

# COMPUTATIONAL APPROACHES IN DRUG DISCOVERY AND PRECISION MEDICINE

EDITED BY: Zunnan Huang, Xiao Jun Yao and Ruo-Xu Gu  
PUBLISHED IN: *Frontiers in Chemistry*





# frontiers

## 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 978-2-88966-601-0

DOI 10.3389/978-2-88966-601-0

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

# COMPUTATIONAL APPROACHES IN DRUG DISCOVERY AND PRECISION MEDICINE

Topic Editors:

**Zunnan Huang**, Guangdong Medical University, China

**Xiao Jun Yao**, Macau University of Science and Technology, China

**Ruo-Xu Gu**, Shanghai Jiao Tong University, China

**Citation:** Huang, Z., Yao, X. J., Gu, R.-X., eds. (2021). Computational Approaches in Drug Discovery and Precision Medicine. Lausanne: Frontiers Media SA.  
doi: 10.3389/978-2-88966-601-0

# Table of Contents

- 04 Editorial: Computational Approaches in Drug Discovery and Precision Medicine**  
Zunnan Huang, Xiao Jun Yao and Ruo-Xu Gu
- 06 Uncovering the Resistance Mechanism of Mycobacterium tuberculosis to Rifampicin Due to RNA Polymerase H451D/Y/R Mutations From Computational Perspective**  
Qianqian Zhang, Xiaoli An, Hongli Liu, Shuo Wang, Tong Xiao and Huanxiang Liu
- 19 Conformation Transition of Intracellular Part of Glucagon Receptor in Complex With Agonist Glucagon by Conventional and Accelerated Molecular Dynamics Simulations**  
Qifeng Bai, Shuoyan Tan, Horacio Pérez-Sánchez, Haixia Feng, Liya Feng, HuanXiang Liu and Xiaojun Yao
- 28 Identification of Drug-Disease Associations Using Information of Molecular Structures and Clinical Symptoms via Deep Convolutional Neural Network**  
Zhanchao Li, Qixing Huang, Xingyu Chen, Yang Wang, Jinlong Li, Yun Xie, Zong Dai and Xiaoyong Zou
- 42 SPVec: A Word2vec-Inspired Feature Representation Method for Drug-Target Interaction Prediction**  
Yu-Fang Zhang, Xiangeng Wang, Aman Chandra Kaushik, Yanyi Chu, Xiaoqi Shan, Ming-Zhu Zhao, Qin Xu and Dong-Qing Wei
- 53 The Use of Computational Approaches in the Discovery and Mechanism Study of Opioid Analgesics**  
Bangyi Zhao, Wei Li, Lijie Sun and Wei Fu
- 63 Molecular Determinants Elucidate the Selectivity in Abscissic Acid Receptor and HAB1 Protein Interactions**  
Jing-Fang Yang, Chun-Yan Yin, Di Wang, Chen-Yang Jia, Ge-Fei Hao and Guang-Fu Yang
- 72 Diterpenoids and Triterpenoids From Frankincense are Excellent Anti-psoriatic Agents: An in silico Approach**  
Sobia Ahsan Halim, Ajmal Khan, Rene Csuk, Ahmed Al-Rawahi and Ahmed Al-Harrasi
- 91 Computational Approaches in Preclinical Studies on Drug Discovery and Development**  
Fengxu Wu, Yuquan Zhou, Langhui Li, Xianhuan Shen, Ganying Chen, Xiaoqing Wang, Xianyang Liang, Mengyuan Tan and Zunnan Huang
- 123 Virtual Screening and Bioactivity Evaluation of Novel Androgen Receptor Antagonists From Anti-PCa Traditional Chinese Medicine Prescriptions**  
Wenya Han, Yuqi Shi, Jie Su, Zhennan Zhao, Xin Wang, Jiazhong Li and Huanxiang Liu





# Editorial: Computational Approaches in Drug Discovery and Precision Medicine

Zunnan Huang<sup>1\*</sup>, Xiao Jun Yao<sup>2\*</sup> and Ruo-Xu Gu<sup>3\*</sup>

<sup>1</sup>Key Laboratory of Big Data Mining and Precision Drug Design of Guangdong Medical University, Guangdong Medical University, Dongguan, China, <sup>2</sup>Dr. Neher's Biophysics Laboratory for Innovative Drug Discovery, Macau University of Science and Technology, Macau, China, <sup>3</sup>Department of Theoretical and Computational Biophysics, Max-Planck Institute for Biophysical Chemistry, Göttingen, Germany

**Keywords:** drug discovery, precision medicine, molecular modeling, data mining, artificial intelligence

## Editorial on the Research Topic

### Computational Approaches in Drug Discovery and Precision Medicine

During the past decades, computational approaches have been highly involved in all stages of drug discovery (Terstappen and Reggiani, 2001; Duarte et al., 2019) and precision medicine (Barbolosi et al., 2016; Delavan et al., 2018), from screening of leading compounds to preclinical trials. These methods, including both structure-based molecular modeling techniques (Kalyaanamoorthy and Chen, 2011) and artificial intelligence (Fleming, 2018; Williams et al., 2018; Chan et al., 2019), accelerate the discovery of drug candidates, guide the repurposing of existing drugs, improve our understanding of biomolecular nanomachines, and reduce the use of experimental animals. To present state-of-the-art computational studies in this field, we launched a research topic in *Frontiers in Chemistry* entitled “Computational Approaches in Drug Discovery and Precision Medicine.” This research topic included nine articles, including two reviews and seven original research articles, which covered theoretical predictions of ligand–protein interactions, drug resistance mechanism, and drug selectivity mechanism in protein–ligand binding, and the prediction of preclinical properties of ligands as well.

Two manuscripts reported case study of virtual screening of drug candidates. Han et al. constructed natural product database by analyzing the ingredients of traditional Chinese medicine prescriptions for treating prostate cancer. Molecular docking and wet experiments were then performed to screen possible ligands targeting androgen receptor, a protein involved in the pathogenesis of prostate cancer. Halim et al. docked active components of frankincense, macrocyclic diterpenoid derivatives, and boswellic acids to multiple proteins which are known drug targets of psoriasis in order to screen possible candidates for treating psoriasis. In addition, the review article by Zhao et al. summarized recent structure-based and ligand-based virtual screening of analgesics targeting opioid receptors, the computational guided studies of subtype selectivity of  $\sigma$  and  $\kappa$  opioid receptors, as well as the activation mechanism.

Molecular modeling has become an important complimentary to experiments in the study of ligand–protein interaction mechanisms. In this regard, Yang et al. investigated subtype selectivity mechanism of pyrabactin, an abscisic acid (ABA)-mimicking ligand, for the ABA receptors, using sequence and structural comparison and free energy calculations. Zhang et al. explored how the mutations of *Mycobacterium tuberculosis* RNA polymerase develop drug resistance to rifampicin.

During the past years, artificial intelligence has been increasingly involved in drug development. In this research topic, two manuscripts employed the application of machine learning to predict the

## OPEN ACCESS

### Edited and reviewed by:

Michael Kassiou,  
The University of Sydney, Australia

### \*Correspondence:

Zunnan Huang  
zn\_huang@gdmu.edu.cn  
Xiao Jun Yao  
xjyao@must.edu.mo  
Ruo-Xu Gu  
guruoxu85@gmail.com

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 09 December 2020

**Accepted:** 23 December 2020

**Published:** 12 February 2021

### Citation:

Huang Z, Yao XJ and Gu R-X (2021)  
Editorial: Computational Approaches  
in Drug Discovery and  
Precision Medicine.  
Front. Chem. 8:639449.  
doi: 10.3389/fchem.2020.639449

ligand–protein interactions. Zhang et al. proposed a method to represent the SMILES strings of ligands and protein sequences to lower-dimensional vectors, which revealed the hidden biophysical and biochemical patterns of ligands and proteins. Their methods are able to predict drug–target interactions in an efficient and reliable way, and will be helpful for repurposing of drugs. Li et al. developed a deep learning–based technique to predict the association between drugs and diseases, using the information of molecular structures and disease symptoms.

One reason for the failure of drug development is undesired pharmacokinetics and toxicity of the ligands. *In silico* evaluation of the ADMET properties of preclinical drugs is able to reduce the cost of drug development significantly. Wu et al. systematically reviewed the current methods, databases, and software developed for the *in vitro* prediction of ADMET properties. They also described related application of these methods, discussed the challenges and limitations of this field, and proposed suggestions.

Molecular dynamics simulation has been an effective tool to study the mechanism of the biological function of proteins. Bai et al. investigate the conformational transition of glucagon receptor (GCGR) in the apo, glucagon bound, and antagonist bound forms by conventional and accelerated molecular dynamics simulations. Their work provided an example of how different ligands modulate the conformational transition

of a protein between different states, and proved that accelerated molecular dynamics simulation is a reliable way to explore the conformational space of proteins.

To summarize, this research topic provides the readers an overview of how computational methods accelerate drug development and increase our understanding of ligand–protein interactions. We sincerely hope the articles in this research topic can draw people's attention and promote the researches in the field of computational studies in drug discovery and precision medicine.

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENTS

We thank all the authors, reviewers and editors who contributed to this Research Topic. ZH thanks the National Natural Science Foundation of China (Grant No. 31770774) for financial support.

## REFERENCES

- Barbolosi, D., Ciccolini, J., Lacarelle, B., Barlési, F., and André, N. (2016). Computational oncology—mathematical modelling of drug regimens for precision medicine. *Nat. Rev. Clin. Oncol.* 13, 242–254. doi:10.1038/nrclinonc.2015.204
- Chan, H. C. S., Shan, H., Dahoun, T., Vogel, H., and Yuan, S. (2019). Advancing drug discovery via artificial intelligence. *Trends Pharmacol. Sci.* 40, 592–604. doi:10.1016/j.tips.2019.06.004
- Delavan, B., Roberts, R., Huang, R., Bao, W., Tong, W., and Liu, Z. (2018). Computational drug repositioning for rare diseases in the era of precision medicine. *Drug Discov. Today* 23, 382394. doi:10.1016/j.drudis.2017.10.009
- Duarte, Y., Márquez-Miranda, V., Miossec, M. J., and González-Nilo, F. (2019). Integration of target discovery, drug discovery and drug delivery: a review on computational strategies. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* 11, e1554. doi:10.1002/wnan.1554
- Fleming, N. (2018). How artificial intelligence is changing drug discovery. *Nature* 557, S55. doi:10.1038/d41586-018-05267-x
- Kalyaanamoorthy, S., and Chen, Y. P. (2011). Structure-based drug design to augment hit discovery. *Drug Discov. Today* 16, 831–839. doi:10.1016/j.drudis.2011.07.006
- Terstappen, G. C., and Reggiani, A. (2001). In silico research in drug discovery. *Trends Pharmacol. Sci.* 22, 23–26. doi:10.1016/S0165-6147(00)01584-4
- Williams, A. M., Liu, Y., Regner, K. R., Jotterand, F., Liu, P., and Liang, M. (2018). Artificial intelligence, physiological genomics, and precision medicine. *Physiol. Genom.* 50, 237–243. doi:10.1152/PHYSIOLGENOMICS.00119.2017

**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.

Copyright © 2021 Huang, Yao and Gu. 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.



# Uncovering the Resistance Mechanism of *Mycobacterium tuberculosis* to Rifampicin Due to RNA Polymerase H451D/Y/R Mutations From Computational Perspective

Qianqian Zhang<sup>1</sup>, Xiaoli An<sup>2</sup>, Hongli Liu<sup>1</sup>, Shuo Wang<sup>1</sup>, Tong Xiao<sup>1</sup> and Huanxiang Liu<sup>1\*</sup>

<sup>1</sup> School of Pharmacy, Lanzhou University, Lanzhou, China, <sup>2</sup> State Key Laboratory of Applied Organic Chemistry and Department of Chemistry, Lanzhou University, Lanzhou, China

## OPEN ACCESS

### Edited by:

Zunnan Huang,  
Guangdong Medical University, China

### Reviewed by:

Bing He,  
University of Michigan, United States  
Guohui Li,  
Dalian Institute of Chemical Physics  
(CAS), China

### \*Correspondence:

Huanxiang Liu  
hxliu@lzu.edu.cn

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 10 July 2019

**Accepted:** 12 November 2019

**Published:** 03 December 2019

### Citation:

Zhang Q, An X, Liu H, Wang S, Xiao T  
and Liu H (2019) Uncovering the  
Resistance Mechanism of  
*Mycobacterium tuberculosis* to  
Rifampicin Due to RNA Polymerase  
H451D/Y/R Mutations From  
Computational Perspective.  
*Front. Chem.* 7:819.  
doi: 10.3389/fchem.2019.00819

Tuberculosis is still one of the top 10 causes of deaths worldwide, especially with the emergence of multidrug-resistant tuberculosis. Rifampicin, as the most effective first-line antituberculosis drug, also develops resistance due to the mutation on *Mycobacterium tuberculosis* (Mtb) RNA polymerase. Among these mutations, three mutations at position 451 (H451D, H451Y, H451R) are associated with high-level resistance to rifampicin. However, the resistance mechanism of Mtb to rifampicin is still unclear. In this work, to explore the resistance mechanism of Mtb to rifampicin due to H451D/Y/R mutations, we combined the molecular dynamics simulation, molecular mechanics generalized-Born surface area calculation, dynamic network analysis, and residue interactions network analysis to compare the interaction change of rifampicin with wild-type RNA polymerase and three mutants. The results of molecular mechanics generalized-Born surface area calculations indicate that the binding free energy of rifampicin with three mutants decreases. In addition, the dynamic network analysis and residue interaction network analysis show that when H451 was mutated, the interactions of residue 451 with its adjacent residues such as Q438, F439, M440, D441, and S447 disappeared or weakened, increasing the flexibility of binding pocket. At the same time, the disappearance of hydrogen bonds between R613 and rifampicin caused by the flipping of R613 is another important reason for the reduction of binding ability of rifampicin in H451D/Y mutants. In H451R mutant, the mutation causes the binding pocket change too much so that the position of rifampicin has a large movement in the binding pocket. In this study, the resistance mechanism of rifampicin at the atomic level is proposed. The proposed drug-resistance mechanism will provide the valuable guidance for the design of antituberculosis drugs.

**Keywords:** tuberculosis, rifampicin, drug resistance, molecular dynamics simulation, molecular mechanics generalized-Born surface area, dynamic network analysis

## INTRODUCTION

Tuberculosis (TB), an infectious disease caused by *Mycobacterium tuberculosis* (Mtb), is the leading cause from a single infectious agent worldwide. Mtb, a pathogenic bacterium species of the family *Mycobacteriaceae*, can attack the lung of people and spread in the population through the droplets from the throat of an active TB infection patient (Mishra and Suroliya, 2018). Millions of people continue to fall sick with TB each year. The 2018 WHO Global Tuberculosis Report (Organization, 2018) estimated 10.0 million new cases of TB and 1.6 million deaths in 2017.

Although it spreads widely, TB is preventable and curable. The 6-month short-course regimen with a combination of four anti-TB drugs (rifampicin, isoniazid, pyrazinamide, and ethambutol) has been used as a standard treatment for active, drug-susceptible TB patients over the past decades (Service and Council, 1981; Chang et al., 2018; Seid et al., 2018; Tiberi et al., 2018). Furthermore, the risk of relapse is generally below 5% reported among drug-susceptible TB patients after treated with standard 6-month regimens in clinical trials (Chang et al., 2006). However, the multidrug-resistant tuberculosis (MDR-TB), which at least resists to both rifampicin and isoniazid, had been emerged in the early 1990s due to multiple factors (He et al., 2008; Sandgren et al., 2009; Ahmad et al., 2018).

Rifampicin, one of the most effective anti-TB drugs, has been used as the first-line treatment in drug-susceptible TB patients, and it is also effective against initial isoniazid resistance (Mitchison and Nunn, 1986). Unfortunately, rifampicin resistance in Mtb arises due to the residues' mutations on its molecular target, *Mycobacterium tuberculosis* RNA polymerase (Mtb-RNAP). More than 95% of the rifampin-resistant strains have mutations in a small region defined "rifampicin resistance-determining region" in Mtb-RNAP (Morlock et al., 2000; Zaw et al., 2018). The most common mutation in rifampicin resistance-determining region are S456, H451, and D441, corresponding to S531, H526, and D516 in *Escherichia coli*, respectively. Studies have shown that 70% of rifampicin-resistant clinical isolates have point mutation in two residues (S456 and H451) (Morlock et al., 2000), and H451 is most usually substituted for Asp (D), Tyr (Y), and Arg (R) (Telenti et al., 1993; Caws et al., 2006; Ma et al., 2006; Chikaonda et al., 2017; Wu and Hilliker, 2017). As early as 1995, the *in vitro* activity experiment of rifampicin by Bodmer et al. (1995) had been demonstrated that H451D/Y/R mutations could cause high-level resistance to rifampicin. After more than 20 years, the level and frequency of resistance to rifampicin are also increasing.

In 2017, there is about 558,000 new cases of rifampicin-resistant tuberculosis (RR-TB), of which 82% are MDR-TB and about 230,000 deaths from MDR/RR-TB (Organization, 2018). Currently, although MDR/RR-TB can be cured with the second-line drugs (e.g., fluoroquinolone and an injectable aminoglycoside), poor efficiency, high toxicity, and expensive price of these drugs make it still difficult for many MDR-TB patients. In some cases, more severe extensively drug-resistant TB may occur, and it will not respond to the most effective second-line anti-TB drugs (Sotgiu et al., 2015; Jeon, 2017; Tiberi

et al., 2018). Obviously, the development of new anti-TB drugs is urgent, and exploring the resistance mechanism of rifampicin is of great significance for the discovery of effective drugs.

In this work, in order to uncover the resistance mechanism of Mtb to rifampicin due to the mutation of Mtb-RNAP at position 451, three independent molecular dynamics (MD) simulations for the wild-type Mtb-RNAP and H451D/Y/R mutants were carried out. Based on the obtained trajectories, the molecular mechanics generalized-Born surface area (MM-GBSA) method was applied to calculate the binding free energy of rifampicin with Mtb-RNAP. Furthermore, dynamic network analysis combined with residue interaction network (RIN) analysis was used to show the detailed changes of interactions among the residues surrounding the binding pocket. With the structural and energy analysis, a possible rifampicin-resistant mechanism was also proposed. Compared with the traditional experimental method, MD simulations can show the intuitive and dynamics interaction change process between rifampicin and Mtb-RNAP due to the point mutation. Together with the energy analysis and the dynamics network analysis, the present study show the essential reason of Mtb-RNAP resistant to rifampicin, which can provide the useful guidance for the further drug design against drug resistance.

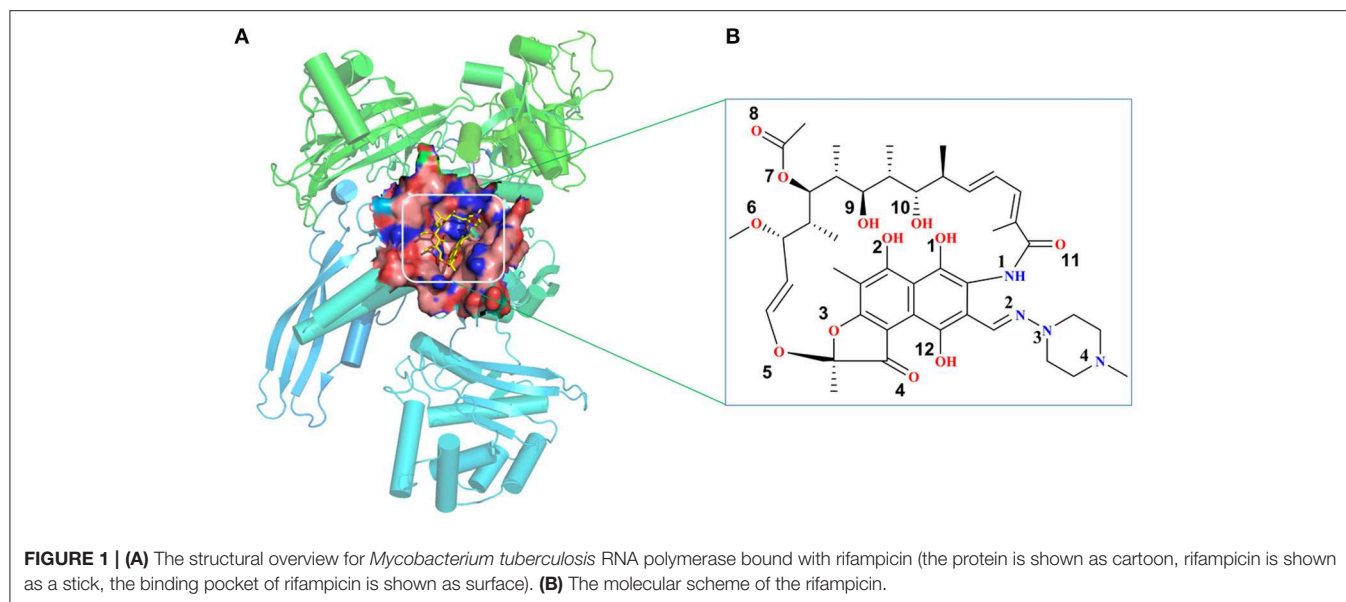
## MATERIALS AND METHODS

### Systems Preparation

The initial atomic coordinate of the wild-type Mtb-RNAP with rifampicin was obtained from Protein Data Bank (Protein Data Bank ID: 5UHB). The crystal structure of Mtb-RNAP reported by Lin et al. (2017) reveals that Mtb-RNAP is composed of six chains, for the A, B chains encoded by the *rpoA* gene, the C chain encoded by the *rpoB* gene (Miller et al., 1994), and the D, E, F chains encoded by *rpoC*, *rpoZ*, and *rpoD*, respectively. Rifampicin binds at the active site of the C chain (shown in **Figure 1**) and inhibits the DNA-directed RNA synthesis of Mtb (McClure and Cech, 1978; Campbell et al., 2001; Somoskovi et al., 2001). Considering that the speed to simulate the whole Mtb-RNAP (~3,826 residues) is too slow, only the C chain complexed with rifampicin was extracted and used as the initial structure of simulations. Furthermore, the deletion of other chains will make the residues of the interface between the two chains unstable, which is inconsistent with that in the multimer. Thus, to simulate the state of interface in the multimer, some relatively flexible and far from the active site amino acid residues were deleted. The three-dimensional structures of three mutants (H451D/Y/R) were obtained by mutating H451 residue in wild type.

To generate the force field parameters for the ligand, the Gaussian 09 program (Frisch et al., 2009) was used to optimize the structure of rifampicin and calculate the electrostatic potential at the Hartree-Fock level with 6-31G\* basis set. Then, the restraint electrostatic potential protocol (Bayly et al., 1993; Cieplak et al., 1995; Fox and Kollman, 1998) was employed to fit the atomic partial charges. The general amber force field (gaff) (Wang et al., 2004) generated by the antechamber program in the Amber14 package (Case et al., 2014) was applied to describe





the ligand. The standard ff99SB force field (Hornak et al., 2006) was used to describe the protein. Then, the LEaP module was applied to add all missing hydrogen atoms and a certain amount of sodium counter-ions to neutralize the unbalanced charges and maintain the systems electro-neutrality. Finally, a rectangular periodic water box of TIP3P (Jorgensen et al., 1983) was added to each system with the water molecules extended 10-Å distance around the complex. The size of the periodic boundary box is  $92.6 \times 114.4 \times 114.9$  Å. The whole system has a total of  $\sim 100,000$  atoms per periodic cell.

## Molecular Dynamics Simulations

All MD simulations were performed with Amber14 package (Case et al., 2014). The process of energy minimization, heating, and equilibration was carried out with the Particle Mesh Ewald Molecular Dynamics module. Initially, the energy minimization of each solvated complex includes three steps. For each step, energy minimization was carried out by the steepest descent method for the first 2,500 steps and conjugated gradient method for the subsequent 2,500 steps. In the first step, all the atoms of the complex were restrained with a force constant of  $2.0 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$  to only minimize the solvent and ion molecules. After that, the protein backbone atoms were fixed with a restraint force of  $2.0 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$  in the second step. Finally, all atoms in the system were minimized without any restraint. After energy minimization, all systems were heated up from 0 to 310 K in the canonical (NVT) ensemble over 100 ps by restraining the protein and the ligand with a  $2.0 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$  force constant and using a Langevin thermostat with a coupling coefficient of 2.0/ps. After heating, five steps MD pre-equilibration at 310 K were performed in the NPT ensemble by restraining all the atoms of the complex with decreasing restraints from 2.0 to 1.5, to 1.0, to 0.5, to 0.1. Then, 50-ns equilibration MD simulation without any restraints was performed to eliminate collisions between atoms. Finally, 500-ns production MD simulation was carried out without any

restraints on each system in the NPT ensemble at the temperature of 310.0 K and pressure of 1 atm. During the simulations, all the bonds involving hydrogen atoms were restrained with SHAKE algorithm (Ryckaert et al., 1977) to avoid too fast vibration of the hydrogen atoms. In addition, periodic boundary conditions were employed, and the long-range Coulombic interactions were treated using the Particle Mesh Ewald (Darden et al., 1993). The time step was set to 2 fs.

