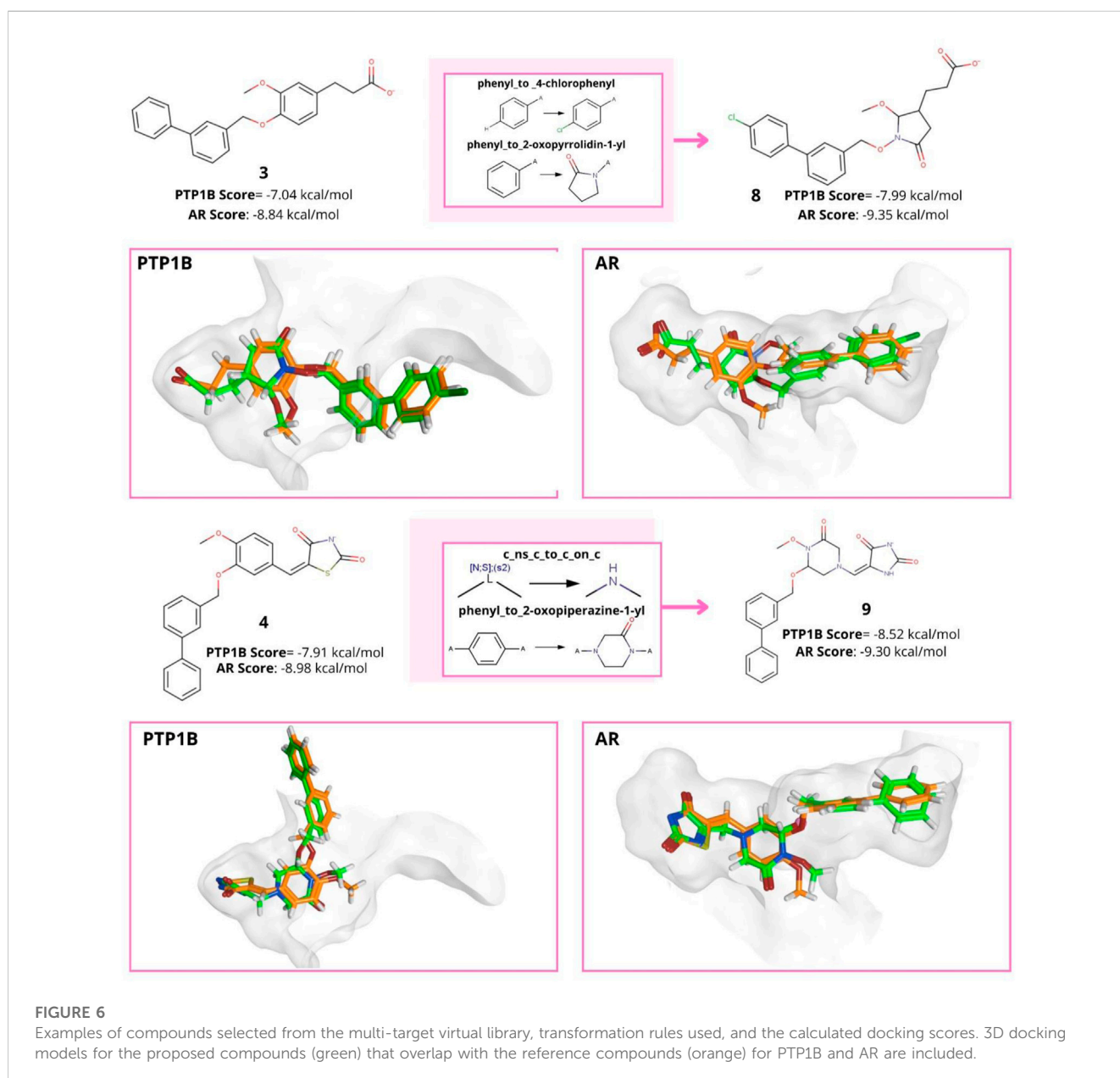


TABLE 5 Estimation of selected ADME-Tox properties of 816 compounds with ADMETlab 2.0<sup>a,b</sup>.

ADME-tox properties	Empirical decision	Approved drugs	Multi-target from literature	Multi-target CompMedChem <sup>c</sup>
% Lipinski	MW ≤ 500; logP ≤ 5; Hacc ≤ 10; Hdon ≤ 5 < 2 violations: Accepted; ≥ 2 violations: Rejected	92.85%	83.51%	100%
LogD at pH = 7.4	Compounds in the range from 1 to 3 log mol/L will be considered proper	1.5807 ± 1.9413	3.4817 ± 1.1394	2.3117 ± 1.257
Caco-2 Permeability	> -5.15: excellent; otherwise: poor	-5.3635 ± 0.5697	-5.0198 ± 0.3816	-5.1287 ± 0.282
Volume Distribution (VD)	0.04–20: excellent; otherwise: poor	0.7733 ± 0.6735	0.6091 ± 0.4483	0.3805 ± 0.1896
CYP1A2-inhibitor	The output value is the probability of being inhibitor, within the range of 0–1	0.1543 ± 0.2344	0.4157 ± 0.2926	0.1384 ± 0.1618
CYP2C9 inhibitor		0.3849 ± 0.3621	0.7215 ± 0.2534	0.287 ± 0.2952
CYP2C19 inhibitor		0.2601 ± 0.3034	0.5659 ± 0.3208	0.2154 ± 0.2288
CYP2D6-inhibitor	0–0.3: excellent; 0.3–0.7: medium; 0.7–1: poor	0.1736 ± 0.2457	0.2659 ± 0.2953	0.1115 ± 0.1707
CYP3A4-inhibitor		0.3034 ± 0.3448	0.3790 ± 0.3219	0.0992 ± 0.1289
Clearance (CL)	>15: excellent	4.4716 ± 3.9932	4.0985 ± 3.3766	6.2131 ± 3.2198
	5–15: medium			
	<5: poor			
Half-life time (T <sub>1/2</sub> )	0–0.3: excellent; 0.3–0.7: medium; 0.7–1: poor	0.4080 ± 0.2759	0.3083 ± 0.2511	0.7309 ± 0.1447
hERG Blockers	0–0.3: excellent; 0.3–0.7: medium; 0.7–1.0: poor	0.2179 ± 0.2446	0.2705 ± 0.2587	0.1631 ± 0.1621
Human hepatotoxicity (H-HT)	0–0.3: excellent; 0.3–0.7: medium; 0.7–1.0: poor	0.5874 ± 0.3218	0.6796 ± 0.2888	0.6023 ± 0.2631
Drug - Induced liver Injury (DILI)	0–0.3: excellent; 0.3–0.7: medium; 0.7–1.0: poor	0.6480 ± 0.3965	0.9073 ± 0.195	0.7124 ± 0.3434

<sup>a</sup>Compounds that showed the potential to interact with PTP1B and AR, selected in Section 3.5.

<sup>b</sup>The color coding represents if the criteria in the empirical decision column are met (green), partially met (yellow), and not (red). The color coding in the table is the same as the one used in ADMETlab 2.0.

<sup>c</sup>Compounds that showed the potential to interact with PTP1B and AR, selected in Section 3.5

likelihood of drug-induced liver injury (DILI) would remain a challenge to optimize.

## 4 Conclusion

Designing multi-target compounds is an attractive approach to develop therapeutic treatments for complex diseases such as T2DM and MetS. Herein, we collected from the literature and analyzed ninety-one multi-target compounds for which *in vivo* antidiabetic activity has been reported, with a total of twenty target combinations. Following an enumeration based on transformation rules, we expanded the relevant chemical space of two of these multi-target hit compounds. More than 450 transformation rules were applied, of which 280 are made openly available to the scientific community. We concluded that the compounds generated with transformation rules have similar physicochemical properties to antidiabetic drugs and multi-target compounds reported in literature. Of the 84,778 generated compounds with valid structures, 85% are predicted to be synthetically feasible. The enumerated compounds are also attractive considering structural and shape diversity.

To focus on attractive and synthetically feasible compounds to perform virtual screening, various drug-likeness and quality filters were applied, yielding a multi-target virtual library with 2037 compounds. After a docking-based virtual screening with PTP1B and AR, 816 multi-

target compounds were selected. Compounds in this library have favorable ADME properties, making the library an attractive source of promising candidates for further research and development.

In line with open and democratization of science, the newly designed multi-target focused library is freely available as a valuable source of starting points for chemical synthesis, biological evaluation, or further computational analysis such as virtual screening or reference libraries in library design.

## Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: [https://figshare.com/projects/Design\\_of\\_a\\_multi-target\\_focused\\_library\\_for\\_antidiabetic\\_targets\\_using\\_a\\_comprehensive\\_set\\_of\\_chemical\\_transformation\\_rules/175194](https://figshare.com/projects/Design_of_a_multi-target_focused_library_for_antidiabetic_targets_using_a_comprehensive_set_of_chemical_transformation_rules/175194).

## Author contributions

FS-G: Writing—original draft, Writing—review and editing, Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software. GN-V: Conceptualization, Investigation, Writing—review and editing, Supervision. JM-F: Funding acquisition,

Project administration, Resources, Supervision, Writing—original draft, Writing—review and editing, Formal Analysis, Conceptualization.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. We thank the support of DGAPA, UNAM, Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), Grant No. IV200121.

## Acknowledgments

FS-G thanks Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT), Mexico for the Ph.D. scholarship 848061. The technical assistance and enthusiastic participation of Jenny Serrano López, Vanesa Garduño Ibarra, Leonardo Saravia Fuentes and Diana Prado Romero (School of Chemistry, UNAM), is greatly acknowledged. We also acknowledge the innovation space UNAM-HUAWEI the computational resources to use their supercomputer under project-7 “*Desarrollo y aplicación de algoritmos de inteligencia artificial para el diseño de fármacos aplicables al tratamiento de diabetes mellitus y cáncer*”.

## References

- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., and Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res.* 4, 63. doi:10.12688/wellcomeopenres.15191.1
- Ammazzalorso, A., Maccallini, C., Amoia, P., and Amoroso, R. (2019). Multitarget PPAR $\gamma$  agonists as innovative modulators of the metabolic syndrome. *Eur. J. Med. Chem.* 173, 261–273. doi:10.1016/j.ejmech.2019.04.030
- Artasensi, A., Pedretti, A., Vistoli, G., and Fumagalli, L. (2020). Type 2 diabetes mellitus: a Review of multi-target drugs. *Molecules* 25, 1987. doi:10.3390/molecules25081987
- Baell, J. B., and Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 53, 2719–2740. doi:10.1021/jm901137j
- Balakumar, P., Rose, M., Ganti, S. S., Krishan, P., and Singh, M. (2007). PPAR dual agonists: are they opening Pandora's Box? *Pharmacol. Res.* 56, 91–98. doi:10.1016/j.phrs.2007.03.002
- Balakumar, P., Mahadevan, N., and Sambathkumar, R. (2019). A contemporary overview of ppar $\gamma$  dual agonists for the management of diabetic dyslipidemia. *Curr. Mol. Pharmacol.* 12, 195–201. doi:10.2174/1874467212666190111165015
- Benson, S. C., Pershadsingh, H. A., Ho, C. I., Chittiboyina, A., Desai, P., Pravenec, M., et al. (2004). Identification of telmisartan as a unique angiotensin II receptor antagonist with selective PPAR $\gamma$  modulating activity. *Hypertension* 43, 993–1002. doi:10.1161/01.HYP.0000123072.34629.57
- 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
- Bhattarai, B. R., Kafle, B., Hwang, J.-S., Khadka, D., Lee, S.-M., Kang, J.-S., et al. (2009). Thiazolidinedione derivatives as PTP1B inhibitors with antihyperglycemic and antiobesity effects. *Bioorg. Med. Chem. Lett.* 19, 6161–6165. doi:10.1016/j.bmcl.2009.09.020
- Bhattarai, B. R., Kafle, B., Hwang, J.-S., Ham, S. W., Lee, K.-H., Park, H., et al. (2010). Novel thiazolidinedione derivatives with anti-obesity effects: dual action as PTP1B inhibitors and PPAR- $\gamma$  activators. *Bioorg. Med. Chem. Lett.* 20, 6758–6763. doi:10.1016/j.bmcl.2010.08.130
- Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., and Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. *Nat. Chem.* 4, 90–98. doi:10.1038/nchem.1243
- Blöcher, R., Lamers, C., Wittmann, S. K., Merk, D., Hartmann, M., Weizel, L., et al. (2016). N-Benzylbenzamides: a novel merged scaffold for orally available dual soluble epoxide hydrolase/peroxisome proliferator-activated receptor  $\gamma$  modulators. *J. Med. Chem.* 59, 61–81. doi:10.1021/acs.jmedchem.5b01239
- Boubia, B., Poupardin, O., Barth, M., Binet, J., Peralba, P., Mounier, L., et al. (2018). Design, synthesis, and evaluation of a novel series of indole sulfonamide peroxisome proliferator activated receptor (PPAR)  $\alpha/\gamma/\delta$  triple activators: discovery of lanifibranor, a new antifibrotic clinical candidate. *J. Med. Chem.* 61, 2246–2265. doi:10.1021/acs.jmedchem.7b01285
- Bredael, K., Geurs, S., Clarisse, D., De Bosscher, K., and D'hooghe, M. (2022). Carboxylic acid bioisosteres in medicinal chemistry: synthesis and properties. *J. Chem. Eng.* 2022, 1–21. doi:10.1155/2022/2164558
- Casimiro-García, A., Filzen, G. F., Flynn, D., Bigge, C. F., Chen, J., Davis, J. A., et al. (2011). Discovery of a series of imidazo[4,5-b]pyridines with dual activity at angiotensin II type 1 receptor and peroxisome proliferator-activated receptor- $\gamma$ . *J. Med. Chem.* 54, 4219–4233. doi:10.1021/jm200409s
- Casimiro-García, A., Heemstra, R. J., Bigge, C. F., Chen, J., Ciske, F. A., Davis, J. A., et al. (2013). Design, synthesis, and evaluation of imidazo[4,5-c]pyridin-4-one derivatives with dual activity at angiotensin II type 1 receptor and peroxisome proliferator-activated receptor- $\gamma$ . *Bioorg. Med. Chem. Lett.* 23, 767–772. doi:10.1016/j.bmcl.2012.11.088
- Chadha, N., and Silakari, O. (2017). Identification of low micromolar dual inhibitors for aldose reductase (ALR2) and poly (ADP-ribose) polymerase (PARP-1) using structure based design approach. *Bioorg. Med. Chem. Lett.* 27, 2324–2330. doi:10.1016/j.bmcl.2017.04.038
- Chemaxon (2023). Chemaxon. Available at: <https://www.chemaxon.com> (Accessed May 22, 2023).
- Chen, X., Chen, Z., Xu, D., Lyu, Y., Li, Y., Li, S., et al. (2021). De novo design of g protein-coupled receptor 40 peptide agonists for type 2 diabetes mellitus based on artificial intelligence and site-directed mutagenesis. *Front. Bioeng. Biotechnol.* 9, 694100. doi:10.3389/fbioe.2021.694100
- Choung, W., Jung, H. J., Nam, E. H., Yang, D., Yoo, B., Choi, H., et al. (2018). Discovery of the bifunctional modulator of angiotensin II type 1 receptor (AT1R) and PPAR $\gamma$  derived from the AT1R antagonist, Fimasartan. *Fimasartan. Bioorg. Med. Chem. Lett.* 28, 3155–3160. doi:10.1016/j.bmcl.2018.08.036
- Colín-Lozano, B., Estrada-Soto, S., Chávez-Silva, F., Gutiérrez-Hernández, A., Cerón-Romero, L., Giacomani-Martínez, A., et al. (2018). Design, synthesis and in combo antidiabetic bioevaluation of multitarget phenylpropanoic acids. *Molecules* 23, 340. doi:10.3390/molecules23020340
- Darwish, K. M., Salama, I., Mostafa, S., Gomaa, M. S., Khafagy, E.-S., and Helal, M. A. (2018). Synthesis, biological evaluation, and molecular docking investigation of benzhydryl- and indole-based dual PPAR- $\gamma$ /FFAR1 agonists. *Bioorg. Med. Chem. Lett.* 28, 1595–1602. doi:10.1016/j.bmcl.2018.03.051

## Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## 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/fphar.2023.1276444/full#supplementary-material>



- Daylight theory: SMIRKS - A reaction transform language (2023). Daylight theory: SMIRKS - a reaction transform language. Available at: <https://www.daylight.com/dayhtml/doc/theory/theory.smirks.html> (Accessed July 1, 2023).
- Dobrică, E.-C., Găman, M.-A., Cozma, M.-A., Bratu, O. G., Pantea Stoian, A., and Diaconu, C. C. (2019). Polypharmacy in type 2 diabetes mellitus: insights from an internal medicine department. *Medicina* 55, 436. doi:10.3390/medicina55080436
- Domínguez-Mendoza, E. A., Galván-Ciprés, Y., Martínez-Miranda, J., Miranda-González, C., Colín-Lozano, B., Hernández-Núñez, E., et al. (2021). Design, synthesis, and *in silico* multitarget pharmacological simulations of acid bioisosteres with a validated *in vivo* antihyperglycemic effect. *Molecules* 26, 799. doi:10.3390/molecules26040799
- Dubois, M. A. J., Croft, R. A., Ding, Y., Choi, C., Owen, D. R., Bull, J. A., et al. (2021). Investigating 3,3-diaryloxetanes as potential bioisosteres through matched molecular pair analysis. *RSC Med. Chem.* 12, 2045–2052. doi:10.1039/d1md00248a
- Dulsat, J., López-Nieto, B., Estrada-Tejedor, R., and Borrell, J. I. (2023). Evaluation of free online ADMET tools for academic or small biotech environments. *Molecules* 28, 776. doi:10.3390/molecules28020776
- Ertl, P., and Schuffenhauer, A. (2009). Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* 1, 8. doi:10.1186/1758-2946-1-8
- Estrada-Soto, S., Cerón-Romero, L., Navarrete-Vázquez, G., Rosales-Ortega, E., Gómez-Zamudio, J., Cruz, M., et al. (2022). PPARα/γ, adiponectin, and GLUT4 overexpression induced by moronic acid methyl ester influenced glucose and triglyceride levels of experimental diabetic mice. *Can. J. Physiol. Pharmacol.* 100, 295–305. doi:10.1139/cjpp-2021-0526
- Fang, Y., Zhang, S., Wu, W., Liu, Y., Yang, J., Li, Y., et al. (2020). Design and synthesis of tetrahydropyridopyrimidine derivatives as dual GPR119 and DPP-4 modulators. *Bioorg. Chem.* 94, 103390. doi:10.1016/j.bioorg.2019.103390
- Galicia-García, U., Benito-Vicente, A., Jebari, S., Larrea-Sebal, A., Siddiqi, H., Uribe, K. B., et al. (2020). Pathophysiology of type 2 diabetes mellitus. *Int. J. Mol. Sci.* 21, 6275. doi:10.3390/ijms21176275
- Gilardi, F., Giudici, M., Mitro, N., Maschi, O., Guerrini, U., Rando, G., et al. (2014). LT175 is a novel PPARα/γ ligand with potent insulin-sensitizing effects and reduced adipogenic properties. *J. Biol. Chem.* 289, 6908–6920. doi:10.1074/jbc.M113.506394
- Gutiérrez-Hernández, A., Galván-Ciprés, Y., Domínguez-Mendoza, E. A., Aguirre-Vidal, Y., Estrada-Soto, S., Almanza-Pérez, J. C., et al. (2019). Design, synthesis, antihyperglycemic studies, and docking simulations of benzimidazole-thiazolidinedione hybrids. *J. Chem. Chem. Eng.* 2019, 1–8. doi:10.1155/2019/1650145
- Hanf, R., Millatt, L. J., Cariou, B., Noel, B., Rigou, G., Delataille, P., et al. (2014). The dual peroxisome proliferator-activated receptor alpha/delta agonist GFT505 exerts anti-diabetic effects in db/db mice without peroxisome proliferator-activated receptor gamma-associated adverse cardiac effects. *Diab. Vasc. Dis. Res.* 11, 440–447. doi:10.1177/1479164114548027
- He, B. K., Ning, Z. Q., Li, Z. B., Shan, S., Pan, D. S., Ko, B. C. B., et al. (2012). *In vitro* and *in vivo* characterizations of chiglitazar, a newly identified PPAR pan-agonist. *PPAR Res.* 2012, 546548. doi:10.1155/2012/546548
- He, Y.-L., Haynes, W., Meyers, C. D., Amer, A., Zhang, Y., Mahling, P., et al. (2019). The effects of licoglitazone, a dual SGLT1/2 inhibitor, on body weight in obese patients with or without diabetes. *Diabetes Obes. Metab.* 21, 1311–1321. doi:10.1111/dom.13654
- Herrera-Rueda, M. Á., Tlahuext, H., Paoli, P., Giacoman-Martínez, A., Almanza-Pérez, J. C., Pérez-Sánchez, H., et al. (2018). Design, synthesis, *in vitro*, *in vivo* and *in silico* pharmacological characterization of anti-diabetic N-Boc-L-tyrosine-based compounds. *Biomed. Pharmacother.* 108, 670–678. doi:10.1016/j.biopha.2018.09.074
- Hidalgo-Figueroa, S., Ramírez-Espinosa, J. J., Estrada-Soto, S., Almanza-Pérez, J. C., Román-Ramos, R., Alarcón-Aguilar, F. J., et al. (2013). Discovery of thiazolidine-2,4-dione/biphenylcarbonitrile hybrid as dual PPAR α/γ modulator with anti-diabetic effect: *in vitro*, *in silico* and *in vivo* approaches. *Chem. Biol. Drug Des.* 81, 474–483. doi:10.1111/cbdd.12102
- Hidalgo-Figueroa, S., Rodríguez-Luévano, A., Almanza-Pérez, J. C., Giacoman-Martínez, A., Ortiz-Andrade, R., León-Rivera, I., et al. (2021). Synthesis, molecular docking, dynamic simulation and pharmacological characterization of potent multifunctional agent (dual GPR40-PPARγ agonist) for the treatment of experimental type 2 diabetes. *Eur. J. Pharmacol.* 907, 174244. doi:10.1016/j.ejphar.2021.174244
- Holstein, A., Beil, W., and Kovacs, P. (2012). CYP2C metabolism of oral anti-diabetic drugs—impact on pharmacokinetics, drug interactions and pharmacogenetic aspects. *Expert Opin. Drug Metab. Toxicol.* 8, 1549–1563. doi:10.1517/17425255.2012.722619
- Hu, L., Zhou, Z., Deng, L., Ren, Q., Cai, Z., Wang, B., et al. (2020). HWL-088, a new and highly effective FFA1/PPARδ dual agonist, attenuates nonalcoholic steatohepatitis by regulating lipid metabolism, inflammation and fibrosis. *J. Pharm. Pharmacol.* 72, 1564–1573. doi:10.1111/jphp.13342
- Huan, Y., Jiang, Q., Li, G., Bai, G., Zhou, T., Liu, S., et al. (2017). The dual DPP4 inhibitor and GPR119 agonist HBK001 regulates glycemic control and beta cell function *ex vivo* and *in vivo*. *Sci. Rep.* 7, 4351. doi:10.1038/s41598-017-04633-5
- Huang, H., Meegalla, S. K., Lanter, J. C., Winters, M. P., Zhao, S., Littrell, J., et al. (2019). Discovery of a GPR40 superagonist: the impact of aryl propionic acid α-fluorination. *ACS Med. Chem. Lett.* 10, 16–21. doi:10.1021/acsmchemlett.8b00444
- Hye Khan, M. A., Kolb, L., Skibba, M., Hartmann, M., Blöcher, R., Proschak, E., et al. (2018). A novel dual PPAR-γ agonist/sEH inhibitor treats diabetic complications in a rat model of type 2 diabetes. *Diabetologia* 61, 2235–2246. doi:10.1007/s00125-018-4685-0
- Ibrahim, M. K., Eissa, I. H., Abdallah, A. E., Metwaly, A. M., Radwan, M. M., and ElSohly, M. A. (2017a). Design, synthesis, molecular modeling and anti-hyperglycemic evaluation of novel quinoxaline derivatives as potential PPARγ and SUR agonists. *Bioorg. Med. Chem.* 25, 1496–1513. doi:10.1016/j.bmc.2017.01.015
- Ibrahim, M. K., Eissa, I. H., Alesawy, M. S., Metwaly, A. M., Radwan, M. M., and ElSohly, M. A. (2017b). Design, synthesis, molecular modeling and anti-hyperglycemic evaluation of quinazolin-4(3H)-one derivatives as potential PPARγ and SUR agonists. *Bioorg. Med. Chem.* 25, 4723–4744. doi:10.1016/j.bmc.2017.07.015
- Kamata, S., Honda, A., and Ishii, I. (2023). Current clinical trial status and future prospects of PPAR-targeted drugs for treating nonalcoholic fatty liver disease. *Biomolecules* 13, 1264. doi:10.3390/biom13081264
- Korn, M., Ehrt, C., Ruggiu, F., Gastreich, M., and Rarey, M. (2023). Navigating large chemical spaces in early-phase drug discovery. *Curr. Opin. Struct. Biol.* 80, 102578. doi:10.1016/j.sbi.2023.102578
- Krishnan, N., Krishnan, K., Connors, C. R., Choy, M. S., Page, R., Peti, W., et al. (2015). PTP1B inhibition suggests a therapeutic strategy for Rett syndrome. *J. Clin. Invest.* 125, 3163–3177. doi:10.1172/JCI80323
- Kumar, M., Kumar Singh, P., Choudhary, S., and Silakari, O. (2022). Hydantoin based dual inhibitors of ALR2 and PARP-1: design, synthesis, *in-vitro* and *in-vivo* evaluation. *Bioorg. Chem.* 129, 106108. doi:10.1016/j.bioorg.2022.106108
- Kumari, S., Carmona, A. V., Tiwari, A. K., and Trippier, P. C. (2020). Amide bond bioisosteres: strategies, synthesis, and successes. *J. Med. Chem.* 63, 12290–12358. doi:10.1021/acs.jmedchem.0c00530
- Kuo, G.-H., Gaul, M. D., Liang, Y., Xu, J. Z., Du, F., Hornby, P., et al. (2018). Synthesis and biological evaluation of benzocyclobutane-C-glycosides as potent and orally active SGLT1/SGLT2 dual inhibitors. *Bioorg. Med. Chem. Lett.* 28, 1182–1187. doi:10.1016/j.bmcl.2018.02.057
- Lamotte, Y., Faucher, N., Sançon, J., Pineau, O., Sautet, S., Fouchet, M.-H., et al. (2014). Discovery of novel indazole derivatives as dual angiotensin II antagonists and partial PPARγ agonists. *Bioorg. Med. Chem. Lett.* 24, 1098–1103. doi:10.1016/j.bmcl.2014.01.004
- Landrum, G. RDKit. Available at: <http://rdkit.org/> (Accessed July 11, 2023).
- Lapuerta, P., Zambrowicz, B., Strumph, P., and Sands, A. (2015). Development of sotagliflozin, a dual sodium-dependent glucose transporter 1/2 inhibitor. *Diab. Vasc. Dis. Res.* 12, 101–110. doi:10.1177/1479164114563304
- Lei, L., Liu, Q., Liu, S., Huan, Y., Sun, S., Chen, Z., et al. (2015). Antidiabetic potential of a novel dual-target activator of glucokinase and peroxisome proliferator activated receptor-γ. *Metabolism* 64, 1250–1261. doi:10.1016/j.metabol.2015.06.014
- Li, Y., Tian, K., Qin, A., Zhang, L., Huo, L., Lei, L., et al. (2014). Discovery of novel urea derivatives as dual-target hypoglycemic agents that activate glucokinase and PPARγ. *Eur. J. Med. Chem.* 76, 182–192. doi:10.1016/j.ejmech.2014.02.024
- Li, Z., Zhou, Z., Deng, F., Li, Y., Zhang, D., and Zhang, L. (2018). Design, synthesis, and biological evaluation of novel pan agonists of FFA1, PPARγ and PPARδ. *Eur. J. Med. Chem.* 159, 267–276. doi:10.1016/j.ejmech.2018.09.071
- Li, Z., Chen, Y., Zhou, Z., Deng, L., Xu, Y., Hu, L., et al. (2019a). Discovery of first-in-class thiazole-based dual FFA1/PPARδ agonists as potential anti-diabetic agents. *Eur. J. Med. Chem.* 164, 352–365. doi:10.1016/j.ejmech.2018.12.069
- Li, Z., Hu, L., Wang, X., Zhou, Z., Deng, L., Xu, Y., et al. (2019b). Design, synthesis, and biological evaluation of novel dual FFA1 (GPR40)/PPARδ agonists as potential anti-diabetic agents. *Bioorg. Chem.* 92, 103254. doi:10.1016/j.bioorg.2019.103254
- Li, Z., Zhou, Z., Hu, L., Deng, L., Ren, Q., and Zhang, L. (2020). ZLY032, the first-in-class dual FFA1/PPARδ agonist, improves glucolipid metabolism and alleviates hepatic fibrosis. *Pharmacol. Res.* 159, 105035. doi:10.1016/j.phrs.2020.105035
- Li, Z., Ren, Q., Zhou, Z., Cai, Z., Wang, B., Han, J., et al. (2021). Discovery of the first-in-class dual PPARδ/γ partial agonist for the treatment of metabolic syndrome. *Eur. J. Med. Chem.* 225, 113807. doi:10.1016/j.ejmech.2021.113807
- Lillich, F. F., Imig, J. D., and Proschak, E. (2020). Multi-target approaches in metabolic syndrome. *Front. Pharmacol.* 11, 554961. doi:10.3389/fphar.2020.554961
- Liu, K., Zhao, X., Qi, X., Hou, D.-L., Li, H.-B., Gu, Y.-H., et al. (2021). Design, synthesis, and biological evaluation of a novel dual peroxisome proliferator-activated receptor alpha/delta agonist for the treatment of diabetic kidney disease through anti-inflammatory mechanisms. *Eur. J. Med. Chem.* 218, 113388. doi:10.1016/j.ejmech.2021.113388
- Maccari, R., and Ottanà, R. (2015). Targeting aldose reductase for the treatment of diabetes complications and inflammatory diseases: new insights and future directions. *J. Med. Chem.* 58, 2047–2067. doi:10.1021/jm500907a