## Molecular Mechanics Generalized-Born Surface Area Calculation

To uncover the effects of mutations on the binding affinity of rifampicin to Mtb-RNAP, the MM-GBSA method was applied to estimate the binding free energy of each complex, which has been successfully used in a lot of researches (Pan et al., 2011; Yang et al., 2012). Here, we extracted 1,000 snapshots at 100-ps interval from the last 100-ns trajectory for each system. The binding free energy was calculated from the equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

where  $G_{\text{complex}}$ ,  $G_{\text{receptor}}$ , and  $G_{\text{ligand}}$  are the free energy of complex, protein, and ligand, respectively. The free energy for each molecular species was calculated based on an average over the extracted snapshots. Each of them can be estimate with the following equations:

$$G = E_{\text{MM}} + G_{\text{sol}} - TS \quad (2)$$

$$E_{\text{MM}} = E_{\text{int}} + E_{\text{ele}} + E_{\text{vdw}} \quad (3)$$

$$E_{\text{int}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} \quad (4)$$

$$G_{\text{sol}} = G_{\text{GB}} + G_{\text{SA}} \quad (5)$$

$$G_{\text{SA}} = \gamma^* \text{SASA} + \beta \quad (6)$$

where  $E_{\text{MM}}$  is the gas-phase energy calculated using the Amber ff03 molecular mechanics force field.  $E_{\text{int}}$  is the internal energy,

including the energy of bond ( $E_{bond}$ ), angle ( $E_{angle}$ ), and torsion ( $E_{torsion}$ ).  $E_{ele}$  and  $E_{vdw}$  are the Coulomb and van der Waals energy, respectively.  $G_{sol}$  is the solvation free energy and can be decomposed into polar solvation free energy ( $G_{GB}$ ) and non-polar solvation free energy ( $G_{SA}$ ).  $G_{GB}$  was calculated by solving the GB equation and the dielectric constants for solute as well as solvent were set to 1.0 and 80.0, respectively (Rocchia et al., 2001).  $G_{SA}$  was estimated by the solvent accessible surface area determined using a water probe radius of 1.4 Å. The surface tension constant  $\gamma$  was set to 0.0072 kcal/(mol·Å<sup>2</sup>), and the non-polar contribution to the solvation free energy term  $\beta$  was set to 0 (Sitkoff et al., 1994).  $T$  and  $S$  are the temperature and the total solute entropy, respectively. The entropy contributions can be estimated by normal mode analysis (Pearlman et al., 1995). However, here, we did not calculate the entropy contributions since our aim is not to obtain the absolute Gibbs energy but to identify the key residues of binding pocket and the detailed interaction features. In addition, previous studies have proven that it is sufficient to compare the binding ability of receptors and ligands based on the values of enthalpy changes ( $\Delta H_{bind}$ ) (Aruksakunwong et al., 2006; Xue et al., 2012).

Moreover, in order to identify the key residues responsible for the binding of rifampicin, the MM-GBSA binding free energy decomposition process was used to decompose the interaction energy to each residue by considering molecular mechanics and solvation energy without considering the contribution of entropy.

## Dynamic Network Analysis

Dynamic network analysis, as an effective method to extract information from the obtained molecular dynamics trajectories, has been successfully applied in protein misfolding (Zhou et al., 2019) and protein–protein interaction analysis (Sethi et al., 2009; Bai et al., 2014). Here, in order to observe the dynamic changes of the residues interaction network, 2,000 snapshots were extracted from the last 50-ns trajectory for each system. In the network, one node represents one residue, and the position of each node is defined at the center of C $\alpha$  atom of residue. The edge represents the interactions of two residues. Furthermore, the edge weight ( $W_{ij}$ ) between two nodes ( $i, j$ ) was defined with the following equation:

$$W_{ij} = -\log(|C_{ij}|) \quad (7)$$

where  $C_{ij}$  represents the pairwise correlations, which is calculated by Carma program (Glykos, 2006), a plugin in VMD (Humphrey et al., 1996). Finally, the NetworkView module in VMD was used to visualize the residue interaction network.

## Residue Interaction Network Analysis

RIN uses a network diagram to simplify the inter-residue interaction, which considers the residues as nodes and physico-chemical interactions such as covalent and non-covalent bonds as edges. RIN method has been successfully used to analyze the effects of mutations on drug resistance (Xue et al., 2012, 2013, 2014). In this work, the residue interaction network generator 2.0 (RING-2.0) (Piovesan et al., 2016) software

was applied to generate the network for the representative structures. The calculation process of RING-2.0 is described as follows: (i) the calculation of the secondary structure elements by incorporating the DSSP algorithm (Kabsch and Sander, 1983); (ii) hydrogen atom placement based on geometric criteria; (iii) hydrogen bond calculation; and (iv) the calculation of van der Waals interactions. Moreover, Cytoscape (Shannon et al., 2003) and the plugin RINalyzer (Doncheva et al., 2011) were used to visualize the residue interaction network.

## RESULTS

### H451D/Y/R Mutations Increased the Flexibility of the Active Pocket

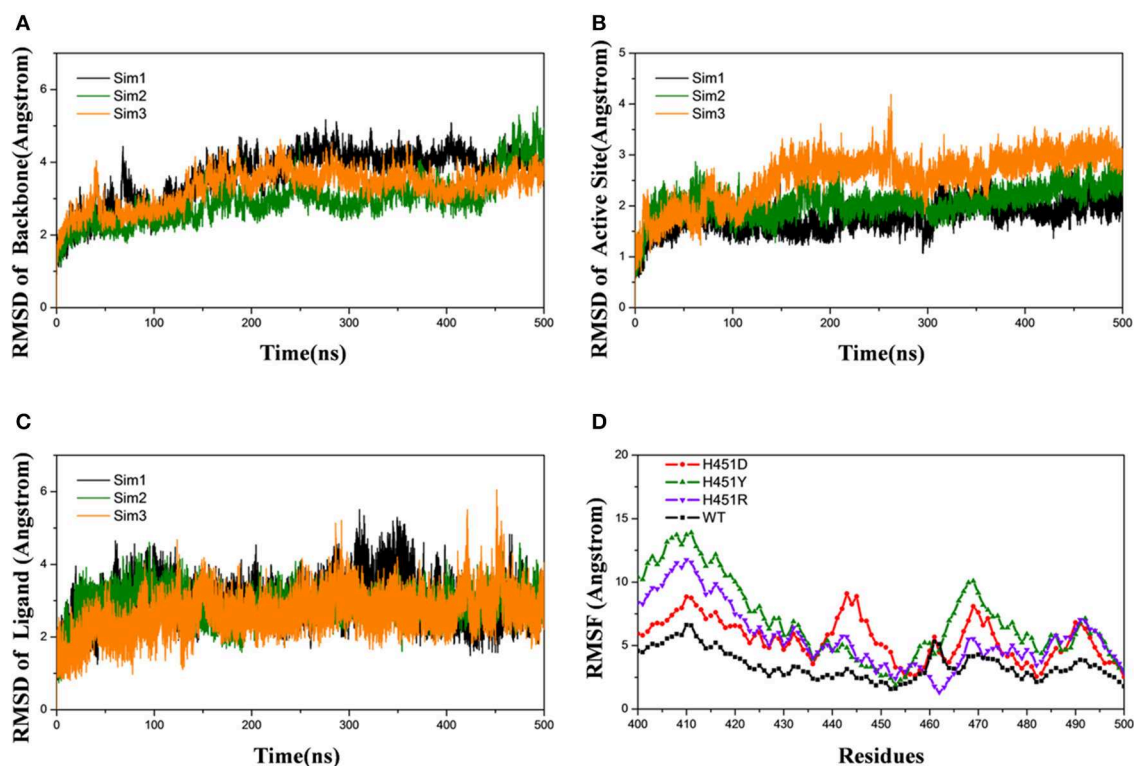
Firstly, the root-mean-square-deviations (RMSD) value for the protein backbone atoms, the active pocket, and the heavy atoms of rifampicin relative to the initial structure were calculated to monitor the equilibrium of each system. As shown in **Figure 2** and **Supplementary Figures 1–3**, three parallel MD simulations have similar fluctuations, suggesting each parallel trajectory can produce reproducible results. Thus, the following analysis was based on one of three parallel MD simulations. As can be seen from the RMSDs of wild-type Mtb-RNAP and three mutants, each system achieves equilibrium after 100 ns. Therefore, the last 100-ns trajectory for each system was used for the following structural and energetic analysis. Additionally, from the monitoring of the RMSD value of the heavy atoms of ligand, we can justify roughly if the ligand can bind to the target stably. From **Figure 2C** and **Supplementary Figures 1–3** the RMSDs of rifampicin in mutants were larger than that in wild type, indicating that rifampicin had a large fluctuation in mutants.

The root-mean-square-fluctuation (RMSF) of each residue was calculated based on the last 100-ns trajectory for each system, and the corresponding results were shown in **Figure 2D**. It can be seen from **Figure 2D** that the RMSF values of the residues have similar trends for all systems. However, H451D/Y/R mutations will increase the RMSF values relative to the wild-type Mtb-RNAP, which indicates that these three mutations increased the flexibility of the binding pocket and weakened the interaction of rifampicin with Mtb-RNAP.

### H451D/Y/R Mutations Weaken the Binding Ability of *Mycobacterium tuberculosis* RNA Polymerase With Rifampicin

To explore the effects of three mutations on the binding of Mtb-RNAP with rifampicin, the binding free energy calculation was performed based on MM-GBSA method. As shown in **Table 1**, the enthalpy changes ( $\Delta H_{bind}$ ) of the wild-type Mtb-RNAP and H451D/Y/R mutants with rifampicin are −43.89, −31.20, −35.55, and −28.58 kcal/mol, respectively. As expected, the binding affinity of Mtb-RNAP to rifampicin reduced obviously due to H451D/Y/R mutations.

By assessing the contributions of individual energy terms, we found that the non-polar interactions (sum of van der



**FIGURE 2 |** Root-mean-square-deviations (RMSDs) for the wild-type system of three independent molecular dynamics simulations: Sim1 (black), Sim2 (green), Sim3 (orange): **(A)** RMSDs for the backbone atoms of protein versus time. **(B)** RMSDs for the backbone atoms of active pocket vs. time. **(C)** RMSDs for the heavy atoms of rifampicin vs. time. **(D)** Root-mean-square-fluctuation for the backbone atoms of active pocket vs. residue number of wild type (black), H451D (red), H451Y (green), and H451R (purple) in Sim1.

**TABLE 1 |** The calculated binding free energy and the detailed contribution of different energy terms (kcal/mol).

Contribution	Wild type	H451D	H451Y	H451R
$\Delta E_{ele}$	$-64.92 \pm 0.17$	$-26.97 \pm 0.21$	$-47.82 \pm 0.27$	$-23.79 \pm 0.22$
$\Delta E_{vdw}$	$-52.49 \pm 0.12$	$-45.07 \pm 0.16$	$-50.01 \pm 0.16$	$-46.03 \pm 0.15$
$\Delta E_{MM}^a$	$-117.41 \pm 0.19$	$-72.04 \pm 0.26$	$-97.84 \pm 0.35$	$-69.83 \pm 0.27$
$\Delta G_{SA}$	$-6.62 \pm 0.01$	$-5.34 \pm 0.02$	$-6.23 \pm 0.01$	$-5.70 \pm 0.02$
$\Delta G_{GB}$	$80.14 \pm 0.14$	$46.18 \pm 0.17$	$68.52 \pm 0.23$	$46.94 \pm 0.19$
$\Delta G_{sol}$	$73.52 \pm 0.14$	$40.84 \pm 0.16$	$62.29 \pm 0.22$	$41.24 \pm 0.18$
$\Delta G_{polar}^b$	$15.22 \pm 0.11$	$19.21 \pm 0.14$	$20.7 \pm 0.18$	$23.15 \pm 0.14$
$\Delta G_{nonpolar}^c$	$-59.11 \pm 0.06$	$-50.41 \pm 0.08$	$-56.24 \pm 0.08$	$-51.73 \pm 0.08$
$\Delta H_{bind}$	$-43.89 \pm 0.12$	$-31.2 \pm 0.17$	$-35.55 \pm 0.22$	$-28.58 \pm 0.16$
MIC (mg/L) <sup>d</sup>	0.25–0.5	>8.0	>8.0	>8.0

<sup>a</sup>  $\Delta E_{MM} = \Delta E_{ele} + \Delta E_{vdw}$ .

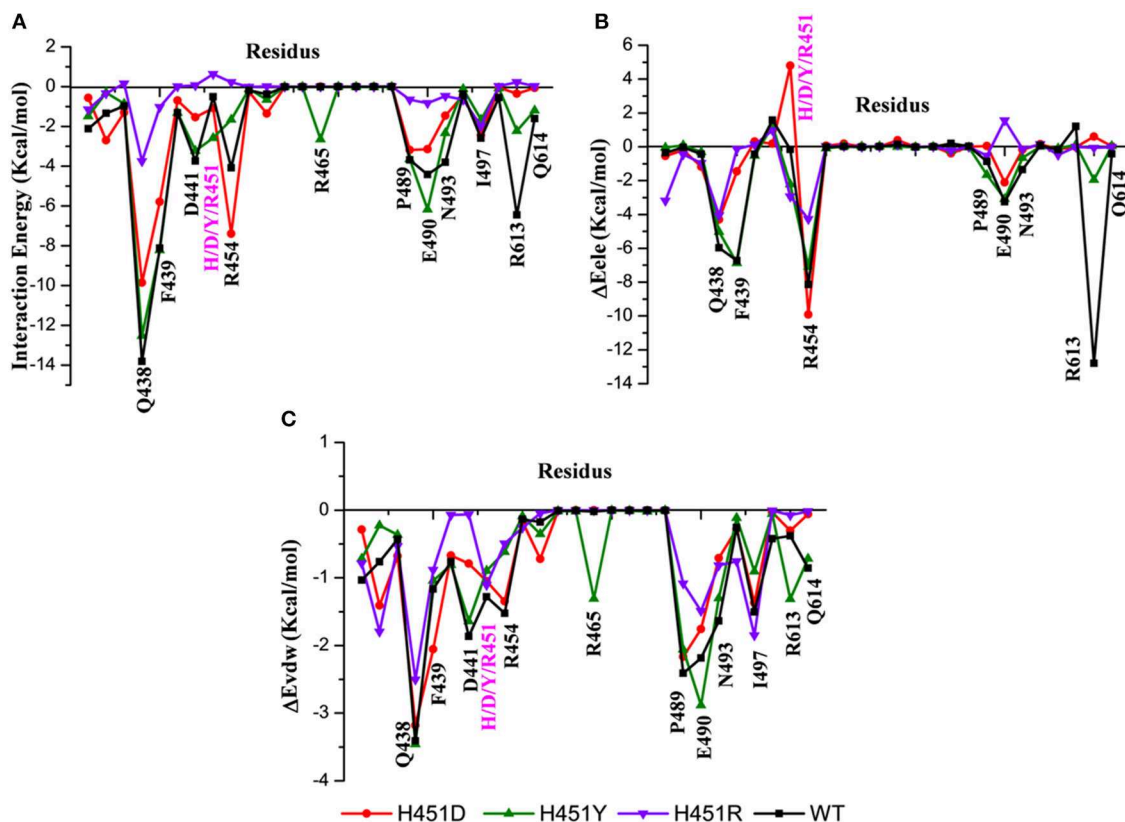
<sup>b</sup>  $\Delta G_{polar} = \Delta E_{ele} + \Delta G_{GB}$ .

<sup>c</sup>  $\Delta G_{nonpolar} = \Delta E_{vdw} + \Delta G_{SA}$ .

<sup>d</sup> In vitro activity of rifampicin in wild-type and rifampicin-resistant *Mycobacterium tuberculosis*.

Waals interaction energy  $\Delta E_{vdw}$  and non-polar interaction energy in solvation free energy  $\Delta G_{SA}$  are the driving force for the binding of rifampicin. However, the energy contributions of  $E_{vdw}$  and  $\Delta G_{SA}$  decrease in the mutants. Relative to non-polar interactions, the polar interaction (sum of electrostatic interaction energy  $\Delta E_{ele}$  and polar interaction energy in solvation free energy  $\Delta G_{GB}$ ) seems

like unfavorable for the binding of rifampicin and more apparent in the mutants (15.22, 19.21, 20.70, and 23.15 kcal/mol for the wild type, H451D, H451Y, and H451R mutants, respectively). Although the contributions of intermolecular electrostatic interactions ( $\Delta E_{ele}$ ) are very favorable, their contributions are compensated by the large desolvation penalties.



**FIGURE 3 | (A)** The total energy contributions of the key residues for binding of rifampicin: wild type (black), H451D (red), H451Y (green), H451R (purple). **(B)** The electrostatic energy contributions of the key residues for binding of rifampicin. **(C)** The van der Waals energy contributions of the key residues for binding of rifampicin.

### Key Residues Responsible for the Reduced Binding Ability of H451D/Y/R Mutants

By decomposing the binding free energy of the wild-type Mtb-RNAP with rifampicin, 10 key residues such as Q438, F439, D441, R454, P489, E490, N493, I497, R613, and Q614 (**Figure 3A**) with energy contributions over 2 kcal/mol are identified. It can be seen that relative to the wild-type Mtb-RNAP, the contributions of Q438, F439, D441, E490, N493, R613, and Q614 have an obvious decrease in H451D mutant. For H451Y mutant, the reduced energy contribution of residues Q438, D441, R454, N493, R613, and Q614 should be responsible for the reduced binding affinity of rifampicin to Mtb-RNAP. Moreover, the energy reduction in H451R mutant is more obvious relative to that in H451D/Y mutants. Actually, the profile of each residue's energy contribution in three mutants shares some similar features. For example, in all three mutants, the residues Q438, D441, N493, R613, and Q614 have obvious reduced contribution for the binding of rifampicin, suggesting that the drug-resistance mechanisms due to three mutations have some similarities.

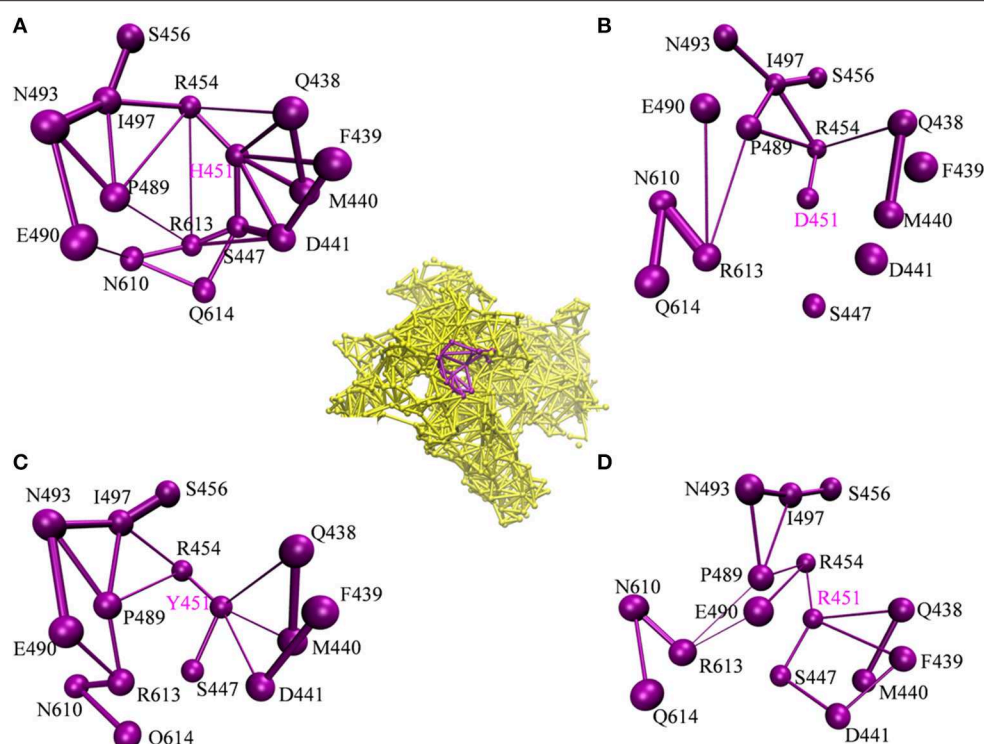
To further explore the origin of the reduced residues' contribution, we compared the electrostatic contribution and van der Waals contribution of the key residues (shown in **Figures 3B,C**, respectively). From **Figure 3B**, we can obtain when H451 is mutated to D451, the electrostatic energy

contribution of D451 (4.80 kcal/mol) is detrimental for rifampicin binding. Moreover, the contributions of electrostatic energy of Q438, F439, E490, and R613 decrease after H451D mutation. For H451Y mutant, the electrostatic energy contribution of three residues (Q438, R454, and R613) also have an obvious reduction. In H451R mutant, the contributions of electrostatic interactions of almost all key residues reduced. In addition, **Figure 3C** shows that the reduction of energy contribution of D441, P489, E490, N493, and Q614 is mainly from the loss of the van der Waals interaction contribution in H451D/Y/R mutants.

### The Dynamic Network Analysis and Residue Interaction Network Analysis Reveal That H451D/Y/R Mutations Weaken the Interaction of Mutated Residue With Its Adjacent Residues

To investigate how H451D/Y/R mutations change the binding pocket and further lead to the resistance of Mtb-RNAP to rifampicin, the dynamic network analysis was further carried out based on the 2,000 snapshots extracted from the equilibrium phase for each system. The obtained results are shown in **Figure 4**. The strength of the total interactions (including van der Waals interaction and hydrogen bond interaction and so on) is indicated by the edge thickness. Moreover, the type of interactions is shown by the two-dimensional RIN interactively





**FIGURE 4 |** The picture of dynamic network analysis for the active pocket of rifampicin binding: **(A)** WT; **(B)** H451D; **(C)** H451Y; **(D)** H451R; The purple spheres represent the residues and the sticks represent the total interactions. The strength of interactions of two residues is indicated by the thickness degree of stick, the thicker of the stick, and the stronger interaction of two residues.

based on the representative structure of each system (**Figure 5**). Here, the representative structures were extracted by clustering analysis, and the conformation with the lowest RMSD to the cluster center was selected.

In the wild-type Mtb-RNAP, H451 residue can be identified as the central node with six first neighbors in the interaction network according to **Figures 4A, 5A**. H451 not only can form strong van der Waals interactions with R454, Q438, F439, M440, and D441 but also form hydrogen bonds with D441, S447. Moreover, S447 acts as a key node in the network; in addition to forming hydrogen bond with H451, it also can form strong interactions with D441 (van der Waals interaction), R613 (hydrogen bond interaction), and Q614. I497, as another key node in the other side of the network, can form van der Waals interactions with R454, P489, and N493. Finally, R454 as a joint node can connect the two sides of the networks together by forming the van der Waals interactions with I497, H451, and the hydrogen bond with Q438. In addition, the weak “triangular shape” interaction network formed by R454, R613, and P489 residues that make the binding pocket more compact and coherent.

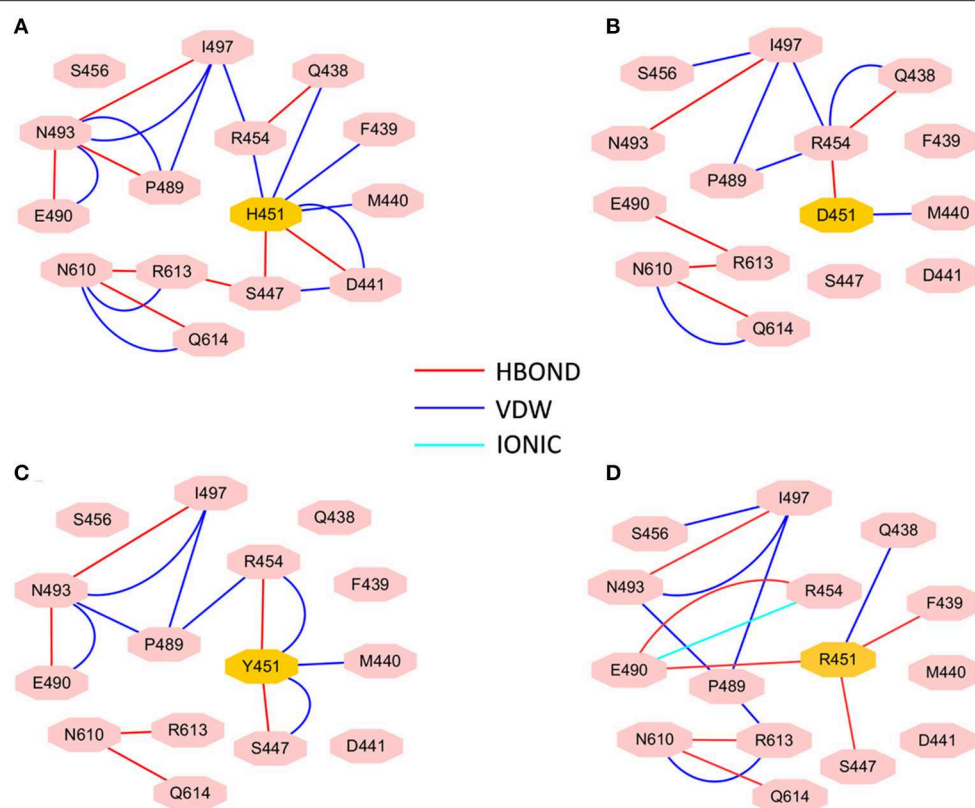
However, in H451D mutant, it is evident that the mutation causes the disappearance of the interactions between D451 and Q438, F439, M440, D441, and S447 mentioned previously (**Figures 4B, 5B**). For H451Y/R mutants, the corresponding interactions are also obviously weakened (**Figures 4C,D**). Overall, in H451D/Y/R mutants, the “triangular shape”

interaction network is broken. Based on these results, it can be concluded that the mutations on 451 indeed reduced the interaction connection of the residues in the binding pocket and then result in the active pocket more flexible and open.

### The Comparison of Binding Modes of Rifampicin With the Wild-Type *Mycobacterium tuberculosis* RNA Polymerase and Three Mutants

To show the detailed rifampicin-resistance mechanism to Mtb-RNAP H451D/Y/R mutants intuitively, further structural analysis was performed. The representative structures are depicted in **Figure 6**. As shown in **Figures 3C, 6A**, the van der Waals interactions of Q438, D441, H451, R454, P489, E490, N493, I497, and Q614 with rifampicin are pivotal for the binding of rifampicin to the wild-type Mtb-RNAP. In addition, the calculation of hydrogen bond occupancy was carried out to monitor the formation of hydrogen bonds between rifampicin and Mtb-RNAP over the whole MD simulations. Based on the result in **Table 2**, we can see that the O8, O11, O9/O2, and O12 atoms of rifampicin can form stable hydrogen bonds with the side chains of Q438, F439, and R613 with high occupancy rate. The formation of these hydrogen bonds makes F439, R613, and Q438 have large electrostatic energy contributions (**Figure 3B**).