- Madariaga-Mazón, A., Naveja, J. J., Medina-Franco, J. L., Noriega-Colima, K. O., and Martínez-Mayorga, K. (2021). DiaNat-DB: a molecular database of antidiabetic compounds from medicinal plants. *RSC Adv.* 11, 5172–5178. doi:10.1039/d0ra10453a
- Mahindroo, N., Huang, C.-F., Peng, Y.-H., Wang, C.-C., Liao, C.-C., Lien, T.-W., et al. (2005). Novel indole-based peroxisome proliferator-activated receptor agonists: design, SAR, structural biology, and biological activities. *J. Med. Chem.* 48, 8194–8208. doi:10.1021/jm0506930
- Maier, J. A., Martínez, 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
- Makhoba, X. H., Viegas, C., Jr, Mosa, R. A., Viegas, F. P. D., and Poole, O. J. (2020). Potential impact of the multi-target drug approach in the treatment of some complex diseases. *Drug Des. devel. Ther.* 14, 3235–3249. doi:10.2147/DDDT.S257494
- Medina-Franco, J. L., Chávez-Hernández, A. L., López-López, E., and Saldívar-González, F. I. (2022). Chemical multiverse: an expanded view of chemical space. *Mol. Inf.* 41, e2200116. doi:10.1002/minf.202200116
- MolVS Molecule validation and standardization — MolVS 0.1.1 documentation. Available at: <https://molvs.readthedocs.io/en/latest/> (Accessed July 11, 2023).
- Otake, K., Azukizawa, S., Takahashi, K., Fukui, M., Shibabayashi, M., Kamemoto, H., et al. (2011). 2-Acyl-tetrahydroisoquinoline-3-carboxylic acids: lead compounds with triple actions, peroxisome proliferator-activated receptor  $\alpha/\gamma$  agonist and protein-tyrosine phosphatase 1B inhibitory activities. *Chem. Pharm. Bull.* 59, 876–879. doi:10.1248/cpb.59.876
- Otake, K., Azukizawa, S., Fukui, M., Kunishiro, K., Kamemoto, H., Kanda, M., et al. (2012). Novel (S)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acids: peroxisome proliferator-activated receptor  $\gamma$  selective agonists with protein-tyrosine phosphatase 1B inhibition. *Bioorg. Med. Chem.* 20, 1060–1075. doi:10.1016/j.bmc.2011.11.035
- Otake, K., Azukizawa, S., Takeda, S., Fukui, M., Kawahara, A., Kitao, T., et al. (2015). Novel 2,7-substituted (S)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acids: peroxisome proliferator-activated receptor  $\gamma$  partial agonists with protein-tyrosine phosphatase 1B inhibition. *Chem. Pharm. Bull.* 63, 998–1014. doi:10.1248/cpb.c15-00508
- OTAVACHemicals Ltd - synthetic organic compounds for research and drug discovery. Available at: <https://www.otavachemicals.com> (Accessed February 23, 2023).
- Packer, M. (2023). Dual SGLT1 and SGLT2 inhibitor sotagliflozin achieves FDA approval: landmark or landmine? *Nat. Cardiovasc. Res.* 2, 705–707. doi:10.1038/s44161-023-00306-x
- Park, M. H., Kim, D. H., Kim, M. J., Lee, E. K., An, H. J., Jeong, J. W., et al. (2016). Effects of MHY908, a new synthetic PPAR $\alpha/\gamma$  dual agonist, on inflammatory responses and Insulin resistance in aged rats. *J. Gerontol. A Biol. Sci. Med. Sci.* 71, 300–309. doi:10.1093/gerona/glv043
- Piemontese, L., Fracchiolla, G., Carrieri, A., Parente, M., Laghezza, A., Carbonara, G., et al. (2015). Design, synthesis and biological evaluation of a class of bioisosteric oximes of the novel dual peroxisome proliferator-activated receptor  $\alpha/\gamma$  ligand LT175. *Eur. J. Med. Chem.* 90, 583–594. doi:10.1016/j.ejmech.2014.11.044
- Press, N. J., Taylor, R. J., Fullerton, J. D., Tranter, P., McCarthy, C., Keller, T. H., et al. (2012). Solubility-driven optimization of phosphodiesterase-4 inhibitors leading to a clinical candidate. *J. Med. Chem.* 55, 7472–7479. doi:10.1021/jm300459a
- Press, N. J., Taylor, R. J., Fullerton, J., Tranter, P., McCarthy, C., Keller, T. H., et al. (2015). Discovery and optimization of 4-(8-(3-Fluorophenyl)-1,7-naphthyridin-6-yl) trans-cyclohexanecarboxylic acid, an improved PDE4 inhibitor for the treatment of chronic obstructive pulmonary disease (COPD). *J. Med. Chem.* 58, 6747–6752. doi:10.1021/acs.jmedchem.5b00902
- Probst, D., and Reymond, J.-L. (2020). Visualization of very large high-dimensional data sets as minimum spanning trees. *J. Cheminform.* 12, 12. doi:10.1186/s13321-020-0416-x
- Rarey, M., Nicklaus, M. C., and Warr, W. (2022). Special issue on reaction informatics and chemical space. *J. Chem. Inf. Model.* 62, 2009–2010. doi:10.1021/acs.jcim.2c00390
- Ren, Q., Deng, L., Zhou, Z., Wang, X., Hu, L., Xie, R., et al. (2020). Design, synthesis, and biological evaluation of novel dual PPAR $\alpha/\delta$  agonists for the treatment of T2DM. *Bioorg. Chem.* 101, 103963. doi:10.1016/j.bioorg.2020.103963
- Ruddigkeit, L., van Deursen, R., Blum, L. C., and Reymond, J.-L. (2012). Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* 52, 2864–2875. doi:10.1021/ci300415d
- Ruiz, F. X., Cousido-Siah, A., Porté, S., Domínguez, M., Crespo, I., Rechlin, C., et al. (2015). Structural determinants of the selectivity of 3-benzyluracil-1-acetic acids toward human enzymes aldose reductase and AKR1B10. *ChemMedChem* 10, 1989–2003. doi:10.1002/cmdc.201500393
- Schattenberg, J. M., Pares, A., Kowdley, K. V., Heneghan, M. A., Caldwell, S., Pratt, D., et al. (2021). A randomized placebo-controlled trial of elafibranor in patients with primary biliary cholangitis and incomplete response to UDCA. *J. Hepatol.* 74, 1344–1354. doi:10.1016/j.jhep.2021.01.013
- Seo, S. M., Ihm, S. H., Yi, J.-E., Jeong, S. H., and Kim, B.-S. (2022). Comparative efficacy and safety of fimasartan in patients with hypertension: a network meta-analysis of randomized controlled trials. *J. Clin. Hypertens.* 24, 971–983. doi:10.1111/jch.14536
- Shah, N., Abdalla, M. A., Deshmukh, H., and Sathyapalan, T. (2021). Therapeutics for type-2 diabetes mellitus: a glance at the recent inclusions and novel agents under development for use in clinical practice. *Ther. Adv. Endocrinol. Metab.* 12, 20420188211042145. doi:10.1177/20420188211042145
- Skoraczynski, G., Kitlas, M., Miasojedow, B., and Gambin, A. (2023). Critical assessment of synthetic accessibility scores in computer-assisted synthesis planning. *J. Cheminform.* 15, 6–9. doi:10.1186/s13321-023-00678-z
- Sodano, T. M., Combee, L. A., and Stephenson, C. R. J. (2020). Recent advances and outlook for the isosteric replacement of anilines. *ACS Med. Chem. Lett.* 11, 1785–1788. doi:10.1021/acsmchemlett.9b00687
- Song, H.-P., Tian, K., Lei, L., Shen, Z.-F., Liu, S.-X., Zhang, L.-J., et al. (2011). Novel N-(pyrimidin-4-yl)thiazol-2-amine derivatives as dual-action hypoglycemic agents that activate GK and PPAR $\gamma$ . *Yao Xue Xue Bao* 1, 166–171. doi:10.1016/j.apsb.2011.07.002
- Subbaiah, M. A. M., and Meanwell, N. A. (2021). Bioisosteres of the phenyl ring: recent strategic applications in lead optimization and drug design. *J. Med. Chem.* 64, 14046–14128. doi:10.1021/acs.jmedchem.1c01215
- Tassopoulou, V.-P., Tzara, A., and Kourounakis, A. P. (2022). Design of improved antidiabetic drugs: a journey from single to multitarget Agents. *ChemMedChem* 17, e202200320. doi:10.1002/cmdc.202200320
- Veith, H., Southall, N., Huang, R., James, T., Fayne, D., Artemenko, N., et al. (2009). Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nat. Biotechnol.* 27, 1050–1055. doi:10.1038/nbt.1581
- Vilar, S., Cozza, G., and Moro, S. (2008). Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr. Top. Med. Chem.* 8, 1555–1572. doi:10.2174/156802608786786624
- Walters, W. P. (2019). Virtual chemical libraries. *J. Med. Chem.* 62, 1116–1124. doi:10.1021/acs.jmedchem.8b01048
- Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* 28, 31–36. doi:10.1021/ci00057a005
- 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
- Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., et al. (2021). ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* 49, W5–W14. doi:10.1093/nar/gkab255
- Xu, G., Gaul, M. D., Kuo, G.-H., Du, F., Xu, J. Z., Wallace, N., et al. (2018). Design, synthesis and biological evaluation of (2S,3R,4R,5S,6R)-5-fluoro-6-(hydroxymethyl)-2-aryltetrahydro-2H-pyran-3,4-diols as potent and orally active SGLT dual inhibitors. *Bioorg. Med. Chem. Lett.* 28, 3446–3453. doi:10.1016/j.bmcl.2018.09.025
- Xu, G., Du, F., Kuo, G.-H., Xu, J. Z., Liang, Y., Demarest, K., et al. (2020). 5,5-Difluoro- and 5-fluoro-5-methyl-hexose-based c-glucosides as potent and orally bioavailable SGLT1 and SGLT2 dual inhibitors. *Bioorg. Med. Chem. Lett.* 30, 127387. doi:10.1016/j.bmcl.2020.127387
- Yang, T., Li, Z., Chen, Y., Feng, D., Wang, G., Fu, Z., et al. (2021). DrugSpaceX: a large screenable and synthetically tractable database extending drug space. *Nucleic Acids Res.* 49, D1170–D1178. doi:10.1093/nar/gkaa920
- Zhang, C., Crawford, J. J., Landry, M. L., Chen, H., Kenny, J. R., Khojasteh, S. C., et al. (2020). Strategies to mitigate the bioactivation of aryl amines. *Chem. Res. Toxicol.* 33, 1950–1959. doi:10.1021/acs.chemrestox.0c00138
- Zhou, Z., Cai, Z., Zhang, C., Yang, B., Chen, L., He, Y., et al. (2022). Design, synthesis, and biological evaluation of novel dual FFA1 and PPAR $\delta$  agonists possessing phenoxycetic acid scaffold. *Bioorg. Med. Chem.* 56, 116615. doi:10.1016/j.bmc.2022.116615



## OPEN ACCESS

## EDITED BY

Rafael Peláez,  
University of Salamanca, Spain

## REVIEWED BY

Parvin Razzaghi,  
Institute for Advanced Studies in Basic Sciences  
(IASBS), Iran  
Karim Abbasi,  
Sharif University of Technology, Iran

## \*CORRESPONDENCE

Chaokun Yan,  
✉ ckyan@henu.edu.cn

RECEIVED 13 November 2023

ACCEPTED 18 January 2024

PUBLISHED 07 February 2024

## CITATION

Luo H, Zhu C, Wang J, Zhang G, Luo J and Yan C  
(2024), Prediction of drug–disease associations  
based on reinforcement symmetric metric  
learning and graph convolution network.  
*Front. Pharmacol.* 15:1337764.  
doi: 10.3389/fphar.2024.1337764

## COPYRIGHT

© 2024 Luo, Zhu, Wang, Zhang, Luo and Yan.  
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.

# Prediction of drug–disease associations based on reinforcement symmetric metric learning and graph convolution network

Huimin Luo<sup>1,2</sup>, Chunli Zhu<sup>1,2</sup>, Jianlin Wang<sup>1,2</sup>, Ge Zhang<sup>1,2</sup>,  
Junwei Luo<sup>3</sup> and Chaokun Yan<sup>1,2,4\*</sup>

<sup>1</sup>School of Computer and Information Engineering, Henan University, Kaifeng, China, <sup>2</sup>Henan Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng, China, <sup>3</sup>College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China, <sup>4</sup>Academy for Advanced Interdisciplinary Studies, Henan University, Zhengzhou, China

Accurately identifying novel indications for drugs is crucial in drug research and discovery. Traditional drug discovery is costly and time-consuming. Computational drug repositioning can provide an effective strategy for discovering potential drug-disease associations. However, the known experimentally verified drug-disease associations is relatively sparse, which may affect the prediction performance of the computational drug repositioning methods. Moreover, while the existing drug-disease prediction method based on metric learning algorithm has achieved better performance, it simply learns features of drugs and diseases only from the drug-centered perspective, and cannot comprehensively model the latent features of drugs and diseases. In this study, we propose a novel drug repositioning method named RSML-GCN, which applies graph convolutional network and reinforcement symmetric metric learning to predict potential drug-disease associations. RSML-GCN first constructs a drug–disease heterogeneous network by integrating the association and feature information of drugs and diseases. Then, the graph convolutional network (GCN) is applied to complement the drug–disease association information. Finally, reinforcement symmetric metric learning with adaptive margin is designed to learn the latent vector representation of drugs and diseases. Based on the learned latent vector representation, the novel drug–disease associations can be identified by the metric function. Comprehensive experiments on benchmark datasets demonstrated the superior prediction performance of RSML-GCN for drug repositioning.

## KEYWORDS

drug repositioning, drug-disease association prediction, graph convolutional network, metric learning, drug discovery

## 1 Introduction

Due to the high time cost, significant investment, and laborious of the traditional drug discovery process, it is challenging to meet the needs of people facing increasingly prevalent complex diseases such as cancer, diabetes, and cardiovascular disease (Chong and Sullivan, 2007; Tamimi and Ellis, 2009). Therefore, more accurately and effectively

capturing drug-related indications in drug development is of great significance. Drug repositioning, or the new use of old drugs, is an attractive means for discovering the new therapeutic potential for existing drugs that have already been approved by the Food and Drug Administration (FDA) for the treatment of diseases (Novac, 2013), so it has the advantages of reduced drug risk, a shortened clinical evaluation cycle, cost-effectiveness, and efficiency (Pushpakom et al., 2019; Luo et al., 2020). Many computational drug repositioning methods have been proposed to identify candidate indications of drugs (Lotfi Shahreza et al., 2017). These methods can be broadly classified into three major categories: (i) machine learning-based drug repositioning methods; (ii) network-based drug repositioning methods; and (iii) recommendation system-based drug repositioning methods.

Machine learning-based methods mainly utilize support vector machine (SVM) (Napolitano et al., 2013), logistic regression (Gottlieb et al., 2011; Qabaja et al., 2014), Naïve Bayes (Yang and Agarwal, 2011), and random forest (Oh et al., 2014) for classification and prediction tasks in drug repositioning. However, these traditional methods rely significantly on input data with features that have been artificially set up well to represent drug and disease characteristics, which results in a high level of implementation complexity (Yadav and Jadhav, 2019). As an extension of machine learning, deep learning has been popularly used in drug repositioning because it possesses inestimable advantages in automatically capturing nonlinear features from raw data. Zeng et al. (2019) put forward a network-based deep learning method, deepDR, which uses a multimodal deep autoencoder to learn nonlinear features of drugs from the heterogeneous networks. Network-based methods analyze the relationship between entities via message passing in different paths constructed by multiple data on the network structure, which is interpretable. Martínez et al. (2015) designed a heterogeneous network-based prioritization method to predict new drug-related diseases. Luo et al. (2016) proposed a bi-random walk (BiRW) algorithm on the drug–disease heterogeneous network to identify potential drug–disease associations. Recently, deep learning technologies have been successfully applied to drug repositioning and drug combination prediction. For example, Dehghan et al. proposed a novel multimodal deep learning-based approach called TripletMultiDTI, which incorporated multiple sources of information and used a new architecture to predict drug–target interaction affinity labels (Dehghan et al., 2022). Rafiei et al. presented a deep learning approach called DeepTraSynergy, which is designed to predict the synergistic effects of drug combinations in cancer treatment by utilizing various data including drug–target interactions, protein–protein interactions, and cell–target interactions to predict the synergistic effects of drug combinations in cancer treatment (Rafiei et al., 2023).

Recommendation system-based methods perform well in various recommend related domains including social media, e-commerce platforms, and personalized reading (Da'u and Salim, 2020). Similar to the recommendation of preferring items to users, the problem of predicting drug–disease associations can be modeled as the problem of recommending potential drugs as potential treatment to diseases (Yang et al., 2019a; Meng et al., 2022). Recently, recommended methods based on matrix factorization and matrix completion have been applied with considerable success to drug repositioning (Yang et al., 2020).

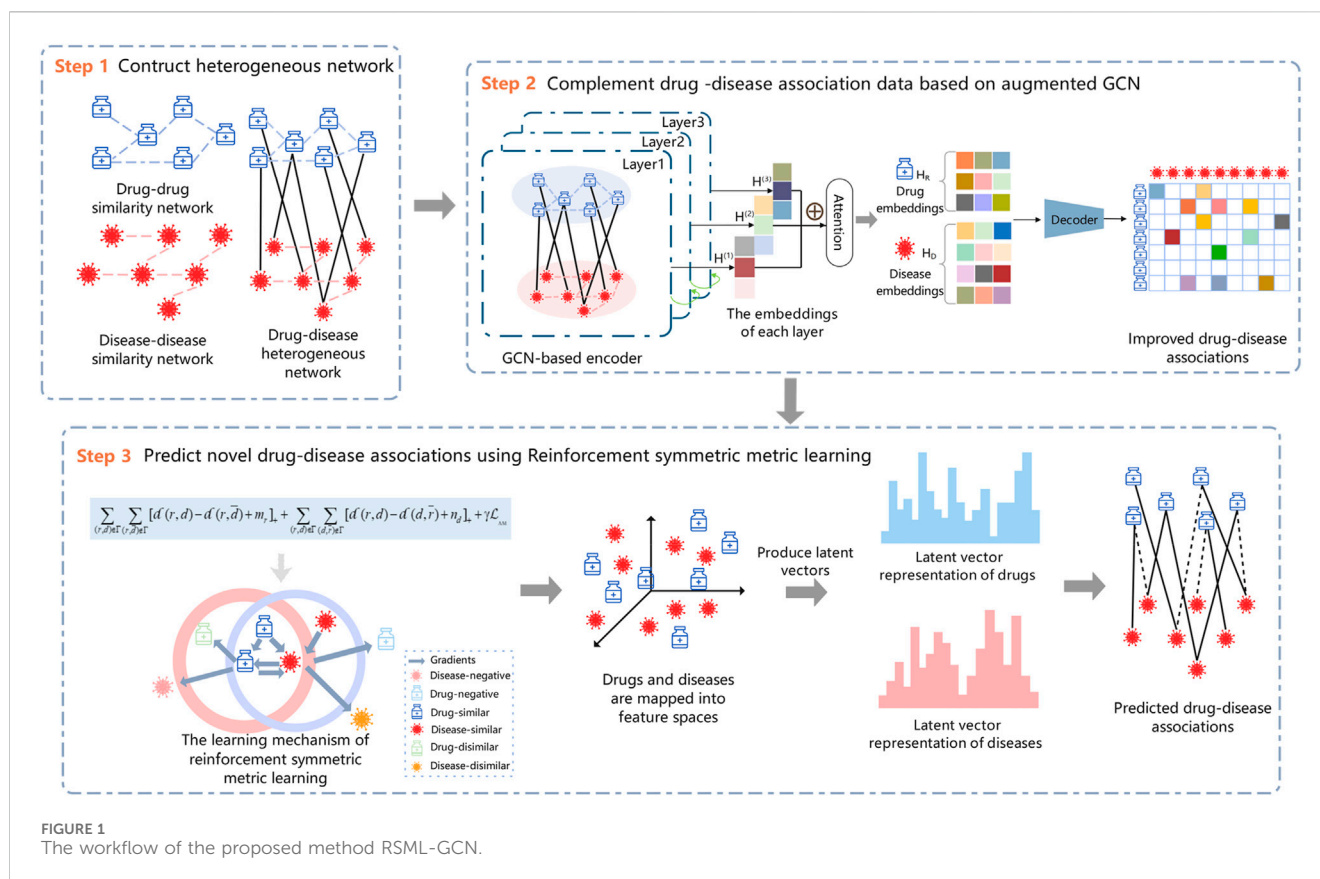
Luo et al. (2018) proposed a drug repositioning recommendation system (DRRS) that uses a fast singular value threshold (SVT) algorithm (Cai et al., 2010) to fill out the unknown entries in the drug–disease adjacency matrix. Yang et al. (2019b) used the generalized matrix factorization method (GMF) involved in the collaborative filtering process to uncover the potential therapeutic relationship between drugs and diseases. Methods based on matrix factorization or matrix completion can be applied flexibly but are inefficient for large-scale data owing to complex matrix operations. In particular, the inner product operation used in the most typical matrix factorization technology violates the triangle inequality rule, potentially leading to suboptimal performance in the recommended models (He et al., 2017). In addition, this simple linear combination overlooks the modeling of the drug–drug and disease–disease relationship in a manner, and only measures the drug–disease relationship. Hence, metric learning is proposed to offset gaps in matrix factorization to enhance the expressiveness of the model. Metric learning methods have been introduced to drug repositioning in the latest studies. For instance, Luo et al. (2021) proposed a collaborative metric learning approach (CMLDR) for drug repositioning. CMLDR projected drugs and diseases into a joint metric space and then predicted the potential drug–disease pairs from the learned vectors by metric learning. While CMLDR has achieved better prediction performance, it concentrated solely on drug-centric learning to learn representations of drugs and diseases based on drug–disease association information.

Graph convolutional network (GCN) (Kipf et al., 2017) extends the convolutional neural network to solve non-Euclidean space problems. It uses structural information on the constructed network by applying convolutional operation to learn network topology preserving node-level feature embeddings to reflect complex biological entity interactions. Recently, GCN has been applied to network analysis to efficiently extract network topology feature. For drug repositioning, GCN can be utilized to extract drug and disease features from the drug–disease heterogeneous network. Then, the extracted features can be further used to calculate drug–disease association scores.

In this study, we proposed a novel computational framework for drug repositioning based on reinforcement symmetric metric learning and GCN. First, in order to alleviate the sparsity problem of drug–disease association data, we utilized Graph Convolutional Network (GCN) on drug–disease heterogeneous network to learn the features of drugs and diseases. The drug–disease association scores can be calculated based on the learned features and are used to further complement the drug–disease association matrix, which can improve the prediction performance of the model. Then, a reinforcement symmetric metric learning method with adaptive margins is proposed, which combines with drug-centric and disease-centric learning simultaneously to learn the vector representation of drugs and diseases to predict new potential drug–disease associations. Finally, we propose to integrate reinforcement symmetric metric learning and GCN model to identify potential therapeutic indications of drugs, which can provide new insights for promoting drug repositioning.

The major contributions of this study are as follows.

- This study proposed a novel framework RSML-GCN, which integrated the symmetric metric learning algorithm and GCN model to identify potential therapeutic indications for drugs, which provides insights into promoting drug repositioning.



- To relieve the problem of the sparsity of drug–disease association data, RSML-GCN applied GCN to complement drug–disease association information.
- The symmetric metric learning algorithm incorporating drug-centric and disease-centric learning is proposed to predict novel potential drug–disease associations.

## 2 Materials and methods

In this study, we model the drug–disease association prediction as a recommendation problem and propose a new drug repositioning approach, RSML-GCN, to predict new therapies for diseases. The method combines GCN and metric learning to construct a novel framework for accurately discovering potential drug–disease associations, as shown in Figure 1. The proposed framework mainly consists of three modules including drug–disease network construction module, drug–disease complementation module and reinforcement symmetric metric learning-based prediction module. First, a drug–disease heterogeneous network is constructed based on the features and association information of drugs and diseases. Then, the low-dimensional embeddings of drugs and diseases are encoded by applying GCN, and a decoder is trained to generate an completed drug–disease association matrix by predicting drug–disease association scores. Finally, the latent representations of drugs and diseases are learned based on the reinforcement symmetric metric learning to predict novel drug–disease associations.

### 2.1 Construction of the drug–disease heterogeneous network

In this work, the similarity of drug pairs is calculated based on the Jaccard similarity coefficient, and the similarity of disease pairs is obtained by calculating the semantic similarity using medical subject descriptors. The detailed calculations are provided in [Supplementary Material](#). A drug similarity network  $R$  and disease similarity network  $D$  can be constructed based on drug similarity and disease similarity, and the edge weight is derived from the similarity value.  $A_r = \{r_1, r_2, \dots, r_M\}$  denotes the set of  $M$  drugs, and  $A_d = \{d_1, d_2, \dots, d_N\}$  denotes the set of  $N$  diseases.  $S_r \in \mathbb{R}^{M \times M}$  denotes the adjacency matrix of the drug similarity network, and  $S_d \in \mathbb{R}^{N \times N}$  denotes the adjacency matrix of the disease similarity network. A drug–disease association network  $S_{rd}$  can be constructed based on the known association information between drugs and diseases. An edge exists between  $r_i$  and  $d_j$  if there is a known association between drug  $r_i$  and disease  $d_j$ . The binary association matrix  $Y \in \{0, 1\}^{M \times N}$  corresponds to  $S_{rd}$ , the entry  $y_{ij}$  of the matrix  $Y$  is 1 if there is an edge between drug  $r_i$  and disease  $d_j$ , otherwise  $y_{ij} = 0$  which does not mean that there is no association between the drug  $r_i$  and disease  $d_j$ , but that there may be a potential association that has not yet been identified. For each drug  $r_i$ , this study aims to identify diseases that are potentially associated with  $r_i$ . The drug–disease heterogeneous network is constructed by integrating three networks: drug–drug similarity network, disease–disease similarity network, and drug–disease association network.



## 2.2 Complement drug–disease associations based on GCN

To solve the problem of the sparse verified drug–disease associations in drug repositioning, we can leverage the related information of drugs and diseases to predict potential indications of drugs to complement the drug–disease association data. GCN learns the low-dimensional representations of nodes from the irregular graph structure, and each of its layers aggregates the neighboring node information of the target node and uses the output of the previous layer as the input of the next layer, which is a process of continuously recursively aggregating neighborhood features. In this work, GCN is introduced by applying the similarity and association information to predict new drug–disease associations, which can complete the drug–disease association matrix from the biological network perspective and be used as a pre-training step to predict the likelihood of drug–disease associations.

First, the adjacency matrix  $G$  corresponding to the drug–disease heterogeneous network is defined.  $S'_r = E_r^{-1/2} S_r E_r^{-1/2}$  and  $S'_d = E_d^{-1/2} S_d E_d^{-1/2}$  are the normalized drug similarity matrix and disease similarity matrix, respectively, where  $E_r = \text{diag}(\Sigma_j S_{r_{ij}})$  and  $E_d = \text{diag}(\Sigma_j S_{d_{ij}})$  ( $S_{r_{ij}}$  or  $S_{d_{ij}}$  is  $(i, j)$ th entry of the similarity matrix) are the degree matrices of the drug and disease similarity matrices, respectively. The introduction of an appropriate degree of similarity contribution can better learn the embedding representation of drugs and diseases. Thus, a similarity penalty factor  $\mu$  is introduced to control the contribution of similarity information, which can be expressed as  $\hat{S}'_r = \mu^* S'_r$ ,  $\hat{S}'_d = \mu^* S'_d$ . Then, the adjacency matrix of the drug–disease heterogeneous network is represented by

$$G = \begin{bmatrix} \hat{S}'_r & Y \\ Y^T & \hat{S}'_d \end{bmatrix} \quad (1)$$

Given the matrix  $G$ , the general process of the convolution operation based on the GCN encoder according to the study of Yu et al. (2020a) can be described as

$$H^{l+1} = f(G, H, W) = \sigma(E^{-1/2} G E^{-1/2} H^l W^l) \quad (2)$$

Here,  $H^{l+1}$  is represented as the embeddings of nodes encoded at layer  $l + 1$ ,  $E$  ( $E = \text{diag}(\Sigma_j G_{ij})$ ) is the degree matrix of the adjacency matrix  $G$ , and  $H^l$  represents the embeddings encoded at layer  $l$ , which is used as the input at layer  $l + 1$ .  $W$  is a learnable weight matrix, and  $\sigma$  is a nonlinear activation function (e.g., RELU activation function).

Following the rule of Eq. 2, the GCN recursively learns node features. After  $l$  layers of iterations ( $l = 1, 2, \dots, L$ ), the GCN captures information about different structures of the heterogeneous network at different layers. To enable the GCN to fully learn the features of the nodes, we use the attention mechanism to connect the embeddings of different layers of GCN learning. Different attention weights are set at different layers. The final embeddings of the obtained drugs and diseases are denoted as  $[H^R \ H^D]^T = \sum \beta_l H^l$ . Here,  $\beta_l$  is initialized to  $1/(l + 1)$ ,  $H^R \in \mathbb{R}^{M \times k}$  and  $H^D \in \mathbb{R}^{N \times k}$  represent the final embeddings of the drugs and diseases, respectively.

To complement the drug–disease association matrix, we feed the final drugs and diseases embeddings into a bilinear decoder (Li et al., 2020b) for link prediction between drugs and diseases. Thus, the reconstruction of the drug–disease association matrix can be represented by  $\tilde{Y} = \rho(H^R W^T H^{D^T})$ , where  $\rho$  is the sigmoid activation

function, and  $W^T$  is the trainable weight matrix. Entry  $y'_{ij}$  in the matrix  $\tilde{Y}$  represents the predicted score between the drug  $r_i$  and the disease  $d_j$ .

Ultimately, we use a binary cross-entropy loss function as the objective function to optimize the drug–disease association continuously.

$$\text{Loss} = -\frac{1}{N \times M} \left( \eta \times \sum_{(i,j) \in \mathcal{J}^+} \log y'_{ij} + \sum_{(i,j) \in \mathcal{J}^-} \log(1 - y'_{ij}) \right) \quad (3)$$

where  $(i, j)$  indicates the drug–disease pair, and  $\eta = |\mathcal{J}^+|/|\mathcal{J}^-|$  indicates the ratio of the number of positive drug–disease pairs to the number of negative drug–disease pairs to balance positive and negative sample data.