By comparing the binding modes of wild-type and H451D mutant (**Figures 6A,B**), the position of rifampicin has a



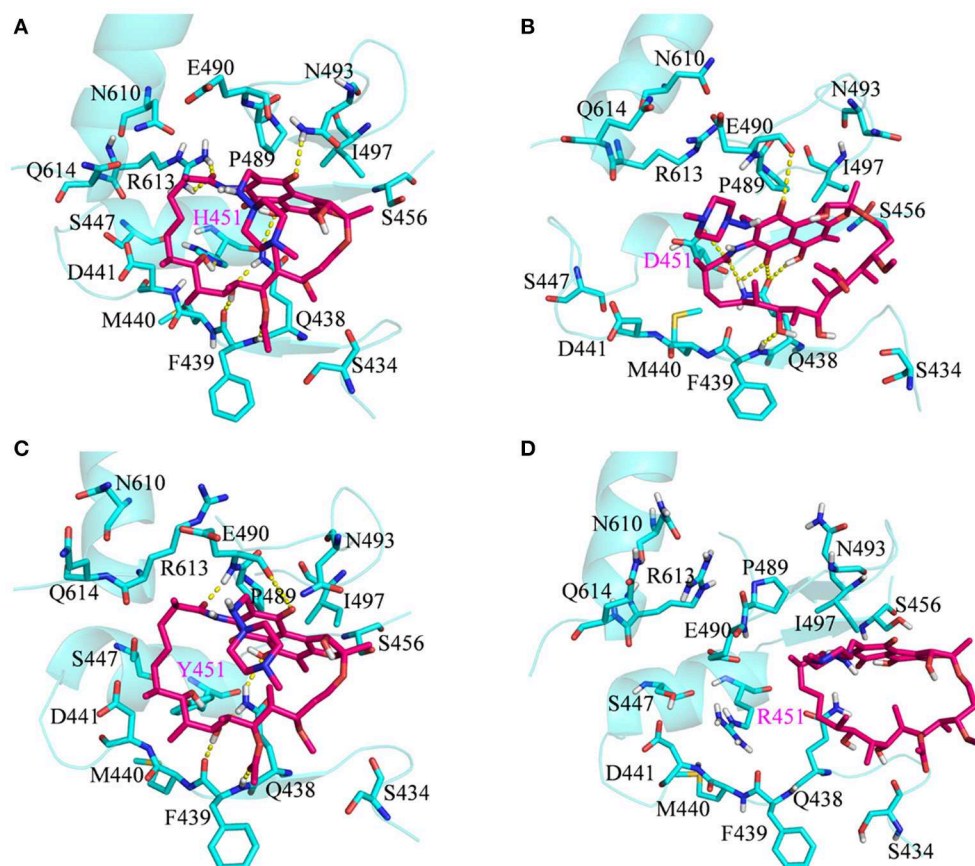
**FIGURE 5 |** RIN for the active pocket of rifampicin binding (A) wild type; (B) H451D; (C) H451Y; (D) H451R. The pink octagons represent residues, and the yellow octagons represent the mutated residues; the edges represent van der Waals (blue) and hydrogen bond (red) interactions and salt bridge interaction (cyan).

clear movement. Combined with the results of RIN analysis (Figures 5A,B), H451D mutation causes the interactions of residue 451 between with some residues of binding pocket disappear, further causing the conformations of some amino acids (such as D441, P489, E490, N493, and Q614) that changed a lot. Finally, the residues D441, P489, E490, N493, and Q614 are away from rifampicin, causing the weakened van der Waals interaction of these residues. From Table 2, some hydrogen bonds between rifampicin and Q438, F439, and R613 disappeared in H451D mutant, which causes the reduction of the electrostatic energy contribution (Figure 3B). Moreover, the H451D mutation leads to the electrostatic repulsion between carboxyl group of D451 and O10 atom of rifampicin and is unfavorable for the binding of rifampicin.

For H451Y mutant, the binding mode of rifampicin with Mtb-RNAP is more similar to that in the wild-type Mtb-RNAP, and the position of rifampicin just has a slight movement (Figures 6A,C). Despite this, the mutation still causes some residues (such as N493, I497, and Q614) away from rifampicin, reducing the nonpolar contribution of these residues. In addition, the result in Table 2 shows that the disappearance of the hydrogen bond between R613 and rifampicin is responsible for the reduction of electrostatic contributions of R613. The reduction of hydrogen bond occupancy rate formed between Q438 and rifampicin causes the loss of electrostatic contribution of Q438.

From the energetic and structural analysis discussed previously, it can be seen that the flipping of R613 greatly affects the binding of rifampicin in H451D/Y mutants. Therefore, it is worthy to explore how the mutation of H451 affects the flipping of R613. For this aim, we superimpose the binding pocket of the wild-type Mtb-RNAP with those of H451D/Y mutants. Figure 7 shows the superimposition results. From Figure 7A, it can be seen that the position of H451 is on H1. When H451 is mutated to D451, no hydrogen bond can be formed between D451 and S447 (Figure 5B), which causes the helix structure of S447 to transform into disordered loops. At the same time, the fluctuation of this loop further causes the disappearance of the hydrogen bond between R613 (located on H2) and S447 (located on H1). Therefore, the R613 residue flips with H2 due to the weaker interaction between H1 and H2. However, the mutation of H451Y enhances the van der Waals interaction of Y451 and S447 (Figure 5C), which increases the distance of S447 and R613 and further to interfere the formation of hydrogen bond. Moreover, the rotation of H2 makes the flipping of R613 more obvious. The flipping of R613 directly causes the disappearance of the hydrogen bonds with rifampicin, further leading that the electrostatic contribution of R613 reduced.

For H451R mutant, there was a large conformational transition of the binding pocket and rifampicin as shown in



**FIGURE 6 |** Three-dimensional representation for the binding mode of rifampicin with wild-type *Mycobacterium tuberculosis* RNA polymerase and three mutants based on the obtained representative structure: **(A)** wild type; **(B)** H451D; **(C)** H451Y; **(D)** H451R. The protein is shown as a cyan cartoon; rifampicin and the key residues are represented as hotpink and cyan carbon sticks, respectively; the intermolecular hydrogen bonds are indicated as dashed yellow lines.

**Figure 6D.** Moreover, R451 has a steric clash with rifampicin and so that rifampicin moves out of the binding pocket. Thereby, the movement of rifampicin causes the energy contribution of some key residues such as R613, E490, S447, D441, M440, and F439 reduce (**Figure 3**). In addition, some hydrogen bonds between rifampicin and F439/R613 also disappear (**Table 2**). In order to observe the changes of the binding pocket more intuitively, the surface map for the binding pocket of the wild-type and H451R mutant is depicted in **Figure 8**. It can be seen that when H451 is replaced by R451, the long side chain of R451 made the binding pocket smaller and cannot accommodate the relatively rigid rifampicin, which leads to the movement of rifampicin.

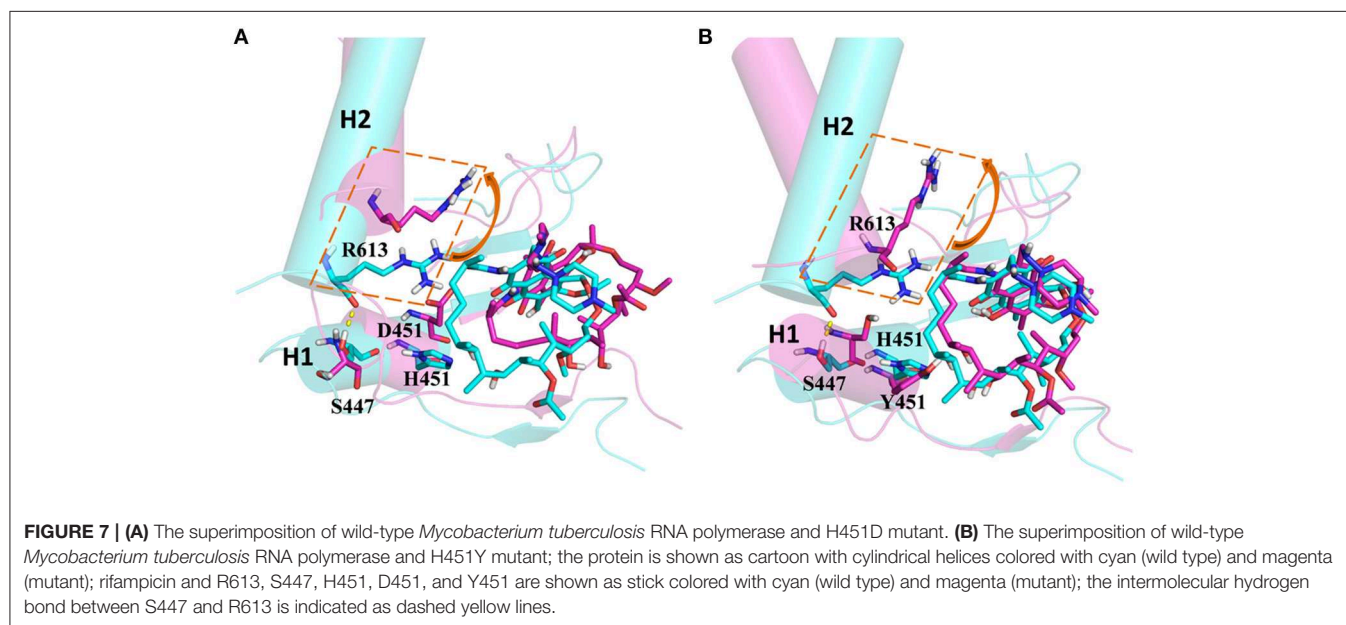
In summary, there are three main binding sites between rifampicin and the active pocket in wild type (**Figure 9A**). The polar pocket formed by the residues N610, R613, Q614, etc. (S1) can act as the hydrogen bond acceptor to form the hydrogen bond with the oxygen atom of rifampicin. The other polar pocket (S2) consists of residues Q438, M440, D441, etc., which also forms stable hydrogen bonds with rifampicin. In addition, the hydrophobic pocket composed of residues such as P489, N493, and I497 just accommodates the naphthalene ring of

rifampicin. However, H451D/Y/R mutants changed the initial binding mode for rifampicin resulting from the change of the side chain size of 451. The N atom from imidazole ring of H451 can form the hydrogen bond interaction with rifampicin in wild-type RNA polymerase, which is interfered by these mutations. Furthermore, the mutations in 451 position impaired the interaction between the residue 451 and other key residues in the active pocket, interfering with the specific orientation of the key residues' side chains, thereby increasing the flexibility of the key residues such as R613. As a result, the binding affinity of rifampicin reduced, and the important interactions between rifampicin and active pocket were disturbed. One of the most apparent changes from the results of binding free energy decomposition was that the reduction of energy contribution from R613, which was mainly because the disappearance of the hydrogen bond between R613 and O atom at the C15 position of rifampicin (**Figure 9B**). Therefore, according to the obtained mechanism, to overcome the drug resistance induced by H451D/Y/R, one possible strategy is to enhance the interaction of the inhibitor with R613 by replacing the carbonyl group at the C15 position in rifampicin with a longer and negatively charged group (R). Such group may recover the hydrogen bond

**TABLE 2** | The occupancy (%) of hydrogen bonds between rifampicin and *Mycobacterium tuberculosis* RNA polymerase.

Complex	Donor	Acceptor	Distance (Å)	Angle (°)	Occupancy (%) <sup>a</sup>
WT	RIF (O9-H6)	Phe439 (O)	2.72	166.09	100.00
	Phe439 (N-H)	RIF (O8)	2.90	163.84	99.92
	Arg613 (NH2-HH22)	RIF (O11)	2.83	158.70	99.82
	Gln438 (NE2-HE22)	RIF (O9)	2.97	158.79	96.27
	Asn493 (ND2-HD21)	RIF (O12)	3.02	147.36	81.44
	Gln438 (NE2-HE21)	RIF (O2)	3.16	156.40	81.37
	Arg613 (NH1-HH12)	RIF (O11)	3.14	137.22	49.60
H451D	Phe439 (N-H)	RIF (O10)	2.30	160.76	98.07
	Arg454 (NH1-HH12)	RIF (O11)	2.89	149.16	86.45
	Gln438 (NE2-HE21)	RIF (O11)	2.30	150.29	77.69
H451Y	Phe439(N-H)	RIF (O8)	2.88	164.64	99.90
	RIF (O9-H6)	Phe439 (O)	2.70	164.10	99.83
	Gln438 (NE2-HE22)	RIF (O9)	2.30	160.02	97.25
	Glu490 (N-H)	RIF (O11)	3.05	163.01	78.20
	Gln438 (NE2-HE21)	RIF (O2)	3.20	157.50	55.74
	Tyr451 (O-H)	RIF (O10)	2.86	158.02	50.49
H451R	Gln438 (NE2-HE21)	RIF (O11)	3.02	154.50	54.26
	RIF (O9-H6)	Ser434 (O)	2.69	162.18	53.93

<sup>a</sup>Only hydrogen bonds that existed more than 50% of the time were shown.



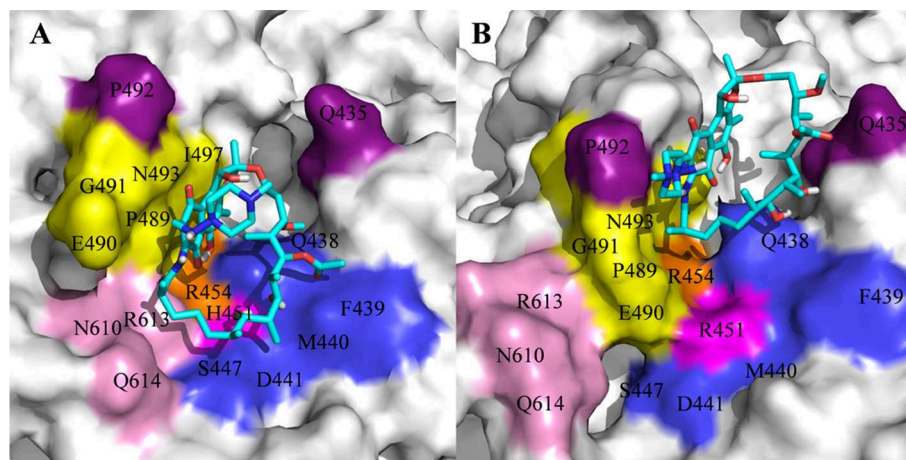
interaction between R613 and inhibitor, which could stabilize the binding of rifampicin.

## CONCLUSION

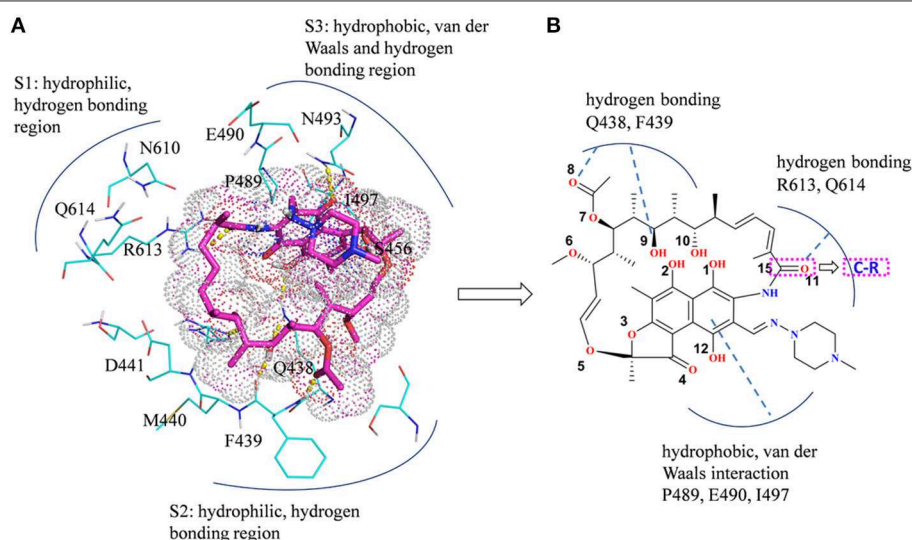
In this study, MD simulations together with MM-GBSA, dynamic network analysis, and RIN analysis were carried out on the complexes of rifampicin with the wild-type Mtb-RNAP and H451D/Y/R mutants to explore the resistant mechanism of

rifampicin. The results from the MM-GBSA calculations are well-consistent with the experimental result. The reduced binding affinity for the studied mutants mainly comes from the loss of van der Waals contribution of D441, P489, E490, N493, I497, and Q614 and electrostatic contribution of Q438, F439, and R613 in mutants. The electrostatic energy contribution of R613 decreases obviously with the disappearance of the hydrogen bonds between R613 and rifampicin, which caused by the conformation flipping of R613 in H451D/Y mutants. The binding modes and dynamic





**FIGURE 8 |** The binding pocket of rifampicin: **(A)** wild type; **(B)** H451R mutant; the pocket is shown as surface, and rifampicin is shown as stick colored with cyan.



**FIGURE 9 | (A)** Three main binding sites between rifampicin and the active pocket. Rifampicin is shown as stick colored with magenta, and the residues are shown as lines colored with cyan. The yellow dotted lines represent the hydrogen bond interactions. **(B)** Structural optimization of rifampicin. R represents a longer negatively charged group.

network analysis show that the weakened interactions among D/Y/R451 with Q438, F439, M440, D441, and S447 increase the flexibility of the binding pocket, thereby reducing the binding affinity of rifampicin to Mtb-RNAP. In addition, these mutations caused the key hydrogen bond interactions between residue 451 and rifampicin disappear. Finally, the position of rifampicin had a clear movement, which changed the stable binding mode of rifampicin. We firstly proposed the atomic level resistance mechanism of Mtb to rifampicin due to H451D/Y/R mutations on Mtb-RNAP. In addition, we also proposed some guidance for the alteration of the rifampicin and the development of new drugs in the future. Though the hypothesis is still unvalidated on the traditional experimental, it can provide some theoretical

underpinnings for the design of new anti-TB drugs to some extent, or a particular aspect.

## DATA AVAILABILITY STATEMENT

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

## AUTHOR CONTRIBUTIONS

QZ and HuL designed the research and wrote the manuscript. QZ, XA, HoL, SW, and TX performed the dynamic simulations and analyzed the data.

## FUNDING

This work was supported by the National Nature Science Foundation of China (Grant No. 21675070).

## REFERENCES

- Ahmad, S., Mokaddas, E., and Al-Mutairi, N. M. (2018). Prevalence of tuberculosis and multidrug resistant tuberculosis in the Middle East region. *Exp. Rev. Anti. Infect. Ther.* 16, 709–721. doi: 10.1080/14787210.2018.1519393
- Aruksakunwong, O., Wolschann, P., Hannongbua, S., and Sompornpisut, P. (2006). Molecular dynamic and free energy studies of primary resistance mutations in HIV-1 protease-ritonavir complexes. *J. Chem. Inf. Model.* 46, 2085–2092. doi: 10.1021/ci060090c
- Bai, Q., Shen, Y., Jin, N., Liu, H., and Yao, X. (2014). Molecular modeling study on the dynamical structural features of human smoothed receptor and binding mechanism of antagonist LY2940680 by metadynamics simulation and free energy calculation. *Biochim. Biophys. Acta* 1840, 2128–2138. doi: 10.1016/j.bbagen.2014.03.010
- Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97, 10269–10280. doi: 10.1021/j100142a004
- Bodmer, T., Zürcher, G., Imboden, P., and Telenti, A. (1995). Mutation position and type of substitution in the  $\beta$ -subunit of the RNA polymerase influence in vitro activity of rifamycins in rifampicin-resistant *Mycobacterium tuberculosis*. *J. Antimicrob. Chemother.* 35, 345–348. doi: 10.1093/jac/35.2.345
- Campbell, E. A., Korzhova, N., Mustae, A., Murakami, K., Nair, S., Goldfarb, A., et al. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104, 901–912. doi: 10.1016/S0092-8674(01)00286-0
- Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., et al. (2014). *AMBER 14*. San Francisco, CA: University of California.
- Caws, M., Duy, P. M., Tho, D. Q., Lan, N. T., Hoa, D. V., and Farrar, J. (2006). Mutations prevalent among rifampin- and isoniazid-resistant *Mycobacterium tuberculosis* isolates from a hospital in Vietnam. *J. Clin. Microbiol.* 44, 2333–2337. doi: 10.1128/JCM.00330-06
- Chang, K. C., Leung, C. C., Yew, W. W., Chan, S. L., and Tam, C. M. (2006). Dosing schedules of 6-month regimens and relapse for pulmonary tuberculosis. *Am. J. Respir. Crit. Care Med.* 174, 1153–1158. doi: 10.1164/rccm.200605-637OC
- Chang, K. C., Nuermberger, E., Sotgiu, G., and Leung, C. C. (2018). New drugs and regimens for tuberculosis. *Respirology* 23, 978–990. doi: 10.1111/resp.13345
- Chikaonda, T., Ketseoglou, I., Nguluwe, N., Krysiak, R., Thengolose, I., Nyakwawa, F., et al. (2017). Molecular characterisation of rifampicin-resistant *Mycobacterium tuberculosis* strains from Malawi. *Afr. J. Lab. Med.* 6:463. doi: 10.4102/ajlm.v6i2.463
- Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995). Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* 16, 1357–1377. doi: 10.1002/jcc.540161106
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- Doncheva, N. T., Klein, K., Domingues, F. S., and Albrecht, M. (2011). Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* 36, 179–182. doi: 10.1016/j.tibs.2011.01.002
- Fox, T., and Kollman, P. A. (1998). Application of the RESP methodology in the parametrization of organic solvents. *J. Phys. Chem.* 102, 8070–8079. doi: 10.1021/jp9717655
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09*. Wallingford, CT: Gaussian Inc.
- Glykos, N. M. (2006). Software news and updates carma: a molecular dynamics analysis program. *J. Comput. Chem.* 27, 1765–1768. doi: 10.1002/jcc.20482
- He, G. X., Zhao, Y. L., Jiang, G. L., Liu, Y. H., Xia, H., Wang, S. F., et al. (2008). Prevalence of tuberculosis drug resistance in 10 provinces of China. *BMC Infect. Dis.* 8:166. doi: 10.1186/1471-2334-8-166
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi: 10.1002/prot.21123
- 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
- Jeon, D. (2017). WHO treatment guidelines for drug-resistant tuberculosis, 2016 update: applicability in South Korea. *Tuberc. Respir. Dis.* 80, 336–343. doi: 10.4046/trd.2017.0049
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi: 10.1002/bip.360221211
- Lin, W., Mandal, S., Degen, D., Liu, Y., Ebright, Y. W., Li, S., et al. (2017). Structural basis of *Mycobacterium tuberculosis* transcription and transcription inhibition. *Mol. Cell* 66:e168. doi: 10.1016/j.molcel.2017.03.001
- Ma, X., Wang, H., Deng, Y., Liu, Z., Xu, Y., Pan, X., et al. (2006). rpoB Gene mutations and molecular characterization of rifampin-resistant *Mycobacterium tuberculosis* isolates from Shandong Province, China. *J. Clin. Microbiol.* 44, 3409–3412. doi: 10.1128/JCM.00515-06
- McClure, W. R., and Cech, C. L. (1978). On the mechanism of rifampicin inhibition of RNA synthesis. *J. Biol. Chem.* 253, 8949–8956.
- Miller, L. P., Crawford, J. T., and Shinnick, T. M. (1994). The rpoB gene of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* 38, 805–811. doi: 10.1128/AAC.38.4.805
- Mishra, A., and Suroli, A. (2018). *Mycobacterium tuberculosis*: surviving and indulging in an unwelcoming host. *IUBMB Life* 70, 917–925. doi: 10.1002/iub.1882
- Mitchison, D. A., and Nunn, A. J. (1986). Influence of initial drug resistance on the response to short-course chemotherapy of pulmonary tuberculosis. *Am. Rev. Respir. Dis.* 133, 423–430.
- Morlock, G. P., Plikaytis, B. B., and Crawford, J. T. (2000). Characterization of spontaneous, *in vitro*-selected, rifampin-resistant mutants of *Mycobacterium tuberculosis* strain H37Rv. *Antimicrob. Agents Chemother.* 44, 3298–3301. doi: 10.1128/AAC.44.12.3298-3301.2000
- Organization, W. H. (2018). *Global Tuberculosis Report 2018*. World Health Organization.
- Pan, D., Sun, H., Bai, C., Shen, Y., Jin, N., Liu, H., et al. (2011). Prediction of zanamivir efficiency over the possible 2009 Influenza A (H1N1) mutants by multiple molecular dynamics simulations and free energy calculations. *J. Mol. Model.* 17, 2465–2473. doi: 10.1007/s00894-010-0929-8
- Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, I. I., DeBolt, T. E., et al. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91, 1–41. doi: 10.1016/0010-4655(95)00041-D
- Piovesan, D., Minervini, G., and Tosatto, S. C. (2016). The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* 44, W367–W374. doi: 10.1093/nar/gkw315
- Rocchia, W., Alexov, E., and Honig, B. (2001). Extending the applicability of the nonlinear Poisson–Boltzmann equation: multiple dielectric constants and multivalent ions. *J. Phys. Chem. B* 105, 6507–6514. doi: 10.1021/jp010454y

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2019.00819/full#supplementary-material>

- Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5
- Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B. K., Church, G. M., and Murray, M. B. (2009). Tuberculosis drug resistance mutation database. *PLoS Med.* 6:e1000002. doi: 10.1371/journal.pmed.1000002
- Seid, M. A., Ayalew, M. B., Muche, E. A., Gebreyohannes, E. A., and Abegaz, T. M. (2018). Drug-susceptible tuberculosis treatment success and associated factors in Ethiopia from 2005 to 2017: a systematic review and meta-analysis. *BMJ Open* 8:e022111. doi: 10.1136/bmjopen-2018-022111
- Service, H. K. C., and Council, B. M. R. (1981). Controlled trial of four thrice-weekly regimens and a daily regimen all given for 6 months for pulmonary tuberculosis. *Lancet* 1, 171–174.
- Sethi, A., Eargle, J., Black, A. A., and Luthey-Schulten, Z. (2009). Dynamical networks in tRNA: protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6620–6625. doi: 10.1073/pnas.0810961106
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sitkoff, D., Sharp, K. A., and Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98, 1978–1988. doi: 10.1021/j100058a043
- Somaskovi, A., Parsons, L. M., and Salfinger, M. (2001). The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in *Mycobacterium tuberculosis*. *Respir. Res.* 2, 164–168. doi: 10.1186/rr54
- Sotgiu, G., D'Ambrosio, L., Centis, R., Mura, I., Castiglia, P., Spanevello, A., et al. (2015). The multidrug-resistant tuberculosis threat: old problems and new solutions. *J. Thorac. Dis.* 7, E354–E360. doi: 10.3978%2Fj.issn.2072-1439.2015.09.21
- Telenti, A., Imboden, P., Marchesi, F., Lowrie, D., Cole, S., Colston, M. J., et al. (1993). Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet* 341, 647–650. doi: 10.1016/0140-6736(93)90417-F
- Tiberi, S., du Plessis, N., Walzl, G., Vjecha, M. J., Rao, M., Ntumi, F., et al. (2018). Tuberculosis: progress and advances in development of new drugs, treatment regimens, and host-directed therapies. *Lancet Infect. Dis.* 18, e183–e198. doi: 10.1016/S1473-3099(18)30110-5
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035
- Wu, E. Y., and Hilliker, A. K. (2017). Identification of rifampicin resistance mutations in *Escherichia coli*, including an unusual deletion mutation. *J. Mol. Microbiol. Biotechnol.* 27, 356–362. doi: 10.1159/000484246
- Xue, W., Ban, Y., Liu, H., and Yao, X. (2014). Computational study on the drug resistance mechanism against HCV NS3/4A protease inhibitors vaniprevir and MK-5172 by the combination use of molecular dynamics simulation, residue interaction network, and substrate envelope analysis. *J. Chem. Inf. Model.* 54, 621–633. doi: 10.1021/ci400060j
- Xue, W., Jin, X., Ning, L., Wang, M., Liu, H., and Yao, X. (2013). Exploring the molecular mechanism of cross-resistance to HIV-1 integrase strand transfer inhibitors by molecular dynamics simulation and residue interaction network analysis. *J. Chem. Inf. Model.* 53, 210–222. doi: 10.1021/ci300541c
- Xue, W., Liu, H., and Yao, X. (2012). Molecular mechanism of HIV-1 integrase-vDNA interactions and strand transfer inhibitor action: a molecular modeling perspective. *J. Comput. Chem.* 33, 527–536. doi: 10.1002/jcc.22887
- Yang, Y., Shen, Y., Li, S., Jin, N., Liu, H., and Yao, X. (2012). Molecular dynamics and free energy studies on Aurora kinase A and its mutant bound with MLN8054: insight into molecular mechanism of subtype selectivity. *Mol. Biosyst.* 8, 3049–3060. doi: 10.1039/c2mb25217a
- Zaw, M. T., Emran, N. A., and Lin, Z. (2018). Mutations inside rifampicin-resistance determining region of rpoB gene associated with rifampicin-resistance in *Mycobacterium tuberculosis*. *J. Infect. Public Health* 11, 605–610. doi: 10.1016/j.jiph.2018.04.005
- Zhou, S., Shi, D., Liu, X., Yao, X., Da, L. T., and Liu, H. (2019). The pH-induced misfolding mechanism of prion protein: insights from microsecond accelerated molecular dynamics simulations. *ACS Chem. Neurosci.* 10, 2718–2729. doi: 10.1021/acscchemneuro.8b00582

**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.

Copyright © 2019 Zhang, An, Liu, Wang, Xiao and Liu. 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.



# Conformation Transition of Intracellular Part of Glucagon Receptor in Complex With Agonist Glucagon by Conventional and Accelerated Molecular Dynamics Simulations

Qifeng Bai<sup>1†</sup>, Shuoyan Tan<sup>2†</sup>, Horacio Pérez-Sánchez<sup>3</sup>, Haixia Feng<sup>2</sup>, Liya Feng<sup>1</sup>, HuanXiang Liu<sup>2</sup> and Xiaojun Yao<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Laurent G. Désaubry,  
Laboratoire de Cardio-Oncologie et  
Chimie Médicinale (CNRS), France

### Reviewed by:

Linlin Yang,  
Zhengzhou University, China  
Abdullahi Ibrahim Uba,  
Kadir Has University, Turkey

### \*Correspondence:

Qifeng Bai  
baiqf@lzu.edu.cn  
Xiaojun Yao  
xjyao@lzu.edu.cn

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

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 19 September 2019

**Accepted:** 22 November 2019

**Published:** 17 December 2019

### Citation:

Bai Q, Tan S, Pérez-Sánchez H,  
Feng H, Feng L, Liu H and Yao X  
(2019) Conformation Transition of  
Intracellular Part of Glucagon  
Receptor in Complex With Agonist  
Glucagon by Conventional and  
Accelerated Molecular Dynamics  
Simulations. *Front. Chem.* 7:851.  
doi: 10.3389/fchem.2019.00851

<sup>1</sup> Key Lab of Preclinical Study for New Drugs of Gansu Province, Institute of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China, <sup>2</sup> School of Pharmacy, Lanzhou University, Lanzhou, China, <sup>3</sup> Computer Science Department, Universidad Católica San Antonio de Murcia (UCAM), Murcia, Spain

The inactive conformations of glucagon receptor (GCGR) are widely reported by crystal structures that support the precision structure for drug discovery of type 2 diabetes. The previous study shows that the intracellular part is open in the glucagon-bound GCGR (glu-GCGR) and closed in the apo-GCGR by accelerated molecular dynamics (aMD) simulations. However, the crystal structure of GCGR in complex with partial agonist shows that the intracellular part is closed in the inactive conformation. To understand the differences between the studies of aMD simulations and crystal structure, the 2,500 ns conventional molecular dynamics (cMD) simulations are performed on the simulated model of glu-GCGR. The result shows that the transmembrane helices (TMH) 6 of glu-GCGR is outward  $\sim 4$  Å to drive the intracellular part of glu-GCGR open until  $\sim 390$  ns cMD simulations. The (TMH) 6 of glu-GCGR becomes closed after  $\sim 490$  ns cMD simulations, which are consistent with the crystal structure of GCGR in complex with the partial agonist. To further elucidate the activation mechanism of GCGR deeply, the simulated models of glu-GCGR, apo-GCGR, and antagonist-bound GCGR (ant-GCGR) are constructed to perform 10 of parallel 300 ns aMD simulations, respectively. The results show that both of glu-GCGR and apo-GCGR can generate the open conformations of the intracellular part. But the glu-GCGR has the higher percentage of open conformations than apo-GCGR. The ant-GCGR is restricted to generate the open conformations of the intracellular part by antagonist MK-0893. It indicates that the glu-GCGR, apo-GCGR, and ant-GCGR can be distinguished by the aMD simulated method. Free energy landscape shows that the open conformations of the intracellular part of GCGR are in intermediate state. Our results show that aMD simulations enhance the space samplings of open conformations of GCGR via adding extra boost potential. It indicates that the aMD simulations are an effective way for drug discovery of GCGR.

**Keywords:** G-protein-coupled receptors, glucagon receptor, molecular dynamics simulations, accelerated molecular dynamics simulations, conformation transition



## INTRODUCTION

The human glucagon receptor (GCGR), which belongs to one member of class B G-protein-coupled receptors (GPCRs), is a potential drug target of type 2 diabetes (Cho et al., 2012). The first crystal structure of the GCGR, which is resolved without any ligand, is lack of the binding site information for studying the interaction mechanism between a ligand and the GCGR (Siu et al., 2013). The GCGR in complex with antagonist MK-0893, which is determined by X-ray diffraction, supplies an accurate model for antagonist design (Jazayeri et al., 2016). With the research development of GCGR structures, the GCGR with crystal extracellular domain (ECD) is reported for understanding the signal transduction mechanism of class B GPCRs (Zhang H. et al., 2017). The structure of the GCGR in complex with glucagon analog is determined for elucidating a two-domain binding mechanism in which C-terminal and N-terminal bind to ECD and transmembrane domain (TMD) binding pocket, respectively (Zhang H. et al., 2018). The glucagon-like peptide-1 receptor (GLP-1R) and the GCGR that belonged to members of class B GPCRs have very similar structures, but they have different regulated mechanisms on blood sugar level. The GCGR decreases the blood sugar level through its inactive conformation. The inhibited binding position of GCGR antagonist is at the boundary between transmembrane helices (TMH) and lipid membrane. It can restrict the outward movement of TMH to reduce post-prandial and fasting glucose concentrations (Bagger et al., 2011). However, the GLP-1R decreases the blood sugar level through its active conformation. The binding pocket of active conformation of GLP-1R is surrounded by seven TMH (Zhang Y. et al., 2017).

The molecular dynamics (MD) simulations supply a reliable way to study the dynamical conformation transition of GCGR at the atomic level. It has been successfully used to study the dynamical conformation transition and binding mechanism between ligands and GPCRs (Li et al., 2010; Bai et al., 2014b; Bai and Yao, 2016), especially class A GPCRs (Feng et al., 2012; Bai et al., 2014a; Wang et al., 2018; Zhang X. et al., 2018). In the previous reports, the MD simulations indicate that the extracellular domain (ECD) of apo-GCGR can be in open or closed state for elucidating glucagon binding mechanism (Yang et al., 2015). The similar phenomena of closed or open ECD domain also happen in apo-GLP-1R, which is studied by MD simulations and the Markov state model (Zhang et al., 2019). These scientific researches give a computational model for studying the two-domain binding mechanism of class B GPCRs. Besides, the accelerated MD (aMD) simulations are performed on the apo-GCGR and the GCGR in complex with glucagon to study the activation mechanism of the GCGR. It indicates that the GCGR is in the active conformation when the glucagon binds to the pocket surrounded by 7TMH, and the apo-GCGR keeps the inactive state during aMD simulations (Li et al., 2016). This study does not elucidate the dynamical conformation characteristics between the GCGR and an antagonist, which can decrease the blood sugar level to treat type 2 diabetes mellitus. And it shows the different phenomenon with the crystal structure of the GCGR in complex with a partial agonist, which shows the inactive

conformation of the intracellular part of the GCGR (Zhang H. et al., 2018).

To understand the differences between the crystal structure of the GCGR in complex with a partial agonist (Zhang H. et al., 2018) and the previous study of aMD simulations (Li et al., 2016), the 2,500 ns of conventional molecular dynamics (cMD) are performed on the same model of glucagon-bound GCGR (glu-GCGR) as previously reported (Li et al., 2016). The results of cMD simulations show the similar structure of the intracellular part of the GCGR with the crystal structure of the GCGR in complex with a partial agonist. Besides, 10 of parallel 300 ns aMD simulations are carried out on glu-GCGR, apo-GCGR, and antagonist-bound GCGR (ant-GCGR), respectively. It points out the differences of structural characteristics of glu-GCGR, apo-GCGR, and ant-GCGR by statistics of aMD simulated trajectories. Free energy landscape profiles the energy change during the conformational transition process of the intracellular part of the GCGR. Our studies give the deep insight into the activation mechanism of the GCGR by cMD and aMD simulations.

## MATERIALS AND METHODS

### Model Preparation

To repeat previous literature experiments, the simulated model of glucagon-bound GCGR (glu-GCGR) is extracted from the previous report of aMD simulations (Li et al., 2016). The antagonist-bound GCGR (ant-GCGR) is obtained from PDB database (PDB ID: 5EE7) (Jazayeri et al., 2016). The miss residues and disulfide bond are automatically dealt with by Schrödinger software release 2015 (Sastri et al., 2013). The apo-GCGR system is built by removing the antagonist MK-0893 of the ant-GCGR (PDB ID: 5EE7). The ant-GCGR and apo-GCGR are constructed with the solvation-membrane box dimensions of  $85 \times 85 \times 68 \text{ \AA}^3$ . The glu-GCGR is immersed in the solvation-membrane box with the dimensions of  $85 \times 85 \times 115 \text{ \AA}^3$ . The TMH of the GCGR are placed along the z axis and are orthogonal to the 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids of which size is  $85 \text{ \AA} \times 85 \text{ \AA}$ . The TIP3P model (Jorgensen et al., 1983) is chosen for water box and 0.15 M NaCl is used to neutralize the computational model systems. The force field parameters of antagonist MK-0893 are generated by the paramchem server of CHARMM General Force Field (CGenFF) (Vanommeslaeghe et al., 2012).

### Conventional and Accelerated Molecular Dynamics Simulations

The simulated model of glu-GCGR except for the membrane is fixed to perform the 50 ps of minimization and 500 ps of equilibrium simulations. Then, only the ligands and protein are constrained for running 100 ps minimization and 500 ps cMD simulations. In the next step, the total 20 ns of equilibrating cMD simulations is performed on the full released systems. At last, 2,500 ns of production cMD simulations are run on glu-GCGR. All cMD simulations are carried out under the constant temperature of 310 K and constant pressure of 1 bar.

The NAMD (Phillips et al., 2005) soft package (version 2.11) is used to run the cMD simulations on the basis of CHARMM 27 force field (Mackerell et al., 1998) in the infinitely periodical water-lipid box. The energy minimizations employ the conjugate gradient method in the NAMD software. The non-bonded cutoff of 12 Å is chosen to calculate the electrostatic interaction based on the particle-mesh Ewald (PME) method (Darden et al., 1993). The Langevin barostat and Langevin thermostat are used to control constant temperature and pressure of simulated system of the glu-GCGR (Feller et al., 1995). The simulated step is set to 2f for the simulated system of the glu-GCGR. And the frames of cMD simulations are stored every 10 ps.

The accelerated MD (aMD) simulations have been successfully applied into the studies of protein conformational transition (Miao et al., 2013; Bai et al., 2014c). It can sample the molecular configuration space efficiently via adding boost potential  $\Delta V(r)$ . The modified potential  $\Delta V^*(r)$  is equal to the sum of original potential  $V(r)$  and boost potential  $\Delta V(r)$ . The modified potential  $\Delta V^*(r)$  and boost potential  $\Delta V(r)$  are shown in Equation (1):

$$\begin{aligned} V^*(r) &= V(r) + \Delta V(r) \\ \Delta V(r) &= \begin{cases} 0, & V(r) \geq E \\ \frac{(E-V(r))^2}{\alpha + E - V(r)}, & V(r) < E \end{cases} \end{aligned} \quad (1)$$

where  $E$  is the threshold energy, and  $\alpha$  is the acceleration factor. The 10 of parallel dual-boost aMD simulations run 300 ns on the simulated systems of glu-GCGR, apo-GCGR, and ant-GCGR after 20 ns equilibrating cMD simulations. The values of threshold energy and acceleration factor are calculated (Miao et al., 2013) as follows (see Equation 2):

$$\begin{aligned} E_{\text{total}} &= V_{\text{total\_avg}} + 0.16 * N_{\text{atoms}} \\ \alpha_{\text{total}} &= 0.16 * N_{\text{atoms}} \\ E_{\text{dihed}} &= V_{\text{dihed\_avg}} + 0.3 * V_{\text{dihed\_avg}} \\ \alpha_{\text{dihed}} &= 0.3 * V_{\text{dihed\_avg}} / 5 \end{aligned} \quad (2)$$

in which  $N_{\text{atom}}$  is the total atom number of the entire system. The values of  $V_{\text{total\_avg}}$  and  $V_{\text{dihed\_avg}}$  are obtained from the average total potential energy and the average dihedral energy in cMD simulation, respectively. Total 9,000 ns aMD simulations are performed on the simulated systems of glu-GCGR, apo-GCGR, and ant-GCGR. And the trajectory snapshots are saved every 10 ps for each system. The residues L253 and Y343 are chosen to analyze the open and closed states of glu-GCGR, apo-GCGR, and ant-GCGR.

## Free Energy Landscape and Weak Interaction Analysis

The free energy landscape can show the energy change during the conformational transition process of the intracellular part of the GCGR. The free energy landscape is calculated as shown in Equation (3):

$$\Delta G = -k_B T \ln g(x, y) \quad (3)$$

where  $k_B$  and  $T$  represent the Boltzmann constant and the temperature, respectively.  $x$  and  $y$  represent the two principal components that are the distance between the C $\alpha$  atoms of L253 and Y343 of the GCGR and the root mean square deviation (RMSD) of the GCGR, respectively.

The graphical topological analysis named independent gradient model (IGM) (Lefebvre et al., 2017) is used to produce the interactive isosurfaces between L253 and Y343 of the GCGR. The weak interaction is computed by Equation (4):

$$\delta g^{\text{inter}} = \left| \nabla \rho^{\text{IGM,inter}} \right| - \left| \nabla \rho^{\text{inter}} \right| \quad (4)$$

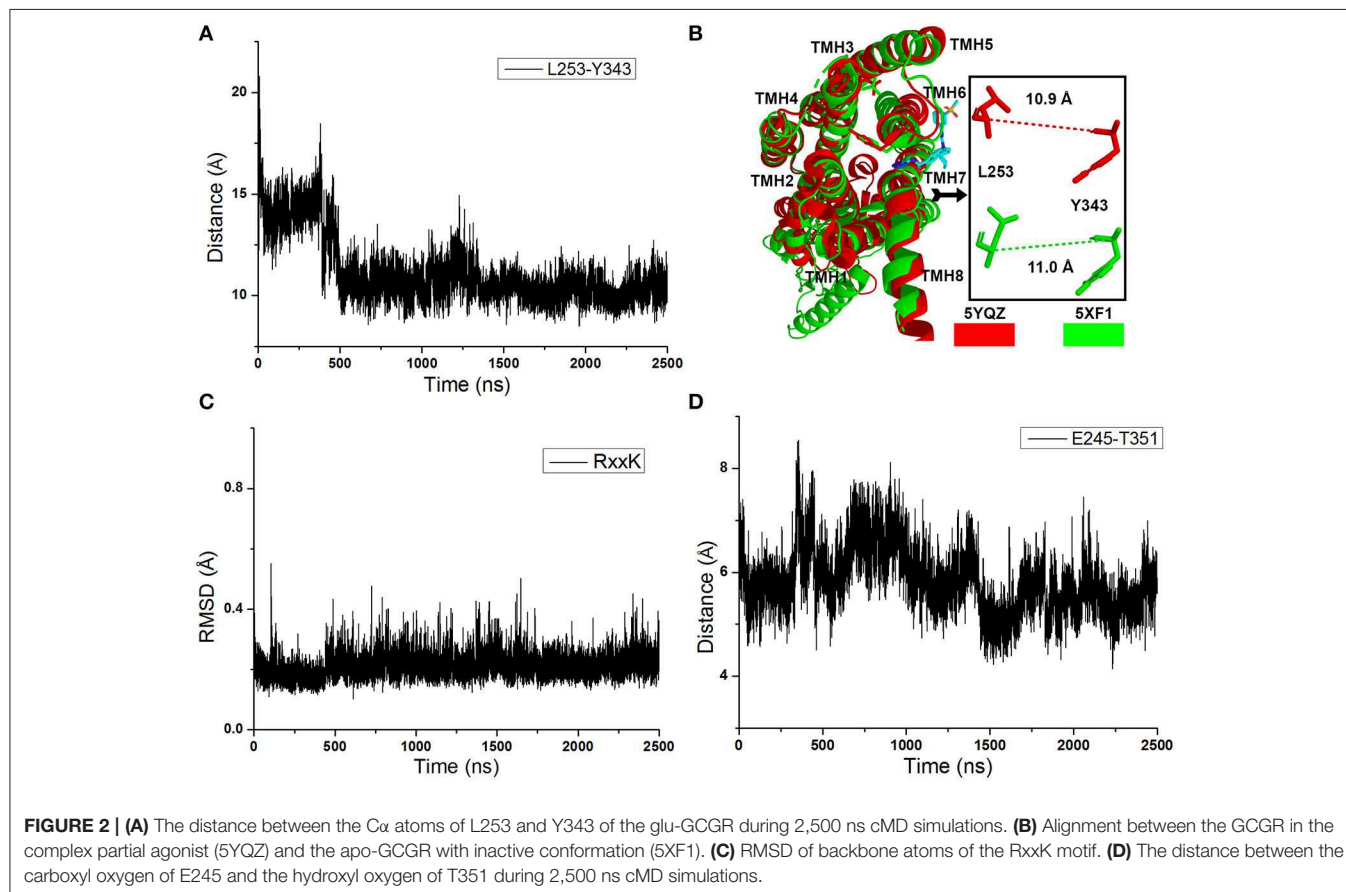
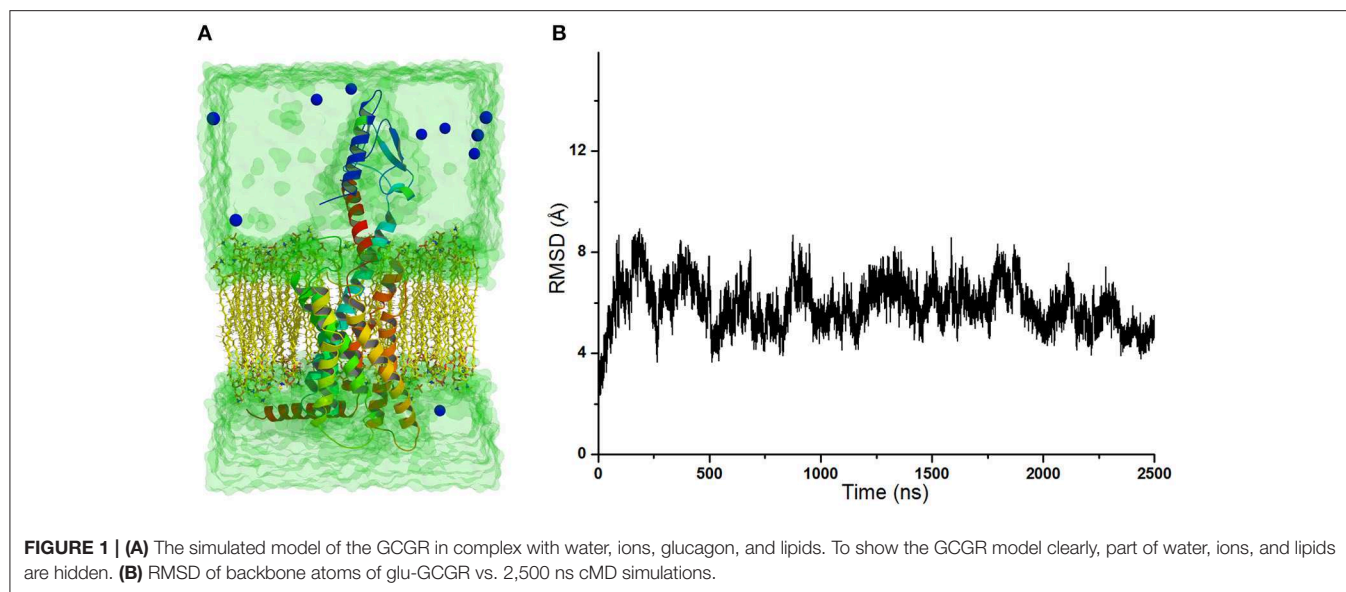
in which  $\nabla \rho^{\text{IGM,inter}}$  is the sum of absolute values of every atomic electron density of the intermolecular fragment.  $\nabla \rho^{\text{inter}}$  is the sum of every atomic electron density of the intermolecular fragment. Here, the Multiwfn software (Lu and Chen, 2012) is used to calculate the IGM. The VMD software (Humphrey et al., 1996) is employed to draw the isosurfaces between residues L253 and Y343 based on the calculated results of Multiwfn. About 91,000 grids are generated to profile isosurfaces between residues L253 and Y343. The isosurfaces are generally divided as follows: blue represented attractive interaction such as hydrogen bonding, red represented repulsive interaction such as molecular steric effect, and green represented van der Waals interactions.

## RESULTS AND DISCUSSIONS

### The Conformational Change of GCGR During 2,500 ns cMD Simulations

The crystal structure of GCGR in complex with the partial agonist peptide indicates that the intracellular part keeps the inactive conformation (Zhang H. et al., 2018). However, the results of accelerated molecular dynamics (aMD) simulations show that the active conformations are predominant in the intracellular parts of the GCGR in complex with agonist glucagon (Li et al., 2016). To understand these different results, the simulated model of the GCGR is built in complex with glucagon, ions, water, and lipids (see **Figure 1A**). The total 2,500 ns of production cMD simulations are carried out on this simulated system. The RMSD of backbone atoms of glu-GCGR keeps the relative equilibrium phase during 2,500 ns cMD simulations (see **Figure 1B**). The distance between the C $\alpha$  atoms of residues L253 and Y343 is chosen as a key principal component to mark the active or inactive state (Li et al., 2016). In our experiment, we also measure the distance between the C $\alpha$  atoms of residues L253 and Y343 based on 2,500 ns cMD simulated trajectories.