We complement the drug–disease association information to alleviate the data sparsity problem by adopting GCN to implement pre-training on the drug–disease heterogeneous network. An entry of 1 in the drug–disease association matrix indicates that the disease is an indication for the drug and is a known association confirmed in clinical trials. In contrast, an entry of 0 means that there may be a potential association that has not yet been identified. GCN is utilized to preprocess unknown drug–disease associations to obtain more promising association information for subsequent prediction tasks. A threshold  $\theta$  is set to screen highly confident drug indications. Specifically, we retain the original value if the drug–disease prediction score is greater than or equal to  $\theta$ . Otherwise, we set it to 0 because a more considerable value suggests a stronger association between the drug and disease. Then, a preprocessed complemented drug–disease association matrix is obtained.

## 2.3 Reinforcement symmetric metric learning

Previous studies based on metric learning have considered drug-centric metrics (Hsieh et al., 2017; Park et al., 2018), neglecting to model drug–disease relationships from the disease perspective, which may lead to biased learning of latent vector representation of drugs and diseases, and limit the predictive performance of the model. Therefore, we take the drug- and disease-centric metrics into account for our reinforcement symmetric metric learning algorithm, which not only considers the relationships between drugs and diseases, but also implicitly establishes drug–drug and disease–disease relationships, thus enhancing the representation learning of drugs and diseases.

The goal of metric learning is to learn a metric function that pulls similar entities closer together and pushes dissimilar ones farther apart (Park et al., 2018; Wu et al., 2020). For example, when identifying possible favorite items for users in the recommendation system, metric learning assigns smaller distances to users and items with existing interactions and larger distances to users and items with unknown interactions. Similarly, it can be applied to the issue of predicting potential possible indications for drugs. The metric learning algorithms project drugs and diseases into the unified vector space and encode the latent vectors of drugs and diseases based on associations between drugs and diseases. This way, distances between drugs and diseases with known associations are closer than that between drugs and diseases without associations or with unknown associations. The likelihood of drug–disease associations is measured by the position of drugs and diseases in the unified metric vector space. Unvalidated diseases are

sorted in descending order by prediction scores for a given drug, and top-k disease recommendations can be obtained.

### 2.3.1 Problem formalization

In this work, the problem of recommending new indications for drugs is formulated as below.  $A_r$  and  $A_d$  denote the set of drugs and diseases, respectively, as described above. All known drug–disease associations can be designated as  $\Gamma = \{(r, d) | r \in A_r, d \in A_d\}$ , and  $N_i^+ = \{d_j | d_j \in \Gamma \text{ and } y_{ij} = 1\}$  represents the set of diseases with known associations with drug  $r_i$ .  $N_i^- = \{d_j | d_j \notin N_i^+ \text{ and } y_{ij} = 0\}$  represents the set of diseases without known associations with drug  $r_i$ .

Based on the completed drug–disease associations, the metric learning projects drugs and diseases into a unified n-dimensional metric vector space. In the unified metric vector space,  $\alpha_r \in \mathbb{R}^n$  is the latent vector of drug  $r$  and  $\beta_d \in \mathbb{R}^n$  is the latent vector of disease  $d$ . The association probability of drug  $r$  and disease  $d$  is measured by a simple and efficient Euclidean distance, defined as

$$\mathbf{d}(r, d) = \|\alpha_r - \beta_d\|_2^2, \quad (4)$$

where  $\|\cdot\|_2$  represents the  $L_2$ -normalization. The calculated Euclidean distance for known drug–disease associations should be smaller than that without known associations.

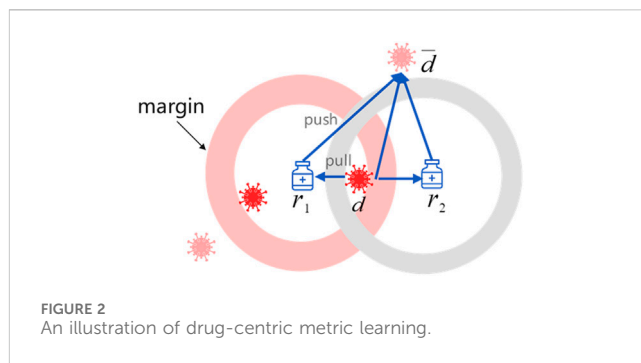
### 2.3.2 The drug-centric metric

Drug-centric metric learning is defined based on the completed drug–disease association matrix. For a given triple  $(r, d, \bar{d})$ ,  $(r, d) \in \Gamma$  represents a known association, which is considered a positive sample, and  $(r, \bar{d}) \notin \Gamma$  represents a negative sample, which is an unknown drug–disease pair that is randomly selected. Metric learning is a similarity measure based on distance, where a closer distance means two entities are more similar. Thus, the measure of similarity can be used for the measure of correlation. Distance and correlation are two opposite concepts in drug–disease association prediction. A closer distance indicates a more possible therapeutic behavior of the drug for the disease. To ensure better learning of latent vectors of drugs and diseases, we set a margin (safe distance)  $m$  and let  $m > 0$  (Li et al., 2020a). We use the following formula to ensure that the distance between drug  $r$  and negative disease  $\bar{d}$  is larger than the distance between drug  $r$  and positive disease  $d$ :

$$\mathbf{d}(r, d) + m \leq \mathbf{d}(r, \bar{d}) \quad (5)$$

Figure 2 illustrates the drug-centric metric learning method in a two-dimensional space, where the margin is designed to separate positive and negative pairs. Specifically, drugs and diseases are represented as latent vectors in a drug–disease metric space. If the predicted drug associated with one disease, the gradient direction moves inward to limit the disease within the safe margin, otherwise, the gradient direction moves outward to keep the disease away from the drug until it exceeds the safety margin. Note that the positive disease is inside the ball centered on drug  $r$ . However, the negative disease is outside this ball centered on drug  $r$ . This guarantees that distances between drugs and positive diseases are smaller than that between drugs and negative diseases, and maximizes the correlation between drugs and associated diseases.

As a result, we adopt triple loss (Schroff et al., 2015) as the objective function for drug-centric metric learning:



$$\mathcal{L}_R = \sum_{(r,d) \in \Gamma} \sum_{(r,\bar{d}) \notin \Gamma} [\mathbf{d}(r, d) - \mathbf{d}(r, \bar{d}) + m]_+ \quad (6)$$

where  $[x]_+ = \max(x, 0)$  denotes the standard hinge loss, which is a widely used loss function in the field of recommendation systems.

### 2.3.3 The disease-centric metric

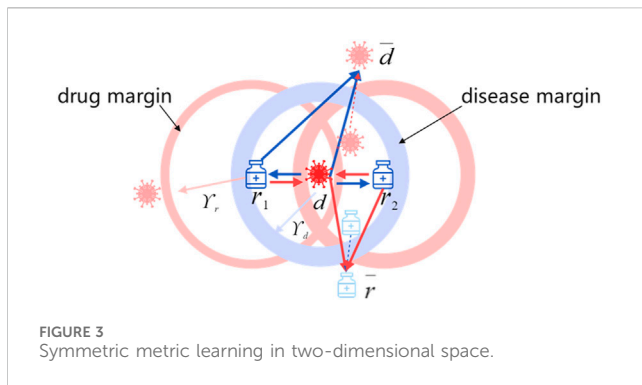
Drug-centric metric learning considers drug–disease associations from the drug perspective, thus bringing diseases associated with the targeted drug closer and having no association farther away. It is not sufficient to accurately locate the positions of drugs and diseases in the unified metric vector space to obtain their latent vectors only from the drug perspective. Moreover, drugs and diseases can be projected into the unified metric space based on the assumption that similar diseases are related to similar drugs (Xuan et al., 2019). Consequently, we introduce the disease-centric metric to explore the relationship between drugs and diseases from the disease perspective. Similarly to the drug-centric metric, for targeted disease, drugs with known associations with it are positioned close to it, or else far away.  $d$  and  $\bar{r}$  are uncorrelated according to the assumption of the distance metric, so they should not be closer together and should meet  $\mathbf{d}(d, \bar{r}) > \mathbf{d}(d, r)$ . Likewise, a margin  $n$  is set, and  $n > 0$ . The following equation is used to ensure that the distance between disease  $d$  and negative drug  $\bar{r}$  is larger than the distance between disease  $d$  and positive drug  $r$ :

$$\mathbf{d}(d, r) + n \leq \mathbf{d}(d, \bar{r}). \quad (7)$$

Because the Euclidean distance possesses symmetry, the disease-centric learning strategy can be replaced by  $\mathbf{d}(r, d) + n \leq \mathbf{d}(d, \bar{r})$ . Figure 3 depicts the symmetric metric learning approach centered on drugs and diseases under the explicit treatment relationship. The disease-centric metric predicts the associated drugs from the perspective of disease and uses the safety margin for gradient learning. Obviously, the objective of symmetric metric learning is to push drugs or diseases that are not associated out of the ball and pull drugs or diseases that are associated or have potential associations into the ball. Thus, distances of known drug–disease pairs are smaller than distances between unknown pairs.

Ultimately, the objective function for the disease-centric learning is defined as below:

$$\mathcal{L}_D = \sum_{(r,d) \in \Gamma} \sum_{(d,\bar{r}) \notin \Gamma} [\mathbf{d}(r, d) - \mathbf{d}(d, \bar{r}) + n]_+ \quad (8)$$



In this work, we aimed to identify the relationship between drugs and diseases from the standpoint of drugs and diseases rather than directly utilizing drug-centric metric learning.

### 2.3.4 Adaptive margin

Previous studies (Johannessen Landmark, 2008; Kingsmore et al., 2020) have found that one drug may treat multiple diseases, and that one disease may also be treated with various drugs. Considering the inconsistency of drug–disease and disease–drug association strengths, different margins are introduced for drugs and diseases. To simulate complicated drug–disease relationships better, we learn personalized margins through adaptive training. In the learning process, we set  $m_r$  and  $n_d$  as margins of the drug and disease, respectively. We prefer to use larger  $m_r$  and  $n_d$  to reduce variations. Particularly for drugs or diseases with fewer associations, more significant margins should be given to avoid overfitting, thus pushing drugs and diseases without associations farther to improve the accuracy of recommendations. Adaptive margins in the objective function can be expressed as

$$\mathcal{L}_{AM} = -\left(\frac{1}{|M|} \sum_r m_r + \frac{1}{|N|} \sum_d n_d\right) \quad (9)$$

### 2.3.5 Optimization

The number of unknown associations in the drug and disease-related data is significantly higher than the number of known associations. Therefore, we optimize the model by negative sampling. Based on known drug–disease associations, for each drug (disease), we randomly select  $P$  diseases (drugs) that are not associated with it as negative samples during the training process, and  $P$  is set as the minimum value of the number of drugs and diseases. By combining drug-centric and disease-centric metric learning losses, we obtain the final loss function for RSML-GCN as follows:

$$\begin{aligned} \mathcal{L} = (\mathcal{L}_R + \mathcal{L}_D) + \gamma \mathcal{L}_{AM} = & \sum_{(r,d) \in \Gamma} \sum_{(r,\bar{d}) \notin \Gamma} [\mathbf{d}(r,d) - \mathbf{d}(r,\bar{d}) + m_r]_+ \\ & + \sum_{(r,d) \in \Gamma} \sum_{(d,\bar{r}) \notin \Gamma} [\mathbf{d}(r,d) - \mathbf{d}(d,\bar{r}) + n_d]_+ \\ & + \gamma \mathcal{L}_{AM} \text{ s.t., } m_r \in (0, l], n_d \in (0, l] \end{aligned} \quad (10)$$

where  $l$  is used to prevent margins from being too large to affect the performance of the prediction. Additionally, to avoid the curse of dimensionality caused by the data points spread too widely, we apply

$l_2$ -norm clipping to the latent vectors of drugs and diseases learning, so that they are confined to the Euclidean ball with the size of  $l$  ( $\|\alpha_r\|_2 \leq l$  and  $\|\beta_d\|_2 \leq l$ ). The objective function is then optimized by using the AdaGrad to control the learning rate to update latent vectors continuously until convergence (Duchi et al., 2011). After the training procedure is completed, Euclidean distance is used to compute the association probabilities between drugs and diseases. A complete description about the procedure of RSML-GCN is presented in Algorithm 1.

**Input:** The matrix of known drug–disease associations  $Y \in \{0,1\}^{M \times N}$ ; The drug similarity matrix  $S_r \in \mathbb{R}^{M \times M}$ ; The disease similarity matrix  $S_d \in \mathbb{R}^{N \times N}$ ; Hyper parameters  $k, L, lr_1, n, lr_2$  and  $\gamma$ .

**Output:** The predicted drug–disease association matrix  $\hat{Y}$ .

```

1: Normalize drug similarity matrix  $S'_r$  and normalized
   disease similarity matrix  $S'_d$  and initialize
   drug–disease heterogeneous graph  $G$ .
2: repeat
3:   for  $l = 1, 2, \dots, L$  do
4:     Learn node features  $H^l$  with Eq. 2;
5:   end for
6:   Combine nodes embeddings  $H^l$  with  $\sum \beta_l H^l$ ,
   obtain the final embeddings of drugs  $H^R$ 
   and the final embeddings of diseases  $H^D$ ;
7:   Obtain the prediction matrix  $\tilde{Y}$  with  $\rho(H^R W^T H^D)$ ;
8:   Update parameters by optimizing Eq. 3;
9: until Eq. 3 is converged, get  $\tilde{Y}$ ;
10:  $\tilde{Y}^*$  is obtained by screen  $\tilde{Y}$  using a threshold  $\theta$ ;
11: for  $(r, d) \in$  sampled drug–disease associations in  $\tilde{Y}^*$  do
12:   sample a negative drug–disease  $(r, \bar{d})$  pairs to
   build a triplet  $(r, d, \bar{d})$ ;
13:   Compute  $\mathcal{L}_R$  with Eq. 6;
14:   sample a negative disease–drug  $(d, \bar{r})$  pairs to
   build a triplet  $(d, r, \bar{r})$ ;
15:   Compute  $\mathcal{L}_D$  with Eq. 8;
16: End for
17: While not converged do
18:   Compute gradients;
19:   Update  $\alpha_r$  and  $\beta_d$  with AdaGrad on Eq. 10;
20:   Compute the predict probability;
21:    $P = \|\alpha_r - \beta_d\|_2^2$ ;
22:   Check whether the model converges on the
   validation set;
23: End while
24:  $\hat{Y} = P$ ;
25: Return  $\hat{Y}$ ;

```

Algorithm 1. RSML-GCN Algorithm.

## 3 Results and discussion

### 3.1 Comparison with other methods

To verify the effectiveness of our method in predicting drug–disease associations, we compared RSML-GCN with

five state-of-the-art drug repositioning methods based on recommendation system and GCN including GRGMF (Zhang et al., 2020), DRWBNCF (Meng et al., 2022), LAGCN (Yu et al., 2020b), DRHGCN (Cai et al., 2021) and CMLDR (Luo et al., 2021). These methods are detailed below.

- GRGMF establishes a generalized matrix factorization model that obtains the latent representation of each node by adaptively learning the neighborhood information of each node, and it introduces external similarity information to facilitate the prediction of potential links.
- DRWBNCF is a neural collaborative filtering method that proposes a new weighted bilinear graph convolution operation to integrate the information of the known drug-disease association, drug's and disease's neighborhood, and neighborhood interaction into a unified representation to infer novel potential drug-disease associations.
- LAGCN is a layer attention GCN that uses GCN to learn embeddings of drugs and diseases from the drug-disease heterogeneous network. The learned embeddings are then integrated by an attention mechanism to predict new associations.
- DRHGCN uses GCN to extract inter-domain and intra-domain feature information of drugs and diseases to find new drug indications based on different network topology information of drugs and diseases in different domains.
- CMLDR is a collaborative metric learning algorithm that predicts the association probability of drugs and diseases by applying metric learning. The latent vectors of drugs and diseases are learned based on the known related information of drugs and diseases and used to identify candidate drug-disease associations.

For a fair comparison, we ran these competing methods with the optimal parameters suggested in the original papers on benchmark datasets. The complete evaluation of all methods was performed under 10-fold cross-validation. The specific experimental settings are described in [Supplementary Material](#). Also, we conducted parameter analysis and selected the best parameters as the recommended settings for RSML-GCN in this work.

## 3.2 Parameter setting

Considering that hyperparameters could affect model performance, we further investigate the influence of hyperparameters including that used in GCN, such as the latent vector dimension  $n$ , the marginal value strengths  $\gamma$ , and weight variables. The specific hyperparameter settings are given in [Supplementary Material](#). According to the previous study (Yu et al., 2020a), we set the parameters for GCN with the embedding dimension  $k = 64$ , number of layers  $L = 3$ , initial learning rate  $lr_1 = 0.008$ , node discard rate  $\beta = 0.6$ , regularize discard rate  $\xi = 0.4$ , and penalty factor  $\mu = 6$ . Moreover, we have

investigated the effect of the latent vector dimension  $n$  by varying its value from 30 to 400, and examined the influence of the marginal value strengths  $\gamma$  by varying its value from 0.01 to 100. The optimal parameters were determined by the grid search method, and detailed information is provided in the [Supplementary Material](#). Finally, the latent vector dimension of drugs and diseases in the metric space was fixed at 250, the initial learning rate  $lr_2$  was 0.05, and the batch size was 512. In terms of variables, refer to the settings of Li et al. (2020a), all weight variables followed a uniform distribution  $[-0.01, 0.01]$  and were randomly initialized, and all latent vectors (such as  $\alpha_r$ ,  $\beta_d$ ) that follow a normal distribution (mean: 0.1, variance: 0.03) were randomly initialized. More detailed parameter settings are described in [Supplementary Figures S1–S4](#).

## 3.3 Performance of RSML-GCN in cross-validation

To evaluate the performance of RSML-GCN, we conducted extensive experiments on two benchmark datasets Cdataset and Fdataset in [Supplementary Table S1](#) and compared RSML-GCN with five state-of-the-art association prediction methods. The performance evaluation results of all methods under 10 times 10-fold cross-validation were reported in [Table 1](#). The experimental results show that RSML-GCN had good performance in relevant metrics and was superior to other methods. In terms of the primary metric, AUPR, RSML-GCN achieved the highest average value of 0.7941, which surpasses GRGMF by 33.7%, and the average AUPR values of DRWBNCF, LAGCN, DRHGCN and CMLDR were 0.4992, 0.1562, 0.5480, and 0.2607, respectively. Additionally, RSML-GCN outperformed other methods in terms of AUC, with an average AUC value of 0.9077. This was 0.20% higher than the second-best method, DRHGCN. DRWBNCF, GRGMF, LAGCN and CMLDR have AUCs of 0.8642, 0.8994, 0.7874 and 0.7999, respectively.

We have performed 10 times 10-fold cross-validation and obtained AUC and AUPR values for all methods. The paired  $t$ -test is applied to statistically test the significance between the proposed method and other existing methods in terms of AUPR values, which have been conducted in previous studies. The paired  $t$ -test results including the  $p$ -values are showed in [Table 2](#). It can be observed that RSML-GCN is statistically significantly better than other methods ( $p < 0.05$ ).

The drug-disease prediction problem was formulated as a top-k recommendation problem, where potential therapeutic diseases are recommended for a specific drug. Therefore, we used top-k prediction results as evaluation metrics, specifically precision@K ( $p@K$ ) and recall@K ( $r@K$ ), which are widely used in recommendation domains. The performance of different models in predicting the top-k drug-disease associations on Cdataset was reported in [Supplementary Figure S5](#). RSML-GCN outperformed other models in terms of  $r@5$ ,  $r@10$ ,  $p@5$ , and  $p@10$ . Additionally, in [Supplementary Figure S6](#), we can find that RSML-GCN also achieves excellent performance in the recall



TABLE 1 Results of different methods under 10 iterations of 10-fold cross-validation.

Datasets	DRWBNCF	GRGMF	LAGCN	DRHGCN	CMLDR	RSML-GCN
AUPR						
Cdataset	0.4821	0.5611	0.1946	0.5562	0.1088	0.8580
Fdataset	0.5163	0.6269	0.1178	0.5397	0.4125	0.7302
Avg	0.4992	0.5940	0.1562	0.5480	0.2607	0.7941
AUC						
Cdataset	0.8480	0.8638	0.8358	0.8756	0.7650	0.9309
Fdataset	0.8803	0.9350	0.7389	0.9362	0.8348	0.8846
Avg	0.8642	0.8994	0.7874	0.9059	0.7999	0.9077

TABLE 2 The statistical significance of performance improvements achieved by RSML-GCN.

Paired t-test	Fdataset	Cdataset
RSML-GCN vs. DRWBNCF	8.44E-25	4.11E-29
RSML-GCN vs. GRGMF	3.92E-22	2.48E-32
RSML-GCN vs. LAGCN	1.81E-23	2.0E-28
RSML-GCN vs. DRHGCN	4.36E-17	4.83E-33
RSML-GCN vs. CMLDR	5.11E-29	2.48E-39

and precision values of the top-k predictions on Fdataset, which is much better than collaborative filtering-based, GCN-based, and metric learning-based methods. Notably, the performance indicators of LAGCN in these results were inferior to those of other methods, potentially due to GCN exhibiting over-smoothing issues stemming from dataset imbalances. The prediction results of the matrix factorization method GRGMF were lower than RSML-GCN, indicating that the metric learning method can effectively compensate for the shortcomings of matrix factorization. In contrast, CMLDR yielded significantly lower results than our proposed method, which suggests the usefulness of increasing the disease-centric auxiliary reuse learning for improving the drug-centric metric. The superior performance of RSML-GCN can be attributed to the following aspects. First, deep learning method is utilized to learn the potential representations of drugs and diseases and generate high confident drug-disease associations. This effectively alleviates the sparsity problem of drug-disease association data and improves the performance of subsequent task predictions. Second, we designed a reinforcement metric learning method to learn the metric between drugs and diseases from both drug and disease aspects, which can improve previous metric learning methods. Finally, by integrating the deep learning method and metric learning method, the proposed method can achieve better performance than other drug-disease prediction methods. Furthermore, we have avoided excessive integration of biological data, as improper handling of such data can introduce noise and adversely affect prediction results. These results comprehensively demonstrate the effectiveness of our proposed method in identifying drug-disease associations.

### 3.4 Ablation experiment

To evaluate the model performance of RSML-GCN, we set up a variant of RSML-GCN, named as RSML. In RSML, we used only reinforcement symmetric metric learning to predict drug-disease association scores, which removes the pre-training step of complementing the drug-disease association matrix using GCN. In order to check the contribution of the pre-training component, we compared RSML-GCN with RSML based on Cdataset.

Based on the drug-disease association matrix, the RSML projected drugs and diseases to the unified metric vector space and learned their latent vectors based on the push-pull mechanism. The Euclidean distance was adopted to obtain the potential treatment probabilities of drugs for diseases. As can be seen in [Supplementary Table S2](#), incorporating GCN in RSML-GCN as a pre-training step to complement the drug-disease association matrix resulted in improved predictive performance. The average AUPR of RSML-GCN was 6.45% higher than that of RSML, while maintaining a comparable AUC. Additionally, significant enhancements were observed across all top-k prediction evaluation metrics, as depicted in [Supplementary Figure S7](#). This improvement can be attributed to GCN's ability to integrate similarity information from drug-disease associations, enabling the learning of more comprehensive representations and acquiring more confident drug-disease association information. Consequently, this approach helps address the imbalance between positive and negative samples to serve downstream tasks better and improve the predictive potential of metric learning method. The results generally indicate the reliability of RSML-GCN for predicting drug-related diseases.

### 3.5 Predicting candidates for new drugs or new diseases

To assess the ability of RSML-GCN in predicting potential indications for new drugs, we removed the associated diseases of the test drug and predicted indications for it on Cdataset. To more accurately display the top-k recommendation performance of the model, we selected drugs associated with

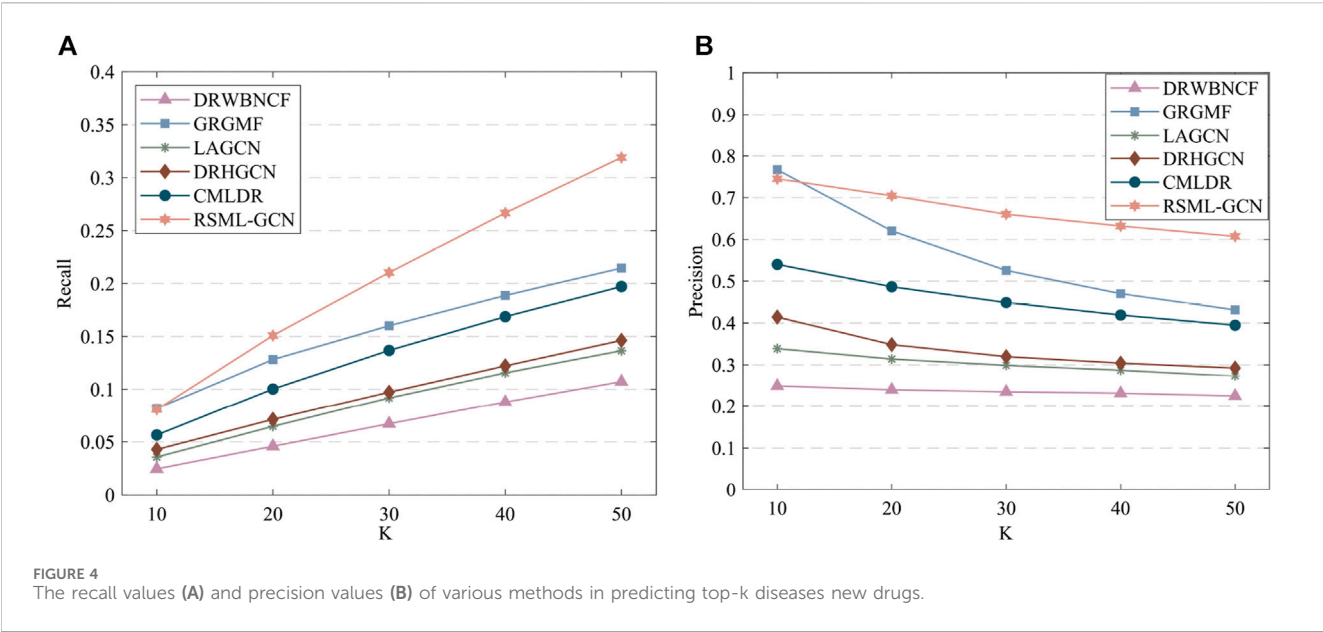


TABLE 3 Results on independent test set.

Methods	DRWBNCF	GRGMF	LAGCN	DRHGCN	CMLDR	RSML-GCN
AUPR	0.0353	0.0140	0.0220	0.0520	0.0459	0.3030
AUC	0.6218	0.5313	0.6215	0.7783	0.5355	0.6842

at least 50 diseases to evaluate the performance of RSML-GCN for new drug prediction. After training, the latent vectors of drugs and diseases in the training samples were learned. For a new drug without any known association, RSML-GCN could obtain latent vectors of the drug by utilizing similarity information from its  $h$ -nearest neighbors in the training set to predict the potential drug-related diseases. In the experiment, empirically,  $h$  was set to 5 to simplify the model.

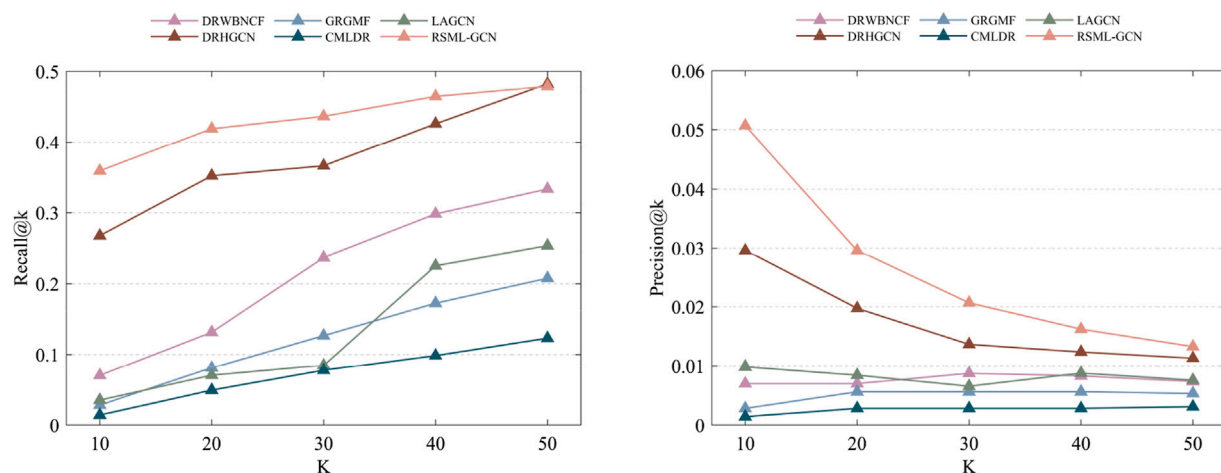
The results of predicting unknown diseases for new drugs are presented in [Supplementary Table S3](#), RSML-GCN exhibited the best performance in the primary metric AUPR (average AUPR = 0.5555), which is higher than GRGMF and CMLDR based on recommendation system by 49.0% and 74.4% (AUPR value), respectively. In terms of AUC, RSML-GCN had an average AUC of 0.6985, which is higher than that of these state-of-the-art prediction methods. The recall and precision of top- $k$  recommendations of RSML-GCN for predicting potential indications for new drugs were reported in [Figure 4](#), which shows the performance of RSML-GCN over other methods for different values of  $K$ . For the average recall value, our RSML-GCN performed better than other methods under most  $K$  values. For example, when  $K = 10$  and  $K = 50$ , RSML-GCN achieved the best average recall values, 0.0807 and 0.3191, respectively. In particular, when  $K = 10$ , DRWBNCF, LAGCN, DRHGCN, and CMLDR obtained recall values of 0.0245, 0.0356, 0.0428 and 0.0565, respectively, the recall values of GRGMF and RSML-GCN were almost comparable.

In addition, when  $K = 10$  and  $K = 50$ , RSML-GCN attained average precision values of 0.7451 and 0.6072, respectively, which is higher than most competitive methods. Overall, the comprehensive results demonstrate that RSML-GCN has an excellent ability to predict related diseases for new drugs.

For a new disease without any known associations, RSML-GCN can use the similarity information of diseases to predict potential candidate drugs for new diseases. We also conducted the experiments, in which all relationships for each disease were removed to predict candidate drugs for new diseases. The results compared with state-of-the-art methods were reported in [Supplementary Table S4](#) and [Supplementary Figure S8](#). RSML-GCN was the second-best, significantly better than DRWBNCF, LAGCN, DRHGCN, and CMLDR. The recall and precision of RSML-GCN also achieved the second-best performance. The reason is that the input of GRGMF contains both drug-drug similarity and disease-disease similarity, while the input of RSML-GCN only contains known drug-disease associations.

### 3.6 Independent test experiments

We also investigated the performance of these prediction methods on the independent test set, another dataset released by [Luo et al. \(2016\)](#) is used to assess the performance of methods. By removing the drugs not included in Fdataset, we



**FIGURE 5**  
The recall and precision values of the top-k recommended drug–disease associations are achieved by different methods on the independent test set.

obtained an independent test set consisting of 89 drug–disease associations involving 71 drugs and 313 diseases. This test set was used to assess the performances of all prediction methods in predicting the drug–disease associations on the Fdataset. Overall, the performance of all the methods moderately deteriorates relative to the 10-fold cross-validations. RSML-GCN remained the best method, which achieved an AUPR value of 0.3030 and an AUC value of 0.6842. DRWBNCF and LAGCN achieved AUC values of 0.6218 and 0.6215, respectively (Table 3). We also show the ability to correctly predict drug–disease associations concerning given top-k thresholds, as shown in Figure 5. Accordingly, RSML-GCN can predict drug–disease associations more accurately than all other five methods on almost every top-rank threshold.

### 3.7 Case study

In this section, we conducted a case study to further evaluate the reliable ability of RSML-GCN to predict novel drug–disease associations. For the analysis, we chose three representative drugs for the treatment of high-incidence diseases, Atorvastatin Calcium, Etoposide, and Riluzole. Atorvastatin Calcium is a commonly used lipid-lowering drug in the clinic, which is mainly used to treat mixed hyperlipidemia and hypercholesterolemia (Egom and Hafeez, 2016). These diseases have a high incidence, are difficult to diagnose and treat, and can potentially induce Cardio-cerebrovascular disease (Yao et al., 2019). Therefore, the analysis of Atorvastatin Calcium is of great significance. Etoposide is a cell cycle specific antitumor drug that is primarily effective against small cell lung cancer (Mascaux et al., 2000), acute leukemia, and malignant lymphoma. Given cancer is complicated and difficult to cure, it is valuable to analyze whether Etoposide can treat other similar diseases in drug reuse. Riluzole is a central nervous system drug that plays a pivotal role in the treatment of Alzheimer’s disease, Parkinson’s

disease, and brain injury, which have a serious impact on patients. Therefore, it is necessary to analyze the new therapeutic potential for this drug to treat a variety of neurological degenerative diseases. Specifically, we applied RSML-GCN to predict candidate diseases for three drugs. For each of the three drugs, all predicted candidate disease scores were ranked by priority, and then we excluded all known drug–disease associations from the primary dataset to generate a new top-ranked list of drug–disease associations. Finally, we used highly reliable sources and clinical trials (i.e., DrugBank (DB) (Law et al., 2013), CTD (Davis et al., 2016), PubChem (Kim et al., 2015), DrugCentral (Avram et al., 2020), and ClinicalTrials) as references to examine the predicted drug–disease associations. Table 4 presents the predicted results of the top 10 candidate diseases for three drugs. The results show Atorvastatin Calcium can also be shown to treat lung disease, left ventricular dysfunction, and is also associated with kidney failure, which are supported by CTD, ClinicalTrials, and DrugCentral. The discovery of Etoposide can be verified in all clinical trials, which shows that Etoposide not only has a good therapeutic effect on a variety of tumors but also can be used to treat Exanthema and drug eruption. In addition, Riluzole was also found to be related to heart failure, drug-induced liver injury, and arrhythmia. To sum up, most of our predictions can be verified by reliable sources and clinical trials. The case study results further demonstrate the effectiveness of RSML-GCN in predicting novel drug–disease associations.

## 4 Conclusion

In this study, we proposed a new framework for drug–disease association prediction by incorporating GCN and reinforced symmetric metric learning, named RSML-GCN. Firstly, in order to alleviate the sparsity problem of drug–disease association data, the GCN was applied to capture the structure of network topology on the heterogeneous network constructed by the biological knowledge

TABLE 4 The top-10 candidate diseases predicted by RSML-GCN for three drugs.

Drug	Rank	Disease	Evidences	Rank	Disease	Evidences
Atorvastatin Calcium	1	Liver Diseases	CTD/ClinicalTrials/DrugCentral	6	Headache	CTD
	2	Ventricular Dysfunction, Left	CTD	7	Hyperalgesia	CTD
	3	Liver Neoplasms	CTD/ClinicalTrials/DrugCentral	8	Renal Insufficiency	CTD/ClinicalTrials/DrugCentral
	4	Vomiting	CTD	9	Edema	CTD/ClinicalTrials
	5	Dizziness	NA	10	Weight Gain	CTD
Etoposide	1	Exanthema	CTD	6	Carcinoma, Squamous Cell	CTD/ClinicalTrials
	2	Drug Eruptions	CTD	7	Skin Neoplasms	CTD/ClinicalTrials
	3	Uterine Cervical Neoplasms	CTD/ClinicalTrials	8	Leukemia	CTD/ClinicalTrials/DB/PubChem/DrugCentral
	4	Carcinoma, Transitional Cell	CTD/ClinicalTrials	9	Lung Diseases, Interstitial	CTD/ClinicalTrials/DrugCentral
	5	Lymphoma, Large	CTD/ClinicalTrials/DB/PubChem/DrugCentral	10	Cerebellar Diseases	CTD/ClinicalTrials
Riluzole	1	Heart Failure	CTD	6	Drug-Related Side Effects and Adverse Reactions	CTD
	2	Chemical and Drug Induced Liver Injury	CTD/DrugCentral	7	Myocardial Infarction	CTD
	3	Acute Kidney Injury	CTD	8	Hypotension	CTD/ClinicalTrials
	4	Arrhythmias, Cardiac	CTD/ClinicalTrials	9	Rhabdomyolysis	NA
	5	Kidney Diseases	NA	10	Brady-cardia	CTD



and known association information of drugs and diseases to complement the missing drug–disease association information, which improves the prediction performance of the model. Secondly, the current metric learning algorithm only learns in a single way centered on drugs, ignoring the influence of diseases. Therefore, a reinforcement symmetric metric learning algorithm combined with drug-centric and disease-centric learning was developed to project drugs and diseases into a unified metric space, and learn their latent vector representations based on push–pull mechanisms to identify potential indications for known drugs and new drugs. Based on the assumption that similar drugs can treat similar diseases, the disease-centric metric learning mechanism was introduced symmetrically, which improved on the previous approach. Moreover, the adaptive margin strategy helped the model select the appropriate margin for different drugs and diseases. Thirdly, this study proposes a new framework integrating reinforcement symmetric metric learning algorithm and GCN model to identify potential therapeutic indications of drugs, which provides new insights for promoting drug repositioning. The results of extensive experiments demonstrated that RSML-GCN performed well and outperformed other drug–disease association prediction methods.

RSML-GCN only utilized drug–disease association data and the single feature information of the drug and the disease to predict potential associations. However, there exists various drug and disease related biological data, and the use of multiple data may help to learn potential indications for drugs. Therefore, in the future of work, more biological data including genes, targets, or miRNAs can be considered and integrated to build a more comprehensive heterogeneous network with multiple relationship types. In addition, the metric learning algorithm only uses known drug–disease association information as input. Future research should design an effective way to integrate related biological data into its learning process to predict potential drug–disease associations.

## 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.

## References

- Avram, S., Bologa, C. G., Holmes, J., Bocci, G., Wilson, T. B., Nguyen, D.-T., et al. (2020). DrugCentral 2021 supports drug discovery and repositioning. *Nucleic Acids Res.* 49 (D1), D1160–D1169. doi:10.1093/nar/gkaa997
- Cai, J. F., Candès, E. J., and Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* 20 (4), 1956–1982. doi:10.1137/080738970
- Cai, L., Lu, C., Xu, J., Meng, Y., Wang, P., Fu, X., et al. (2021). Drug repositioning based on the heterogeneous information fusion graph convolutional network. *Brief. Bioinforma.* 22 (6), bbab319. doi:10.1093/bib/bbab319
- Chong, C. R., and Sullivan, D. J. (2007). New uses for old drugs. *Nature* 448 (7154), 645–646. doi:10.1038/448645a
- Da'u, A., and Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. *Artif. Intell. Rev.* 53 (4), 2709–2748. doi:10.1007/s10462-019-09744-1
- Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., King, B. L., McMorran, R., et al. (2016). The comparative toxicogenomics database: update 2017. *Nucleic Acids Res.* 45 (D1), D972–D978. doi:10.1093/nar/gkw838
- Dehghan, A., Razzaghi, P., Abbasi, K., and Gharaghani, S. (2022). TripletMultiDTI: multimodal representation learning in drug-target interaction prediction with triplet loss function. *Expert Syst. Appl.* 232, 120754. doi:10.1016/j.eswa.2023.120754
- Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* 12 (7).
- Egom, E. E. A., and Hafeez, H. (2016). Biochemistry of statins. *Adv. Clin. Chem.* 73, 127–168. doi:10.1016/bs.acc.2015.10.005
- Gottlieb, A., Stein, G. Y., Rupp, E., and Sharan, R. (2011). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* 7 (1), 496. doi:10.1038/msb.2011.26

## Author contributions

HL: Conceptualization, Methodology, Writing–original draft, Writing–review and editing. CZ: Data curation, Methodology, Writing–original draft, Writing–review and editing. JW: Data curation, Formal Analysis. GZ: Methodology, Writing–original draft. JL: Writing–review and editing. CY: Writing–original draft, Writing–review and editing, Conceptualization.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (Grant Nos. 61802113, 61802114), and the Science and Technology Development Plan Project of Henan Province (Grant No. 212102210091).

## 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/fphar.2024.1337764/full#supplementary-material>

- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. (2017). "Neural collaborative filtering," in Proceedings of the 26th international conference on world wide web, Australia, April, 2017, 173–182.
- Hsieh, C.-K., Yang, L., Cui, Y., Lin, T.-Y., Belongie, S., and Estrin, D. (2017). "Collaborative metric learning," in Proceedings of the 26th International Conference on World Wide Web, Australia, April, 2017, 193–201.
- Johannessen Landmark, C. (2008). Antiepileptic drugs in non-epilepsy disorders: relations between mechanisms of action and clinical efficacy. *CNS Drugs* 22 (1), 27–47. doi:10.2165/00023210-200822010-00003
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2015). PubChem substance and compound databases. *Nucleic Acids Res.* 44 (D1), D1202–D1213. doi:10.1093/nar/gkv951
- Kingsmore, K. M., Grammer, A. C., and Lipsky, P. E. (2020). Drug repurposing to improve treatment of rheumatic autoimmune inflammatory diseases. *Nat. Rev. Rheumatol.* 16 (1), 32–52. doi:10.1038/s41584-019-0337-0
- Kipf, T. N., and Welling, M. (2017). "Semi-supervised classification with graph convolutional networks," in International Conference on Learning Representations (ICLR), Toulon, France, April, 2017.
- Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., et al. (2013). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42 (D1), D1091–D1097. doi:10.1093/nar/gkt1068
- Li, M., Zhang, S., Zhu, F., Qian, W., Zang, L., Han, J., et al. (2020a). "Symmetric metric learning with adaptive margin for recommendation," in Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, February, 2020, 4634–4641.
- Li, Z., Li, J., Nie, R., You, Z.-H., and Bao, W. (2020b). A graph auto-encoder model for miRNA-disease associations prediction. *Brief. Bioinforma.* 22 (4), bbaa240. doi:10.1093/bib/bbaa240
- Lotfi Shahreza, M., Ghadiri, N., Mousavi, S. R., Varshosaz, J., and Green, J. R. (2017). A review of network-based approaches to drug repositioning. *Briefings Bioinforma.* 19 (5), 878–892. doi:10.1093/bib/bbx017
- Luo, H., Li, M., Wang, S., Liu, Q., Li, Y., and Wang, J. (2018). Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics* 34 (11), 1904–1912. doi:10.1093/bioinformatics/bty013
- Luo, H., Li, M., Yang, M., Wu, F.-X., Li, Y., and Wang, J. (2020). Biomedical data and computational models for drug repositioning: a comprehensive review. *Briefings Bioinforma.* 22 (2), 1604–1619. doi:10.1093/bib/bbz176
- Luo, H., Wang, J., Li, M., Luo, J., Peng, X., Wu, F.-X., et al. (2016). Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* 32 (17), 2664–2671. doi:10.1093/bioinformatics/btw228
- Luo, H., Wang, J., Yan, C., Li, M., Wu, F. X., and Pan, Y. (2021). A novel drug repositioning approach based on collaborative metric learning. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 18 (2), 463–471. doi:10.1109/TCBB.2019.2926453
- Martínez, V., Navarro, C., Cano, C., Fajardo, W., and Blanco, A. (2015). DrugNet: network-based drug-disease prioritization by integrating heterogeneous data. *Artif. Intell. Med.* 63 (1), 41–49. doi:10.1016/j.artmed.2014.11.003
- Mascaux, C., Paesmans, M., Berghmans, T., Branle, F., Lafitte, J. J., Lemaitre, F., et al. (2000). A systematic review of the role of etoposide and cisplatin in the chemotherapy of small cell lung cancer with methodology assessment and meta-analysis. *Lung Cancer* 30 (1), 23–36. doi:10.1016/S0169-5002(00)00127-6
- Meng, Y., Lu, C., Jin, M., Xu, J., Zeng, X., and Yang, J. (2022). A weighted bilinear neural collaborative filtering approach for drug repositioning. *Briefings Bioinforma.* 23 (2), bbab581. doi:10.1093/bib/bbab581
- Napolitano, F., Zhao, Y., Moreira, V. M., Tagliaferri, R., Kere, J., D'Amato, M., et al. (2013). Drug repositioning: a machine-learning approach through data integration. *J. Cheminformatics* 5 (1), 30. doi:10.1186/1758-2946-5-30
- Novac, N. (2013). Challenges and opportunities of drug repositioning. *Trends Pharmacol. Sci.* 34 (5), 267–272. doi:10.1016/j.tips.2013.03.004
- Oh, M., Ahn, J., and Yoon, Y. (2014). A network-based classification model for deriving novel drug-disease associations and assessing their molecular actions. *PLoS ONE* 9 (10), e111668. doi:10.1371/journal.pone.0111668
- Park, C., Kim, D., Xie, X., and Yu, H. (2018). "Collaborative translational metric learning," in 2018 IEEE International Conference on Data Mining (ICDM), Singapore, November, 2018, 367–376.
- Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2019). Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18 (1), 41–58. doi:10.1038/nrd.2018.168
- Qabaja, A., Alshalalfa, M., Alanazi, E., and Alhaji, R. (2014). Prediction of novel drug indications using network driven biological data prioritization and integration. *J. Cheminformatics* 6 (1), 1. doi:10.1186/1758-2946-6-1
- Rafiei, F., Zeraati, H., Abbasi, K., Ghasemi, J. B., Parsaeian, M., and Masoudi-Nejad, A. (2023). DeepTraSynergy: drug combinations using multimodal deep learning with transformers. *Bioinformatics* 39 (8), btad438. doi:10.1093/bioinformatics/btad438
- Schroff, F., Kalenichenko, D., and Philbin, J. (2015). "FaceNet: a unified embedding for face recognition and clustering," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June, 2015, 815–823.
- Tamimi, N. A., and Ellis, P. J. N. C. P. (2009). Drug development: from concept to marketing. *Nephron Clin. Pract.* 113 (3), c125–c131. doi:10.1159/000232592
- Wu, H., Zhou, Q., Nie, R., and Cao, J. (2020). Effective metric learning with co-occurrence embedding for collaborative recommendations. *Neural Netw.* 124, 308–318. doi:10.1016/j.neunet.2020.01.021
- Xuan, P., Cao, Y., Zhang, T., Wang, X., Pan, S., and Shen, T. (2019). Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics* 35 (20), 4108–4119. doi:10.1093/bioinformatics/btz182
- Yadav, S. S., and Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* 6 (1), 113. doi:10.1186/s40537-019-0276-2
- Yang, L., and Agarwal, P. (2011). Systematic drug repositioning based on clinical side-effects. *PLoS one* 6, e28025. doi:10.1371/journal.pone.0028025
- Yang, M., Luo, H., Li, Y., and Wang, J. (2019a). Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* 35 (14), 455–463. doi:10.1093/bioinformatics/btz331
- Yang, M., Wu, G., Zhao, Q., Li, Y., and Wang, J. (2020). Computational drug repositioning based on multi-similarities bilinear matrix factorization. *Brief. Bioinforma.* 22 (4), bbab267. doi:10.1093/bib/bbaa267
- Yang, X., Zamit, I., Liu, Y., and He, J. (2019b). Additional Neural Matrix Factorization model for computational drug repositioning. *BMC Bioinform* 20 (1), 423. doi:10.1186/s12859-019-2983-2
- Yao, Q., Zhang, X., Huang, Y., Wang, H., Hui, X., and Zhao, B. (2019). Moxibustion for treating patients with hyperlipidemia: a systematic review and meta-analysis protocol. *Med. Baltim.* 98 (48), e18209. doi:10.1097/md.00000000000018209
- Yu, Z., Huang, F., Zhao, X., Xiao, W., and Zhang, W. (2020a). Predicting drug-disease associations through layer attention graph convolutional network. *Briefings Bioinforma.* 22 (4), bbab243. doi:10.1093/bib/bbaa243
- Yu, Z., Huang, F., Zhao, X., Xiao, W., and Zhang, W. (2020b). Predicting drug-disease associations through layer attention graph convolutional network. *Brief. Bioinforma.* 22 (4), bbab243. doi:10.1093/bib/bbaa243
- Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R., and Cheng, F. (2019). deepDR: a network-based deep learning approach to *in silico* drug repositioning. *Bioinformatics* 35 (24), 5191–5198. doi:10.1093/bioinformatics/btz418
- Zhang, W., Yue, X., Lin, W., Wu, W., Liu, R., Huang, F., et al. (2018). Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinforma.* 19 (1), 233. doi:10.1186/s12859-018-2220-4
- Zhang, Z.-C., Zhang, X.-F., Wu, M., Ou-Yang, L., Zhao, X.-M., and Li, X.-L. (2020). A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks. *Bioinformatics* 36 (11), 3474–3481. doi:10.1093/bioinformatics/btaa157



## OPEN ACCESS

## EDITED BY

Rafael Peláez,  
University of Salamanca, Spain

## REVIEWED BY

Stanislav Yanev,  
Bulgarian Academy of Sciences (BAS), Bulgaria  
Afzal Basha Shaik,  
Jawaharlal Nehru Technological University,  
Kakinada, India

## \*CORRESPONDENCE

Concepción Alonso,  
✉ concepcion.alonso@ehu.eus  
David Tejedor,  
✉ dtejedor@ipna.csic.es

RECEIVED 24 November 2023

ACCEPTED 19 February 2024

PUBLISHED 04 March 2024

## CITATION

Martin-Encinas E, Fuertes M,  
Delgado-Hernández S, García-Tellado F,  
Tejedor D and Alonso C (2024), Synthesis,  
biological and computational evaluation of  
novel cyanomethyl vinyl ether derivatives.  
*Front. Pharmacol.* 15:1344042.  
doi: 10.3389/fphar.2024.1344042

## COPYRIGHT

© 2024 Martín-Encinas, Fuertes, Delgado-Hernández, García-Tellado, Tejedor and Alonso. 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.

# Synthesis, biological and computational evaluation of novel cyanomethyl vinyl ether derivatives

Endika Martín-Encinas<sup>1</sup>, María Fuertes<sup>1</sup>,  
Samuel Delgado-Hernández<sup>2</sup>, Fernando García-Tellado<sup>2</sup>,  
David Tejedor<sup>2\*</sup> and Concepción Alonso<sup>1\*</sup>

<sup>1</sup>Department of Organic Chemistry I, Faculty of Pharmacy and Lascaray Research Center (Lascaray Research Center), University of the Basque Country/Euskal Herriko Unibertsitatea (UPV/EHU), Vitoria-Gasteiz, Spain, <sup>2</sup>Instituto de Productos Naturales y Agrobiología, Consejo Superior de Investigaciones Científicas (Spanish National Research Council) Avda, La Laguna, Spain

This work explores the biological evaluation of novel cyanomethyl vinyl ether derivatives as antiproliferative agents. Tubulin, crucial to microtubule structure and function, is a target for cancer therapies. *In vitro* cytotoxicity assessments revealed significant activity in SKOV3 ovarian carcinoma cells and A549 lung carcinoma cells. Structure-Activity Relationship (SAR) analysis indicated that the *E* isomer and specific substitutions influenced the biological activity. Computational assays predicted favorable ADME properties, highlighting potential as anticancerous agents. Molecular docking studies demonstrated that compound **12E**, with the *E* geometry of the double bond and fused polyaromatic rings such as phenanthrene, has robust interaction with tubulin, suggesting enhanced stability due to diverse amino acid interactions. Comparative spatial distributions with colchicine further indicated potential mechanistic similarities.

## KEYWORDS

cancer, chemotherapy, cyanomethyl vinyl ethers, cytotoxicity, tubulin inhibitors

## 1 Introduction

Cancer represents one of the greatest challenges facing our society today, with nearly 20 million new cases and 10 million deaths being observed by 2022 (GLOBOCAN, 2022). In fact, the development of new drugs against this group of diseases is being addressed for years (Mattiuzzi et al., 2019). Tubulin is a protein critical to the structure and function of microtubules, which are essential components of the cell's cytoskeleton. It consists of two subunits, alpha and beta, that assemble to form the building blocks of microtubules. These microtubules play a fundamental role in various cellular processes, including cell division, intracellular transport, and cell shape maintenance (Howard et al., 2009). For all this, tubulin is an important target for cancer therapies, as inhibiting its function can disrupt the division of rapidly dividing cancer cells, leading to their death.

Microtubules (MTs) are crucial cellular polymers composed of tubulin dimers, playing pivotal roles in intracellular trafficking, cell morphology, and mitotic spindle assembly (Goodson et al., 2018). Tubulin, a 50 kDa GTP-binding protein, comprises six family members in eukaryotic cells, with  $\alpha$  and  $\beta$  tubulins forming cytoplasmic microtubules and  $\gamma$ ,

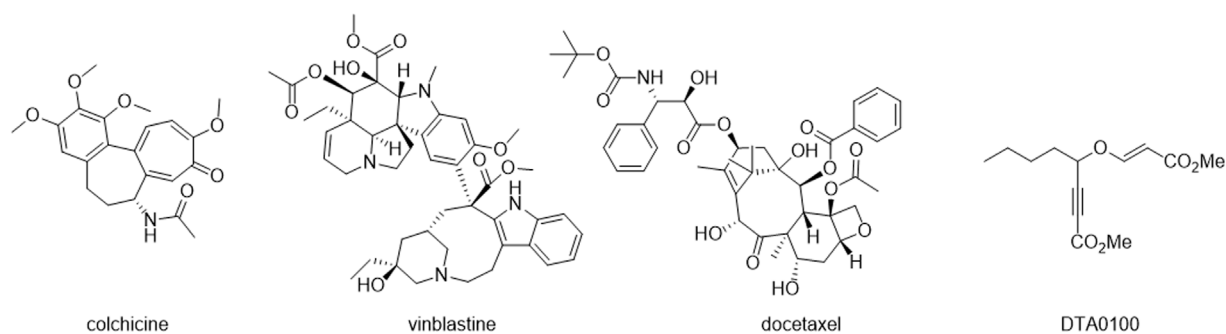


FIGURE 1  
Chemical structures of colchicine, vinblastine, docetaxel and DTA0100.

$\delta$ , and  $\epsilon$  tubulins localizing to the centrosome. Widely distributed,  $\alpha$ - $\beta$  tubulin heterodimers bind GTP, and  $\beta$  tubulin hydrolyzes GTP during microtubule polymerization.  $\gamma$  tubulin is involved in nucleating microtubule growth, while  $\delta$ ,  $\epsilon$ , and  $\zeta$  tubulins are specific to cilia, flagella, and basal bodies (Findeisen et al., 2014). In humans, 23 functional tubulin genes contribute to the structural complexity of microtubules, characterized by the assembly of  $\alpha$ - $\beta$  tubulin heterodimers into linear protofilaments. Protofilaments laterally associate to form the pseudo-helical structure of microtubules, with most having 13 protofilaments, though variations exist across species (Chaaban et al., 2017).

Microtubule-targeting agents (MTAs) form a diverse group of compounds capable of binding tubulin, influencing microtubule (MT) dynamics by either stabilizing or destabilizing the MT polymer (Steinmetz et al., 2018). Classified into two main categories, MT-stabilizing agents (MSAs) and MT-destabilizing agents (MDAs), these drugs bind to one of seven sites in tubulin dimers. MSAs and MDAs exhibit contrasting effects on MT polymer mass at high concentrations, inducing depolymerization or stabilization. They alter the monomer-to-polymer ratio, impacting cellular functions without significant changes in total MT polymer at low concentrations. Notably, drugs targeting vinca, colchicine, and taxane sites are extensively studied MTAs, historically employed as medicines (Matthew et al., 2021). Colchicine (Figure 1), derived from autumn crocus, is one of the earliest reported MTAs, historically used for gout treatment. Over the past century, MTAs have found applications as herbicides, anti-parasitics, anti-fungal agents, have been explored for neurodegenerative disease and cancer treatment (Graham et al., 1953; Loeffler et al., 1977; Kovacs et al., 2023) and also for antiprotozoal activities (Bethencourt-Estrella et al., 2023; Chao-Pellicer et al., 2023).

After that, various types of tubulin inhibitors, such as vinblastine and docetaxel (Figure 1), have been developed and are used to treat different types of cancer (Kingston, 2001). However, despite tubulin inhibitors have demonstrated significant efficacy in the treatment of cancer, it is important to note that tubulin inhibitors are not specific to cancer cells and can affect healthy cells that divide rapidly, such as bone marrow and gastrointestinal cells. This explains some of the side effects associated with these drugs, such as bone marrow suppression,

peripheral neuropathy, and gastrointestinal problems (Dumontet et al., 2010). In this way, computational tools play a pivotal role in discovering new bioactive molecules by enabling precise and rapid calculations of chemical and biological properties. These tools expedite the drug design process by analyzing extensive databases and predicting biological activity, paving the way for the efficient development of compounds with therapeutic potential. (Lipinski et al., 2012; Sliwoski et al., 2014).

Finally, recent investigations have provided compelling evidence regarding the capacity of inhibiting tubulin by compounds with vinyl ethers in their structure. Where, the compound DTA0100 (Figure 1) clearly induced microtubule depolymerization, leading to disturbance of cell cycle kinetics and subsequent apoptosis (Podolski-Renić et al., 2017).

Thereby, in this work we report the biological evaluation of readily available conjugated cyanomethyl vinyl ethers prepared by an organocatalytic multicomponent cyanovinylolation of aldehydes recently described by our group (Delgado-Hernández et al., 2021).

## 2 Materials and methods

### 2.1 Chemistry

#### 2.1.1 General experimental information

All reagents from commercial suppliers were used without further purification. All solvents were freshly distilled before use from appropriate drying agents. All other reagents were recrystallized or distilled when necessary. Analytical TLCs were performed with silica gel 60 F254 plates. Visualization was accomplished by UV light. Column chromatography was carried out using silica gel 60 (230–400 mesh ASTM). NMR spectra were obtained on a Bruker Avance 500 MHz spectrometers and recorded at 25°C. Chemical shifts for  $^1\text{H}$ -NMR spectra are reported in ppm downfield from TMS, chemical shifts for  $^{13}\text{C}$ -NMR spectra are recorded in ppm relative to internal deuterated chloroform ( $\delta = 77.2$  ppm for  $^{13}\text{C}$ ). Coupling constants ( $J$ ) are reported in Hertz. The terms m, s, d, t, q refers to multiplet, singlet, doublet, triplet, quartet.  $^{13}\text{C}$ -NMR were broadband decoupled from hydrogen nuclei. High resolution mass spectra (HRMS) was measured by EI method with a Agilent LC-Q-TOF-MS 6520 spectrometer.



### 2.1.2 Compounds purity analysis

All synthesized compounds were analyzed by HPLC to determine their purity. The analyses were performed on Agilent 1,260 infinity HPLC system (C-18 column, Hypersil, BDS, 5  $\mu$ m, 0.4 mm  $\times$  25 mm). All the tested compounds were dissolved in dichloromethane, and 1  $\mu$ L of the sample was loaded onto the column. Ethanol and heptane were used as mobile phase, and the flow rate was set at 1.0 mL/min. The maximal absorbance at the range of 190–625 nm was used as the detection wavelength. The purity of all the tested compounds is >95%, which meets the purity requirement by the Journal.

### 2.1.3 Synthesis of cyanomethyl vinyl ethers

All compounds were synthesized using the methodology described in reference (Delgado-Hernández, 2021) except compound 3 and 4 which are described below.

#### 2.1.3.1 General procedure for the synthesis of cyanomethyl vinyl ethers 3 and 4

To a solution of benzaldehyde (2.0 mmol), acetone cyanohydrin (2.0 mmol) and the corresponding propiolate derivative (2.0 mmol) in *n*-hexanes (6 mL) was added *N*-methylmorpholine (0.05 mmol) at once and the reaction mixture was stirred for 1 h at room temperature. The solvent was removed under reduced pressure, and the residue was purified by flash column chromatography (silica gel; *n*-hexane/ethyl acetate: 80/20) to give the desired 3-(cyanomethoxy)acrylate.

**Propyl (E)-3-(cyano(phenyl)methoxy)acrylate (3E).** (438.2 mg, 46%). White solid:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 400 MHz): 0.94 (t, 3H,  $^3J_{(\text{HH})} = 7.4$  Hz), 1.62–1.71 (m, 2H), 4.08 (t, 2H,  $^3J_{(\text{HH})} = 6.07$  Hz), 5.53 (d, 1H,  $^3J_{(\text{HH})} = 12.6$  Hz), 5.64 (s, 1H), 7.46–7.53 (m, 5H), 7.55 (d, 1H,  $^3J_{(\text{HH})} = 12.6$  Hz).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 100 MHz):  $\delta = 10.4, 22.0, 66.0, 70.2, 101.6, 115.3, 127.5$  (2C), 129.4 (2C), 130.8, 131.2, 157.9, 166.4 ppm. HRMS (ESI $^-$ ):  $m/z$  [M] $^-$  calculated for  $\text{C}_{14}\text{H}_{15}\text{NO}_3$  244.0974, found 244.0979.