As shown in **Figure 2A**, the distance between the C $\alpha$  atoms of L253 and Y343 increases to  $\sim 15$  Å quickly at the initial cMD simulations. It means the TMH 6 of glu-GCGR is outward  $\sim 4$  Å to drive the intracellular conformation of the glu-GCGR open. The results are consistent with the reported results (Li et al., 2016). Then, the distance between the C $\alpha$  atoms of L253 and Y343 is changed until  $\sim 390$  ns. The distance between the C $\alpha$  atoms of L253 and Y343 becomes  $\sim 11$  Å from  $\sim 490$  ns MD simulations. The inactive crystal conformation of the



GCGR shows that the distance is about 11 Å between the C $\alpha$  atoms of L253 and Y343 (see **Figure 2B**). And the distance between the C $\alpha$  atoms of L253 and Y343 is 10.9 Å in the crystal partial agonist-bound GCGR. The alignment shows that the

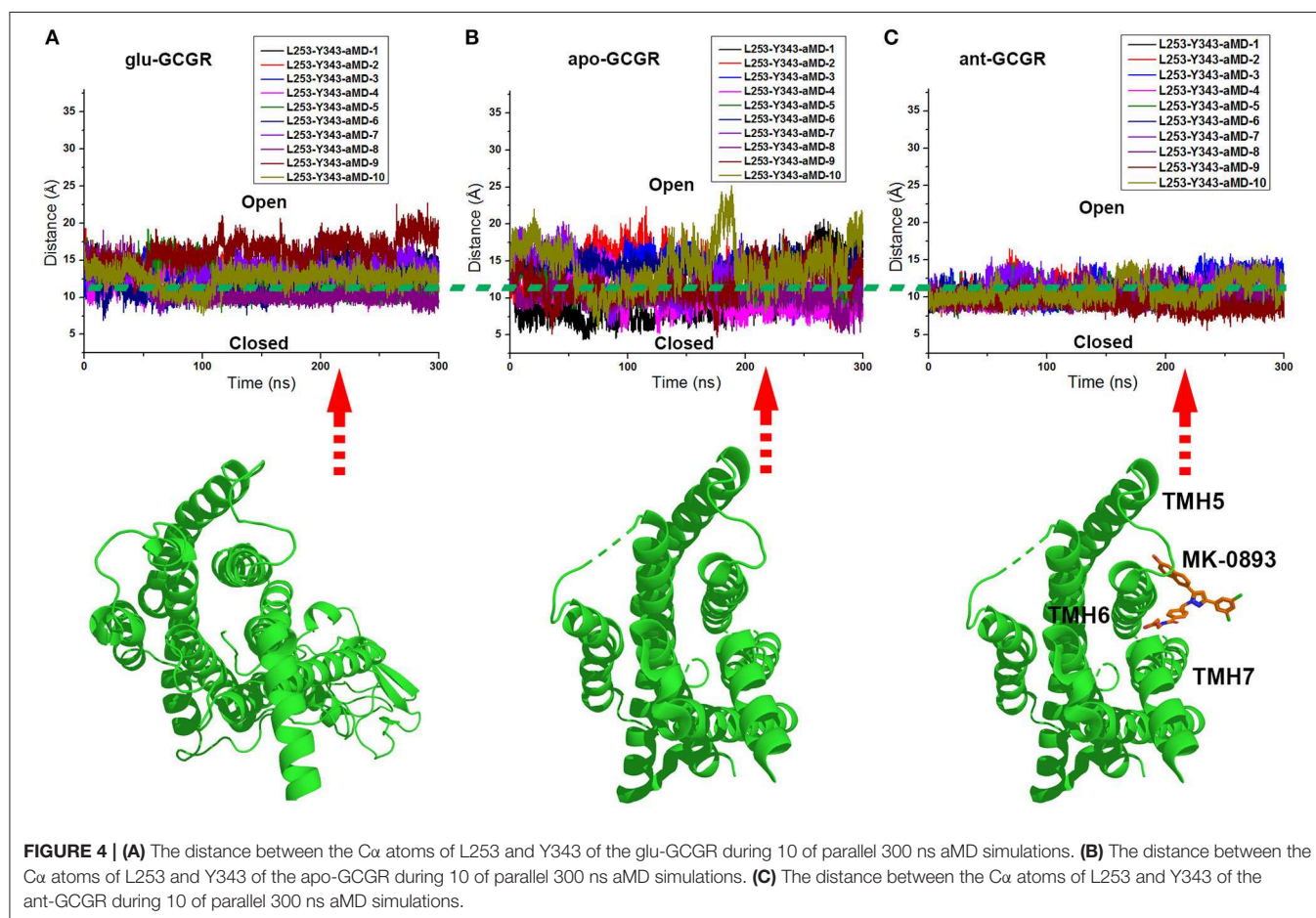
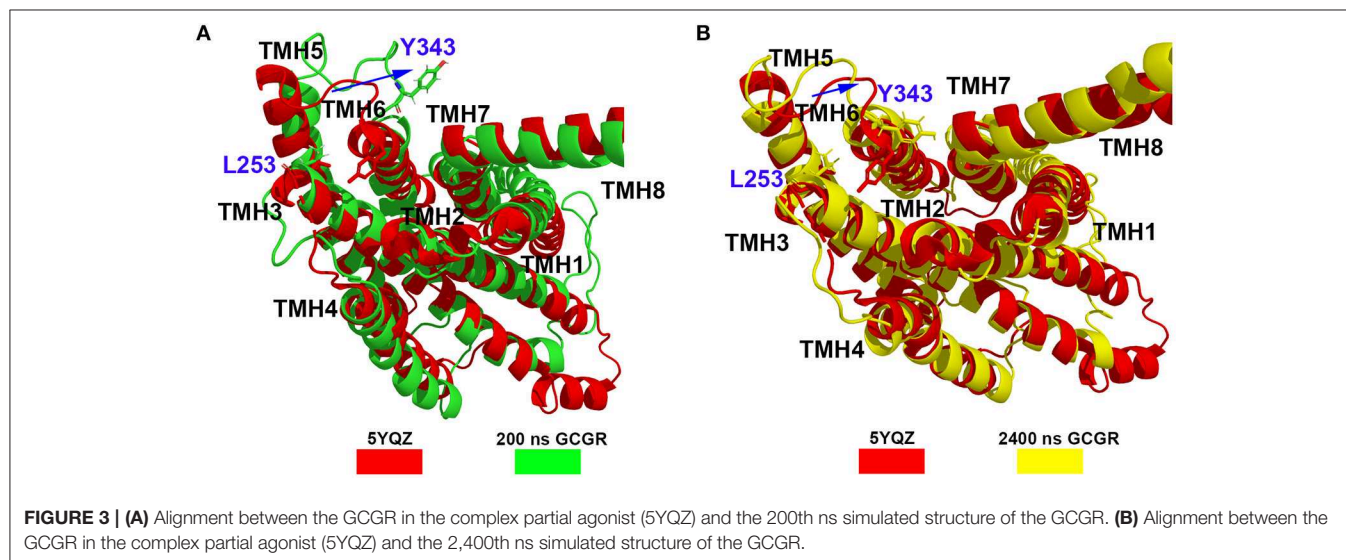
partial agonist-bound GCGR has a similar intracellular structure with the inactive conformation of apo-GCGR (see **Figure 2B**). It indicates the intracellular helices of the GCGR trend to be inactive conformations whether or not the partial agonist binds



to the GCGR. Our cMD simulations show similar results with the crystal structure of the partial agonist-bound GCGR.

Besides, the RxxK motif, which is in the end of TMH6, plays an important role in the activation of class B GPCRs. The RMSD

of backbone atoms of the GCGR RxxK motif also shows a change in ~490 ns MD simulation (see **Figure 2C**). It indicates that the TMH6 changes its conformation. The crystal structure of the GCGR shows that the carboxyl of E245 contains two oxygen



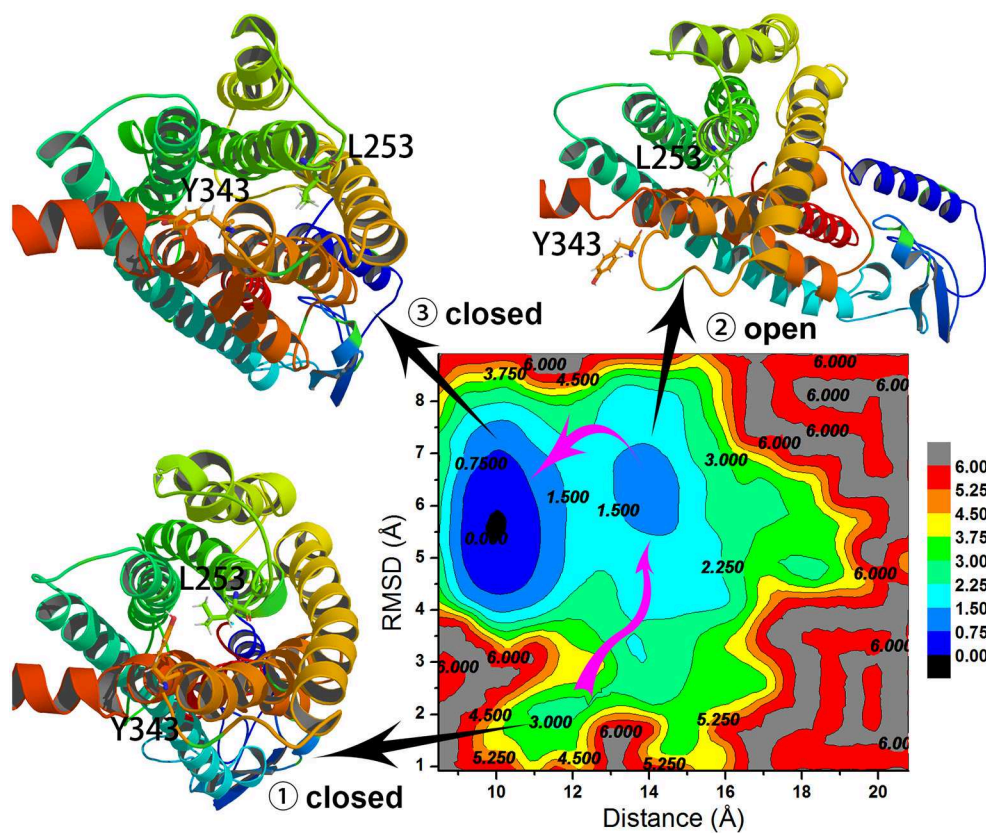
atoms, which are 4.3 and 5.9 Å away from the hydroxyl oxygen of T351, respectively. Due to the rotation of oxygen atoms of carboxyl of E245 in the cMD simulated process, the distance has a fluctuation between the carboxyl oxygen of E245 and hydroxyl oxygen of T351 during 2,500 ns cMD simulations (see **Figure 2D**). The 200th ns conformation, which is in the range of open intracellular helices, and the 2,400th ns conformation, which is in the range of closed intracellular helices, are extracted for comparing with the crystal GCGR (PDB ID: 5YQZ) in complex with the partial agonist (see **Figures 3A,B**). It shows that the intracellular part of TMH6 of the 200th ns GCGR, which contains the Y343, is away from the crystal structure of the GCGR (**Figure 3A**). But the 2,400th ns GCGR has a similar distance between C $\alpha$  atoms of L253 and Y343 with the crystal GCGR. It is only a slight change in the intracellular part of TMH6 by comparing with the crystal GCGR (see **Figure 3B**). These results are consistent with a previous research report (Zhang H. et al., 2018).

### Conformational Change of Intracellular Helices by aMD Simulations

Li et al. (2016) show that the active conformations of glu-GCGR are preponderant during 220 ns aMD simulations. However, our cMD simulated results and the crystal structure of the GCGR in complex with a partial agonist indicate that the intracellular part

of the glu-GCGR is inclined to keep the inactive conformation. These different results are attributed to the aMD simulations. The aMD simulations can reduce the energy barriers to enhance the conformational space sampling. So it infers that the apo-GCGR can also produce the active conformations by aMD simulations if enough aMD simulations are carried out. In this study, the simulated models of the glu-GCGR, apo-GCGR, and ant-GCGR are performed to verify our inference with 10 of parallel 300 ns aMD simulations, respectively.

The x and y axes choose the same range of values for plotting the distance between the C $\alpha$  atoms of L253 and Y343 of the glu-GCGR, apo-GCGR, and ant-GCGR vs. 300 ns aMD simulations. The horizontal dotted line marks about 11 Å, which crosses over the same position of plotted contours of the glu-GCGR, apo-GCGR, and ant-GCGR (see **Figures 4A–C**). As shown in **Figures 4A,B**, some distances between the C $\alpha$  atoms of L253 and Y343 of glu-GCGR and apo-GCGR reach more than 15 Å during 300 ns aMD simulations. These results are different from a previous report that shows the distance between the C $\alpha$  atoms of L253 and Y343 of the apo-GCGR remains around 11 Å during aMD simulations (Li et al., 2016). The main reason is that the aMD simulations are the enhanced stochastic sampling process. Enough number of parallel MD simulations can get more small probability conformations of GCGR. Actually, some of aMD simulations of the apo-GCGR show that the distance between



**FIGURE 5 |** Free energy landscape profiled by the distance between the C $\alpha$  atoms of L253 and Y343 of the GCGR and the RMSD of the GCGR.

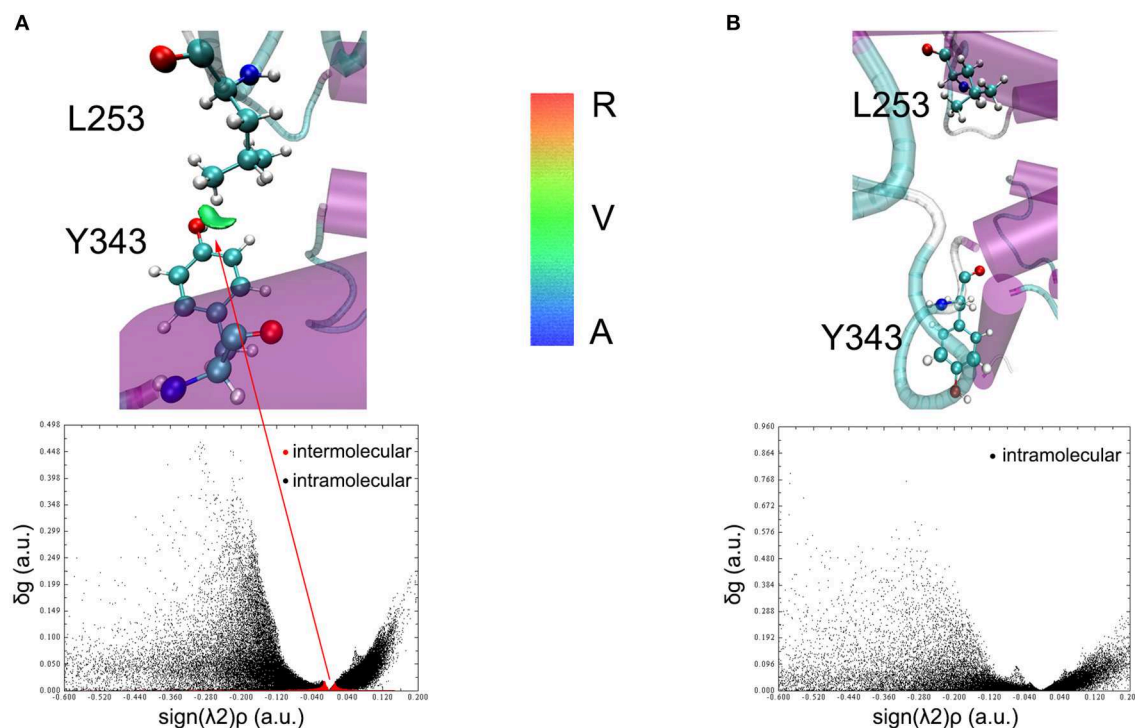
the C $\alpha$  atoms of L253 and Y343 is around 15 Å during 300 ns aMD simulations (see **Figure 4B**). However, it is obvious that the area under the horizontal dotted line in the plotted contour of the glu-GCGR is less than in the plotted contour of the apo-GCGR. Due to the same parameters and the number of aMD simulations, it can infer that the glu-GCGR can generate more active conformations than the apo-GCGR during 300 ns aMD simulations. It indicates that the agonist glucagon can enhance the probability of GCGR active conformations. By compared with the cMD simulated results of the GCGR in complex with an agonist, the aMD simulations show the higher conformational thermodynamic energies of the GCGR in complex with an agonist (see **Figure S1**). Because the antagonist MK-0893 binds to the clefts of TMH5-TMH6 and TMH6-TMH7 (see **Figure 4C**), the distance between the C $\alpha$  atoms of L253 and Y343 of the ant-GCGR is constrained to reach the 15 Å during 300 ns aMD simulations. Most of conformations of the GCGR in complex with the antagonist MK-0893 are kept in the inactive state during 10 of parallel 300 ns aMD simulations.

### Free Energy Landscape Surface Analysis

To understand the closed and open process of the GCGR deeply, the underlying free energy landscape is profiled based on the simulated trajectories. The free energy landscape is drawn by two principal components, which are the distance between the C $\alpha$  atoms of L253 and Y343 of the GCGR and the RMSD of the

GCGR, respectively (see **Figure 5**). To elucidate the function of L253 and Y343 in the closed and open process of the GCGR during MD simulations, the first simulated step frame and the 2,400th ns simulated frame are chosen to calculate the weak interaction between L253 and Y343 of the GCGR, respectively (see **Figures 6A,B**). As shown in **Figure 5**, in the first simulated step, the closed conformation of the GCGR has a high energy barrier of  $\sim 3.0$  kcal mol $^{-1}$  on account of agonist glucagon in the pocket of the GCGR. In this state, the interaction between the residues L253 and Y343 is mainly van der Waals, which is the weak interaction to lock the residues L253 and Y343 of the GCGR (see **Figure 6A**). Then, the conformation of the GCGR opens quickly, which corresponds to  $\sim 1.5$  kcal mol $^{-1}$  energy barrier. In this state, the residues L253 and Y343 of the GCGR are separated from each other. And the van der Waals interaction disappeared between residues L253 and Y343 of the GCGR (see **Figure 6B**). The free energy landscape shows that the open conformation of the GCGR is in the intermediate state. The main reason is that the open intracellular part of the GCGR is not stable and with relative high energy barrier of  $\sim 1.5$  kcal mol $^{-1}$ . Finally, the intracellular conformation of the GCGR falls into the lowest deep energy well (see **Figure 5**). The TMH3 and TMH6 of the GCGR trend to be closed, which correspond to the crystal structure of the GCGR in complex with a partial agonist (Zhang H. et al., 2018).

So far, the differences can be elucidated between the crystal structure of the GCGR in complex with a partial agonist and the



**FIGURE 6 |** The weak interaction analysis between L253 and Y343. **(A)** Gradient isosurfaces between L253 and Y343 in the first simulated step. **(B)** Gradient isosurfaces between L253 and Y343 in the 2,400th ns simulated step. The color bar shows the blue, green, and red scale, which range from  $-0.04$  to  $0.02$  au. Blue scale (A in color bar) represents the strong attractive interaction such as hydrogen bonding. Red scale (R in color bar) represents repulsive interaction such as molecular steric effect. Green scale (V in color bar) indicates van der Waals interactions.



studies of aMD simulations of Li et al. The aMD simulations add boost potential  $\Delta V(r)$  on the original potential  $V(r)$  (see Equation 1). If the extra energy adds to the lowest deep energy well of **Figure 5**, the lowest deep energy well will become the higher deep energy well, which corresponds to the open conformational state of the GCGR. So the aMD simulations can enhance the space samplings of open conformations of the GCGR. In total, if we want to study the drug potency on the targets, an enough number of parallel aMD simulations should be performed on the apo-structure and ligand-bound models. The binding ability of ligands can be reflected by the statistics of pointed features based on the aMD simulated trajectories. The aMD simulations supply a reliable, fast, and precise way for drug discovery.

## CONCLUSIONS

In this study, the 2,500 ns of cMD and 9,000 ns of aMD simulations are performed to elucidate the differences between the crystal structure of the GCGR in complex with a partial agonist and the previous study of aMD simulations (Li et al., 2016). Our results profile the activation mechanism of the GCGR deeply based on a previous study of aMD simulations. It indicates that an enough number of parallel aMD simulations can distinguish the structural characteristics of the apo-GCGR, agonist- and antagonist-bound GCGR. The aMD simulations need less computational time to reach the similar results of cMD simulations. The aMD simulations not only can be used to study

the interaction mechanism between the GCGR and ligands, but also can be further used to estimate the drug potency in other receptor targets.

## DATA AVAILABILITY STATEMENT

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

## AUTHOR CONTRIBUTIONS

QB, ST, HP-S, HF, HL, and XY design the entire experiments. QB, ST, HP-S, HF, and LF performed and analyzed cMD and aMD simulations. The manuscript is prepared with discussions and improvements from all authors.

## FUNDING

The work is supported by the National Natural Science Foundation of China (Grant No. 21605066), Fundamental Research Funds for the Central Universities (Grant No. lzujbky-2018-92) and CERNET Innovation Project (NGII20180704).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2019.00851/full#supplementary-material>

## REFERENCES

- Bagger, J. I., Knop, F. K., Holst, J. J., and Vilsboll, T. (2011). Glucagon antagonism as a potential therapeutic target in type 2 diabetes. *Diabetes Obes. Metab.* 13, 965–971. doi: 10.1111/j.1463-1326.2011.01427.x
- Bai, Q., Perez-Sanchez, H., Zhang, Y., Shao, Y., Shi, D., Liu, H., et al. (2014a). Ligand induced change of beta2 adrenergic receptor from active to inactive conformation and its implication for the closed/open state of the water channel: insight from molecular dynamics simulation, free energy calculation and Markov state model analysis. *Phys. Chem. Chem. Phys.* 16, 15874–15885. doi: 10.1039/C4CP01185F
- Bai, Q., Shi, D., Zhang, Y., Liu, H., and Yao, X. (2014b). Exploration of the antagonist CP-376395 escape pathway for the corticotropin-releasing factor receptor 1 by random acceleration molecular dynamics simulations. *Mol. Biosyst.* 10, 1958–1967. doi: 10.1039/c4mb00037d
- Bai, Q., and Yao, X. (2016). Investigation of allosteric modulation mechanism of metabotropic glutamate receptor 1 by molecular dynamics simulations, free energy and weak interaction analysis. *Sci. Rep.* 6:21763. doi: 10.1038/srep21763
- Bai, Q., Zhang, Y., Li, X., Chen, W., Liu, H., and Yao, X. (2014c). Computational study on the interaction between CCR5 and HIV-1 entry inhibitor maraviroc: insight from accelerated molecular dynamics simulation and free energy calculation. *Phys. Chem. Chem. Phys.* 16, 24332–24338. doi: 10.1039/C4CP03331K
- Cho, Y. M., Merchant, C. E., and Kieffer, T. J. (2012). Targeting the glucagon receptor family for diabetes and obesity therapy. *Pharmacol. Ther.* 135, 247–278. doi: 10.1016/j.pharmthera.2012.05.009
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: an  $N$  [center-dot]  $\log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* 103, 4613–4621. doi: 10.1063/1.470648
- Feng, Z., Hou, T., and Li, Y. (2012). Studies on the interactions between beta2 adrenergic receptor and Gs protein by molecular dynamics simulations. *J. Chem. Inf. Model* 52, 1005–1014. doi: 10.1021/ci200594d
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38, 27–38. doi: 10.1016/0263-7855(96)00018-5
- Jazayeri, A., Dore, A. S., Lamb, D., Krishnamurthy, H., Southall, S. M., Baig, A. H., et al. (2016). Extra-helical binding site of a glucagon receptor antagonist. *Nature* 533, 274–277. doi: 10.1038/nature17414
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Lefebvre, C., Rubez, G., Khartabil, H., Boisson, J. C., Contreras-Garcia, J., and Henon, E. (2017). Accurately extracting the signature of intermolecular interactions present in the NCI plot of the reduced density gradient versus electron density. *Phys. Chem. Chem. Phys.* 19, 17928–17936. doi: 10.1039/C7CP02110K
- Li, Y., Sun, J., Li, D., and Lin, J. (2016). Activation and conformational dynamics of a class B G-protein-coupled glucagon receptor. *Phys. Chem. Chem. Phys.* 18, 12642–12650. doi: 10.1039/C6CP00798H
- Li, Y. Y., Hou, T. J., and Goddard, W. A. 3rd (2010). Computational modeling of structure-function of g protein-coupled receptors with applications for drug design. *Curr. Med. Chem.* 17, 1167–1180. doi: 10.2174/092986710790827807
- Lu, T., and Chen, F. (2012). Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* 33, 580–592. doi: 10.1002/jcc.22885
- Mackerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616. doi: 10.1021/jp973084f

- Miao, Y., Nichols, S. E., Gasper, P. M., Metzger, V. T., and Mccammon, J. A. (2013). Activation and dynamic network of the M2 muscarinic receptor. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10982–10987. doi: 10.1073/pnas.1309755110
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289
- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* 27, 221–234. doi: 10.1007/s10822-013-9644-8
- Siu, F. Y., He, M., De Graaf, C., Han, G. W., Yang, D., Zhang, Z., et al. (2013). Structure of the human glucagon class B G-protein-coupled receptor. *Nature* 499, 444–449. doi: 10.1038/nature12393
- Vanommeslaeghe, K., Raman, E. P., and Mackerell, A. D. Jr. (2012). Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* 52, 3155–3168. doi: 10.1021/ci3003649
- Wang, L., Yuan, Y., Chen, X., Chen, J., Guo, Y., Li, M., et al. (2018). Probing the cooperative mechanism of the mu-delta opioid receptor heterodimer by multiscale simulation. *Phys. Chem. Chem. Phys.* 20, 29969–29982. doi: 10.1039/C8CP06652C
- Yang, L., Yang, D., De Graaf, C., Moeller, A., West, G. M., Dharmarajan, V., et al. (2015). Conformational states of the full-length glucagon receptor. *Nat. Commun.* 6:7859. doi: 10.1038/ncomms8859
- Zhang, H., Qiao, A., Yang, D., Yang, L., Dai, A., De Graaf, C., et al. (2017). Structure of the full-length glucagon class B G-protein-coupled receptor. *Nature* 546, 259–264. doi: 10.2210/pdb5xez/pdb
- Zhang, H., Qiao, A., Yang, L., Van Eps, N., Frederiksen, K. S., Yang, D., et al. (2018). Structure of the glucagon receptor in complex with a glucagon analogue. *Nature* 553, 106–110. doi: 10.1038/nature25153
- Zhang, J., Bai, Q., Perez-Sanchez, H., Shang, S., An, X., and Yao, X. (2019). Investigation of ECD conformational transition mechanism of GLP-1R by molecular dynamics simulations and Markov state model. *Phys. Chem. Chem. Phys.* 21, 8470–8481. doi: 10.1039/C9CP00080A
- Zhang, X., Yuan, Y., Wang, L., Guo, Y., Li, M., Li, C., et al. (2018). Use multiscale simulation to explore the effects of the homodimerizations between different conformation states on the activation and allosteric pathway for the mu-opioid receptor. *Phys. Chem. Chem. Phys.* 20, 13485–13496. doi: 10.1039/C8CP02016G
- Zhang, Y., Sun, B., Feng, D., Hu, H., Chu, M., Qu, Q., et al. (2017). Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature* 546, 248–253. doi: 10.1038/nature22394

**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.

Copyright © 2019 Bai, Tan, Pérez-Sánchez, Feng, Feng, Liu and Yao. 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.