**Propyl (Z)-3-(cyano(phenyl)methoxy)acrylate (3Z).** (323.0 mg, 33%). White solid:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 400 MHz): 0.92 (t, 3H,  $^3J_{(\text{HH})} = 7.4$  Hz), 1.60–1.69 (m, 2H), 4.06 (t, 2H,  $^3J_{(\text{HH})} = 6.07$  Hz), 5.08 (d, 1H,  $^3J_{(\text{HH})} = 7.1$  Hz), 5.81 (s, 1H), 6.59 (d, 1H,  $^3J_{(\text{HH})} = 7.0$  Hz), 7.42–7.44 (m, 3H), 7.53–7.56 (m, 2H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 100 MHz):  $\delta = 10.4, 22.0, 65.7, 72.0, 101.4, 115.6, 127.4$  (2C), 129.4 (2C), 130.6, 131.3, 152.6, 164.4 ppm. HRMS (ESI $^-$ ):  $m/z$  [M] $^-$  calculated for  $\text{C}_{14}\text{H}_{15}\text{NO}_3$  244.0974, found 244.0979.

**Isoropyl (E)-3-(cyano(phenyl)methoxy)acrylate (4E).** (382.52 mg, 52%). Colorless oil:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 400 MHz): 1.25 (s, 6H,  $^3J_{(\text{HH})} = 6.2$  Hz), 5.02–5.09 (m, 1H), 5.50 (d, 1H,  $^3J_{(\text{HH})} = 12.6$  Hz), 5.63 (s, 1H), 7.46–7.53 (m, 5H), 7.54 (d, 1H,  $^3J_{(\text{HH})} = 12.6$  Hz).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 100 MHz):  $\delta = 21.9, 67.7, 70.1, 102.0, 115.3, 127.4$  (2C), 129.4 (2C), 130.7, 131.2, 157.8, 165.9 ppm. HRMS (ESI $^+$ ):  $m/z$  [M + Na] $^+$  calculated for  $\text{C}_{14}\text{H}_{15}\text{NO}_3$  268.0950, found 268.0944.

**Isoropyl (Z)-3-(cyano(phenyl)methoxy)acrylate (4Z).** (259.8 mg, 35%). Light yellowish oil:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ , 400 MHz): 1.23 (s, 6H,  $^3J_{(\text{HH})} = 6.2$  Hz), 5.00–5.07 (m, 1H), 5.06 (d, 1H,  $^3J_{(\text{HH})} = 7.0$  Hz), 5.82 (s, 1H), 6.57 (d, 1H,  $^3J_{(\text{HH})} = 7.0$  Hz), 7.42–7.44 (m, 3H), 7.53–7.56 (m, 2H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ , 100 MHz):  $\delta = 21.8, 67.3, 71.9, 101.9, 115.7, 127.3$  (2C), 129.3 (2C), 130.5, 131.4,

152.4, 163.6 ppm. HRMS (ESI $^+$ ):  $m/z$  [M + Na] $^+$  calculated for  $\text{C}_{14}\text{H}_{15}\text{NO}_3$  268.0950, found 268.0944.

## 2.2 Biology

### 2.2.1 Materials

Reagents and solvents were used as purchased without further purification. All stock solutions of the investigated compounds were prepared by dissolving the powdered materials in appropriate amounts of DMSO. The final concentration of DMSO never exceeded 5% (v/v) in reactions. Under these conditions DMSO was also used in the controls and was not seen to affect tested compounds activity. The solutions were stored at 5°C until they were used.

### 2.2.2 Cytotoxicity assays

Cells were cultured according to the supplier's instructions (ATCC technologies, A-549 (CCL-185), BT-20 (HTB-19), SKOV3 (HTB-77), HCT116 (CCL-247), MRC5 (CCL-171). Cells were seeded in 96-well plates at a density of  $2\text{--}2.5 \times 10^3$  cells per well and incubated overnight in 0.1 mL of media supplied with 10% Fetal Bovine Serum (Lonza) in 5%  $\text{CO}_2$  incubator at 37°C. On day 2, drugs were added and samples were incubated for 48 h. After treatment, 10  $\mu$ L of cell counting kit-8 was added into each well for additional 2 h incubation at 37°C. The absorbance of each well was determined by an Automatic Elisa Reader System at 450 nm wavelength.

## 2.3 Computational assays

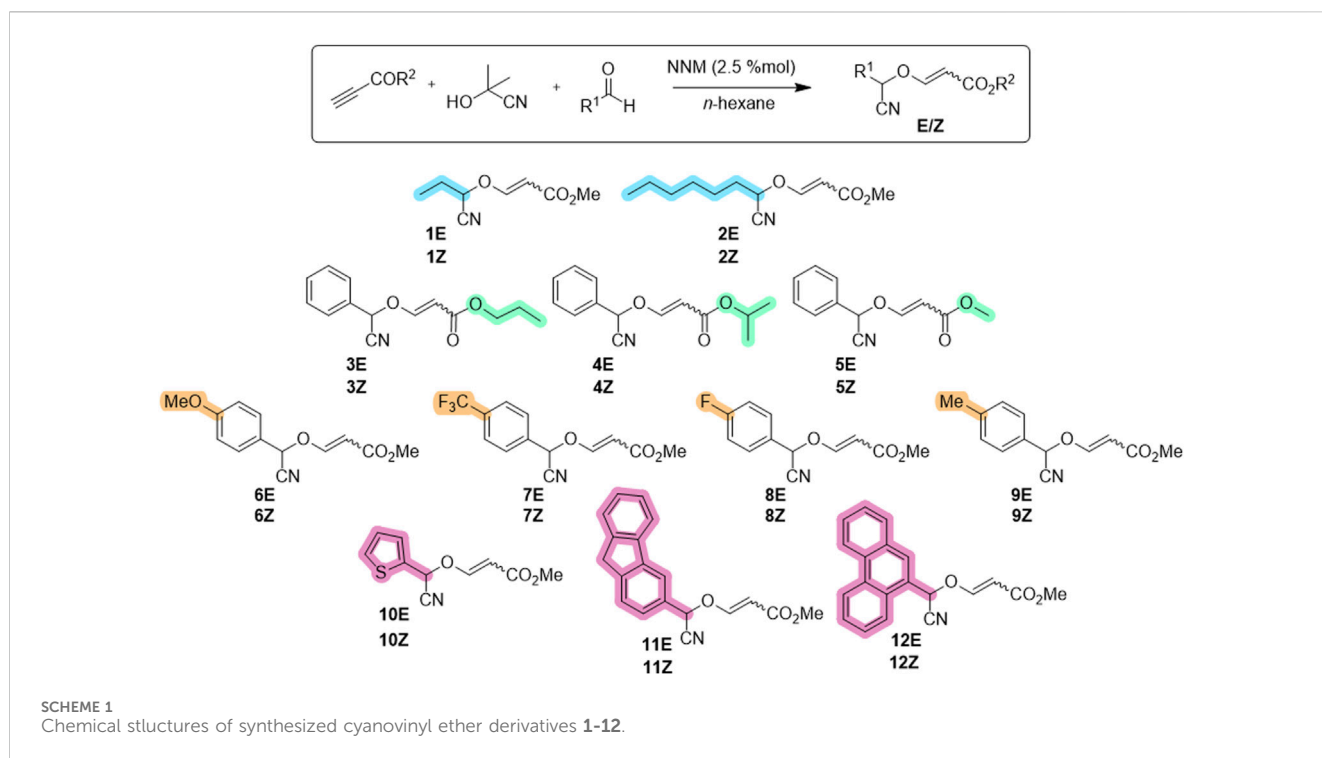
### 2.3.1 In silico ADME

The physicochemical and pharmacokinetics properties of tested compounds were calculated using swissADME (<http://www.swissadme.ch/>) and pkCSM (<http://biosig.unimelb.edu.au/pkcsml/>) online web servers. Chemical structures were imported in swissADME and pkCSM tools to calculate molecular as well as ADME properties of the compounds (Daina et al., 2017; Bakchi et al., 2022).

### 2.3.2 Docking studies

Tubuline complex for the docking in the Protein Data Bank (PDB). The X-ray structure code 1SA0 (3.58 Å resolution) was chosen, a tubuline domain complex with colchicine as a ligand. Maestro (Frisch et al., 2016) graphic interface was used, and the Glide 6.9 application (Schrodinger Release, 2015-1: Glide, 2015a) in XP mode (extraprecision) (Friesner et al., 2006) was chosen for the docking. The grid was set up in a box of  $20 \times 20 \times 20$  Å, centered in the geometric center of colchicine ( $\beta$ -Chain). The colchicine region in the active site was selected as the target for the screening. The protein complex was optimized and minimized using the Protein Preparation Wizard panel of Schrodinger Suites 2015.1 (Schrodinger Release, 2015-1: Protein Preparation Wizard, 2015b). Likewise, the structures of the different ligands to be interacted with protein and the ligand initially present in the complex, colchicine, were prepared. The binding orders and the protonation states of residue were corrected. The complex previously indicated and used for the different docking processes.





## 3 Results

### 3.1 Chemistry

The compounds used for this study were chosen from the set of compounds synthesized in our previous study or were synthesized for the first time to include carboxylic esters with longer alkyl chains (3–4). Thus, the reaction was implemented for the synthesis of these 3-substituted 3-(cyanomethoxy)acrylates, using aldehydes as substrates, acetone cyanohydrin as the cyanide anion source, and commercially available methyl propiolate or readily synthesized alkyl propiolates as the source of the vinyl component. The multicomponent reaction is catalyzed by *N*-methylmorpholine (2.5 mol%) to deliver the desired 3-(cyanomethoxy)acrylates in excellent yields (84%–95% for isomer mixtures; 32%–56% for separated isomers) for their biological evaluation (Scheme 1).

### 3.2 Biology

#### 3.2.1 *In Vitro* cytotoxicity

The cytotoxicity of the new synthesized 3-(cyanomethoxy)acrylates was investigated *in vitro* by testing the antiproliferative activities against two human cancer cell lines: A549 (carcinomic human alveolar basal epithelial cells), SKOV3 (human ovarian carcinoma), BT20 (human breast carcinoma), HCT-116 (human colorectal carcinoma) and MRC-5 (non-malignant lung fibroblasts). The cell counting kit (CCK8) assay was employed to assess growth inhibition and, the cell proliferation inhibitory activities of the compounds are listed in Table 1 as  $IC_{50}$  values (Table 1).

Firstly, selective cytotoxicity was observed in SKOV3 and BT20 cell lines, where most of the compounds showed

cytotoxicity. Notably, the best results were observed for compounds 1E and 7E with  $IC_{50}$  values of  $2.92 \pm 0.28 \mu M$  and  $2.62 \pm 0.53 \mu M$ , respectively in the SKOV3 cell line (Table 1, entries 2 and 14) and for compound 9E in the BT20 line with  $IC_{50}$  of  $4.98 \pm 2.77 \mu M$  (Table 1, entry 18).

On the other hand, only four of the compounds studied showed cytotoxic activity against the A549 line, where compound 12E obtained the best result with an  $IC_{50}$  value of  $9.89 \pm 0.25 \mu M$  (Table 1, entry 24). Precisely this compound 12E showed interesting cytotoxicity in all the cell lines studied. Besides the above mentioned result on the A549 cell line, it also showed second best result among all the studied compounds on the BT20 cell line with an  $IC_{50} = 9.89 \pm 0.25$  and in HCT116 cell line with an  $IC_{50} = 16.10 \pm 1.20$  (Figure 2; Table 1, entry 24).

And remarkably, except for compound 12E, which showed cytotoxicity in all cell lines, none of the tested compounds showed cytotoxicity in the non-cancerous line MRC5. Such behavior against the non-cancerous line MRC5 is relevant, as this selectivity is not found in colchicine, with  $IC_{50}$  values =  $5.72 \pm 2.1 \mu M$  in the non-cancerous cell line MRC5 (Table 1, entry 1).

### 3.3 Structure-activity relationship (SAR)

Structure-Activity Relationship (SAR) analysis is a fundamental approach that investigates the links between the structure of chemical compounds and their observed biological or physicochemical properties. By systematically studying the relationships between structural features and activity, SAR analysis provides valuable insights for rational compound design and optimization. The structure-activity relationship of the compounds observed in this study is described (Figure 3).

TABLE 1 Antiproliferative activities of tested compounds.

Entry	Compound	IC50 (μM)				
		SKOV3	BT20	A549	HCT116	MRC5
1	<b>COL</b>	11.58 ± 2.33	21.98 ± 1.30	2.39 ± 1.40	9.32 ± 0.2	5.72 ± 2.1
2	<b>1E</b>	2.92 ± 0.28	7.07 ± 1.28	>50	>50	>50
3	<b>1Z</b>	>50	>50	>50	>50	>50
4	<b>2E</b>	41.58 ± 2.46	15.83 ± 2.90	>50	>50	>50
5	<b>2Z</b>	>50	17.28 ± 3.44	>50	>50	>50
6	<b>3E</b>	>50	ND	>50	>50	>50
7	<b>3Z</b>	>50	ND	>50	>50	>50
8	<b>4E</b>	>50	ND	>50	>50	>50
9	<b>4Z</b>	>50	ND	>50	>50	>50
10	<b>5E</b>	3.84 ± 1.56	8.31 ± 1.26	18.85 ± 1.20	>50	>50
11	<b>5Z</b>	29.81 ± 2.09	19.92 ± 3.42	>50	>50	>50
12	<b>6E</b>	12.44 ± 1.18	27.79 ± 5.31	>50	>50	>50
13	<b>6Z</b>	31.33 ± 4.42	12.61 ± 1.76	>50	>50	>50
14	<b>7E</b>	2.62 ± 0.53	10.02 ± 2.17	29.34 ± 1.10	>50	>50
15	<b>7Z</b>	>50	>50	>50	>50	>50
16	<b>8E</b>	>50	ND	>50	>50	>50
17	<b>8Z</b>	>50	ND	>50	>50	>50
18	<b>9E</b>	8.08 ± 2.08	4.98 ± 2.77	21.38 ± 0.83	>50	>50
19	<b>9Z</b>	>50	20.24 ± 3.78	>50	>50	>50
20	<b>10E</b>	>50	ND	>50	>50	>50
21	<b>10Z</b>	>50	ND	>50	>50	>50
22	<b>11E</b>	28.85 ± 2.15	21.67 ± 0.67	>50	>50	>50
23	<b>11Z</b>	38.29 ± 2.31	>50	>50	>50	>50
24	<b>12E</b>	7.35 ± 1.42	7.0 ± 1.96	9.89 ± 0.25	16.10 ± 1.20	14.39 ± 0.86
25	<b>12Z</b>	11.93 ± 1.88	11.94 ± 1.10	>50	>50	>50

ND, not determined. COL, colchicine. Bold values are numbers of synthesized compounds.

Based on the obtained biological results and taking into account the structural differences of the tested compounds, it has been observed that the *E* isomer of the tested derivatives significantly increases the biological activity, being even, in some cases, crucial for the compound to exhibit cytotoxicity.

Furthermore, it has been observed that substituents in the ether group (R<sup>2</sup>) can significantly modify the activity of the tested derivatives, where the methyl group (R<sup>2</sup> = Me) appears to be crucial for the biological activity of the compound (Table 1, entries 6, 8 and 10).

Finally, the nature of the substituents in a position respect to the cyano group (R<sup>1</sup>) has no great effect on the biological activity, observing cytotoxicity with both aromatic and aliphatic groups (Table 1, entries 2 and 10). Furthermore, in the case of aromatic groups, on the one hand, derivatives with a carbon atom at position 4 of the phenyl group show increased cytotoxicity (Table 1, entries

12, 14 and 18), while on the other hand, polyaromatic substituents show loss of selectivity against non-tumorigenic cells (Table 1, entry 24).

### 3.4 Computational assays

#### 3.4.1 In silico ADME

Computer based ADME studies, also known as computer-based studies, play a crucial role in the drug discovery and development process. ADME stands for Absorption, Distribution, Metabolism, and Excretion, which are key factors that determine the pharmacokinetics and efficacy of a drug. In silico ADME studies involve the use of computational models and algorithms to predict the ADME properties of a drug candidate before it is tested *in vitro* or *in vivo*. These studies not only save time and resources but also

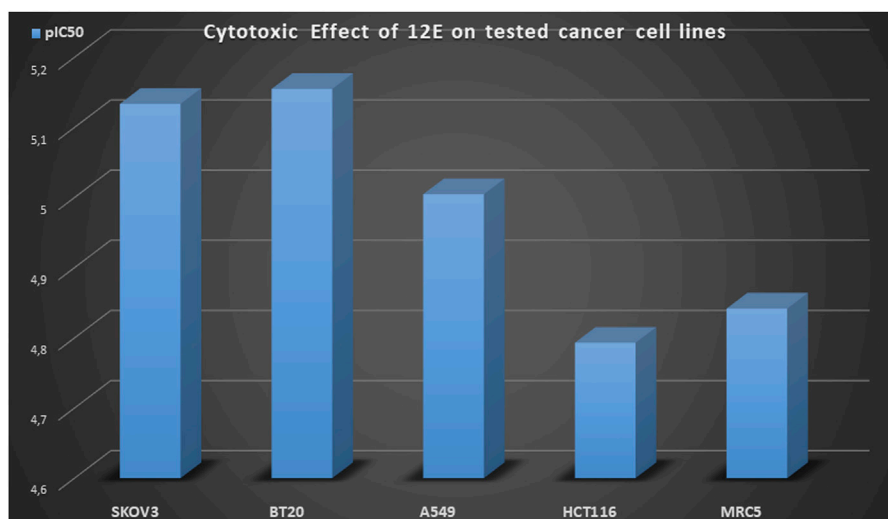


FIGURE 2  
Cytotoxic effect (pIC<sub>50</sub> values) of cyanovinyl ether derivative **12E**.

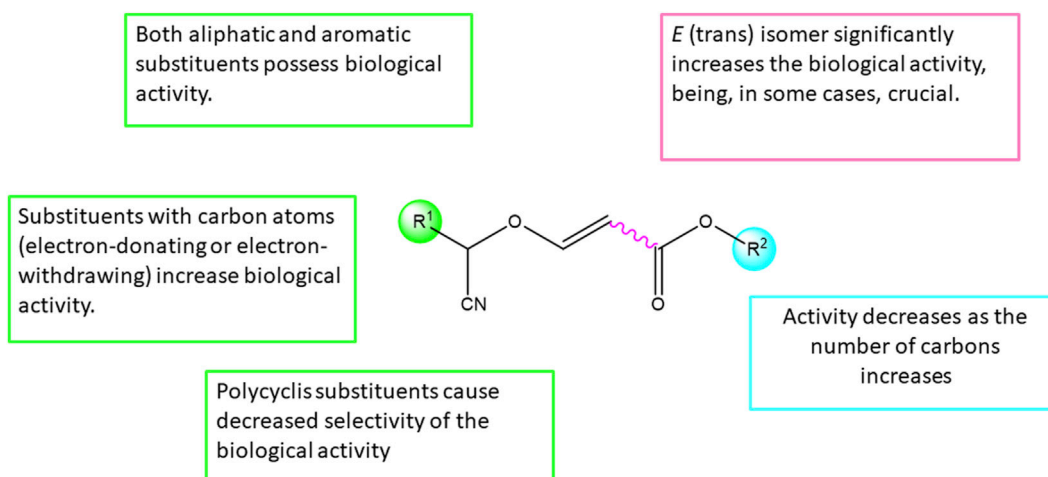


FIGURE 3  
Structure-Activity Relationship (SAR) of synthesized cyanovinyl ethers.

help in the identification of potential safety concerns associated with a drug candidate.

On the one hand, molecular properties such as partition coefficient (Log Po/w), molecular weight, hydrogen bond donors and acceptors, topological polar surface area (TPSA), violation of Lipinski rule of five were assessed (Table 2).

In general, a drug candidate with a LogP value between 0 and 5 is considered to have favorable ADME properties. A LogP value that is too low may indicate poor lipid solubility, which can affect the absorption and distribution of the drug in the body. On the other hand, a LogP value that is too high may indicate poor aqueous solubility, which can lead to poor bioavailability and potential toxicity due to the accumulation of the drug in fatty tissues. In the case of our compounds, they have a LogP value between 2.0 and 3.18, so they would fall within the favorable range mentioned above (Table 2).

TPSA, or the Topological Polar Surface Area, is another commonly used parameter in drug discovery and development to predict the ADME properties of a drug candidate. TPSA is a measure of the polar surface area of a compound, which is important for its interaction with biological targets and its ability to cross biological membranes. In general, a drug candidate with a TPSA value between 20 and 140 Å<sup>2</sup> is considered to have favorable ADME properties. As shown in Table 2, all the compounds tested in this work fall within the favorable range for good oral bioavailability.

Finally, the Lipinski's Rule of Five, a widely used rule in drug discovery to assess the drug-like properties of a compound, was performed to study our compounds. It is recommended that the orally active drug candidate should not have more than one violation of the Lipinski's rule. As shown in Table 2, all compounds tested meet the established criteria.

TABLE 2 ADME properties of tested compounds.

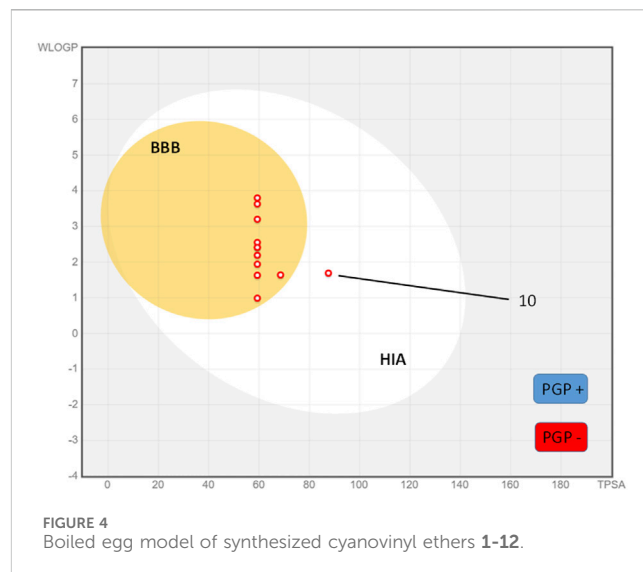
Entry	Compound	LogP	Mol. Wt	H-donor	H-acceptor	TPSA	Lipinski #violations
1	<b>COL</b>						
2	<b>1</b>	2.00	169.18	0	4	59.32	0
3	<b>2</b>	3.06	225.28	0	4	59.32	0
4	<b>3</b>	2.81	245.27	0	4	59.32	0
5	<b>4</b>	2.50	245.27	0	4	59.32	0
6	<b>5</b>	2.39	217.22	0	4	59.32	0
7	<b>6</b>	2.57	247.25	0	5	68.55	0
8	<b>7</b>	2.55	285.22	0	7	59.32	0
9	<b>8</b>	2.30	235.21	0	5	59.32	0
10	<b>9</b>	2.52	231.25	0	4	59.32	0
11	<b>10</b>	2.29	223.25	0	4	87.56	0
12	<b>11</b>	2.98	305.33	0	4	59.32	0
13	<b>12</b>	3.18	307.34	0	4	59.32	0

COL, colchicine. Bold values are numbers of synthesized compounds.

On the other hand, the absorption of a drug is a critical factor in determining its efficacy and safety (Supplementary Table S1). It refers to the process by which a drug enters the bloodstream and reaches its target site of action. Understanding the factors that influence drug absorption is essential for optimizing drug therapy and minimizing adverse effects. Because of this, the absorption of drug was evaluated based on aqueous solubility, intestinal absorption and permeability showed in Supplementary Table S1. Thus, the aqueous solubility (Log S) of all compounds ranges from  $-4.944$  to  $-0.757$  log mol/L, which shows the moderate solubility in water of the synthesized compounds. In addition, all of compounds show intestinal absorption above 95% that is interesting due to the most orally administered drugs are primarily absorbed through small intestine due to its large surface area. Finally, the predicted Caco-2 permeability, the logarithm of the apparent permeability coefficient (log Papp  $>8 \times 10^{-6}$  cm/s), was studied. In our case, all of the synthesized compounds have high Caco-2 permeability taking into account that their predicted value is  $>0.90$ .

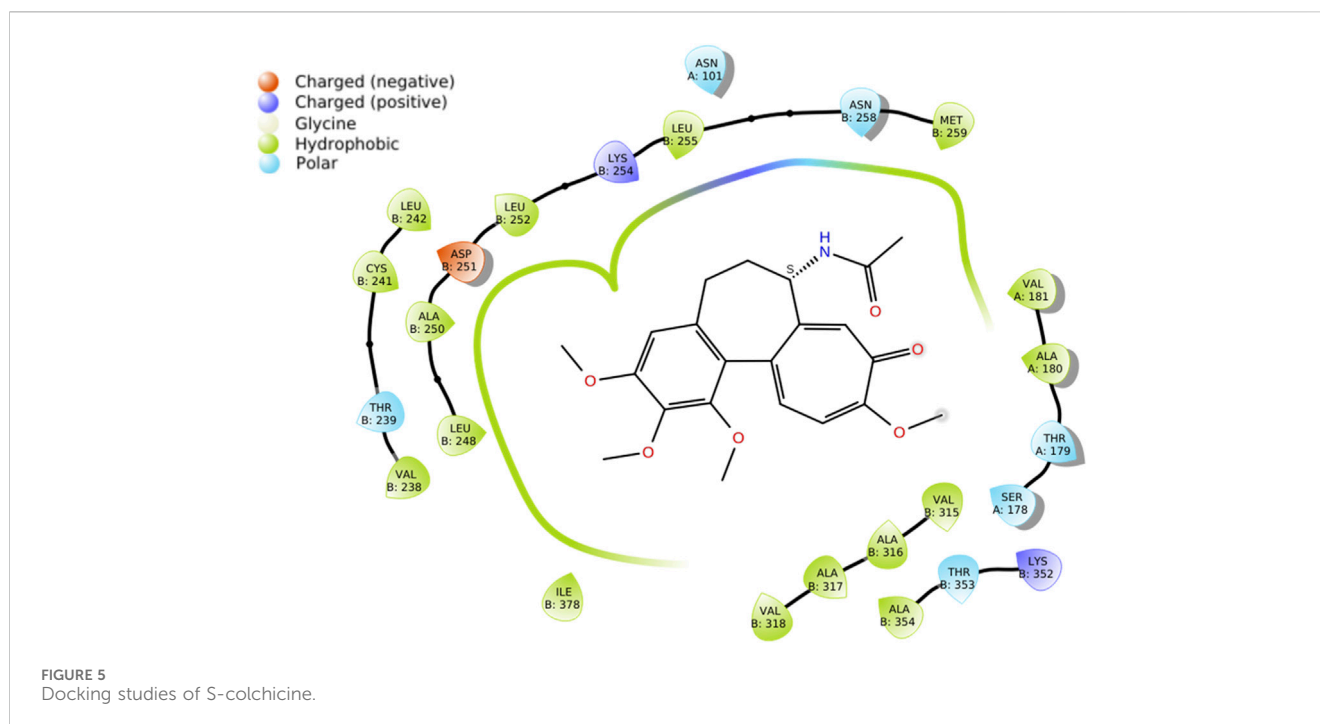
Lastly, the synthesized compounds volume of distribution (VDss), blood-brain barrier permeability (BBB permeability) and the fraction of unbound was further assessed (Supplementary Table S1). On the one hand, taking into account that the predicted value of Log VDss  $>0.45$  L/kg indicates higher volume of drug distribution, we can say that all synthesized compounds have low volume of distribution in tissues. On the other hand, blood-Brain Barrier (BBB) property is crucial for the effectiveness of drugs in treating central nervous system disorders. The compound is said to be easily permeable through BBB if the predicted value of log BB is  $>0.3$  and poorly distributed if the value is  $<-1$ . It is interesting to note that all compounds show excellent BBB parameters.

Cytochrome P450 (CYP) enzymes are a family of heme-containing enzymes involved in the metabolism of a wide range of endogenous compounds and xenobiotics. These enzymes play a



crucial role in drug metabolism, and their activity can influence the efficacy and safety of many therapeutic agents. The compounds were studied as possible CYP2D6, CYP3A4, CYP1A2, CYP2C19, and CYP2C9 enzyme inhibitors (Supplementary Table S1). It is noteworthy that the vast majority of compounds (3, 5, 7, 8, 9, 10, 11 and 12) show the ability to inhibit the CYP1A2 enzyme, and in addition, only compounds 11 and 12 show the ability to inhibit the cytochrome CYP2C9 and CYP2C19 which are primarily expressed in the liver and plays a critical role in the metabolism of numerous drugs, including nonsteroidal anti-inflammatory drugs (NSAIDs), anticoagulants, and antidiabetic agents. Summarizing, the total clearance is primarily a combination of hepatic as well as renal clearance and is measured by the proportionality constant CLtot in





log(mL/min/kg). The predicted value of all the synthesized compounds shows lower CL<sub>tot</sub> ranging from 0.427 to 1.863.

The boiled egg model is a simple and inexpensive method that has been used to predict the ADME parameters of drugs. This model involves the use of a boiled egg as a surrogate for the human body, and the measurement of the drug's permeability through the eggshell and its distribution within the egg. Figure 4 shows the results of the compounds synthesized in a boiled egg projection where the egg yolk corresponds to both the blood-brain barrier (BBB) and human gastrointestinal absorption (HIA) compounds and the egg white only to the HIA-positive compounds. As can be seen in the boiled egg, all compounds tested have positive BBB permeability and HIA, except compound 10 which only show positive HIA.

On the other hand, the representation shows a classification of the compounds by color according to their relationship with the P-glycoprotein (PGP). Point colored in blue are for molecules predicted to be substrate of the P-glycoprotein (PGP+) and hence actively pumped up from the brain or to the gastrointestinal lumen. If predicted non-substrate of the P-glycoprotein (PGP-), the related point is in red. When a compound is identified as a substrate for P-glycoprotein (PGP), it signifies the molecular recognition by PGP of said compound as a species earmarked for active extracellular transport. This discernment carries profound ramifications for the pharmacokinetic profile of pharmaceutical agents. If a drug emerges as a PGP substrate, its processes of absorption, distribution, and elimination stand susceptible to modulation by the dynamic activity of this integral membrane protein (Gottesman et al., 2002). In the case of our tested compounds, it should be noted that all derivatives present the necessary conditions to be PGP-substrate candidates (Figure 4). Concerning our tested compounds, it is worth noting that none meet the necessary criteria to be

considered substrates of P-glycoprotein (PGP). As a result, these compounds will not be expelled from the cell, potentially increasing their cytotoxic effects.