# Identification of Drug-Disease Associations Using Information of Molecular Structures and Clinical Symptoms via Deep Convolutional Neural Network

Zhanchao Li<sup>1,2\*</sup>, Qixing Huang<sup>1</sup>, Xingyu Chen<sup>1</sup>, Yang Wang<sup>3</sup>, Jinlong Li<sup>1</sup>, Yun Xie<sup>1</sup>, Zong Dai<sup>3</sup> and Xiaoyong Zou<sup>3\*</sup>

<sup>1</sup> School of Chemistry and Chemical Engineering, Guangdong Pharmaceutical University, Guangzhou, China, <sup>2</sup> School of Chemistry, Sun Yat-Sen University, Guangzhou, China, <sup>3</sup> Key Laboratory of Digital Quality Evaluation of Chinese Materia Medica of State Administration of Traditional Chinese Medicine, Guangzhou, China

## OPEN ACCESS

### Edited by:

Xiao Jun Yao,  
Macau University of Science and  
Technology, Macau

### Reviewed by:

Tingjun Hou,  
Zhejiang University, China  
Xuemei Pu,  
Sichuan University, China

### \*Correspondence:

Zhanchao Li  
zhanchao8052@gdpu.edu.cn  
Xiaoyong Zou  
ceszxy@mail.sysu.edu.cn

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 29 July 2019

**Accepted:** 18 December 2019

**Published:** 10 January 2020

### Citation:

Li Z, Huang Q, Chen X, Wang Y, Li J,  
Xie Y, Dai Z and Zou X (2020)  
Identification of Drug-Disease  
Associations Using Information of  
Molecular Structures and Clinical  
Symptoms via Deep Convolutional  
Neural Network. *Front. Chem.* 7:924.  
doi: 10.3389/fchem.2019.00924

Identifying drug-disease associations is helpful for not only predicting new drug indications and recognizing lead compounds, but also preventing, diagnosing, treating diseases. Traditional experimental methods are time consuming, laborious and expensive. Therefore, it is urgent to develop computational method for predicting potential drug-disease associations on a large scale. Herein, a novel method was proposed to identify drug-disease associations based on the deep learning technique. Molecular structure and clinical symptom information were used to characterize drugs and diseases. Then, a novel two-dimensional matrix was constructed and mapped to a gray-scale image for representing drug-disease association. Finally, deep convolution neural network was introduced to build model for identifying potential drug-disease associations. The performance of current method was evaluated based on the training set and test set, and accuracies of 89.90 and 86.51% were obtained. Prediction ability for recognizing new drug indications, lead compounds and true drug-disease associations was also investigated and verified by performing various experiments. Additionally, 3,620,516 potential drug-disease associations were identified and some of them were further validated through docking modeling. It is anticipated that the proposed method may be a powerful large scale virtual screening tool for drug research and development. The source code of MATLAB is freely available on request from the authors.

**Keywords:** convolutional neural network, deep learning, drug-disease associations, fingerprint, symptoms

## INTRODUCTION

Traditional drug development usually follows this paradigm of one drug, one gene, one disease, which is an expensive and time-consuming process with stunningly high failure rate. By conservative estimates, it takes about 15 years and \$0.8–1.5 billion to bring a drug to market (Dudley et al., 2011; Yu et al., 2015), and during the development stage, almost 90% of the small molecules cannot pass the Phase I clinical trial and finally be eliminated (Wu et al., 2019). On the other hand, disease burden is increasing globally due to the growth of population, outbreak

of infectious disease and emergence of antibiotic resistance (Shameer et al., 2018). In order to circumvent this dilemma, drug repositioning has become a promising alternative strategy for drug research and development (Wu et al., 2017).

Drug repositioning, also known as drug repurposing, drug reprofiling, and drug redirecting, which aims to find potential new indication for existing drug and apply the drug to the treatment of disease other than the drug's originally intended disease (Luo et al., 2016). It offers a possible way to greatly save time and cost, especially improve success rate, because of existing pharmacological and toxicological properties, as well as safety on known drug (Wang et al., 2014; Sun et al., 2017). However, successful drug repurposing stories are rare and rather random events (Sun et al., 2017). Well-known examples are sildenafil (trade name Viagra) and minoxidil, which were originally used to treat angina and hypertension. At present, they have been repurposed for the treatment of erectile dysfunction and hair loss, due to accidental discovery with a bit of luck (Bovac, 2013; Varothai and Bergfeld, 2014; Wu et al., 2019). Weakness of drug reprofiling is that it relies mainly on prior knowledge and clinical trials (Park et al., 2017), which is unfeasible in general and much too expensive to be applied on a large scale (Sun et al., 2017). For example, there are 2,593 approved small molecule drugs in DrugBank (Wishart et al., 2018) and 19,941 disease entries in MalaCards (Rappaport et al., 2017), resulting in more than 50 million of drug-disease combinations. Undoubtedly, it is almost impossible to effectively validate all possible associations of drug-disease through laboratory works and clinical trials. Therefore, it is urgent to develop *in silico* drug redirecting approaches for discovering new indications for approved drugs on a large scale.

Fortunately, with the accumulation of drug and disease related data, as well as the development of machine learning, numerous theoretical methods have been proposed to find new indications of drugs by identifying potential drug-disease associations. These computational approaches can be roughly divided into three mainstreams: drug-based, disease-based, and network-based. The former two are according to the assumption that drugs having similar structures/properties are inclined to be associated with diseases having similar pathogenesis/symptoms, and vice versa (Liu et al., 2016; Wu et al., 2017; Shameer et al., 2018). For example, Gottlieb et al. (2011) utilized multiple drug-drug and disease-disease similarity measures for the prediction of drug repurposing using the logistic regression classifier. Zhu and Zhu (2015) introduced a method to identify repositioned drug for breast cancer by integrating the breast cancer survival data with the drug sensitivity information. By integrating information of drug chemical substructure, target domain and annotation, a novel method was presented to predict drug-disease associations based on the Laplacian regularized sparse subspace learning (Liang et al., 2017). Based on the drug features and disease semantic information, Zhang et al. (2018a) proposed a similarity constrained matrix factorization method for the prediction of drug-disease associations. Khalid and Sezerman (2018) combined the biological pathways, binding site structural similarities, disease-disease similarity with logistic regression classifier to predict approved and novel drug-disease associations. Wang et al. (2013) trained a support vector machine model

to identify potential drug-disease interactions by integrating molecular structure, molecular activity and phenotype data. A support vector machine model was also built by Moghadam et al. (2016) to recognize novel drug indications through adopting kernel fusion technique and various features of drug and disease. By considering information of drug chemical structures, drug targets and gene expression patterns, Napolitano et al. (2013) also developed a support vector machine classifier to predict novel drug-disease associations.

The last one is based on the principle of "guilt-by-association" that drugs treating with same disease share structure/network properties and the diseases treated with the same drug also share phenotype/network properties (Wu et al., 2017). For instance, Zhao and Li (2012) defined a network-based gene closeness profile to relate drug to disease. Then, a Bayesian partition method was utilized to elucidate drug-disease associations by identifying drug-gene-disease co-modules. Huang et al. (2013) combined three different networks with edge weights of drug, genomic and disease phenotype, and developed network propagation approach to infer the drug-disease associations. Oh et al. (2014) constructed an integrative genetic network including protein-protein interaction network and gene regulatory network. Then, the distance between topology drug-module and disease-module were adopted as features for the prediction of novel drug-disease associations based on the random forest algorithm. A causal network connecting drug-target-pathway-gene-disease was built by Yang et al. (2014) who calculated the association scores between drugs and diseases by evaluating a drug's effects on multiple targets and pathway. Finally, probabilistic matrix factorization models were learned to identify therapeutic associations. Based on the propagation flow algorithm, Martinez et al. (2015) developed DrugNet for prioritization of drug-disease relationships through a network of interconnected drugs, proteins and diseases. A novel methodology was proposed by Yu et al. (2015) to discover the drug-disease associations by constructing heterogeneous network consisting of drugs, protein complexes and diseases. By building a heterogeneous network including drug-drug similarity network, disease-disease similarity network and known drug-disease association network, Liu et al. (2016) proposed a two-pass random walks with restart to predict new indications for approved drugs. Yu et al. (2016) represented a cluster method for prediction of new drug indications by using the identified drugs and disease modules based on the constructed drug network and disease network. Wu et al. (2017) constructed a novel weighted drug-disease pair network, where a node is a drug-disease pair and a weighted edge represents the node-node relation. Then, a semi-supervised graph cut algorithm was adopted to identify the potential drug-disease treatment interactions. Drug-disease associations were formulated as a bipartite network, Zhang et al. (2018b) presented the network topological similarity-based inference method to predict unobserved drug-disease associations based on the linear neighborhood similarity. Wu et al. (2019) introduced a method to detect drug-disease treatment relations by using drug-disease, drug-protein and disease-protein interaction data based on the random forest algorithm. By considering network similarities of drugs and

diseases, Cui et al. (2019) proposed a novel method to predict drug-disease interactions based on the Gaussian interaction profile kernels and  $L_{2,1}$ -norm.

Despite progresses in the past decade on identification of drug-disease associations, accurate prediction of treatment relations is still far from satisfaction. The first one recognized limitation, especially for the drug-based and disease-based methods, is the lack of a uniform or universal definition for calculating similarity, resulting in dramatical change of similarity score from one method to another. The second one, especially for network-based methods, is only fit for drugs or diseases included in the built network or dataset. Hence, some methods usually fail to discover the novel drugs and new indications. The third one is the restriction of only semantic similarity for assessing disease similarity based on the human phenotype ontology (Groza et al., 2015) and disease ontology (Bello et al., 2018). However, precise semantic relationships are not often captured (Zhu et al., 2013). Others or new attributes, such as symptoms (i.e., clinical manifestations information) may be adopted to characterize the disease, because symptoms are the most directly observable characteristics of a disease and the basis of clinical disease classification (Zhou et al., 2014), as well as diseases with similar symptoms usually share common genetic mechanisms (Xu et al., 2013). In addition, deep learning has been widely used in various research fields as a modern machine learning technique (Bai et al., 2019; Li et al., 2019; Steuer et al., 2019). However, its effectiveness for drug-disease associations prediction has not been evaluated.

In view of these reasons, a novel computational method was developed to identify drug-disease associations based on the information of molecular structures and clinical symptoms through deep learning method. Instead of using the information of drug related side effects, activity, target protein and their interactions, as well as disease-related human phenotype ontology, disease ontology, gene ontology and disease genes, the chemical fingerprints and disease symptoms were only utilized for enhancing the generalization ability. A novel two-dimensional matrix was constructed to characterize drug-disease association by considering the information of drug and disease, simultaneously. Finally, deep convolutional neural network was employed to construct model for identifying potential drug-disease associations.

## MATERIALS AND METHODS

### Collection of Drug-Disease Associations

In order to construct a comprehensive and high-quality dataset of drug-disease associations, firstly, we downloaded the information of drug-disease associations contained in the file CTD\_chemiclas\_diseases.tsv from the Comparative Toxicogenomics Database (CTD, Ver. Feb, 2017) (Davis et al., 2019), which is a robust, publicly available database and provides manually curated information about chemical, gene, protein, disease and their relationships. Secondly, removed drug-disease pairs without annotation “therapeutic” in the field of Direct Evidence and with annotation “drug combination” in the field of Chemical Name, meaning that the obtained drugs itself have therapeutic effects on diseases, rather than exert

functions by combining with other drugs. Thirdly, deleted drug-disease associations in which drugs had no information of CID numbers and SMILES (canonical simplified molecular input line entry system) strings in the PubChem database (Kim et al., 2016). Fourthly, canceled drug-disease pairs in which diseases were un-included in the work of Zhou et al. (2014). Finally, 26,521 drug-disease associations containing 4,501 drugs and 2,093 diseases were obtained (**Supporting Information 1**). These retrieved drug-disease associations were considered as positive examples.

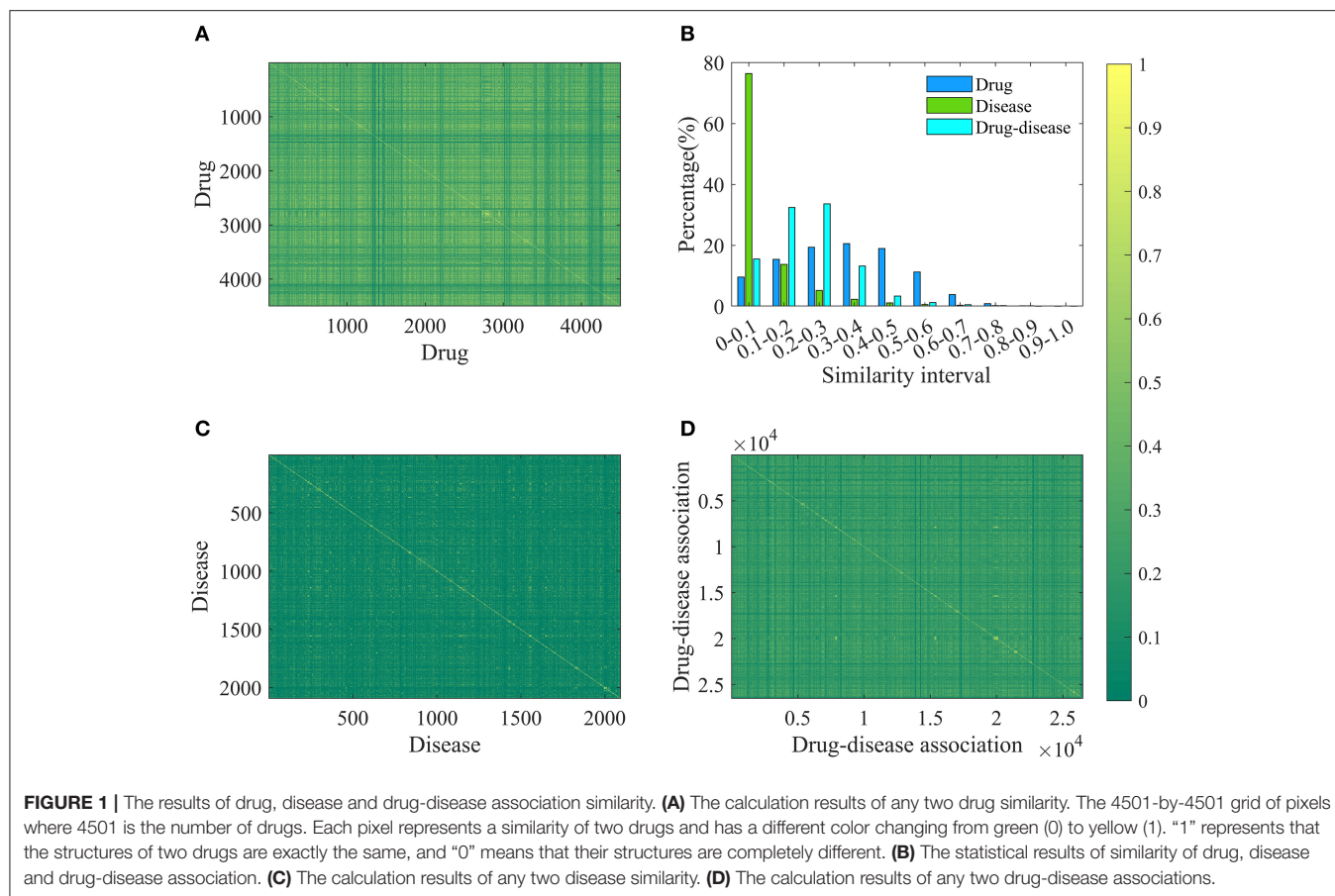
The goal of current research is to identify potential therapeutic relationships from tremendous combinations between drugs and diseases based on deep learning method. This is a binary classification problem, therefore, it is necessary to build negative examples (i.e., drug-disease non-association pairs). Unfortunately, there is no database dedicated to collecting drugs without treatment relationships for diseases due to lack of research and application value. Consequently, we had to use the following strategy to produced negative samples: (1) Randomly selected drug and disease from positive samples to form new drug-disease association pair. (2) Eliminated the new association pair if it existed in the downloaded file CTD\_chemiclas\_diseases.tsv, otherwise, considered it as a negative sample. (3) Repeated steps (1) and (2), until the number of negative samples equals the number of positive samples.

Finally, a benchmark dataset with equal size of true drug-disease associations and false drug-disease associations was established. The “1:1” ratio can overcome the limitation of a larger number of negative examples and lead to unbiased prediction.

### Characterization of Drug-Disease Associations

In order to increase the applicability of the current method, the Pubchem molecular fingerprint descriptor was calculated to characterize drug molecule by using the information of SMILES format and PaDEL-descriptor software (Yap, 2011). This fingerprint descriptor is a binary feature vector with 881 dimensions, in which every element corresponds to one specific chemical substructure and is encoded as either 1 or 0 to show clearly whether the substructure is contained in the drug molecular. For simplicity, the molecular fingerprint descriptor of drug  $i$  is represented by  $F_{i,n}$  ( $n = 1, 2, \dots, 881$ ). The molecular fingerprint is a simple but effective descriptor in the wide use of quantitative structure-activity relationship (Banerjee and Preissner, 2018; Zheng et al., 2019). Based on the fingerprint descriptor, the similarity of any two drug molecules was evaluated by calculating the Jaccard similarity coefficient (Levandowsky and Winter, 1971; Fuxman Bass et al., 2013; Li et al., 2016). Similarity values and statistical results are shown in **Figures 1A,B**. Clearly, the similarity values are in the range of 0–0.9956, about 10, 15, 19, 20, 18, and 11% are located in the range of [0–0.1], [0.1–0.2], [0.2–0.3], [0.3–0.4], [0.4–0.5], and [0.5–0.6], suggesting that drug molecular structures are very diverse and complex.

For each disease, the symptom information was retrieved from the human symptoms-disease network (Zhou et al.,



2014), resulting in 105,892 connections between 2,093 diseases and 322 symptom terms. The associations between diseases and symptoms were acquired based on the co-occurrence of disease terms and symptom terms in the MeSH metadata field of PubMed and quantified using the term frequency-inverse document frequency. Therefore, a disease can be characterized by a feature vector with 322 dimensions, in which every element corresponding to one specific symptom and is encoded as a value larger than or equal to zero to explain the strength of the association between disease and symptom. For convenience of description, the symptomatic feature of disease  $j$  is characterized by  $D_j, m$  ( $m = 1, 2, \dots, 322$ ). This representation is reasonable and based on this fact that many symptoms are not always present for a disease and occur with varying frequency. For any two diseases, we calculated the cosine value of the included angle between corresponding two symptom feature vectors to assess disease diversity. The statistical results and cosine values were shown in **Figures 1B,C**. It is clear that most (about 76%) of the cosine values are lower than 0.1, revealing that diseases in the benchmark dataset belong to various categories.

Different from previous studies, a novel gray-scale image method was proposed to characterize drug-disease relationships by considering both drug and disease properties. For drug  $i$  and disease  $j$ , a two-dimensional matrix  $FD_{ij}$  with  $881 \times 322$  was constructed. Value of element located in the  $n$ -th row and  $m$ -th

column was calculated according to the following Equation (1):

$$FD_{ij}(n, m) = F_i(n) + D_j(m) \quad (1)$$

Then, the matrix was mapped to a gray-scale image to characterize the relationship between drug  $i$  and disease  $j$ . The rationality of the method is rooted in the paradigm of “structure determines function” and the fact that the clinical manifestation of disease is symptoms, which are widely used in disease diagnosis, treatment and classification research. Therefore, we utilize molecular fingerprint descriptors to characterize drug chemical structure and provide information on its functions, as well as adopt symptom features to represent disease and provide information on its pathological mechanisms, respectively. The introduced method is helpful to elucidate the relationship between drug and disease at the level of molecular structure and clinical phenotype.

For any two drug-disease associations  $Dr_1-Di_1$  and  $Dr_2-Di_2$ , their similarity is defined by the following Equation (2):

$$Sim_{Dr_1-Di_1, Dr_2-Di_2} = \frac{Jac(Dr_1, Dr_2) + Cos(Di_1, Di_2)}{2} \quad (2)$$

Where,  $Jac(Dr_1, Dr_2)$  and  $Cos(Di_1, Di_2)$  mean the Jaccard similarity coefficient and cosine value, respectively. According to the definition, the similarity value of two drug-disease pairs



is always located in the range of 0 and 1, and the higher value means the more similar. The statistical results and similarity values between any two drug-disease associations were shown in **Figures 1B,D**. Clearly, about 98% of the similarity values are in the range of 0–0.5, revealing that the benchmark dataset is complex and contains various drug-disease associations.

## Construction and Assessment of Model

The goal of this work is to identify whether an unknown drug-disease association has a therapeutic relationship or not, which is a two-class classification problem. Hence, deep convolution neural network was utilized to discriminate potential drug-disease associations owing to the success in image recognition and biomedicine (Esteva et al., 2017; Pelt and Sethian, 2018; Sullivan et al., 2018). The architecture and parameters of deep convolution neural network were optimized based on experience, and listed in **Table 1**. In addition, we used the optimizer of stochastic gradient descent with momentum 0.9. Initial learning rate was 0.01 and reduced the learning rate by a factor of 0.1 every 10 epochs. Maximum number of epochs for training was set to 50 and used a mini-batch with 128 observations at each iteration. The default values were used for all other parameters and the program was executed based on the MATLAB software.

In order to evaluate the performance of current method, 20,000 positive and negative samples were randomly chosen from the benchmark dataset to construct a training set, and the remaining positive and negative samples were used to build a test set. In addition to accuracy (AC), sensitivity (SE), specificity (SP), precision (PR) and Matthew's correlation coefficient (MCC), we also utilize receiver operating characteristic curve (ROC), precision recall (RE) curve (PRC) and corresponding area (ROCA and PRCA) to estimate the predictive ability of the model.

**TABLE 1** | The architecture and parameters of deep convolution neural network.

Layer	Size
Image input	881 × 322
Convolutional	32 filters with 5 × 5, stride 2 × 2
ReLU	–
Max pooling	2 × 2, stride 2 × 2
Convolutional	64 filters with 5 × 5, stride 2 × 2
ReLU	–
Max pooling	2 × 2, stride 2 × 2
Convolutional	128 filters with 5 × 5, stride 2 × 2
ReLU	–
Max pooling	2 × 2, stride 2 × 2
Fully connected	500, dropout = 0.5
Fully connected	500, dropout = 0.5
Fully connected	500, dropout = 0.5
Fully connected	500, dropout = 0.5
Fully connected	2
Softmax	–
Classification	2

Flowchart of the current method is shown in **Figure 2**, and detailed steps were described as follows:

- Step 1. Retrieved drug-disease associations from the CTD database.
- Step 2. Obtained the SMILES strings of drug molecules from the PubChem compound database and symptoms of diseases from the human symptoms-disease network, respectively.
- Step 3. Produced molecular fingerprint descriptors and disease symptom features to characterize drugs and diseases, respectively.
- Step 4. Generated two-dimensional matrixes and map it to gray-scale images to characterize drug-disease associations.
- Step 5. Divided the benchmark dataset into training set and test set to build model and evaluate performance, respectively.

## RESULTS AND DISCUSSION

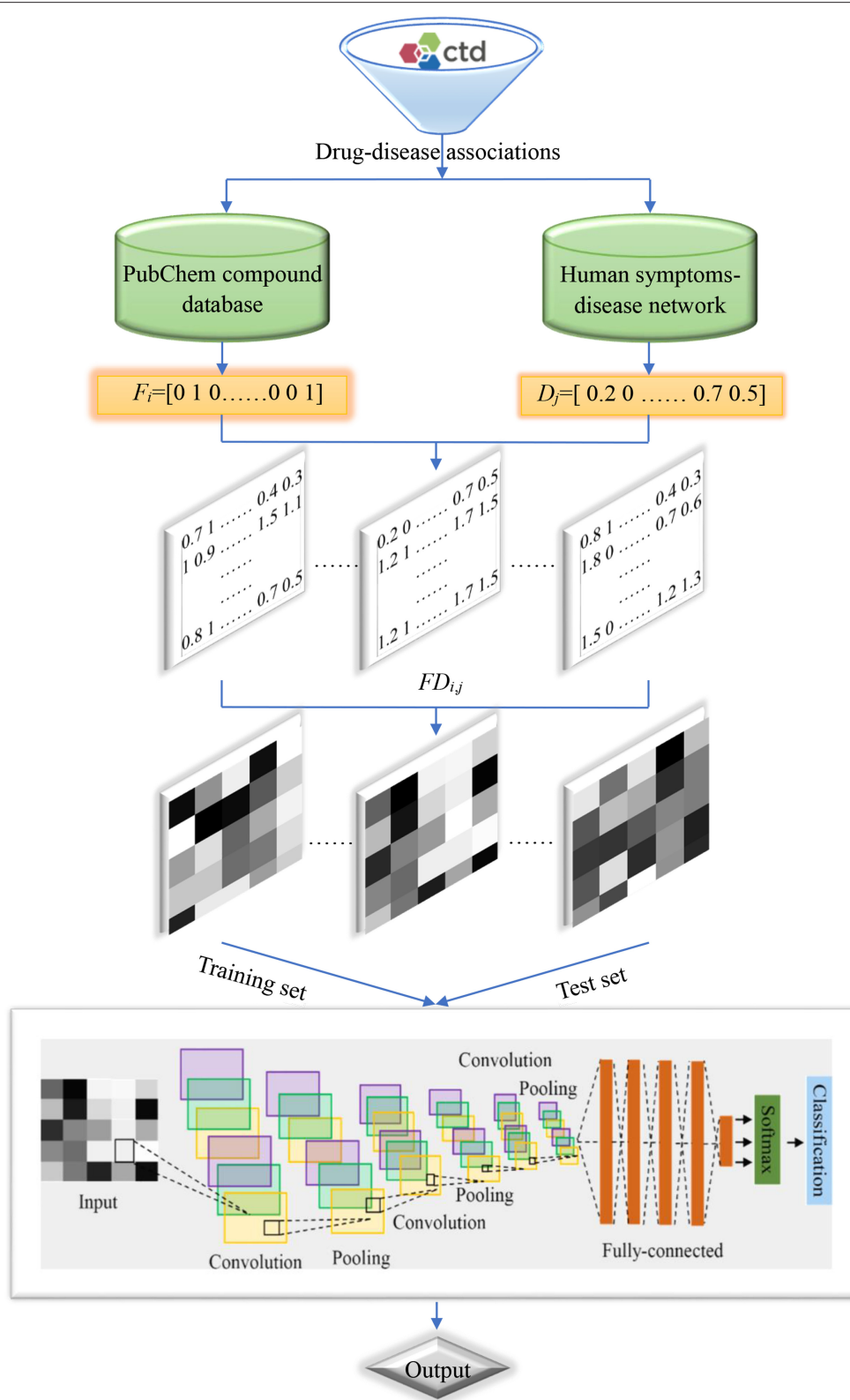
### Performance Evaluation of Current Method

In order to evaluate the performance for the negative sample random generation method, parallel experiments are performed 10 times for generating the negative samples, building the model and evaluating the performance. The statistical results of AC, SE, SP, PR, and MCC, as well as ROC and PRC derived from the training set and test set are shown in **Figure 3** and listed in **Table 2**, respectively. For training set, average values of AC, SE, SP, PR, MCC, ROCA, and PRCA are 89.90, 88.96, 90.85, 90.67%, 0.7982, 0.9637 and 0.9651, with the relative standard deviations 0.30, 0.44, 0.16, 0.19, 0.66, 0.19, and 0.19%. For test set, average values and the corresponding relative standard deviations are 86.51 and 0.21%, 86.23 and 0.36%, 86.79 and 0.19%, 86.72 and 0.17%, 0.7302 and 0.50%, 0.9360 and 0.14%, 0.9352 and 0.17%, respectively. The AC, SE, SP, and PR from the training set and test set are higher than 85%. Meanwhile, the relative standard deviations are lower than 1%. These results reveal that the developed method can effectively capture information of drug-disease associations, and also has a strong robustness for generating negative samples and an outstanding ability to identify drug-disease associations.

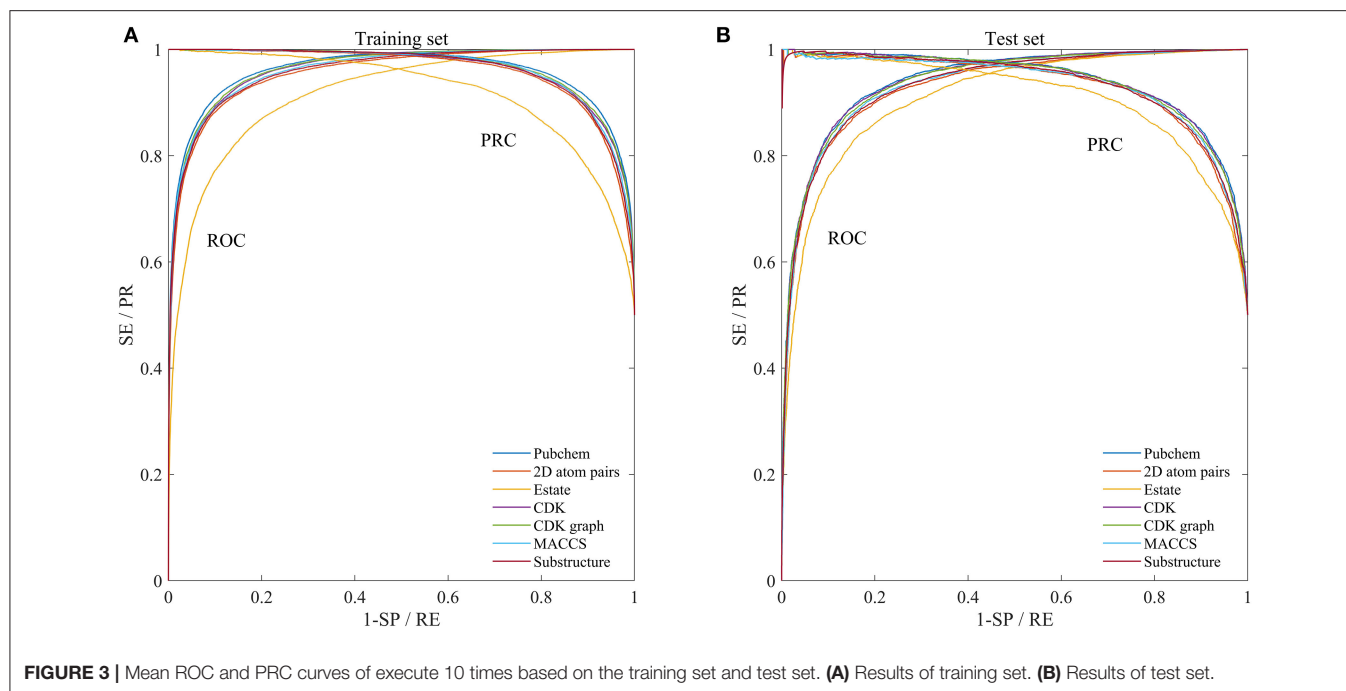
### Comparison of Molecular Fingerprint Descriptors

In addition to the Pubchem fingerprint descriptor, we also calculated six kinds of fingerprint descriptors such as 2D atom pairs, Estate, CDK, CDK graph, MACCS, and Substructure (their detailed description can refer to the help file of PaDEL-descriptor). Then the model was constructed and evaluated based on the benchmark dataset. The statistical results, ROC and PRC were illustrated in **Figure 3** and listed in **Table 2**, respectively.

We can see that the Estate descriptor achieve the lowest average AC, SE, SP, PR, MCC, AUCR, and AUCP for both the training set and test set, which may be caused by the fact that the descriptor has only a 79-dimensional feature vector and cannot adequately describe the molecular structure information. For 2D atom pairs, CDK, CDK graph, MACCS and Substructure, AC from the training set and the test set are about 89 and 86%, about 0.9 and 0.5% lower than



**FIGURE 2 |** Flowchart of the current method.



**FIGURE 3 |** Mean ROC and PRC curves of execute 10 times based on the training set and test set. **(A)** Results of training set. **(B)** Results of test set.

those of Pubchem, respectively. For AC from the training set and test, statistical hypothesis tests including Wilcoxon rank sum test and two-sample Kolmogorov-Smirnov test between Pubchem and other descriptors were performed, and the corresponding results were listed in **Table 3**. For Wilcoxon rank sum test, most of the  $p$ -values are  $<1.8 \times 10^{-4}$ , only  $p$ -value between the Pubchem and the CDK graph derived from test set is  $1.309 \times 10^{-3}$ . All  $p$ -values indicate that significant differences existed in the AC from Pubchem and other six descriptors. For two-sample Kolmogorov-Smirnov test, the lowest and highest are  $3.286 \times 10^{-5}$  and  $9.050 \times 10^{-3}$ , respectively. All these  $p$ -values show significant differences. Therefore, Pubchem molecular fingerprint descriptor is the optimal feature for characterizing molecular structure in current research.

### Proportion of Positive and Negative Samples

In order to overcome the problem of classification hyperplane skewness, a very common phenomenon in the field of machine learning, the ratio between positive and negative samples was set to 1:1. In fact, the number of negative samples is much larger than that of positive samples for identifying drug-disease associations. To assess the effect of positive and negative sample ratios on the performance of current method, we constructed a series of datasets in which the ratio was set to 1:2, 1:3, ..., 1:10. Then, 3/4 of the positive and negative samples were randomly choose as the training set for building model, and the remaining positive and negative samples were considered as the test set for evaluating performance. The whole process was repeated five times, and the statistical mean results were display in **Figure 4**. For convenience

of comparison, the statistical results in section of performance evaluation of current method also exhibited in **Figure 4**.

As shown in **Figures 4A,C**, average values of AC and SP increase slowly as the ratio changes from 1:1 to 1:10 for training sets and test sets. However, average values of SE and PR are slowly decreasing. For **Figures 4B,D**, we can see that average values of MCC and AUCP are also slowly decreasing as the ratios increase. The average values of AUCR fluctuate within a very small range. This result indicates that as the ratio improves, the number of negative samples in the training set dramatically increases and provide more negative sample information for training model, which makes the model easier to identify negative samples, but more difficult to identify positive samples. Although AC takes into account the prediction results of positive and negative samples simultaneously, its value is mainly determined by the prediction result of negative samples. Therefore, average values of AC improve as the ratios increase. On the contrary, average values of MCC and AUCP decrease. We also note that AUCR is insensitive to the ratio between positive and negative samples in the current research. Hence, it is reasonable to set the ratio of positive and negative samples to 1:1, which can ensure the model has high sensitivity, because the aim of our research is to identify potential drug-disease associations.

### Identification Power of New Indications for Existing Drugs

Finding new indications for marketable drugs can help pharmaceutical companies reduce costs and time. Our approach ability for drug repositioning was further estimated through generating new training set and test set based on the step-by-step strategy: (1) Randomly selected a positive sample (i.e., drug-disease association  $Dr_1-Di_1$ ) to enter the training set. (2) Chose

**TABLE 2 |** The statistical results of performed 10 times based on the training set and independent test set (Rsd: relative standard deviation).

Dataset	Fingerprint	AC(%)/ Rsd(%)	SE(%)/ Rsd(%)	SP(%)/ Rsd(%)	PR(%)/ Rsd(%)	MCC/ Rsd(%)	AUCR/ Rsd(%)	AUCP/ Rsd(%)
Training set	Pubchem	89.90/ 0.29	88.96/ 0.44	90.85/ 0.16	90.67/ 0.19	0.7982/ 0.66	0.9637/ 0.19	0.9651/ 0.19
	2D atom pairs	88.73/ 0.18	87.36/ 0.23	90.10/ 0.17	89.82/ 0.17	0.7749/ 0.42	0.9531/ 0.08	0.9560/ 0.09
	Estate	83.51/ 0.27	80.22/ 0.40	86.79/ 0.18	85.86/ 0.21	0.6716/ 0.66	0.9106/ 0.25	0.9161/ 0.22
	CDK	88.66/ 0.78	87.97/ 1.0	89.37/ 0.58	89.22/ 0.63	0.7733/ 1.8	0.9557/ 0.59	0.9571/ 0.57
	CDK graph	89.28/ 0.78	88.55/ 0.33	90.02/ 0.23	89.87/ 0.24	0.7858/ 0.63	0.9606/ 0.18	0.9617/ 0.20
	MACCS	88.46/ 0.75	86.77/ 0.99	90.15/ 0.54	89.80/ 0.60	0.7696/ 1.7	0.9531/ 0.50	0.9520/ 0.43
	Substructure	88.83/ 0.10	87.38/ 0.14	90.28/ 0.07	89.99/ 0.08	0.7769/ 0.66	0.9558/ 0.06	0.9590/ 0.06
	Pubchem	86.51/ 0.21	86.23/ 0.36	86.79/ 0.19	86.72/ 0.18	0.7302/ 0.50	0.9360/ 0.14	0.9352/ 0.16
Test set	2D atom pairs	85.57/ 0.10	84.57/ 0.21	86.58/ 0.16	86.31/ 0.13	0.7116/ 0.24	0.9257/ 0.08	0.9261/ 0.09
	Estate	83.04/ 0.32	80.07/ 0.62	86.00/ 0.10	85.12/ 0.15	0.6619/ 0.79	0.9041/ 0.23	0.9057/ 0.21
	CDK	86.02/ 0.40	85.69/ 0.76	86.35/ 0.31	86.26/ 0.29	0.7204/ 0.96	0.9309/ 0.39	0.9304/ 0.22
	CDK graph	86.09/ 0.22	85.75/ 0.34	86.43/ 0.18	86.34/ 0.18	0.7218/ 0.53	0.9330/ 0.15	0.9326/ 0.22
	MACCS	85.17/ 0.38	83.92/ 0.79	86.43/ 0.35	86.08/ 0.29	0.7037/ 0.50	0.9217/ 0.25	0.9172/ 0.24
	Substructure	85.60/ 0.16	84.74/ 0.25	86.46/ 0.29	86.22/ 0.25	0.7121/ 0.38	0.9278/ 0.08	0.9283/ 0.10
	Pubchem	86.51/ 0.21	86.23/ 0.36	86.79/ 0.19	86.72/ 0.18	0.7302/ 0.50	0.9360/ 0.14	0.9352/ 0.16
	2D atom pairs	85.57/ 0.10	84.57/ 0.21	86.58/ 0.16	86.31/ 0.13	0.7116/ 0.24	0.9257/ 0.08	0.9261/ 0.09

**TABLE 3 |** The *p* values of hypothesis tests between Pubchem and other molecular fingerprint descriptors based on the AC.

	Pubchem	
	Wilcoxon rank sum (Train/Test)	Kolmogorov-Smirnov (Train/Test)
2D atom pairs	1.827e-04/	1.888e-05/
	1.817e-04	1.888e-05
Estate	1.827e-04/	1.888e-05/
	1.817e-04	1.888e-05
CDK	2.165e-05/	3.286e-05/
	7.361e-04	9.050e-03
CDK graph	3.791e-04/	1.216e-03/
	1.309e-03	6.899e-03
MACCS	1.827e-04/	1.888e-05/
	1.827e-04	1.888e-05
Substructure	1.827e-04/	1.888e-05/
	1.827e-04	1.888e-05

all positive samples including disease Di<sub>1</sub> into the training set. (3) Repeated the steps 1 and 2, until the number of positive samples chosen reached 3/4 of all positive samples. The remaining 1/4 was entered into the test set. (4) Randomly selected a negative sample (i.e., drug-disease non-association pair NDr<sub>1</sub>-NDi<sub>1</sub>) into

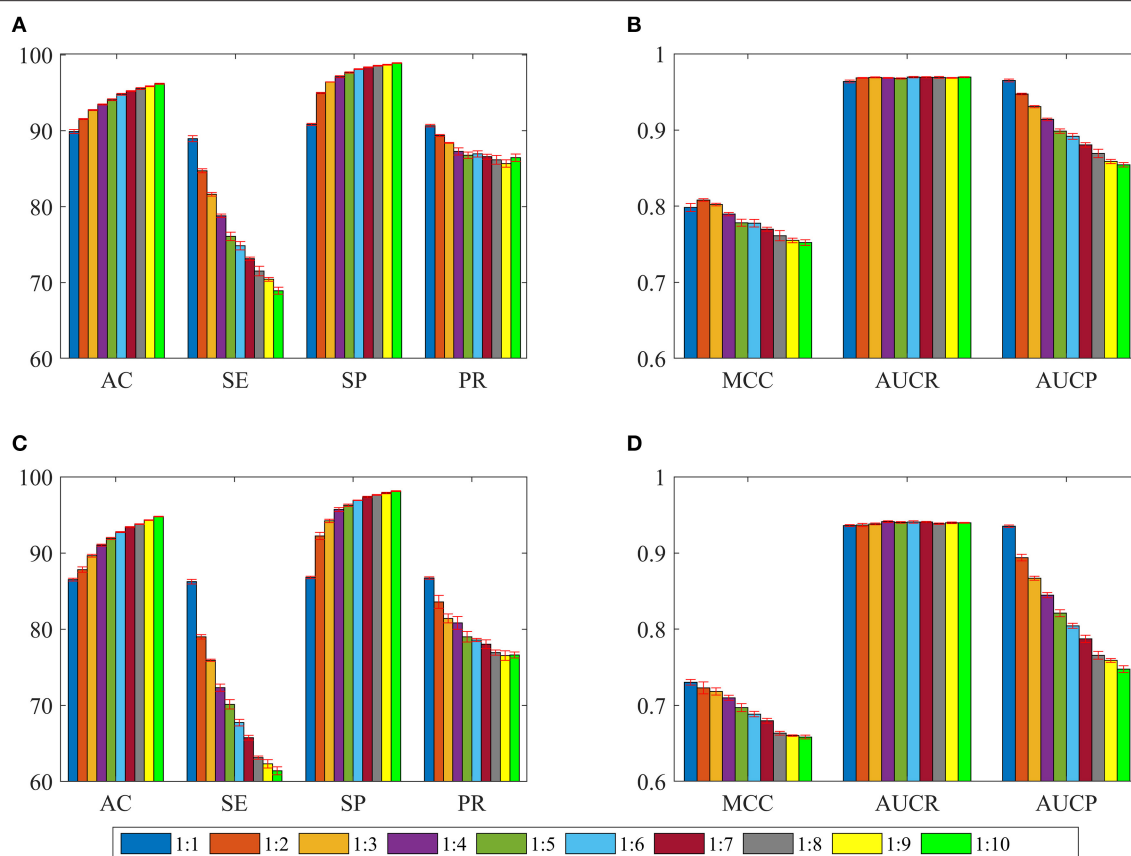
the training set. (5) All negative samples containing disease NDi<sub>1</sub> were also entered into the training set. (6) Repeated steps 4 and 5 until the number of negative samples selected achieved 3/4 of all negative samples. The remaining 1/4 was contained in the test set.

Based on the strategy, a disease is either involved in the training set or in the test set, which can guarantee disease information in the test set not existing in the training set. All of the above steps were repeated 10 times, 10 new training sets and corresponding test sets were then generated, and their prediction results were illustrated in **Figures 5A,B**.

For training sets, AC, SE, SP, and PR are located in the range of [87.82–88.47%], [86.39–87.64%], [88.52–89.65%], [88.39–89.39%], respectively. The MCC, AUCR, and AUCP change from 0.7567 to 0.7694, 0.9493 to 0.9547, 0.9512 to 0.9665, respectively. The corresponding relative standard deviations are 0.26, 0.42, 0.37, 0.33, 0.60, 0.18, and 0.20%, respectively. The average values are 88.14, 87.18, 89.10, 88.89%, 0.7630, 0.9519, and 0.9539, only about 1.76, 1.78, 1.75, 1.78%, 0.035, 0.012, and 0.011 lower than those of the training set derived from the benchmark dataset with the Pubchem descriptor (listed in **Table 2**).

For test sets, the minimum and maximum values of AC, SE, SP, PR, MCC, AUCR, and AUCP are 74.46 and 77.60%, 65.50 and 71.40%, 82.72 and 85.22%, 89.30 and 82.31%, 0.4960 and 0.5568, 0.8147 and 0.8486, 0.8279 and 0.8541, respectively. The





**FIGURE 4 |** The statistical average results of various dataset, in which the ratios between positive and negative samples are 1:1, 1:2, 1:3, ..., 1:10. The panels in (A–D) represent the mean values of AC, SE, SP, PR, MCC, AUCR, and AUCP, respectively. The red vertical bars indicate the standard deviations. (A) The statistical results of AC, SE, SP, and PR from training sets. (B) The statistical results of AUCR and AUCP from training sets. (C) The statistical results of AC, SE, SP, and PR derived from independent test sets. (D) The statistical results of AUCR and AUCP derived from independent test sets.

corresponding relative standard deviations are 1.4, 2.7, 1.1, 1.2, 3.8, 1.4, and 1.1%, respectively. The average values are 76.28, 68.60, 83.97, 81.05%, 0.5320, 0.8340 and 0.8414, only about 10.23, 17.63, 2.82, 5.67%, 0.1982, 0.1290, and 0.0938 lower than those of the test set from the benchmark dataset with the Pubchem descriptor (listed in Table 2).

These results uncover that our method still obtains high predictive accuracy even though both training test and test sets are constructed rigorously, indicating that it has ability for identifying new drug indications.

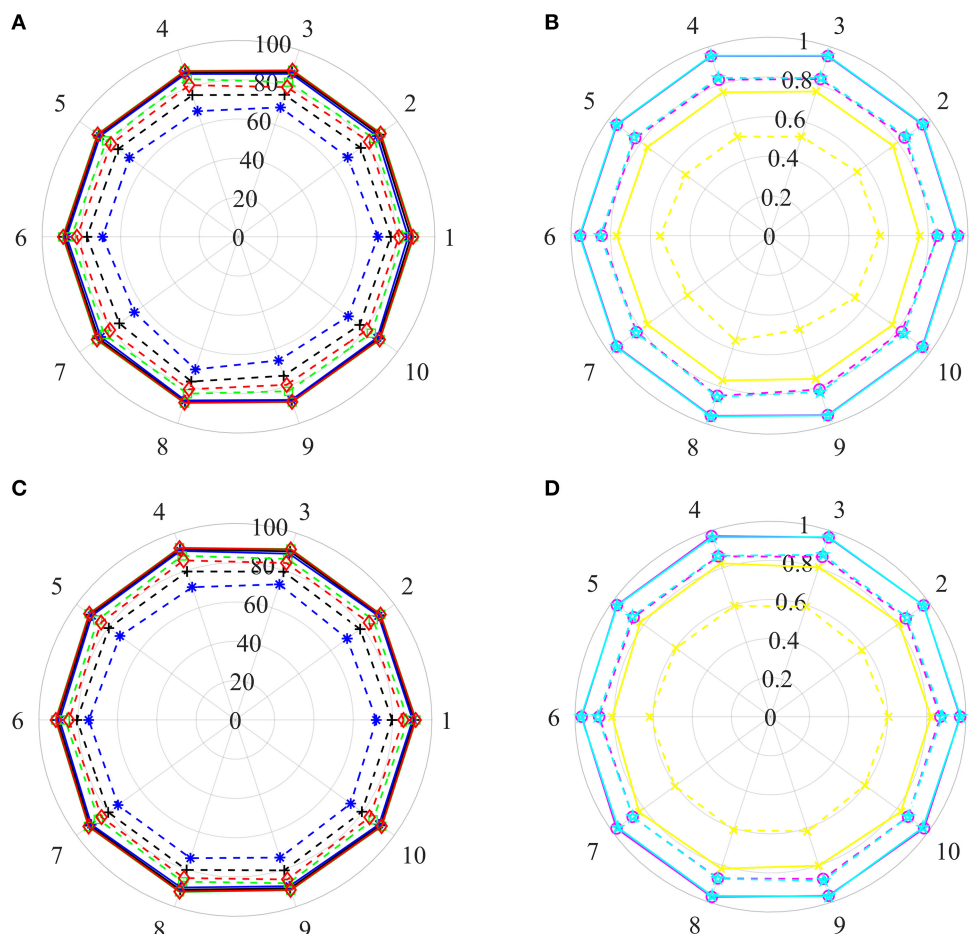
## Recognition Ability of Potential Drug Molecules

Pharmaceutical companies are more interested in which drug or compound is effective on a new disease, i.e., whether this novel disease is associated with known or potential drug molecule. To this end, we appraise the performance of our method for identifying potential drug molecules or lead compounds by generating a series of training test and test sets base on the step-by-step strategy mentioned above. In Steps 2 and 5, selected all positive examples including drug Dr<sub>1</sub> and all negative

examples comprising drug ND<sub>r1</sub> into the training set and test set, respectively. The process was executed 10 times, 10 training sets and test sets were then obtained, and their prediction results were shown in Figures 5C,D.

For training sets, values of Acc, Sen, Spe, Pre, MCC, AUCR, and AUCP fluctuate from 90.08 to 91.50%, 88.91 to 90.63%, 91.16 to 92.37%, 90.96 to 92.23%, 0.8018 to 0.8301, 0.9638 to 0.9730, 0.9648 to 0.9725, respectively. The corresponding relative standard deviations and average values are 0.54 and 90.85%, 0.68 and 89.91%, 0.44 and 91.80%, 0.46 and 91.64%, 1.2% and 0.8172, 0.33% and 0.9691 and 0.25% and 0.9671, respectively. These results are very close to the results of the training set derived from the benchmark dataset with the Pubchem descriptor (listed in Table 2).

For test sets, values of Acc, Sen, Spe, Pre, MCC, AUCR, and AUCP are in the range of [78.51–80.61%], [70.34–74.61%], [86.14–87.84%], [83.96–85.51%], [0.5780, 0.6171], [0.8556, 0.8764] and [0.8642, 0.8829], respectively. the corresponding relative standard deviations and average values are 0.79 and 79.77%, 1.9 and 72.69%, 0.72 and 86.84%, 0.67 and 84.68%, 2.0 and 0.6015, 0.74% and 0.8675, 0.72% and 0.8732. Although the average Acc of test sets is about 10% lower than that of training



**FIGURE 5 |** The prediction results of various training sets and independent test sets based on the step-by-step strategy. In polar coordinate system, radius is 100 (A,C) and 1 (B,D), respectively. The solid and dashed lines represent the training set and the independent test set, respectively. Black+plus sign, blue+asterisk, green+ tetragonum, red+diamond, yellow+cross, magenta+circle and cyan+ pentagon mean the AC, SE, SP, PR, MCC, AUCR, and AUCP, respectively. (A,C) The results of AC, SE, SP, and PR. (B,D) The results of MCC, AUCR, and AUCP.

sets, it is reasonable because the drug information in the test sets are excluded from the training sets. Therefore, these results suggest that current method can recognize candidate drugs or lead compounds with a high prediction accuracy.