### 3.4.2 Molecular docking

Molecular docking has emerged as a powerful computational tool in drug discovery and structural biology. This technique plays a crucial role in predicting the binding mode and affinity between a protein target and small molecule ligands. By simulating the interactions at the atomic level, molecular docking enables researchers to understand the underlying molecular mechanisms and optimize the design of potential therapeutic compounds.

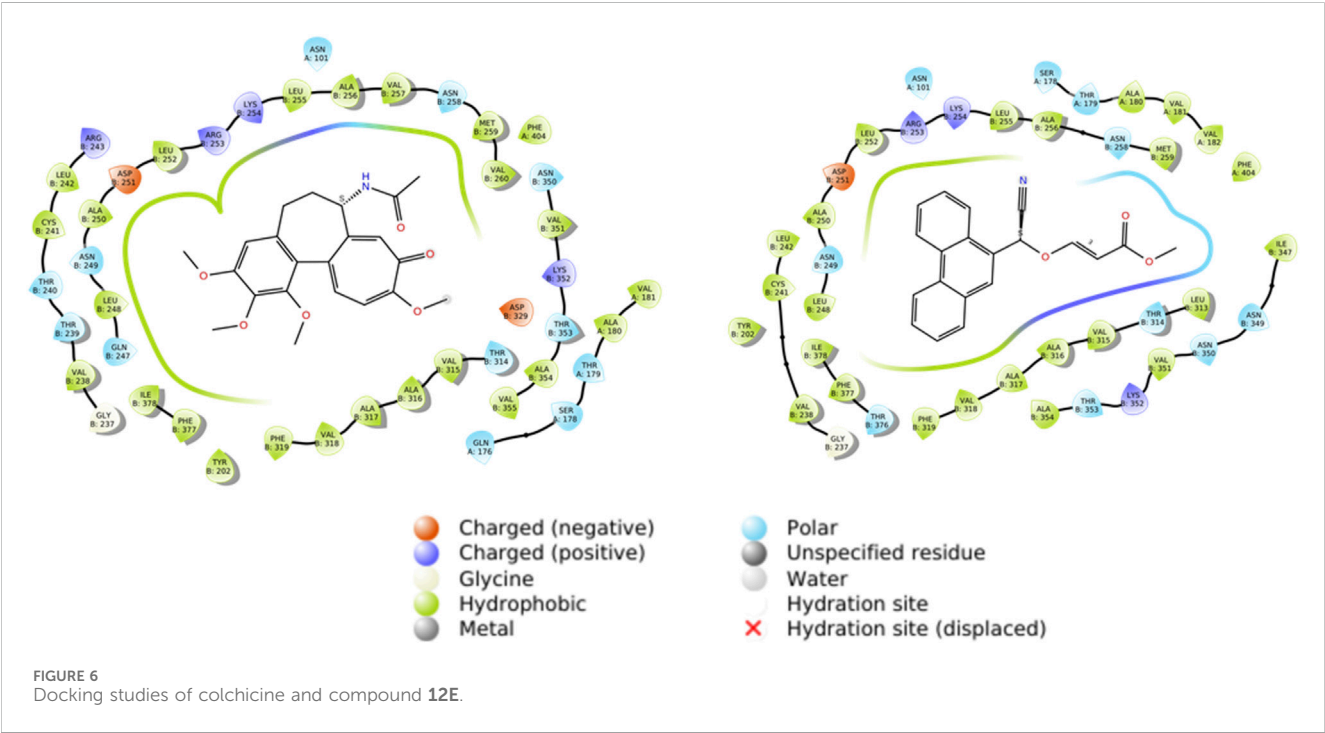
The crystallographic data of tubulin in complex with modify colchicine was retrieved from the Protein Data Bank (PDB ID: 1SA0). As it can be observed in Figure 5 docking studies of S-colchicine demonstrated robust interactions with crucial amino acid residues, such as Leu248, Asp251, Leu255, and Val260, within the colchicine-binding site of tubulin. Consequently, these amino acid residues played a pivotal role in facilitating inhibitor binding.

Among all the compounds synthesized in this work, a molecular docking study was carried out to investigate its plausible binding pattern and its interaction with the key amino acids in the active site of the protein. The results obtained from molecular docking studies (Table 3) are presented specifically focused on the 'gScore' and 'geMode' parameters. On the one hand, 'gScore' parameter serves as the comprehensive affinity score calculated through docking algorithms, reflecting the overall interaction strength between a target molecule and a ligand with a low 'gScore' correlates with an increased likelihood of a stable interaction. On the other hand, the 'geMode' pertains to the global energy model, offering an assessment of the quality and stability of the resulting docking configuration. This parameter provides insights into the energetics governing the ligand-receptor interaction.

TABLE 3 Docking studies of tested compounds.

Entry	Compound	<i>E</i> Isomer		<i>Z</i> Isomer	
		gScore (kcal/mol)	geMode (kcal/mol)	gScore (kcal/mol)	geMode (kcal/mol)
1	<b>1</b>	−3.685	−27.533	−3.227	−26.279
2	<b>2</b>	1.113	−31.899	−3.102	−31.099
3	<b>3</b>	−4.064	−41.695	−5.415	−38.099
4	<b>4</b>	−4.110	−41.695	−5.484	−43.077
5	<b>5</b>	−4.872	−37.238	−4.712	−37.384
6	<b>6</b>	−5.922	−40.470	−5.459	−41.437
7	<b>7</b>	−5.920	−40.470	−5.866	−41.968
8	<b>8</b>	−5.176	−36.886	−4.572	−35.867
9	<b>9</b>	−5.564	−39.392	−5.492	−36.957
10	<b>10</b>	−4.357	−38.398	−4.039	−32.436
11	<b>11</b>	−5.878	−49.124	−7.247	−51.923
12	<b>12</b>	−7.733	−66.097	−6.844	−52.137
13	COL	−7.578 gScore (kcal/mol)		−57.983 geMode (kcal/mol)	

COL, colchicine. Bold values are numbers of synthesized compounds.



Thus, a molecular docking study examined the interaction of compounds **1–12** with tubulin, taking into account both *E* and *Z* isomeric forms. There were observed variances in the gScore and geMode values across various compounds, as shown in Table 3.

In the first place, it is noteworthy that the average energy of geMode for *E* isomer (−39.931 kcal/mol) is lower than the average

energy of compounds with *Z* isomerism (−37.742 kcal/mol). In other words, compounds with *E* isomerism tend to form a more stable protein-inhibitor complex, in general. For example, compound **8**, in its *E* and *Z* forms, exhibited a substantial difference in gScore (−5.176 vs. −4.571 kcal/mol) and geMode (−36.886 vs. −35,867 kcal/mol), highlighting a pronounced isomer-dependent effect on binding affinity.

TABLE 4 Type of interaction in molecular docking.

Residues (Chain B)	Bond type	Bond distance (Å)
ASP 251	Hydrophobic	9.11
LEU 252	Hydrophobic	7.57
ARG 233	Hydrophobic	8.61
LYS 254	Hydrophobic	9.98
PHE 319	Hydrophobic	10.3
VAL 318	Hydrophobic	9.44
ALA 317	Hydrophobic	10.4
LEU 313	Polar	6.98
MET 259	Polar	7.01
ASN 258	Polar	8.12

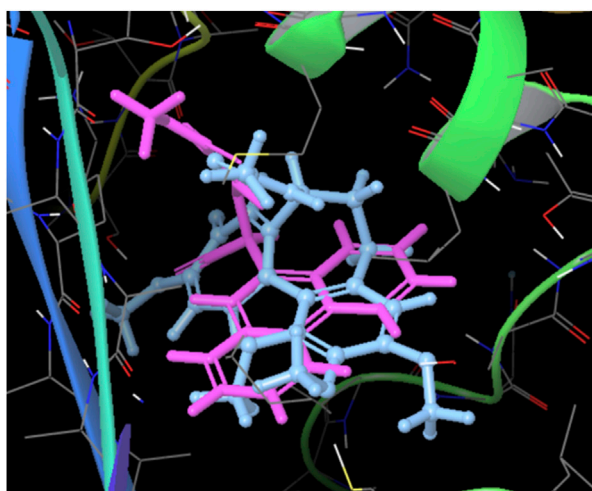


FIGURE 7  
Superposed spatial distributions of colchicine and compound 12E in the formation of complexes.

In this context, it is noteworthy to emphasize that among the range of compounds evaluated, the one that stands out for exhibiting the most intense and sustained interaction with the protein is compound **12E** (Table 3, entry 12). This compound demonstrates a remarkable gScore of  $-7.733$  kcal/mol, indicative of an exceptionally high affinity for the target protein. The pronounced stability of this interaction suggests that the *E* isomer of **12E** is highly conducive to form a robust protein-inhibitor complex, potentially translating into enhanced efficacy as an inhibitory agent. Additionally, it is worth noting that this compound exhibited cytotoxic activity across all tested cell lines.

The stability values exhibited by the tubulin-compound **12E** complex may be attributed to the formation of tight interactions with various amino acids of the protein. As observed in Figure 6, both colchicine and compound **12E** show hydrophobic interactions with residues such as Leu252, Asp251, and Leu255 of tubulin, among others (Table 4). However, in addition to these

interactions, compound **12E** presents weak polar interactions with residues like Met259 and Leu313 (Table 4).

Additionally, it is worth highlighting that both molecules, colchicine and compound **12E**, engage in the formation of complexes that exhibit comparable spatial distributions, as illustrated in Figure 7. The spatial congruence in their complex formations suggests a potential similarity in their binding mechanisms and affinities for specific regions on the tubulin structure. This could potentially explain the enhanced stability of the tubulin-compound **12E** complex.

## 4 Discussion

In the realm of antiproliferative agents, this groundbreaking study on cyanomethyl vinyl ether derivatives not only unveils novel compounds but also propels the frontier of cancer research. The synthesis, employing a cutting-edge organocatalytic cyanovinylation methodology, not only underscores the accessibility but also highlights the versatility of these compounds in the pursuit of effective antiproliferative interventions.

Biological evaluations have unearthed compelling cytotoxic activities, particularly noteworthy in the SKOV3 ovarian carcinoma cells, where compounds **5E** and **7E** exhibited pronounced efficacy. The nuanced exploration of the Structure-Activity Relationship (SAR) provides crucial insights, emphasizing the pivotal role of the *E* isomer and specific substitutions in steering the biological activity of these derivatives.

Concluding from the SAR study, the geometry of the double bond of the cyanovinyl ethers is definitely important, being the *E* stereochemistry of the double bond a relevant structural feature giving the most biologically active compounds, even crucial in several cases for activity to be observed. Another important structural point to note in these compounds is also the R<sup>2</sup> substituent, more interesting those with low number of carbon atoms, in particular the most active ones with methyl group. With respect to the R<sup>1</sup> substituent (in position  $\alpha$  with respect to the cyano group), the influence of groups with fused aromatic rings such as phenanthrene is noteworthy. Therefore, **12E** derivative would be most interesting one from a biological point of view.

Computational assays forecast a promising future for these compounds, positioning them as potential anticancerous agents. The Absorption, Distribution, Metabolism, and Excretion (ADME) studies paint a favorable picture, suggesting not only efficacy but also safety in their pharmacokinetic profile. The meticulous adherence to Lipinski's Rule of Five further accentuates the drug-like properties of these compounds.

Molecular docking studies delve into the atomic-level interactions, revealing the profound affinity of compound **12E** with tubulin (with the best gScore value). The diverse amino acid interactions, including weak polar interactions with Met259 and Leu313, and robust polar interactions with Thr314 and Val315, underscore the multifaceted stability of the tubulin-compound **12E** complex. This interactional richness potentially translates into enhanced efficacy as an inhibitory agent.

Moreover, the spatial congruence in complex formations with colchicine provides a glimpse into potential mechanistic similarities,

enriching our understanding of their binding mechanisms and affinities for specific regions on the tubulin structure.

In summary, this work not only introduces a promising cohort of compounds with substantial cytotoxic potential but also advances our comprehension of their structural-functional dynamics. These findings lay a robust foundation for future drug development endeavors, heralding a new chapter in the pursuit of efficacious antiproliferative therapies.

## Data availability statement

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

## Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

## Author contributions

EM-E: Investigation, Writing—original draft. Maria Fuertes: Investigation, Writing—review and editing. SD: Investigation, Writing—review and editing. FG-T: Supervision, Writing—review and editing. DT: Funding acquisition, Investigation, Supervision, Writing—review and editing. Concepcion CA: Data curation, Funding acquisition, Supervision, Writing—review and editing.

## References

- Bakchi, B., Dileep Krishna, A., Sreecharan, E., Jaya Ganesh, V. B., Niharika, M., Maharshi, S., et al. (2022). An overview on applications of SwissADME web tool in the design and development of anticancer, antitubercular and antimicrobial agents: a medicinal chemist's perspective. *J. Molec. Struct.* 1259, 132712–132744. doi:10.1016/j.molstruc.2022.132712
- Bethencourt-Estrella, C. J., Delgado-Hernandez, S., Lopez-Arencibia, A., San Nicolas-Hernandez, D., Tejedor, D., Garcia-Tellado, F., et al. (2023). *In vitro* activity and mechanism of cell death induction of cyanomethyl vinyl ethers derivatives against *Trypanosoma cruzi*. *Int. J. Parasitol. Drugs Drug Resist.* 22, 72–80. doi:10.1016/j.ijpddr.2023.05.001
- Chaaban, S., and Brouhard, G. J. (2017). A microtubule bestiary: structural diversity in tubulin polymers. *Mol. Biol. Cell* 28, 2924–2931. doi:10.1091/mbc.116-05-0271
- Chao-Pellicer, J., Arberas-Jiménez, I., Delgado-Hernández, S., Sifaoui, I., Tejedor, D., Garcia-Tellado, F., et al. (2023). Cyanomethyl vinyl ethers against *Naegleria fowleri*. *ACS Chem. Neurosci.* 14, 2123–2133. doi:10.1021/acscchemneuro.3c00110
- 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, 42717–42730. doi:10.1038/srep42717
- Delgado-Hernández, S., García-Tellado, F., and Tejedor, D. (2021). Cyanovinylolation of aldehydes: organocatalytic multicomponent synthesis of conjugated cyanomethyl vinyl ethers. *Molecules* 26, 4120. doi:10.3390/molecules26144120
- Dumontet, C., and Jordan, M. A. (2010). Microtubule-binding agents: a dynamic field of cancer therapeutics. *Nat. Rev. Drug Discov.* 9, 790–803. doi:10.1038/nrd3253
- Findeisen, P., Mühlhausen, S., Dempewolf, S., Hertzog, J., Zietlow, A., Carlomagno, T., et al. (2014). Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. *Genome Biol. Evol.* 6, 2274–2288. doi:10.1093/gbe/evu187
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* 49, 6177–6196. doi:10.1021/jm051256o
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). *Gaussian 16, revision A. 03*. Wallingford CT: Gaussian, Inc.
- GLOBOCAN (2022). Globocan/factsheets/populations, Available at: <https://gco.iarc.who.int/media/globocan/factsheets/populations/900-world-fact-sheet.pdf> (accessed January 30 2024).
- Goodson, H. V., and Jonasson, E. M. (2018). Microtubules and microtubule-associated proteins. *Cold Spring Harb. Perspect. Biol.* 10, a022608. doi:10.1101/cshperspect.a022608
- Gottesman, M. M., Fojo, T., and Bates, S. E. (2002). Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat. Rev. Cancer* 2, 48–58. doi:10.1038/nrc706
- Graham, W., and Roberts, J. B. (1953). Intravenous colchicine in the management of gouty arthritis. *Ann. Rheum. Dis.* 12, 16–19. doi:10.1136/ard.12.1.16
- Howard, J., and Hyman, A. A. (2009). Dynamics and mechanics of the microtubule plus end. *Nature* 458, 903–908. doi:10.1038/nature01600
- Kingston, D. G. (2001). Taxol, a molecule for all seasons. *Chem. Commun. (Camb.)*, 2019–2024.
- Kovács, E., Ali, H., Minorics, R., Traj, P., Resch, V., Paragi, G., et al. (2023). Synthesis and antiproliferative activity of steroidal diaryl ethers. *Molecules* 28, 1196. doi:10.3390/molecules28031196
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* 64, 4–17. doi:10.1016/j.addr.2012.09.019

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Financial support from Grants PID2021-122558OB-I00, PID2022-136566NB-I00 and PDC2022-133706-I00 funded by MCIN/AEI/10.13039/501100011033 and by the “European Union Next GenerationEU/PRTR” and by the Basque Government (GV, IT1701-22) is gratefully 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/fphar.2024.1344042/full#supplementary-material>



- Loeffler, L. J., Sajadi, Z., and Hall, I. H. (1977). Antineoplastic agents 1 N-Protected vinyl, 1,2-dihaloethyl, and cyanomethyl esters of phenylalanine. *J. Med. Chem.* 20 (12), 1578–1584. doi:10.1021/jm00222a008
- Matthew, S., Chen, Q.-Y., Ratnayake, R., Fermain, C. S., Lucena-Agell, D., Bonato, F., et al. (2021). Gatorbulin-1, a distinct cyclodepsipeptide chemotype, targets a seventh tubulin pharmacological site. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2021847118. doi:10.1073/pnas.2021847118
- Mattiuzzi, C., and Lippi, G. (2019). Current cancer epidemiology. *J. Epidemiol. Glob. H.* 9, 217–222. doi:10.2991/jeqh.k.191008.001
- Podolski-Renić, A., Banković, J., Dinić, J., Rios-Luci, C., Fernandes M., X., Ortega, N., et al. (2017). DTA0100, dual topoisomerase II and microtubule inhibitor, evades paclitaxel resistance in P-glycoprotein overexpressing cancer cells. *Eur. J. Pharm. Sci.* 105, 159–168. doi:10.1016/j.ejps.2017.05.011
- Schrodinger Release (2015a). *Schrodinger Release 2015-1: Glide, version 6.9*, schrodinger. New York: LLC.
- Schrodinger Release (2015b). *Schrodinger Release 2015-1: protein preparation wizard, epik 3.1*, schrodinger. New York: LLC.
- Sliwoski, G., Kothiwale, S., Meiler, J., Lowe, E. W., Jr., and Barker, L. (2014). Computational methods in drug discovery. *E. Pharmacol. Rev.* 66 (1), 334–395. doi:10.1124/pr.112.007336
- Steinmetz, M. O., and Prota, A. E. (2018). Microtubule-targeting agents: strategies to hijack the cytoskeleton. *Trends Cell Biol.* 28, 776–792. doi:10.1016/j.tcb.2018.05.001



## OPEN ACCESS

## EDITED BY

Lei Xi,  
Virginia Commonwealth University,  
United States

## REVIEWED BY

Luca Sancineto,  
University of Perugia, Italy  
Arpan Acharya,  
University of Nebraska Medical Center,  
United States

## \*CORRESPONDENCE

Francisco Castillo,  
✉ francisco.castillo@medinaandalucia.es  
David Ramírez,  
✉ dramirez@udec.cl

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 23 February 2024

ACCEPTED 29 May 2024

PUBLISHED 10 July 2024

## CITATION

Castillo F, Ramírez D, Ramos MC, Martínez-Arribas B, Domingo-Contreras E, Mackenzie TA, Peña-Varas C, Lindemann S, Montero F, Annang F, Vicente F, Genilloud O, González-Pacanowska D and Fernandez-Godino R (2024), Repurposing the Open Global Health Library for the discovery of novel Mpro destabilizers with scope as broad-spectrum antivirals.  
*Front. Pharmacol.* 15:1390705.  
doi: 10.3389/fphar.2024.1390705

## COPYRIGHT

© 2024 Castillo, Ramírez, Ramos, Martínez-Arribas, Domingo-Contreras, Mackenzie, Peña-Varas, Lindemann, Montero, Annang, Vicente, Genilloud, González-Pacanowska and Fernandez-Godino. 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.

# Repurposing the Open Global Health Library for the discovery of novel Mpro destabilizers with scope as broad-spectrum antivirals

Francisco Castillo<sup>1\*†</sup>, David Ramírez<sup>2\*†</sup>, María C. Ramos<sup>1</sup>, Blanca Martínez-Arribas<sup>3</sup>, Elisabeth Domingo-Contreras<sup>1</sup>, Thomas A. Mackenzie<sup>1</sup>, Carlos Peña-Varas<sup>2,4</sup>, Sven Lindemann<sup>5</sup>, Fernando Montero<sup>1,6</sup>, Fredderick Annang<sup>1</sup>, Francisca Vicente<sup>1</sup>, Olga Genilloud<sup>1</sup>, Dolores González-Pacanowska<sup>3</sup> and Rosario Fernandez-Godino<sup>1</sup>

<sup>1</sup>Fundación MEDINA, Granada, Spain, <sup>2</sup>Departamento de Farmacología, Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción, Chile, <sup>3</sup>Instituto de Parasitología y Biomedicina Lopez-Neyra, Consejo Superior de Investigaciones Científicas, Granada, Spain, <sup>4</sup>Doctorado en Biotecnología Molecular, Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción, Chile, <sup>5</sup>Strategic Innovation, Merck Healthcare KGaA, Darmstadt, Germany, <sup>6</sup>Department of Physical Chemistry and Institute of Biotechnology, Universidad de Granada, Granada, Spain

The SARS coronavirus 2 (SARS-CoV-2) epidemic remains globally active. The emergence of new variants of interest and variants of concern (VoCs), which are potentially more vaccine-resistant and less sensitive to existing treatments, is evident due to their high prevalence. The prospective spread of such variants and other coronaviruses with epidemic potential demands preparedness that can be met by developing fast-track workflows to find new candidates that target viral proteins with a clear *in vitro* and *in vivo* phenotype. Mpro (or 3CLpro) is directly involved in the viral replication cycle and the production and function of viral polyproteins, which makes it an ideal target. The biological relevance of Mpro is highly conserved among betacoronaviruses like HCoV-OC43 and SARS-CoV-2, which makes the identification of new chemical scaffolds targeting them a good starting point for designing broad-spectrum antivirals. We report an optimized methodology based on orthogonal cell-free assays to identify small molecules that inhibit the binding pockets of both SARS-CoV-2-Mpro and HCoV-OC43-Mpro; this blockade correlates with antiviral activities in HCoV-OC43 cellular models. By using such a fast-tracking approach against the Open Global Health Library (Merck KGaA), we have found evidence of the antiviral activity of compound OGH98. *In silico* studies dissecting intermolecular interactions between OGH98 and both proteases and comprising docking and molecular dynamics simulations (MDSS) concluded that the binding mode was primarily governed by conserved H-bonds with their C-terminal amino acids and that the rational design of OGH98 has potential against VoCs proteases resistant to current therapeutics.

## KEYWORDS

COVID-19, betacoronavirus, high-throughput sequencing M-PRO thermal shift, molecular dynamics simulations, molecular docking, antiviral agents

# 1 Introduction

As coronaviruses continue their global spread, new variants of concern (VoCs) are constantly being detected by genetic surveillance. This has raised a need for the identification of effective therapeutics that overcome the decreased success and increased resistance to existing antivirals (Tan et al., 2022). Mpro mediates viral replication, and it is directly linked to the infection spread in host organisms, which makes it a straightforward target. Rational design studies that led to the validation of the Mpro inhibitor PF-07321332 offer proof of such a target's druggability. PF-07321332 (nirmatrelvir) is a peptidomimetic drug designed and developed by Pfizer that blocks Mpro viral proteases of beta and alphacoronaviruses, as demonstrated by a low-throughput enzymatic confirmatory assay (Pang et al., 2023). After a couple of years of research, that pan-inhibitory activity demonstrated effective translation well into animal models and human patients when delivered orally (Owen et al., 2021). Therefore, the direct implementation of state-of-the-art cell-free/biochemical assays in high-throughput screening (HTS) format as primary assays could lead to a more rapid discovery of alternative drug precursors, which may be particularly useful against currently emerging Mpro variants resistant to current inhibitors such as Paxlovid (Ip et al., 2023). Finding alternative drug discovery workflows to identify new pharmacophores to block Mpro is a need that has also been flagged by structure–activity studies with phenylbenziselenazol-3(2H)-one (ebselen) derivatives, in which Mpro mutations of concern at the so-called *gatekeeper residues* leading to Mpro hyperactivity were studied by Sahoo et al. (2023).

A medium-throughput version of a cell-free FRET enzymatic assay used on a small subset of classic natural products has led to the identification of quercetin as a SARS-CoV-2-Mpro inhibitor, whose mechanism of action relies on the destabilization of the Mpro target according to the thermal shift assay and *in silico* structural biology studies (Abian et al., 2020). Unfortunately, quercetin has shown a marginal *in vivo* effect as an antiviral agent so far, and its mild therapeutic benefits are hard to correlate with Mpro inhibition. Instead, quercetin might inhibit other viral proteins, such as S proteins, or even furin (Di Petrillo et al., 2022). Such mechanistic ambiguity could be overcome with the use of cell-free HTS setups to screen other chemical spaces, such as libraries of synthetic small molecules, to detect inhibitors of higher specificity for the Mpro target. Ideally, this approach should be coupled with secondary assays quickly converging toward bioactive compounds with better broad-spectrum antiviral profiles and absorption, distribution, metabolism, and excretion (ADME) profiles than quercetin.

To test this premise, we have performed a thermal shift assay with the Open Global Health Library (OGHL) (Merck KGaA, Darmstadt, Germany), which is comprised of 250 bioactive synthetic small molecules with demonstrated anti-infective applications (Abraham et al., 2020), but it has never been assayed against coronavirus proteases. The micromolar inhibitory activity of the best compound, OGHL98, a SARS-CoV-2-Mpro destabilizer ( $\Delta T_m = -4.5^\circ\text{C} \pm 0.3^\circ\text{C}$ ), was further validated against SARS-CoV-2-Mpro and HCoV-OC43-Mpro (two proteases that share 48.5% of the amino acid sequence identity) in the respective FRET enzymatic assays. To understand how this compound was blocking both Mpro proteases, we designed a computational pipeline including

molecular docking followed by long-term molecular dynamics simulations (MDSs). According to such *in silico* structural studies, the broad-spectrum inhibitory activity against these betacoronaviral proteases relies on a conserved network of intermolecular hydrogen bonds established between the C-terminal residues of each protease and the 4-(methylcarbamoyl) benzoic acid moiety of the OGHL98 compound. Finally, detectable antiviral activity was confirmed for OGHL98 against the HCoV-OC43 virus in the micromolar concentration range (cytopathic half maximal effective concentration,  $EC_{50}$  value of  $33\ \mu\text{M}$ ; maximum viral RNA inhibition  $>50\%$  at  $7.5\ \mu\text{M}$ ). Future studies are required to confirm the promising ADME/Tox profile predicted in this work and further improve the potency/selectivity of OGHL98 and its benzoic acid moiety using medicinal chemistry tools. More importantly, the orthogonal workflow presented here and based on cost-effective cell-free assays has been demonstrated to be efficient at feeding computational rational design workflows with interesting inhibitors, which delineates a straightforward discovery workflow to be implemented on other cysteine proteases from viruses with pandemic potential.

## 2 Materials and methods

### 2.1 Recombinant production of SARS-CoV-2 and HCoV-OC43 Mpro cysteine proteases

The original pGEX-6p-1 plasmid was donated by Professor Yang's laboratory. Each plasmid encodes an N-terminal GST tag, followed by the SARS-CoV family autocleavage site (TSAVLQSGFRK) that allows the *in vivo* release of the final C-terminal SARS-CoV-2 and HCoV-OC43 Mpro-His<sub>6x</sub> used for our protein studies (Xue et al., 2007). The pGEX-6p-1 plasmid (100 ng) was transformed into 20  $\mu\text{L}$  of BL21-RIPL competent cells (Agilent™). Bacterial cultures with successful transformants were grown in 250 mL of LB/ampicillin (100  $\mu\text{g}/\text{mL}$ )/chloramphenicol (34  $\mu\text{g}/\text{mL}$ ) media at  $37^\circ\text{C}$  overnight. Then, 4 L of LB/ampicillin (100  $\mu\text{g}/\text{mL}$ ) were inoculated (1/100 dilution) and incubated at  $37^\circ\text{C}$  until  $OD_{600}$  0.8. Overexpression was induced with 1 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) at  $18^\circ\text{C}$  overnight. Cells were harvested by centrifugation at  $4^\circ\text{C}$  for 10 min at 7,000 g and re-suspended in lysis buffer (Tris 20 mM, pH 8). Cell lysis was achieved by sonication, from which the debris were removed by centrifugation at  $4^\circ\text{C}$  for 30 min at 10,000 g. Mpro-His<sub>6x</sub> protein was captured from the filtered supernatant using a cobalt HiTrap Column (Cytiva™) and then eluted with 250 mM imidazole. Eluted Mpro-His<sub>6x</sub> was then buffer-exchanged in lysis buffer for further purification using a HiTrap™ Capto™ Q ImpRes Anion Exchanger (Cytiva™). Fractions of purity above 95% were eluted at 300 mM NaCl, which was buffer-exchanged in phosphate-buffered saline (PBS) before use.

### 2.2 OGHL compound library

The Open Global Health Library, comprising 250 synthetic small molecules, was donated by Merck KGaA™ (Darmstadt,

Germany) upon request via <https://www.merckgroup.com/en/research/open-innovation/biopharma-open-innovation-portal/open-global-health-library.html>. All compounds and controls in this study were provided as 10 mM DMSO stocks and had analytically confirmed purities >90%.

## 2.3 Thermal shift assay

SYPRO Orange (Thermo Fisher Scientific™) was employed as an extrinsic fluorescent probe. An assay master mix containing 5x SYPRO and SARS-CoV-2 Mpro 3  $\mu$ M in PBS was dispensed into 384-well microplates containing the compound library, which was assayed in 2% DMSO and 100- $\mu$ M final concentration (Abian et al., 2020). Negative controls contained the master mix and an equivalent volume of DMSO. Positive controls contained the master mix and 200  $\mu$ M quercetin. Unfolding curves were registered from 20°C to 95°C at a 0.5°C/min scan rate in a Bio-Rad™ CFX 384 qPCR real-time thermal cycler using default HEX filter settings. The midpoint unfolding temperature,  $T_m$ , was calculated in each well as the inflection point and compared to the controls. Primary destabilizing hits were considered using a threshold of  $T_m$  shift  $\leq -2.0^\circ\text{C}$ .

## 2.4 Enzymatic inhibitory assay

Primary hits at 2x concentration and 2% DMSO were pre-incubated with 4  $\mu$ M SARS-CoV-2-Mpro or HCoV-OC43 enzyme in the assay buffer (20 mM Tris-HCl, pH 7.3, 100 mM NaCl, 1 mM EDTA, and 1 mM TCEP) for 30 min in a low-volume 384-well plate (Qiao et al., 2021). Enzyme activity was monitored on an EnVision Multilabel Plate Reader (PerkinElmer™) at Ex/Em of 320/405 nm after the addition of an equivalent volume of 10  $\mu$ L of a 2x concentration of the 40  $\mu$ M peptide substrate MCA-AVLQSGFR-Lys (Dnp)-K. The labeled peptide was purchased from JPT™ as a lyophilized powder (purity >95%). The enzymatic reaction was monitored until reaching equilibrium, according to the end-point inhibitory assay, using quercetin as the positive control and DMSO as the negative control. End-point data for each compound were expressed as fluorescence arbitrary units ( $y$ -axis) against the log of compound concentration ( $x$ -axis), from which the  $IC_{50}$  values were obtained.

## 2.5 Antiviral activity in a HCoV-OC43 surrogate model

**Reagents and antibodies.** Quercetin, bovine serum albumin, resazurin (Sigma-Aldrich), and ribavirin (Santa Cruz Biotechnology™) were diluted in 100% DMSO. The mouse monoclonal anti-HCoV-OC43 antibody (MAB 2012) was purchased from Millipore™; the Alexa Fluor 488-conjugated anti-mouse secondary antibody and Hoechst 33342 were purchased from Thermo Fisher Scientific™ (Martínez-Arribas et al., 2023).

**Cell culture.** The cell lines used in this study were obtained from the American Type Culture Collection (ATCC). The human lung

fibroblast cell line MRC-5 (CCL-171, ATCC) was cultured in minimum essential medium (MEM) (Life Technologies™) supplemented with 10% fetal bovine serum (FBS) (Life Technologies™), 100 units/mL penicillin, and 100  $\mu$ g/mL streptomycin (Life Technologies™). Cells were incubated at 37°C in a humidified atmosphere of 5%  $\text{CO}_2$  and were periodically analyzed and confirmed to be *mycoplasma* negative.

**Virus production.** The human betacoronavirus HCoV-OC43 (VR-1588, ATCC) was propagated in MRC-5 human cells. In brief, MRC-5 cells were seeded at 90% confluence and inoculated with HCoV-OC43 in infection media (MEM, 2% inactivated FBS, penicillin/streptomycin). Infected cells were incubated for 2 h at 33°C, rocking the flask every 15 min for virus adsorption, and the culture was completed with infection media after adsorption. Infected cells were incubated at 33°C for 5–7 days until more than 50% of the cells presented a cytopathic effect (CPE), resulting in cell death. The infected culture was subjected to three freeze–thaw cycles and centrifuged at 3,000 g for 10 min at 4°C to spin down cells and cell debris for virus recovery. Viral particles were recovered from the supernatant, aliquoted in cryotubes, rapidly frozen in a dry-ice/ethanol bath, and stored at  $-80^\circ\text{C}$  (Martínez-Arribas et al., 2023).

**Batch infection with HCoV-OC43.** MRC-5 cells at 90% confluence were infected with HCoV-OC43 at a multiplicity of infection (MOI) of 0.1. Virus adsorption was performed for 2 h at 33°C, rocking the cells every 15 min, and then the infected cells were incubated for 24 h at 33°C before seeding into 96-well plates (Martínez-Arribas et al., 2023).

**CPE inhibition and cytotoxicity assays.** Infected cells were washed, trypsinized, and seeded in plates containing the compounds at a cellular concentration of  $2 \times 10^4$  cells/well in infection media. The plates were incubated at 37°C for 96 h in the presence of the compounds. Infection media were aspirated 5 days after infection, and 120  $\mu$ L of infection media containing 20% resazurin was added per well. Infected cells treated with 400  $\mu$ M ribavirin and infected cells with 0.2% DMSO were used as the positive and negative controls, respectively. MRC-5 cells were seeded/well in 96-well plates containing the compounds. After 96 h, the cells were incubated with 20% resazurin for 2 h at 37°C. MRC-5 cells treated with 50  $\mu$ M of tamoxifen were used as the negative control (100% cell death reference), while positive controls corresponded to MRC-5 cells incubated in the presence of 0.2% DMSO. Fluorescence was determined at 550–590 nm using a Tecan™ Infinite Plate Reader (Martínez-Arribas et al., 2023).

**RNA isolation and RT-PCR.** Viral RNA from the supernatants was purified using the Macherey-Nagel NucleoSpin RNA Kit. RT-qPCR was performed in a single step using the One-Step TB Green PrimeScript RT-PCR Kit II (Takara Bio™). The HCoV-OC43 nucleocapsid gene was amplified with the following primers: the forward primer 5' AGCAACCAGGCTGATGTCAATACC-3' and the reverse primer 5' AGCAGACCTTCCTGAGCCTTCAAT-3. A standard curve was generated with purified HCoV-OC43 RNA (Viracell™) (Min et al., 2020; Martínez-Arribas et al., 2023).

**Immunofluorescence of HCoV-OC43.** For HCoV-OC43 detection, 4 days after infection, cells were fixed for 20 min with 4% paraformaldehyde and permeabilized for 10 min with 0.4% Triton X-100. After 1 h of blocking with 5% BSA, the cells were incubated O/N with anti-HCoV-OC43. Cells were washed and

incubated for 1 h with the Alexa Fluor 488-conjugated anti-mouse secondary antibody (Thermo Fisher Scientific™) and then washed and incubated for 20 min with Hoechst 33342 for nuclei staining. Digital images were captured using the Operetta CLS High Content Analysis System (PerkinElmer™) with a ×5 air objective. The number of nuclei and the number of cells positive for HCoV-OC43 staining were determined, and the percentage of infection was expressed as the ratio of HCoV-OC43 positive cells/total nuclei (Martínez-Arribas et al., 2023).

**Data analysis.** CPE inhibition activities of non-cytotoxic compounds were determined using Equation (1):

$$\text{CPE Inhibition (\%)} = \frac{(\text{Fluo}_{\text{well}} - \text{Fluo}_{\text{neg}})}{(\text{Fluo}_{\text{pos}} - \text{Fluo}_{\text{neg}})} \times 100, \quad (1)$$

where  $\text{Fluo}_{\text{well}}$  is the measured fluorescence of each well,  $\text{Fluo}_{\text{pos}}$  is the average fluorescence of the positive control (infected MRC-5 cells 0.2% DMSO), and  $\text{Fluo}_{\text{neg}}$  is the average fluorescence of the negative control (non-infected cells).

**Cytotoxicity:** cellular cytotoxicity was determined using Equation (2):

$$\text{Viability (\%)} = \frac{(\text{Fluo}_{\text{well}} - \text{Fluo}_{\text{neg}})}{(\text{Fluo}_{\text{pos}} - \text{Fluo}_{\text{neg}})} \times 100, \quad (2)$$

where  $\text{Fluo}_{\text{well}}$  is the measured fluorescence of each well,  $\text{Fluo}_{\text{neg}}$  is the average fluorescence of the negative control (cells treated with 50 nM tamoxifen), and  $\text{Fluo}_{\text{pos}}$  is the average fluorescence of the positive control (0.2% DMSO) (Martínez-Arribas et al., 2023).

## 2.6 Computational studies

**OC43 MPro modeling:** given that the structure of HCoV-OC43-Mpro has not been solved yet, a homology model was built using the crystal structure of SARS-CoV-2-Mpro as the template (PDB code: 6LU7, (Jin et al., 2020)). Since OC43-MPro is contained in the OC43 replicase polyprotein 1ab (Uniprot ID: P0C6X6), both the sequences (SARS-CoV-2-1ab and OC43-1ab) were aligned. The MPro fraction that showed the optimal sequence alignment was used to build the final model of HCoV-OC43-Mpro, which was optimized using Prime (Schrödinger Suite) and validated using PROCHECK (Jacobson et al., 2004).

**Molecular docking:** new inhibitors, OGHL98 and OGHL43, and control compounds, aspirochlorine and quercetin, were prepared with Maestro and LigPrep. The prepared ligands were docked to the HCoV-OC43-Mpro model and the SARS-CoV-2-Mpro PDB (code: 6LU7) using Glide and Schrödinger Suite (Halgren et al., 2004). Before docking calculations, proteins were prepared using Maestro (Madhavi Sastry et al., 2013), which removed the original ligands, metals, and water molecules. Hydrogens of ionizing residues at pH 7.4 ± 2.0 were then added, and the missing side chains were modeled by Prime. The minimization of the corresponding protein structures was calculated using OPLS3. The same grid box was defined for both targets using the N3 ligand co-crystallized in SARS-CoV2-Mpro as the center of the corresponding boxes. The docking was then performed with the Glide standard precision (SP) function (Friesner et al., 2006). The top 10 poses per docked ligand were

selected and subjected to rescoring by calculating the binding free energy ( $\Delta G_{\text{bind}}$ ) with Prime (Jacobson et al., 2002; 2004), which was calculated in terms of the molecular mechanics-generalized born surface area (MM-GBSA). This computational method combines molecular mechanics energy and implicit solvation models, which enables rescoring and correlation between the experimental activities ( $\text{IC}_{50}$ ) and the predicted  $\Delta G_{\text{bind}}$ . The corresponding  $\Delta G_{\text{bind}}$  values for each ligand–target complex were calculated, as previously reported (Rojas-Prats et al., 2021).

**Molecular dynamics simulations:** the best post-processed docking solutions between the four inhibitors of interest and both Mpro targets (SARS-CoV-2 and HCoV-OC43) were selected according to their best  $\Delta G_{\text{bind}}$  profiles. Such docking solutions were subjected to MDs using Desmond software (Bowers et al., 2006) and OPLS3e (Roos et al., 2019). To prepare the systems, the ligand–target complexes were solvated with pre-equilibrated water molecules (SPC) in a periodic boundary condition box. Then, the systems were neutralized by adding  $\text{Na}^+$  or  $\text{Cl}^-$  counter-ions at a final concentration of 0.15 M NaCl to simulate the physiological conditions. Next, each system was relaxed using the default Desmond relaxation protocol and then equilibrated for 25 ns with a spring constant force of  $1.0 \text{ kcal} \times \text{mol}^{-1} \times \text{\AA}^{-2}$ , which was applied to the backbone atoms of the proteins and the ligands. The simulations were performed using the NPT ensemble at constant pressure (1 atm), temperature (310 K), and number of atoms using the isothermal–isobaric ensemble and the Nose–Hoover method, with a relaxation time of 1 ps. The MTK algorithm was applied with a time step of 2 fs. Then, the last frame was taken, and a second non-restricted MD was extended until 3  $\mu\text{s}$  if necessary, for which the same conditions described above were applied. Systems were then analyzed using in-house PyMol and VMD scripts.

**Prediction of the ADME/Tox properties:** we computed the physicochemical descriptors, ADME, pharmacokinetic properties, and drug-like nature of the studied compounds using the SwissADME server (Daina, Michielin, and Zoete, 2017a). In brief, 42 descriptors were predicted, including physicochemical, lipophilicity, water solubility, and pharmacokinetic properties. From these descriptors, SwissADME assessed the compounds' acceptability based on a bioavailability score (drug-likeness).

## 3 Results

The compound OGHL98 was characterized as a novel inhibitor of coronaviral cysteine proteases using protein thermal shift assays and FRET enzymatic assays. To complement the *in vitro* characterization of OGHL98 beyond these cell-free setups, biological activity tests were performed using a model betacoronavirus, HCoV-OC43. The specific molecular interaction profiles with coronaviral proteases used in the enzymatic assays, HCoV-OC43-Mpro and SARS-CoV-2-Mpro, were studied using a computational pipeline. The pipeline included molecular modeling and docking, followed by binding free energy calculations and long-term molecular dynamics simulations. The latest highlighted the intermolecular contacts established by the main pharmacophoric core of OGHL98. We also predicted a promising ADME/Tox profile for OGHL98 to confirm the efficiency of our drug discovery approach in identifying new and developable antiviral inhibitors.



### 3.1 Identification of novel Mpro destabilizers OGHL43 and OGHL98

Overall, the 250 screened compounds from the Open Global Health Library were synthesized to be structurally diverse. As such, these compounds had shown activity in anti-infective screening campaigns against diverse targets such as amebiasis, AMR, Chagas, visceral leishmaniasis, cryptosporidiosis, human African trypanosomiasis, malaria, schistosomiasis, tuberculosis, or soil-transmitted helminthiasis, but they had never been used in screens against viral proteases.

As described in more detail in [Supplementary Figure S1](#), the collection of 250 diverse compounds presented the following overall physicochemical properties: molecular weight range between 200.0 Da and 700.0 Da, partition coefficient (LogP) between -0.1 and 8.0, number of rotatable bonds between 1 and 15, number of hydrogen bond donors between 0 and 5, number of hydrogen bond acceptors between 2 and 12, and total polar surface area between 0 Å<sup>2</sup> and 180 Å<sup>2</sup>.

Such a library of 250 bioactive compounds and two reference compounds (aspirochlorine and quercetin), previously identified as SARS-CoV-2 antivirals ([Singh, Sharma, and Nandi, 2020](#)), were jointly screened against purified SARS-CoV-2-Mpro by thermal shift. The robustness of the screening results was supported by plate quality controls using quercetin ([Abian et al., 2020](#)) and DMSO, which yielded a Z'-factor of 0.7 ([Lilly et al., 2004](#)). Two hits were selected because they presented a negative T<sub>m</sub> shift greater than 2°, which implies a significant destabilization of the protein in the presence of the ligands of interest. Negative shift had already been validated as a feasible mechanism of action to block Mpro with non-covalent small molecules like quercetin ([Abian et al., 2020](#); [Mangiavacchi et al., 2021](#)). Then, the two new small molecules, OGHL98 and OGHL43, and the reference compounds aspirochlorine and quercetin were subjected to a functional enzymatic FRET assay with comparable resolution (Z'-factor = 0.6; assay window >5), in which quercetin, as the positive control compound, presented an expected IC<sub>50</sub> of 28.2 ± 11.4 μM ([Abian et al., 2020](#)). Unfortunately, aspirochlorine (ΔT<sub>m</sub> < -10°C, [Supplementary Figure S2B](#)), one of the strongest destabilizers found in the primary screen and whose antiviral activity had been postulated by other authors ([Singh, Sharma, and Nandi, 2020](#)), could not be orthogonally confirmed as an Mpro inhibitor for interfering with the signal of the FRET enzymatic assay ([Supplementary Figure S3B](#)). More importantly, the FRET assay confirmed the blockade of SARS-CoV-2-Mpro proteolysis by the two new destabilizing molecules, OGHL98 and OGHL43. Furthermore, OGHL98 was defined as the top-performing compound (ΔT<sub>m</sub> = -4.5 ± 0.3; IC<sub>50</sub> = 19.6 ± 5.6 μM; [Figures 1A–C](#)). OGHL43 (ΔT<sub>m</sub> = -3.7 ± 0.3; IC<sub>50</sub> ~ 80 μM; [Supplementary Figures S2A, S3A](#)) was validated as another suitable destabilizer but with lower potency. The dose-response activity of these two new OGHL inhibitors implied a rate of confirmed hits of 0.8% for the whole orthogonal screen, which is in line with the expected efficiency and quality standards of the HTS format for cell-free assays ([Lilly et al., 2004](#)). Altogether, these results confirmed OGHL98 as the novel SARS-CoV-2-Mpro destabilizer that demonstrated the

highest potency in two cell-free orthogonal assays, reporting a specific impact on protein folding and activity.

Complementarily, we recombinantly expressed and purified the HCoV-OC43-Mpro enzyme to test the inhibitory activity of OGHL98 by a FRET enzymatic assay, in which the compound presented an IC<sub>50</sub> of 11.4 ± 3.1 μM ([Figure 1C](#), blue trace). This inhibitory activity against HCoV-OC43-Mpro is comparable to the one observed in the SARS-CoV-2 FRET assay (IC<sub>50</sub> of 19.6 ± 5.6 μM; [Figure 1C](#), orange trace). Jointly, the results pointed toward a broad-spectrum mechanism of action for OGHL98, which was capable of blocking, at low micromolar concentrations, two betacoronaviral proteases with a 48.5% amino acid sequence homology. This mechanism of action would justify potential antiviral activity in biological assays measuring the inhibition of the infection by different betacoronaviruses related to HCoV-OC43 and SARS-CoV-2.

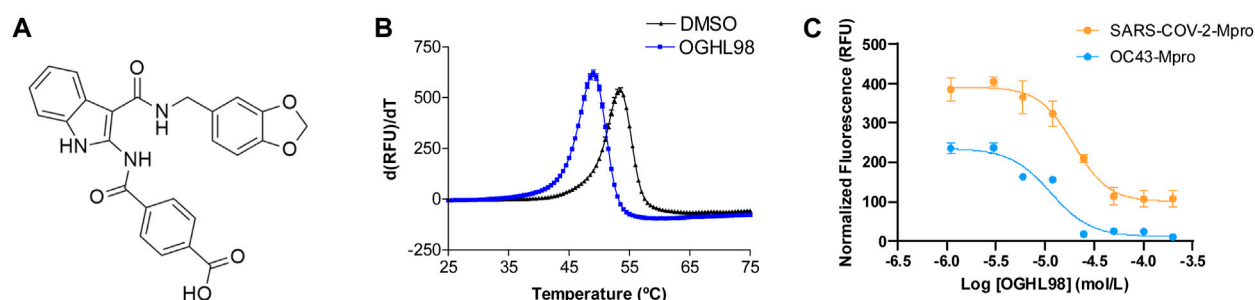
### 3.2 HCoV-OC43 surrogate model confirmed the antiviral activity of OGHL98

The Mpro destabilizers that were identified in the thermal shift assay were further validated *in vitro* using biological assays. This way, we expected to link their already defined mechanism of action to a specific antiviral activity in infected cells.

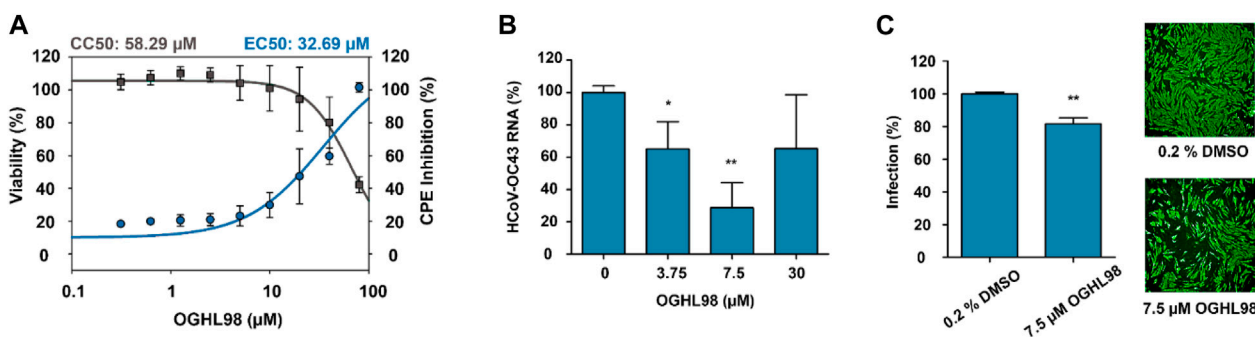
The biological characterization of the best compounds was first addressed through a simple phenotypic assay, in which a successful infection of HCoV-OC43 caused a measurable CPE in the lung cell line MRC-5 ([Smee et al., 2017](#); [Martínez-Arribas et al., 2023](#)). CPE inhibition was assessed 96 h after treatment. The EC<sub>50</sub> values reported for antiviral activity were calculated and compared with the corresponding CC<sub>50</sub> (half maximal cytotoxic concentration) values, which were obtained in parallel in non-infected MRC-5 cells treated with the same compounds so that non-specific cytotoxic effects could be discriminated. OGHL98 presented the highest activity and selectivity according to an EC<sub>50</sub> value of 32.69 μM and a CC<sub>50</sub> value of 58.29 μM ([Figure 2A](#)). The EC<sub>50</sub> value for quercetin was 56.47 μM, and it presented a lower CC<sub>50</sub> value (28.98 μM, [Supplementary Figure S4](#)), thus indicating a lack of selectivity that could explain the ambiguous activities described for this compound (<https://pubchem.ncbi.nlm.nih.gov/source/hsdb/3529>). In addition, aspirochlorine and OGHL43 were deprioritized because they did not exhibit meaningful CPE inhibition ([Supplementary Figure S4](#)).

To link the CPE inhibitory activity of the best compound, OGHL98, to a specific blockade of viral propagation, the RNA levels of HCoV-OC43 were evaluated in the supernatant of infected MRC-5 cells ([Min et al., 2020](#)). At a concentration of 7.5 μM, HCoV-OC43 RNA levels were reduced by more than 50% ([Figure 2B](#)), orthogonally confirming the inhibitory activity observed in the CPE assay.

Finally, we assessed whether the decrease in viral egress induced by OGHL98 corresponded to reduced replication levels of HCoV-OC43 within host cells. For this, we performed a complementary immunofluorescence study with a monoclonal antibody directed against the nucleocapsid of HCoV-OC43 ([Figure 2C](#)). The data confirmed that the infection of MRC-5 cells decreased by 20% after 72 h of treatment with 7.5 μM of OGHL98.



**FIGURE 1**  
Discovery and characterization of OGHL98 as a novel inhibitor of coronaviral proteases. **(A)** Structure of OGHL98. **(B)** Thermal shift validation reporting a  $\Delta T_m = -4.5 \pm 0.3$  compared to DMSO control. **(C)** FRET enzymatic assay reporting an  $IC_{50} = 19.6 \pm 5.6 \mu M$  for SARS-CoV-2-Mpro and  $11.4 \pm 3.1 \mu M$  for HCoV-OC43-Mpro.



**FIGURE 2**  
Characterization of the antiviral activity of OGHL98. **(A)** Evaluation of the OGHL98 inhibitor in MRC-5 cells infected with HCoV-OC43 by dose-response curves:  $CC_{50}$  (black) and  $EC_{50}$  (blue) values for CPE inhibition at 5 days and 96 h post-infection in the presence of the compound. **(B)** HCoV-OC43 RNA levels in infected cells treated with increasing concentrations of OGHL98. **(C)** Representative immunofluorescence images and corresponding bar plots quantifying the percentage of infection of HCoV-OC43 after 72 h in the presence of  $7.5 \mu M$  OGHL98. Control conditions 0.2% DMSO.

### 3.3 *In silico* studies validated the target engagement of novel destabilizers to SARS-CoV-2-Mpro and HCoV-OC43-Mpro

The identification and confirmation of OGHL98, which blocks viral targets directly involved in the viral replication cycle and the production/function of viral polypeptides like the two coronaviral proteases studied here, could be considered a valid starting point to design broad-spectrum antivirals. Nevertheless, the potency, specificity, ADME, and toxicity profile of OGHL98 must be optimized through an iterative process of rational design that commonly begins with the identification of key pharmacophoric factors for proper ligand binding to the target(s) of interest.

To meet the first objective and since the structure of HCoV-OC43-Mpro has not been solved yet, a homology model was built using the crystal structure of SARS-CoV-2-Mpro as the template (PDB code: 6LU7 (Jin et al., 2020)), both of which present 48.5% of the amino acid sequence identity. Given that OC43-Mpro is contained within the OC43 replicase polyprotein 1ab (Uniprot ID: P0C6X6), the two sequences (SARS-CoV-2-1ab and OC43-1ab) were aligned, and the Mpro fraction that presented the best sequence alignment was used to build the final HCoV-OC43-Mpro

model. Accordingly, further computational studies were performed with the resulting Mpro structures (SARS-CoV-2 and HCoV-OC43) to investigate the molecular determinants of the destabilization induced by the binding of our selected compounds (Figures 3, 4; Supplementary Figure S5).

The newly identified compounds (OGHL43 and OGHL98) were docked into both SARS-CoV-2-Mpro and HCoV-OC43-Mpro targets. Quercetin and aspirochlorine were also docked as control compounds. Then, all docking poses were re-scored according to their predicted binding free energy. Top-scoring solutions were selected to assess the most relevant interactions between each compound and key residues of SARS-CoV-2-Mpro (T26, E166, and Q189), as well as the equivalent residues of HCoV-OC43-Mpro (H41, E166, and Q189) (Supplementary Figure S5), which, in all cases, resulted in feasible contacts and geometries and, thus, suggested a specific binding for all cases considered.

The next step was a full characterization of the interactions between each ligand/compound and both targets, SARS-CoV-2-Mpro and HCoV-OC43-Mpro, by molecular dynamics simulations. We ran long-term MDs for both control compounds quercetin and aspirochlorine ( $3 \mu s$ ) and the novel inhibitors OGHL43 and OGHL98 ( $1 \mu s$ ). Root-mean-square deviation (RMSD) profiles for

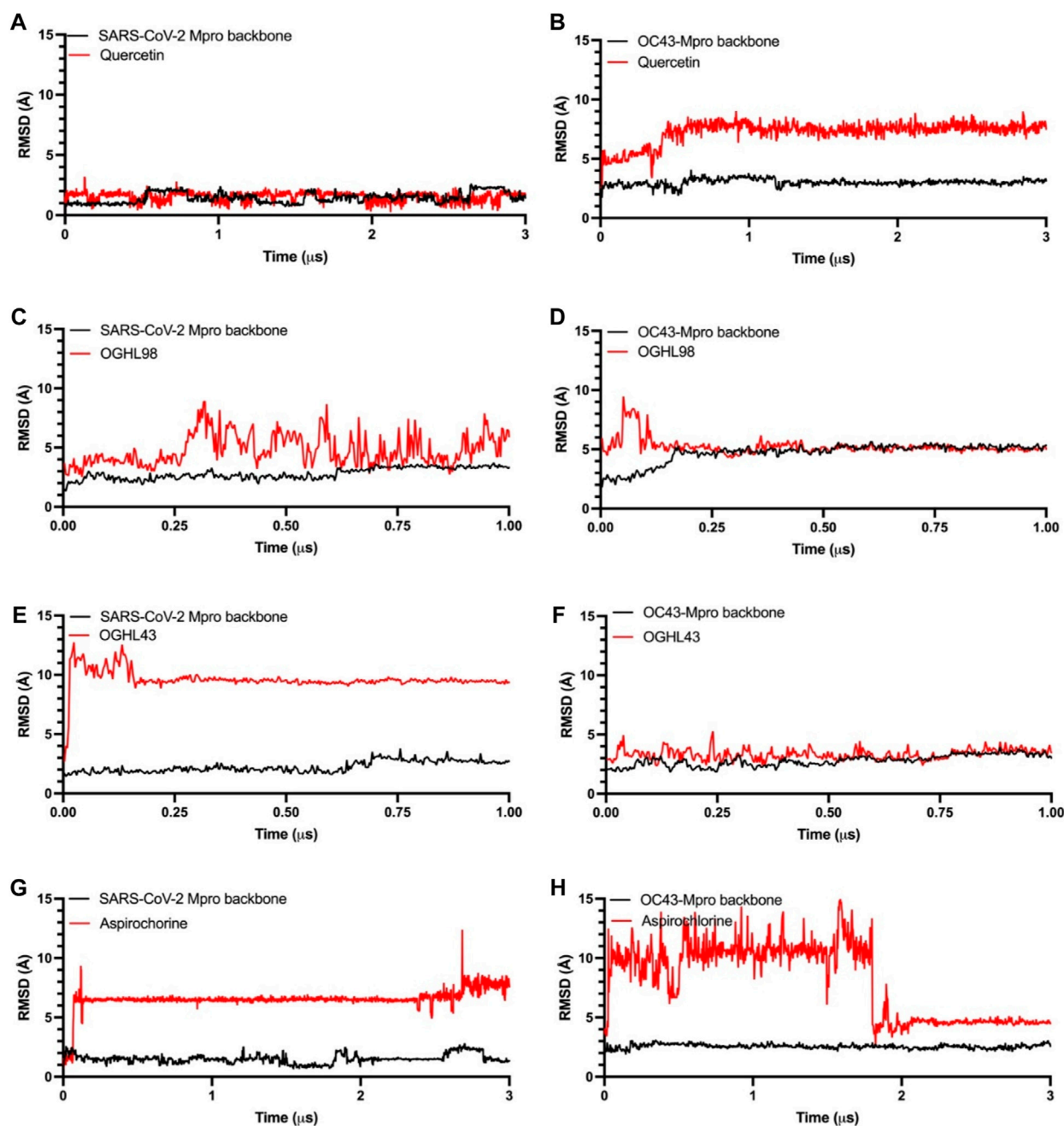
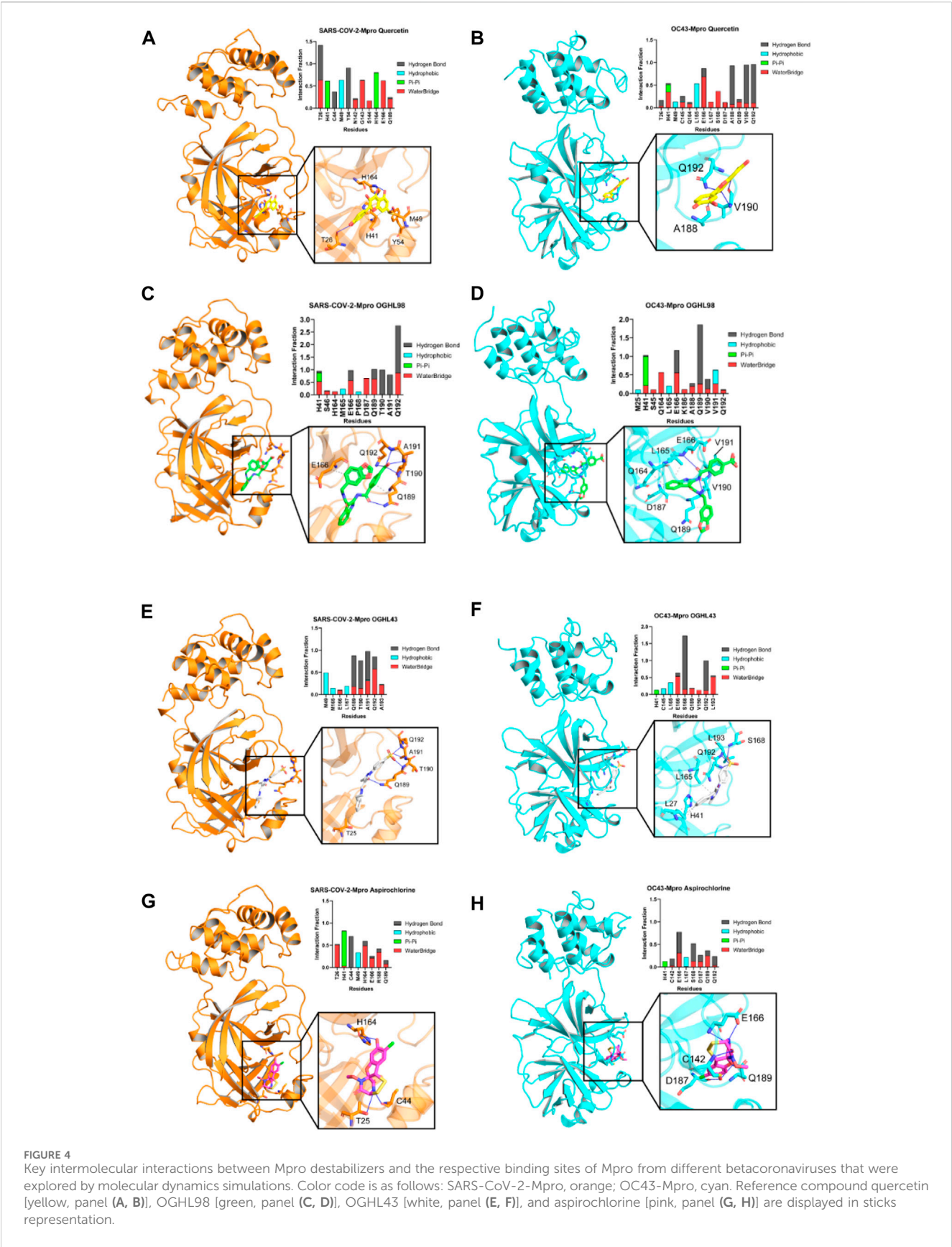


FIGURE 3

Time dependence of the RMSD for Mpro protein backbones (black) and ligand atoms (red) during the unrestrained molecular dynamics simulations. Reference compound quercetin [panel (A, B)], OGHL98 [panel (C, D)], OGHL43 [panel (E, F)], and aspirochlorine [panel (G, H)].

both protein backbones remained stable and reached equilibrium during the whole set of MD trajectories (Figure 3, black traces). In the RMSD profiles following quercetin's (Figures 3A, B, red traces) interaction with SARS-CoV-2-Mpro and HCoV-OC43-Mpro, quercetin was observed to interact stably with the binding site during the whole simulation. In the RMSD profiles between the other control compound, aspirochlorine, and SARS-CoV-2-Mpro (Figure 3G), aspirochlorine rotated  $\sim 180^\circ$  at  $\sim 100$  ns and then moved toward a contiguous region of the binding site that was richer in beta-sheet folding. This rearrangement in aspirochlorine

translated into RMSD changes of  $\sim 6$  Å within the protein, after which aspirochlorine adopted a new conformation that remained stable until the end of the 3  $\mu$ s simulation. When it comes to the aspirochlorine/HCoV-OC43-Mpro trajectory (Figure 3H), the chlorine group of this small molecule altered its initial position when exposed to the solvent, subsequently moved around the perimeter of the binding site of HCoV-OC43-Mpro for the first 1.8  $\mu$ s of the simulation, and then returned to its initial position, but this time with the chlorine group pointing toward HCoV-OC43-Mpro, which was a conformation that remained stable until the end



**FIGURE 4** Key intermolecular interactions between Mpro destabilizers and the respective binding sites of Mpro from different betacoronaviruses that were explored by molecular dynamics simulations. Color code is as follows: SARS-CoV-2-Mpro, orange; OC43-Mpro, cyan. Reference compound quercetin [yellow, panel (A, B)], OGH98 [green, panel (C, D)], OGH43 [white, panel (E, F)], and aspirochlorine [pink, panel (G, H)] are displayed in sticks representation.



of the simulation. As described in the RMSD profiling of aspirochlorine, OGHL43 also rotated  $\sim 180^\circ$  when it interacted with SARS-CoV-2-Mpro at early time points in the trajectory (Figure 3E) and then remained in the binding site of SARS-CoV-2-Mpro during the rest of the simulation, which meant that OGHL43 adopted a binding mode consistent with the one observed in the RMSD interaction profile of the same OGHL43 compound with HCoV-OC43-Mpro (Figure 3F), and it also matched with the corresponding RMSD profiles of OGHL98 (Figures 3C, D). Therefore, from the RMSD profiles of the newly studied ligands, it can be inferred that both OGHL43 and OGHL98 establish stable contacts with both Mpro targets and have common features with the RMSD profiles of the control compounds.

To better understand the intermolecular forces that drive the interactions of each ligand with the binding site of SARS-CoV-2-Mpro, we built interaction profiles where we plotted the corresponding fraction of the trajectory during which each interaction remained stable (Figure 4). For the main reference compound, quercetin, the interaction pattern observed was comparable to that of previous reports that described quercetin as prone to interact with N-terminal  $\beta 3/\alpha A$  residues of SARS-CoV-2-Mpro, such as M49 (interaction fractions  $>0.5$ , Figure 4A). Our analysis also concurred with the literature in the identification of frequent contacts between quercetin and  $\beta 10$ -11 residues of the protein (N142-E166). In addition, the novel compounds OGHL98 (Figure 4C) and OGHL43 (Figure 4E) shared frequent intermolecular hydrogen bonds and water bridges established with two sets of residues, M165-L167 ( $\beta 11$ -12) and Q189-A193, located at the C-terminal of SARS-CoV-2-Mpro. In the case of OGHL98, such intermolecular interactions were mostly driven by the 4-(methylcarbamoyl) benzoic acid moiety, which participated in strong water bridges and hydrogen bonds with key residues Q189-Q192 and with neighboring residues from the linker T190-A191 (Figure 4C) (Abian et al., 2020; Cho et al., 2021).

To further assess whether the molecular determinants detected in the MD simulations with our set of inhibitors and SARS-CoV-2-Mpro could justify the enzymatic inhibition that was also observed *in vitro* for HCoV-OC43-Mpro, we built similar binding profiles for this target (45.8% sequence identity), which is considered to be relatively conserved and for which a relatively homologous binding site had been predicted (Berry, Fielding, and Gamielien, 2015). Notably, intermolecular contacts established between the benzoic moiety of the main inhibitor OGHL98 and the key residues were highly preserved across the whole set of trajectories for the two proteins (Figures 4C, D).

In addition, the rest of the evaluated compounds were found to be prone to establishing favorable hydrogen bonds with a comparable set of C-terminal residues from both HCoV-OC43-Mpro and SARS-CoV-2-Mpro, as shown in the rest of the panels of Figure 4, with reference compound aspirochlorine included (Figures 4G, H). This is important because common intermolecular interaction patterns would justify a common target destabilization mechanism when bound to homologous viral proteases, enabling a match between wet and dry lab results and paving the way for further rational design efforts on OGHL98.

### 3.4 Predictive ADME profiling of OGHL98 using physicochemical descriptors

The SwissADME server was used to predict the ADME, pharmacokinetic properties, and drug-like properties of the best antiviral inhibitor OGHL98, which was compared with the internal reference compounds aspirochlorine and quercetin (Daina, Michielin and Zoete, 2017b) plus the gold-standard inhibitors (Antonopoulou et al., 2022; Sahoo et al., 2023) like covalent inhibitor ebselen or specific inhibitors of Mpro from SARS-CoV-2 like ML-188 (Tables 1, 2).

Importantly, our best inhibitor OGHL98 had a molecular weight (MW)  $< 500$  g/mol, which was within the optimal range for a potential lead drug. All the remaining physicochemical descriptors, like the number of rotatable bonds, hydrogen bond acceptors (HB-A), donors (HB-D), topological polar surface area (TPSA), lipophilicity index, and water solubility, were also in the corresponding acceptable ranges (Table 1, left). A similar trend was observed for the pharmacokinetic properties (Table 1, right). The absence of drug-likeness violations in the complementary analysis, which is summarized in Table 2, further confirmed the effectiveness of our drug-discovery workflow in identifying feasible and developable small-molecule inhibitors.

## 4 Discussion

During the SARS-CoV-2 pandemic, the testing of a small subset of classic natural products led to the identification of quercetin as a SARS-CoV-2-Mpro inhibitor, whose mechanism of action relies on the non-covalent binding and destabilization of this macromolecule, and it was used as a starting point to design quercetin derivatives having a scope as antivirals (Abian et al., 2020; Mangiavacchi et al., 2021). Unfortunately, quercetin has shown a marginal *in vivo* effect as an antiviral agent (Di Petrillo et al., 2022), and its mild therapeutic benefits are hard to correlate with Mpro inhibition. This fact, together with the potential spread of new variants of interest and concern, which are likely to be more vaccine-resistant and less sensitive to existing Mpro inhibitors in the clinic (Ip et al., 2023; Sahoo et al., 2023), raised the question of how feasible it would be to implement cell-free HTS setups to explore alternative chemical spaces like synthetic small molecules and quickly detect specific destabilizers of the Mpro target with better antiviral profiles than quercetin. Ideally, the characterization of these new destabilizers will provide medicinal chemists with new chemical scaffolds so that they can be optimized into broad-spectrum antivirals, which are considered ideal preparedness tools against future pandemics.

A good starting point to meet these needs is the successful implementation of two orthogonal thermal-shift and FRET assays in the HTS format, which have led to the discovery of a novel small molecule called OGHL98 by screening the Open Global Health Library (Merck KGaA). To date, OGHL98 has been described as a mere PDE5 inhibitor with the potential to treat erectile dysfunction and pulmonary arterial hypertension (Ahmed, Geethakumari, and Biswas, 2021). In this work, we have characterized a new activity for OGHL98 by confirming its potential as a developable antiviral against the infection of HCoV-OC43, which is a biosafety level-2 coronavirus. The mechanism of action of OGHL98 involves the



TABLE 1 Physicochemical and pharmacokinetic descriptors of OGHL98 and quercetin calculated with SwissADME.

Compound	Physicochemical properties					Lipophilicity	Water solubility	Pharmacokinetics		
	MW <sup>a</sup>	RB <sup>b</sup>	HB-A <sup>c</sup>	HB-D <sup>d</sup>	TPSA <sup>e</sup>	Consensus log Po/w <sup>f</sup>	Solubility (mol/L)	GI abs <sup>g</sup>	BBB <sup>h</sup>	log Kp (cm/s) <sup>i</sup>
OGHL98	457.43	8	6	4	129.75	3.26	8.92E-08	High	No	−6.34
Quercetin	302.24	1	7	5	131.36	1.23	6.98E-04	High	No	−7.05
Ebselen (IC <sub>50</sub> = 0.67 uM)	274.18	1	1	0	22.00	1.75	1-20E-04	High	Yes	−6.00
Aspirochlorine	360.79	1	5	2	138.7	0.78	1.42e-03	High	No	−7.78
Perampanel (IC <sub>50</sub> = 100–250 uM)	349.38	3	3	0	58.68	3.71	2.33e-05	High	Yes	−5.99
F01 (IC <sub>50</sub> = 54 uM)	286.71	3	3	1	59.06	2.34	6.51e-04	High	Yes	−6.56
ML188 (IC <sub>50</sub> = 2.5 uM)	431.51	9	4	1	108.36	3.27	4.38e-05	High	No	−6.72
ML300 (IC <sub>50</sub> = 4.99 uM)	433.54	9	4	1	75.44	4.06	3.48e-06	High	No	−5.42

<sup>a</sup>Molecular weight (g/mol).  
<sup>b</sup>Number of rotatable bonds.  
<sup>c</sup>Number of hydrogen bond acceptors.  
<sup>d</sup>Number of hydrogen bond donors.  
<sup>e</sup>Topological polar surface area (Ertl, Rohde, and Selzer, 2000).  
<sup>f</sup>Average of iLOGP, XLOGP, WLOGP, MLOGP, and SILICOS-IT predictions (Daina, Michielin, and Zoete, 2017b).  
<sup>g</sup>Gastrointestinal absorption.  
<sup>h</sup>Blood–brain barrier permeation.  
<sup>i</sup>Skin permeation: QSPR model (Potts and Guy, 1992).

TABLE 2 Drug-likeness properties of OGHL98 and quercetin calculated using SwissADME.

Compound	Lipinski # violations <sup>a</sup>	Ghose # violations <sup>b</sup>	Veber # violations <sup>c</sup>	Egan # violations <sup>d</sup>	Muegge # violations <sup>e</sup>
OGHL98	0	0	0	0	0
Quercetin	0	0	0	0	0
Ebselen (IC <sub>50</sub> = 0.67 uM)	0	0	0	0	0
Aspirochlorine	0	0	0	1	0
Perampanel (IC <sub>50</sub> = 100–250 uM)	0	0	0	0	0
F01 (IC <sub>50</sub> = 54 uM)	0	0	0	0	0
ML188 (IC <sub>50</sub> = 2.5 uM)	0	0	0	0	0
ML300 (IC <sub>50</sub> = 4.99 uM)	0	0	0	0	0

<sup>a</sup>Lipinski (Pfizer) filter (Lipinski et al., 2001): MW ≤ 500; MLOGP ≤4.15; N or O ≤ 10; NH or OH ≤ 5.  
<sup>b</sup>Ghose filter (Ghose, Viswanadhan and Wendoloski, 1999): 160 ≤ MW ≤ 480; −0.4 ≤ WLOGP ≤5.6; 40 ≤ MR ≤ 130; 20 ≤ atoms ≤70.  
<sup>c</sup>Veber (GSK) filter (Veber et al., 2002): rotatable bonds ≤10; TPSA ≤140.  
<sup>d</sup>Egan (Pharmacia) filter (Veber et al., 2002): WLOGP ≤5.88; TPSA ≤131.6.  
<sup>e</sup>Muegge (Bayer) filter (Muegge, Heald, and Brittelli, 2001): 200 ≤ MW ≤ 600; −2 ≤ XLOGP ≤5; TPSA ≤150; number of rings ≤7; number of carbon atoms >4; number of heteroatoms >1; number of rotatable bonds ≤15.

enzymatic inhibition of two conserved viral proteases, SARS-CoV-2-Mpro and HCoV-OC43-Mpro, as demonstrated by the cell-free setups. Therefore, this fast-tracking methodology can lead to the development of novel and efficient broad-spectrum antivirals.

It is important to note that the design of OGHL98 analogs that increase EC<sub>50</sub>/CC<sub>50</sub> selectivity is required to meet general pre-clinical standards (Sun et al., 2022). Furthermore, medicinal chemistry efforts will also be required to increase the potency that OGHL98 has shown *in vitro*. In this line, the long-term MD simulations suggest *in silico* optimization of the hydrogen bonds between the nitrogen atoms of OGHL98 (amide groups and indole moiety) and the backbone of the E166 residue in the wild-type Mpro. These contacts might be further improved to efficiently block VoC mutations within this region of Mpro, which is responsible for the resistance against nirmatrelvir or ensitrelvir (Ip et al., 2023). It is important to highlight that more than 50% of FDA-approved drugs are nitrogen-containing molecules, the majority of which are N-heterocyclic small molecules (Kerru et al., 2020), like

OGHL98, which supports the proposed strategy. Alternatively, it might be possible to design OGHL98 analogs with strengthened intermolecular interactions between the carboxylic acid group of the parent compound and the Q189 residue in wild-type Mpro. This second strategy could also be extended to another set of interesting Mpro VoC mutants, such as Q189K, ΔQ189, Q189H, Q189L, Q189P, Q189G, and Q189S (Najjar-Debbiny et al., 2023).

Finally, it is important to note that OGHL98 reported here fulfills all the drug-likeness properties (similar to control compounds) to be a potential drug (Tables 1, 2) according to the predicted physicochemical descriptors and pharmacokinetic properties (Daina, Michielin, and Zoete, 2017a). The computational pipeline and the experimental design advocate for the oral use of OGHL98 analogs of optimized potency, selectivity, and ADME/Tox profile.

## 5 Conclusion

We have demonstrated that the antiviral activity of OGHL98 exceeds the biological performance of the previously reported Mpro destabilizer, quercetin (Abian et al., 2020). This work reveals that OGHL98 acts as an inhibitor of two (SARS-CoV-2 and OC43) coronaviral proteases, which supports the theory that broad-spectrum inhibition against betacoronaviral proteases could be achieved using rational design approaches.

More importantly, the cell-free primary assays combined here to find hits define a cost-effective early drug-discovery workflow amenable for the screening of massive small-molecule libraries to feed the computational rational design campaigns and further validate more potent pharmacophores with double-digit selectivity indexes.

Such a future line of research is of special interest if implemented against hyperactive Mpro variants resistant to current inhibitors or other viral proteases from viruses with pandemic potential, for which no specific inhibitors have been identified so far.

## 6 Scope statement

The emergence of new variants of concern of coronaviruses, potentially more vaccine-resistant and less sensitive to existing treatments, is evident due to their high prevalence. A prospective spread of such VoCs demands a preparedness that can be met by fast-tracking workflows aiming at viral protein targets with a clear *in vitro/in vivo* phenotype. Mpro (or 3CLpro) is directly involved in the viral replication cycle and the production and function of viral polyproteins. These roles are conserved among betacoronaviruses like HCoV-OC43 and SARS-CoV-2, which makes the identification of new inhibitors for them a good starting point for designing broad-spectrum antivirals. We report an optimized methodology based on orthogonal cell-free assays to identify small molecules that inhibit the binding pockets of both SARS-CoV-2-Mpro and HCoV-OC43-Mpro, whose blockade correlates with antiviral activities in HCoV-OC43 cellular models. By using such a fast-tracking approach against the Open Global Health Library (Merck KGaA), we have found evidence of new antiviral activity for compound OGHL98. *In silico* molecular dynamics dissecting intermolecular interactions

between OGHL98 and both proteases concluded that the binding mode was primarily governed by conserved H-bonds with their C-terminal amino acids and that the rational design of OGHL98 has potential against VoC proteases.

## 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 authors.

## Ethics statement

Ethical approval was not required for the studies on humans and animals in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

## Author contributions

FC: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing—original draft, and writing—review and editing. DR: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, resources, software, validation, visualization, writing—original draft, and writing—review and editing. MR: data curation, formal analysis, investigation, methodology, validation, and writing—original draft. BM-A: formal analysis, investigation, and writing—original draft. ED-C: investigation, methodology, validation, and writing—original draft. TM: investigation, validation, and writing—original draft. CP-V: data curation, software, visualization, and writing—original draft. SL: methodology, resources, and writing—review and editing. FM: methodology and writing—original draft. FA: validation and writing—original draft. FV: supervision and writing—review and editing. OG: supervision and writing—review and editing. DG-P: funding acquisition, supervision, and writing—review and editing. RF-G: project administration, resources, supervision, writing—original draft, and writing—review and editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. DG-P's laboratory and research have been funded by the European Commission—Next Generation EU (regulation EU 2020/2094) through CSIC's Global Health Platform (PTI Salud Global), the Instituto de Salud Carlos III Subdirección General de Redes y Centros de Investigación Cooperativa-Red de Investigación Cooperativa en Enfermedades Tropicales (RICET: RD16/0027/0014), and the MCIN/AEI/10.13039/501100011033, and by the grants PID2019-109623RB-I00, MCIN/AEI/10.13039/501100011033, and FEDER Una manera de hacer Europa (2016-79957-R), and PID2022-142971OB-I00, MCIN/AEI/10.13039/

501100011033/FEDER, UE, and by the Junta de Andalucía (BIO-199). DR's research has been funded by the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT), grant number 1220656, and the Agencia Nacional de Investigación y Desarrollo (ANID), grant number COVID0199.

## Acknowledgments

The authors thank Professor Yang's laboratory for generously donating their pGEX-6p-1 plasmid to produce recombinant Mpro. They also thank Susan Weiss' laboratory for generously sharing their virus stock preparation protocol.

## 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.

## References

- Abian, O., Ortega-Alarcon, D., Jimenez-Alesanco, A., Ceballos-Laita, L., Vega, S., Reyburn, H. T., et al. (2020). Structural stability of SARS-CoV-2 3CLpro and identification of quercetin as an inhibitor by experimental screening. *Int. J. Biol. Macromol.* 164, 1693–1703. doi:10.1016/j.ijbiomac.2020.07.235
- Abraham, M., Gagaring, K., Martino, M. L., Vanaershot, M., Plouffe, D. M., Calla, J., et al. (2020). Probing the open global Health chemical diversity library for multistage-active starting points for next-generation antimalarials. *ACS Infect. Dis.* 6 (4), 613–628. doi:10.1021/acinfecdis.9b00482
- Antonopoulou, I., Sapountzaki, E., Rova, U., and Christakopoulos, P. (2022). Inhibition of the main protease of SARS-CoV-2 (Mpro) by repurposing/designing drug-like substances and utilizing nature's toolbox of bioactive compounds. *Comput. Struct. Biotechnol. J.* 20, 1306–1344. doi:10.1016/j.csbj.2022.03.009
- Berry, M., Fielding, B., and Gamielien, J. (2015). Human coronavirus OC43 3CL protease and the potential of ML188 as a broad-spectrum lead compound: homology modelling and molecular dynamic studies. *BMC Struct. Biol.* 15 (1), 8. doi:10.1186/s12900-015-0035-3
- Bowers, K. J., et al. (2006). "Molecular dynamics---Scalable algorithms for molecular dynamics simulations on commodity clusters," in *Proceedings of the 2006 ACM/IEEE conference on supercomputing - SC '06* (New York, New York, USA: ACM Press), 84. doi:10.1145/1188455.1188544
- Cho, E., Rosa, M., Anjum, R., Mehmood, S., Soban, M., Mujtaba, M., et al. (2021). Dynamic profiling of  $\beta$ -coronavirus 3CL M<sup>pro</sup> protease ligand-binding sites. *J. Chem. Inf. Model.* 61 (6), 3058–3073. doi:10.1021/acs.jcim.1c00449
- Daina, A., Michielin, O., and Zoete, V. (2017a). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7 (1), 42717. doi:10.1038/srep42717
- Daina, A., Michielin, O., and Zoete, V. (2017b). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7 (1), 42717. doi:10.1038/srep42717
- Di Petrillo, A., Orrù, G., Fais, A., and Fantini, M. C. (2022). Quercetin and its derivatives as antiviral potentials: a comprehensive review. *Phytotherapy Res.* 36 (1), 266–278. doi:10.1002/ptr.7309
- 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 (20), 3714–3717. doi:10.1021/jm000942e
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision Glide: docking and scoring incorporating a model of hydrophobic enclosure for Protein–Ligand complexes. *J. Med. Chem.* 49 (21), 6177–6196. doi:10.1021/jm051256o
- Ghose, A. K., Viswanadhan, V. N., and Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* 1 (1), 55–68. doi:10.1021/cc9800071
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* 47 (7), 1750–1759. doi:10.1021/jm030644s
- Ip, J. D., Wing-Ho Chu, A., Chan, W. M., Cheuk-Ying Leung, R., Umer Abdullah, S. M., Sun, Y., et al. (2023). Global prevalence of SARS-CoV-2 3CL protease mutations associated with nirmatrelvir or ensitrelvir resistance. *eBioMedicine* 91, 104559. doi:10.1016/j.ebiom.2023.104559
- Jacobson, M. P., Friesner, R. A., Xiang, Z., and Honig, B. (2002). On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* 320 (3), 597–608. doi:10.1016/S0022-2836(02)00470-9
- Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., et al. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins Struct. Funct. Bioinforma.* 55 (2), 351–367. doi:10.1002/prot.10613
- 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 (7811), 289–293. doi:10.1038/s41586-020-2223-y
- Kerru, N., Gummidi, L., Maddila, S., Gangu, K. K., and Jonnalagadda, S. B. (2020). A review on recent advances in nitrogen-containing molecules and their biological applications. *Molecules* 25 (8), 1909. doi:10.3390/molecules25081909
- Lilly, E., Coussens, N. P., Sittampalam, G. S., Guha, R., Brimacombe, K., Grossman, A., et al. (2004). *Assay guidance manual*. Available at: <https://ncats.nih.gov/expertise/preclinical/agm>.
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. *Adv. Drug Deliv. Rev.* 46 (1–3), 3–26. The article was originally published in *Advanced Drug Delivery Reviews* 23 (1997) 3–25. 1'. doi:10.1016/S0169-409X(00)00129-0
- Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Computer-Aided Mol. Des.* 27 (3), 221–234. doi:10.1007/s10822-013-9644-8
- Mangiavacchi, F., Botwina, P., Menichetti, E., Bagnoli, L., Rosati, O., Marini, F., et al. (2021). Seleno-functionalization of quercetin improves the non-covalent inhibition of Mpro and its antiviral activity in cells against SARS-CoV-2. *Int. J. Mol. Sci.* 22 (13), 7048. doi:10.3390/ijms22137048
- Martinez-Arribas, B., Annang, F., Díaz-González, R., Pérez-Moreno, G., Martín, J., Mackenzie, T. A., et al. (2023). Establishment of a screening platform based on human coronavirus OC43 for the identification of microbial natural products with antiviral activity. *Microbiol. Spectr.* 12, e0167923. doi:10.1128/spectrum.01679-23
- Min, J. S., Kim, D. E., Jin, Y. H., and Kwon, S. (2020). Kurarinone inhibits HCoV-OC43 infection by impairing the virus-induced autophagic flux in MRC-5 human lung cells. *J. Clin. Med.* 9 (7), 2230. doi:10.3390/jcm9072230
- Muegge, I., Heald, S. L., and Brittelli, D. (2001). Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* 44 (12), 1841–1846. doi:10.1021/jm015507e
- Najjar-Debbiny, R., Gronich, N., Weber, G., Khoury, J., Amar, M., Stein, N., et al. (2023). Effectiveness of Paxlovid in reducing severe coronavirus disease 2019 and

mortality in high-risk patients. *Clin. Infect. Dis.* 76 (3), e342–e349. doi:10.1093/cid/ciac443

Owen, D. R., Allerton, C. M. N., Anderson, A. S., Aschenbrenner, L., Avery, M., Berritt, S., et al. (2021). An oral SARS-CoV-2 M<sup>pro</sup> inhibitor clinical candidate for the treatment of COVID-19. *Science* 374 (6575), 1586–1593. doi:10.1126/science.abl4784

Pang, X., Xu, W., Liu, Y., Li, H., and Chen, L. (2023). The research progress of SARS-CoV-2 main protease inhibitors from 2020 to 2022. *Eur. J. Med. Chem.* 257, 115491. doi:10.1016/j.ejmech.2023.115491

Potts, R. O., and Guy, R. H. (1992). Predicting skin permeability. *Pharm. Res.* 9 (5), 663–669. doi:10.1023/A:1015810312465

Qiao, Z., Wei, N., Jin, L., Zhang, H., Luo, J., Zhang, Y., et al. (2021). The Mpro structure-based modifications of ebelen derivatives for improved antiviral activity against SARS-CoV-2 virus. *Bioorg. Chem.* 117, 105455. doi:10.1016/j.bioorg.2021.105455

Rojas-Prats, E., Martinez-Gonzalez, L., Gonzalo-Consuegra, C., Liachko, N. F., Perez, C., Ramirez, D., et al. (2021). Targeting nuclear protein TDP-43 by cell division cycle kinase 7 inhibitors: a new therapeutic approach for amyotrophic lateral sclerosis. *Eur. J. Med. Chem.* 210, 112968. doi:10.1016/j.ejmech.2020.112968

Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J. M., Lu, C., et al. (2019). OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* 15 (3), 1863–1874. doi:10.1021/acs.jctc.8b01026

Sahoo, P., Lenka, D. R., Batabyal, M., Pain, P. K., Kumar, S., Manna, D., et al. (2023). Detailed insights into the inhibitory mechanism of new ebelen derivatives against main

protease (M<sup>pro</sup>) of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). *ACS Pharmacol. Transl. Sci.* 6 (1), 171–180. doi:10.1021/acspsci.2c00203

Singh, P., Sharma, A., and Nandi, S. P. (2020). 'Identification of potent inhibitors of COVID-19 main protease enzyme by molecular docking study'. Available at: <https://chemrxiv.org/engage/chemrxiv/article-details/60c74a4e4c89195569ad31e4>.

Smee, D. F., Hurst, B. L., Evans, W. J., Clyde, N., Wright, S., Peterson, C., et al. (2017). Evaluation of cell viability dyes in antiviral assays with RNA viruses that exhibit different cytopathogenic properties. *J. Virological Methods* 246, 51–57. doi:10.1016/j.jviromet.2017.03.012

Sun, D., Gao, W., Hu, H., and Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B* 12 (7), 3049–3062. doi:10.1016/j.apsb.2022.02.002

Tan, C. Y., Chiew, C. J., Lee, V. J., Ong, B., Lye, D. C., and Tan, K. B. (2022). Comparative effectiveness of 3 or 4 doses of mRNA and inactivated whole-virus vaccines against COVID-19 infection, hospitalization and severe outcomes among elderly in Singapore. *Lancet Regional Health - West. Pac.* 29, 100654. doi:10.1016/j.lanwpc.2022.100654

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45 (12), 2615–2623. doi:10.1021/jm020017n

Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., et al. (2007). Production of authentic SARS-CoV Mpro with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.* 366 (3), 965–975. doi:10.1016/j.jmb.2006.11.073

# Frontiers in Pharmacology

Explores the interactions between chemicals and living beings

The most cited journal in its field, which advances access to pharmacological discoveries to prevent and treat human disease.

## Discover the latest Research Topics

[See more →](#)

### Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne, Switzerland  
[frontiersin.org](https://frontiersin.org)

### Contact us

+41 (0)21 510 17 00  
[frontiersin.org/about/contact](https://frontiersin.org/about/contact)



### Frontiers in Pharmacology