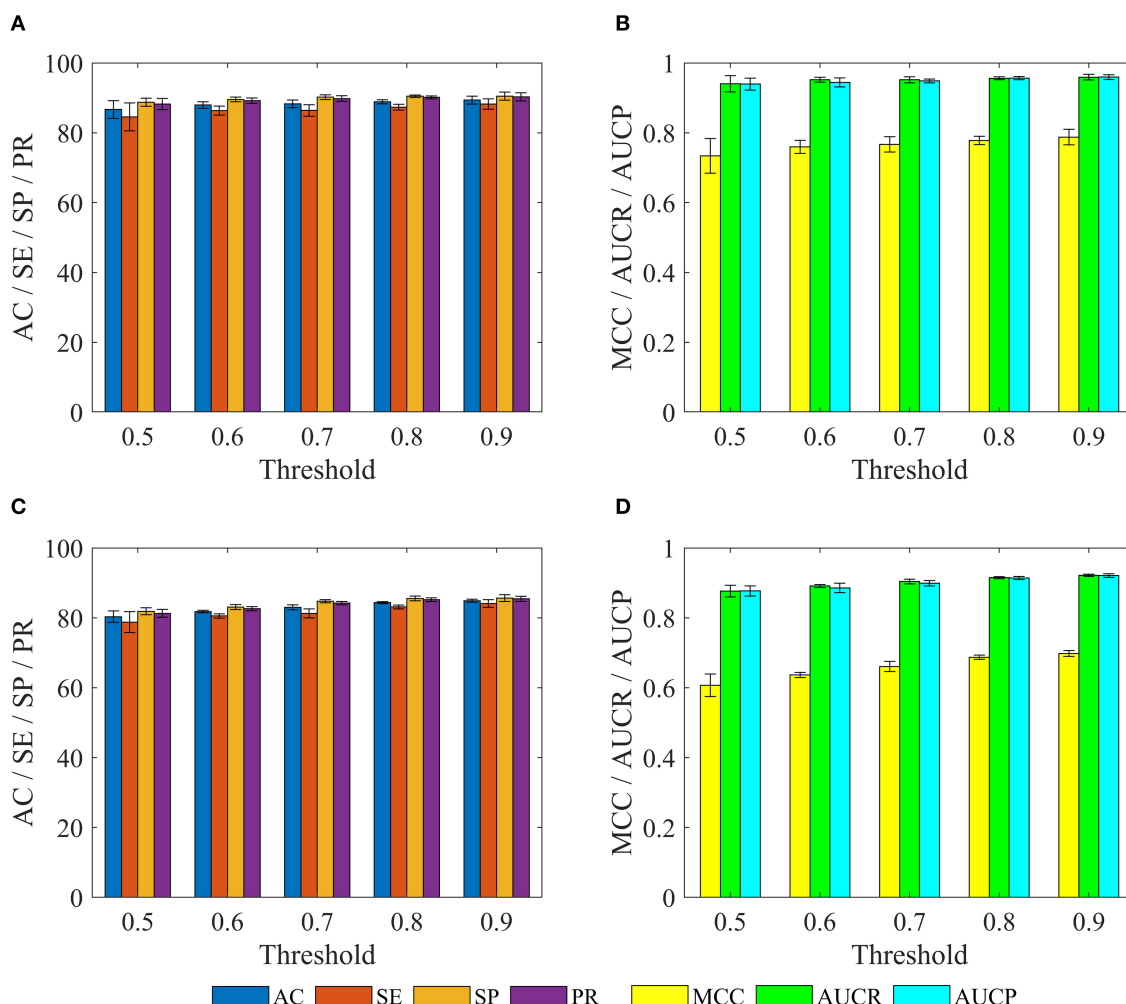
## Prediction Capability of Potential Drug-Disease Associations

We further investigated performance for recognizing potential drug-disease associations by constructing a series of non-redundant benchmark dataset. Here, a non-redundant drug-disease associations database was constructed by randomly winnowing those association pairs that have more than a given threshold (i.e., similarity) to other pairs presented in the benchmark dataset. Then, 3/4 of positive and negative examples were adopted as training set to train model, and the remaining examples were utilized as test set to evaluate performance. The similarity threshold was set to 0.5, 0.6,...,0.9, and the construction of non-redundant dataset was repeated 10 times for each threshold, respectively. Note that thresholds of 0.1, 0.2,

0.3, and 0.4 were not employed, because the number of samples in the non-redundant dataset was too small to be statistically significant. The statistical results of training sets and test sets based on the various thresholds were shown in **Figure 6**.

For training set, we can observe that with the decrease of the threshold from 0.9 to 0.5, average values of AC, SE, SP, PR, MCC, AUCR, and AUCP are also gradually reduced from 89.37 to 86.65%, 88.22 to 84.56%, 90.51 to 88.74%, 90.29 to 88.21%, 0.7876 to 0.7339, 0.9592 to 0.9403, 0.9595 to 0.9396. The current method still obtains average accuracy of 86.65%, even if the threshold is reduced to 0.5.

For test set, average values of AC, SE, SP, and PR decrease gradually with the decrease of threshold from 0.9 to 0.5, but these values are consistently higher than 80% for threshold from 0.9 to 0.6. Meanwhile, average values of MCC, AUCR, and AUCP are always higher than 0.63, 0.89, and 0.88, respectively. Even when the threshold is decreased to 0.5, the current method still achieves the average AC of 80.33%, SE of 78.78%, SP of 81.88%, PR of 81.29%, Mcc of 0.6071, AUCR of 0.8766, AUCP of 0.8769.



**FIGURE 6 |** The statistical results of various training sets and independent test sets based on the different cutoffs. The panels in (A–D) indicate the mean values of AC, SE, SP, PR, MCC, AUCR, and AUCP, respectively. The black vertical bars indicate the standard deviations. (A,B) for training sets. (C,D) for independent test sets.

These results demonstrate that the proposed method still achieves state-of-the-art performance, and has ability to identify novel drug-disease associations.

### Identify Ability of Independent Test Set

After evaluated the performance of the proposed method for identifying new indications, potential drugs and novel drug-disease associations, we further assessed the true predictive power based on the independent test set, which were generated by collecting drug-disease associations information from the CTD database (Ver. Jun, 2019). The independent test set contains 1,323 drug-disease associations (Supporting Information 2), which are not included in the benchmark dataset. In the new associations, drugs and diseases in 377 and 38 drug-disease pairs did not appear in the benchmark dataset, respectively.

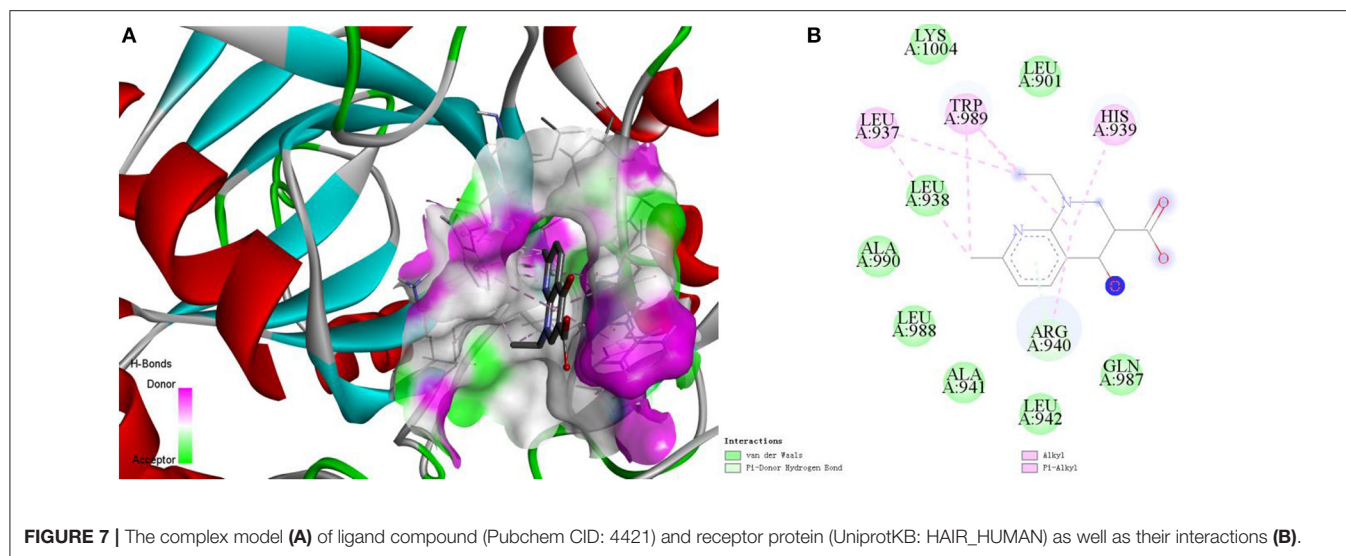
The final identification model was built based on the all drug-disease associations in the benchmark dataset, and then used it to predictive the new drug-disease associations in the independent test set. We find that 973 drug-disease associations

were correctly identified, and the prediction accuracy was 73.54% (973/1,323). For the 377 and 38 associations, 271 and 27 were correctly recognized, accuracy was 71.88% (271/377) and 71.05% (27/38), respectively.

These results reveal that our approach still achieves more than 70% prediction accuracy for these new drug-disease associations, indicating the reliability of the method.

### Large-Scale Prediction of Drug-Disease Associations

We further conducted a comprehensive and large-scale prediction for unknown drug-disease associations by using the final model. In order to generate the unknown associations, we firstly downloaded the information on structure and physicochemical properties of compounds/drugs from the DrugBank dataset. Secondly, deleted those compounds/drugs according to the Lipinski's rule of five (i.e., molecular mass <500 daltons, <5 hydrogen bond donors and 10 hydrogen bond acceptors, octanol-water partition coefficient  $\log P < 5$ ). Thirdly,



randomly combined the compounds/drugs collected from the DrugBank and diseases involved in the benchmark dataset. Finally, 24,266,646 unknown associations were generated. The final model identified 3,620,516 potential associations. We rank all the potential associations according to the probability in descending order, and the results show that 294,354 associations (**Supporting Information 3**) are the most likely to be putative associations because their probability values are higher than 0.99.

Here, we take the one recognized associations as examples to illustrate the practical application of current method. Alopecia, also known as hair loss or baldness, refers to partial or complete loss of hair from part of the head or body. It usually can be classified into four types: male-pattern hair loss, female-pattern hair loss, alopecia areata and telogen effluvium, and the corresponding cause is genetics and male hormones, unclear, autoimmune, physically or psychologically stressful event (Vary, 2015). Although medications minoxidil, finasteride, and dutasteride have been used to treat hair loss, they have limited effects and can only prevent further baldness without regenerating lost hair (Rogers and Avram, 2008; Banka et al., 2013). The current method identified a possible association between the disease and compound nalidixic acid (Pubchem CID: 4421). The compound is a synthetic quinolone and composed of 1,8-naphthyridin-4-one substituted by carboxylic acid, ethyl and methyl groups at positions 3, 1, and 7, respectively. Some studies have shown that lysine-specific demethylase hairless (UniprotKB: HAIR\_HUMAN) is a protein related with hair loss (Klein et al., 2002; Liu et al., 2014). Docking simulations between the compound and the protein was executed by using the AutoDock (Santos-Martins et al., 2014) program and DS visualizer software. The 3-dimensional structural information of protein was downloaded from the SWISS-MODEL Repository. The Lamarckian genetic algorithm was adopted to search the docking conformation. Finally, the optimized docking model with binding energy  $-6.26$  kcal/mol and inhibition constant ( $K_i$ )  $25.66 \mu\text{M}$  was obtained. Complex

model of protein and compound as well as their interactions were displayed in **Figures 7A,B**.

We can observe that Var der Waals interactions exist between the compound and amino acid residues Leu901, Leu938, Ala941, Leu942, Gln987, Leu988, Ala990, and Lys1004. The small molecule is connected to the protein through  $\pi$ -donor hydrogen bond between six-membered ring and residue Arg940. Moreover, there are some hydrophobic interactions such as  $\pi$ -alkyl interactions between six-membered ring and residues His939, Arg940, and Trp989, alkyl interactions between the compound and residues Leu937 and Trp989. Therefore, we can assume that the compound may be used as a drug to treat hair loss through interacting with the target protein, which is worthy of further experimental verification.

## CONCLUSIONS

In this study, clinical manifestations information and molecule fingerprint descriptor are utilized to characterize disease and drug, respectively. A novel two-dimensional matrix is constructed and then map it into a gray-scale image to characterize drug-disease association. Deep convolutional neural network is introduced to construct model to identify potential drug-disease associations. The performance of current method is evaluated by building the benchmark dataset, and the optimal molecule fingerprint descriptor is determined by comparing with other various descriptors. In addition, the prediction ability of our method for identifying new drug indications, lead compounds, potential and true drug-disease associations has also been validated through a series of experiments. Finally, the practical application capability has been demonstrated by molecular simulation experiments. Our work gives a new insight for study of drug-disease associations at the level of disease clinical symptom and drug molecule structure. It is anticipated that the proposed method may be a powerful tool for new drug research and development.



## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://ctdbase.org/>.

## AUTHOR CONTRIBUTIONS

ZL collected the data, experimented, and drafted the manuscript. QH provides experimental ideas and revised manuscripts. XC, YW, JL, YX, and ZD participated in the discussion of this work. ZL and XZ also participated in the discussion and revision of the manuscript. All authors read, commented, and approved the final manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (21675035; 21675180; 21775169),

the Scientific Technology Project of Guangdong Province (2015A030401033; 2016B010108007; 2016A010103039; 2017B020221001), Special Innovation Project of University in Guangdong Province (2018KTSCX109), Special Funds of Key Disciplines Construction from Guangdong and Zhongshan Cooperating, and the Scientific Technology Project of Guangzhou City (201604020145).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2019.00924/full#supplementary-material>

**Supporting Information 1** | The drug-disease associations retrieved from the CTD database.

**Supporting Information 2** | The independent test set.

**Supporting Information 3** | The most likely drug-disease associations.

## REFERENCES

- Bai, F., Hong, D., Lu, Y., Liu, H., Xu, C., and Yao, X. (2019). Prediction of the antioxidant response elements' response of compound by deep learning. *Front. Chem.* 7:385. doi: 10.3389/fchem.2019.00385
- Banerjee, P., and Preissner, R. (2018). BitterSweetForest: a random forest based binary classifier to predict bitterness and sweetness of chemical compounds. *Front. Chem.* 6:93. doi: 10.3389/fchem.2018.00093
- Banka, N., Bunagan, M. J., and Shapiro, J. (2013). Pattern hair loss in men: diagnosis and medical treatment. *Dermatol. Clin.* 31, 129–140. doi: 10.1016/j.det.2012.08.003
- Bello, S. M., Shimoyama, M., Mitaka, E., Lauderkind, S. J. F., Smith, C. L., Eppig, J. T., et al. (2018). Disease ontology: improving and unifying disease annotations across species. *Dis. Model. Mech.* 11:dmm032839. doi: 10.1242/dmm.032839
- Bovav, N. (2013). Challenges and opportunities of drug repositioning. *Trends Pharmacol. Sci.* 34, 267–272. doi: 10.1016/j.tips.2013.03.004
- Cui, Z., Gao, Y. L., Liu, J. X., Wang, J., Shang, J., and Dai, L. Y. (2019). The computational prediction of drug-disease interactions using the dual-network L<sub>2,1</sub>-CMF method. *BMC Bioinformatics* 20:5. doi: 10.1186/s12859-018-2575-6
- Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., McMorran, R., Wieggers, J., et al. (2019). The comparative toxicogenomics database: update 2019. 47, D948–D954. doi: 10.1093/nar/gky868
- Dudley, J. T., Deshpande, T., and Butte, A. J. (2011). Exploiting drug-disease relationships for computational drug repositioning. *Brief Bioinform.* 12, 303–311. doi: 10.1093/bib/bbr013
- Esteve, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. doi: 10.1038/nature21056
- Fuxman Bass, J. I., Diallo, A., Nelson, J., Soto, J. M., Myers, C. L., and Walhout, A. J. (2013). Using networks to measure similarity between genes: association index selection. *Nat. Methods* 10, 1169–1179. doi: 10.1038/nmeth.2728
- Gottlieb, A., Stein, G. Y., Ruppert, E., and Sharan, R. (2011). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* 7:496. doi: 10.1038/msb.2011.26
- Groza, T., Kohler, S., Moldenhauer, D., Vasilevsky, N., Baynam, G., Zemojtel, T., et al. (2015). The human phenotype ontology: semantic unification of common and rare disease. *Am. J. Hum. Genet.* 97, 111–124. doi: 10.1016/j.ajhg.2015.05.020
- Huang, Y. F., Yeh, H. Y., and Soo, V. W. (2013). Inferring drug-disease associations from integration of chemical, genomic and phenotype data using network propagation. *BMC Med. Genomics* 6:S4. doi: 10.1186/1755-8794-6-S3-S4
- Khalid, Z., and Sezerman, O. U. (2018). Computational drug repurposing to predict approved and novel drug-disease associations. *J. Mol. Graph. Model.* 85, 91–96. doi: 10.1016/j.jmgm.2018.08.005
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213. doi: 10.1093/nar/gkv951
- Klein, I., Bergman, R., Indelman, M., and Sprecher, E. (2002). A novel missense mutation affecting the human hairless thyroid receptor interacting domain 2 causes congenital atrichia. *J. Invest. Dermatol.* 119, 920–922. doi: 10.1046/j.1523-1747.2002.00268.x
- Levandosky, M., and Winter, D. (1971). Distance between sets. *Nature* 234, 34–35. doi: 10.1038/234034a0
- Li, F., Wan, X., Xing, J., Tan, X., Li, X., Wang, Y., et al. (2019). Deep neural network classifier for virtual screening inhibitors of (S)-adenosyl-l-methionine (SAM)-dependent methyltransferase family. *Front. Chem.* 7:324. doi: 10.3389/fchem.2019.00324
- Li, J., Sun, Y., Johnson, R. J., Sciaky, D., Wei, C. H., Leaman, R., et al. (2016). BioCreative V CDR task corpus: a resource for chemical disease relation extraction. *Database* 2016:baw068. doi: 10.1093/database/baw068
- Liang, X., Zhang, P., Yan, L., Fu, Y., Peng, F., Qu, L., et al. (2017). LRSSL: predict and interpret drug-disease associations based on data integration using sparse subspace learning. *Bioinformatics* 33, 1187–1196. doi: 10.1093/bioinformatics/btw770
- Liu, H., Song, Y., Guan, J., Luo, L., and Zhuang, Z. (2016). Inferring new indications for approved drugs via random walk on drug-disease heterogeneous networks. *BMC Bioinformatics* 17:539. doi: 10.1186/s12859-016-1336-7
- Liu, L., Lim, H., Casta, A., Kobayashi, Y., Shapiro, L. S., and Christiano, A. M. (2014). Hairless is a histone H3K9 demethylase. *FASEB J.* 28, 1534–1542. doi: 10.1096/fj.13-237677
- 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, 2664–2671. doi: 10.1093/bioinformatics/btw228
- Martinez, 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, 41–49. doi: 10.1016/j.artmed.2014.11.003
- Moghadam, H., Rahgozar, M., and Gharaghani, S. (2016). Scoring multiple features to predict drug disease associations using information fusion and aggregation. *SAR QSAR Environ. Res.* 27, 609–328. doi: 10.1080/1062936X.2016.1209241
- 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. Cheminform.* 5:30. doi: 10.1186/1758-2946-5-30

- 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:e111668. doi: 10.1371/journal.pone.0111668
- Park, S., Lee, D. G., and Shin, H. (2017). Network mirroring for drug repositioning. *BMC Med. Inform. Decis. Mak.* 17:55. doi: 10.1186/s12911-017-0449-x
- Pelt, D. M., and Sethian, J. A. (2018). A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl. Acad. Sci. U.S.A.* 115, 254–259. doi: 10.1073/pnas.1715832114
- Rappaport, N., Twik, M., Plaschkes, I., Nudel, R., Iny Stein, T., Levitt, J., et al. (2017). MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res.* 45, D877–D887. doi: 10.1093/nar/gkw1012
- Rogers, N. E., and Avram, M. R. (2008). Medical treatments for male and female pattern hair loss. *J. Am. Acad. Dermatol.* 59, 547–566. doi: 10.1016/j.jaad.2008.07.001
- Santos-Martins, D., Forli, S., Ramos, M. J., and Olson, A. J. (2014). AutoDock4(Zn): an improved autodock force field for small-molecule docking to zinc metalloproteins. *J. Chem. Inf. Model.* 54, 2371–2379. doi: 10.1021/ci500209e
- Shameer, K., Glicksberg, B. S., Hodos, R., Johnson, K. W., Badgeley, M. A., Readhead, B., et al. (2018). Systematic analyses of drugs and disease indications in RepurposeDB reveal pharmacological, biological and epidemiological factors influencing drug repositioning. *Brief Bioinform.* 19, 656–678. doi: 10.1093/bib/bbw136
- Steuer, A. E., Brockbals, L., and Kraemer, T. (2019). Metabolomic strategies in biomarker research-new approach for indirect identification of drug consumption and sample manipulation in clinical and forensic toxicology? *Front. Chem.* 7:319. doi: 10.3389/fchem.2019.00319
- Sullivan, D. P., Winsnes, C. F., Akesson, L., Hjelmare, M., Wiking, M., Schutten, R., et al. (2018). Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nat. Biotechnol.* 36, 820–829. doi: 10.1038/nbt.4225
- Sun, P., Guo, J., Winnenburg, R., and Baumbach, J. (2017). Drug repurposing by integrated literature mining and drug-gene-disease triangulation. *Drug Discov. Today* 22, 615–619. doi: 10.1016/j.drudis.2016.10.008
- Varothai, S., and Bergfeld, W. F. (2014). Androgenetic alopecia: an evidence-based treatment update. *Am. J. Clin. Dermatol.* 15, 217–230. doi: 10.1007/s40257-014-0077-5
- Vary, J. C. (2015). Selected disorders of skin appendages-acne, alopecia, hyperhidrosis. *Med. Clin. North. Am.* 99, 1195–1211. doi: 10.1016/j.mcna.2015.07.003
- Wang, W., Yang, S., Zhang, X., and Li, J. (2014). Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* 30, 2923–2930. doi: 10.1093/bioinformatics/btu403
- Wang, Y., Chen, S., Deng, N., and Wang, Y. (2013). Drug repositioning by kernel-based integration of molecular structure, molecular activity, and phenotype data. *PLoS ONE* 8:e78518. doi: 10.1371/journal.pone.0078518
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi: 10.1093/nar/gkx1037
- Wu, G., Liu, J., and Wang, C. (2017). Predicting drug-disease interactions by semi-supervised graph cut algorithm and three-layer data integration. *BMC Med. Genomics* 10:79. doi: 10.1186/s12920-017-0311-0
- Wu, G., Liu, J., and Yue, X. (2019). Prediction of drug-disease associations based on ensemble meta paths and singular value decomposition. *BMC Bioinformatics* 20:134. doi: 10.1186/s12859-019-2644-5
- Xu, R., Li, L., and Wang, Q. (2013). Towards building a disease-phenotype knowledge base: extracting disease-manifestation relationship from literature. *Bioinformatics* 19, 2186–2194. doi: 10.1093/bioinformatics/btt359
- Yang, J., Li, Z., Fan, X., and Cheng, Y. (2014). Drug-disease association and drug-repositioning predictions in complex diseases using causal inference-probabilistic matrix factorization. *J. Chem. Inf. Model.* 54, 2562–2569. doi: 10.1021/ci500340n
- Yap, C. W. (2011). PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* 32, 1466–1467. doi: 10.1002/jcc.21707
- Yu, L., Huang, J., Ma, Z., Zhang, J., Zou, Y., and Gao, L. (2015). Inferring drug-disease associations based on known protein complexes. *BMC Med. Genomics* 8:S2. doi: 10.1186/1755-8794-8-S2-S2
- Yu, L., Ma, X., Zhang, L., Zhang, J., and Gao, L. (2016). Prediction of new drug indications based on clinical data and network modularity. *Sci. Rep.* 6:32530. doi: 10.1038/srep32530
- Zhang, W., Yue, X., Huang, F., Liu, R., Chen, Y., and Ruan, C. (2018b). Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network. *Methods* 145, 51–59. doi: 10.1016/j.ymeth.2018.06.001
- Zhang, W., Yue, X., Lin, W., Wu, W., Liu, R., Huang, F., et al. (2018a). Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinformatics* 19:233. doi: 10.1186/s12859-018-2220-4
- Zhao, S., and Li, S. (2012). A co-module approach for elucidating drug-disease associations and revealing their molecular basis. *Bioinformatics* 28, 955–961. doi: 10.1093/bioinformatics/bts057
- Zheng, S., Chang, W., Xu, W., and Lin, Y. F. (2019). e-Sweet: a machine-learning based platform for the prediction of sweetener and its relative sweetness. *Front. Chem.* 7:35. doi: 10.3389/fchem.2019.00035
- Zhou, X., Menche, J., Barabasi, A. L., and Sharma, A. (2014). Human symptoms-disease network. *Nat. Commun.* 5:4212. doi: 10.1038/ncomms5212
- Zhu, L., and Zhu, F. (2015). Identification association of drug-disease by using functional gene module for breast cancer. *BMC Med. Genomics* 8:S3. doi: 10.1186/1755-8794-8-S2-S3
- Zhu, Q., Freimuth, R. R., Pathak, J., Durski, M. J., and Chute, C. G. (2013). Disambiguation of PharmGKB drug-disease relations with NDF-RT and SPL. *J. Biomed. Inform.* 46, 690–696. doi: 10.1016/j.jbi.2013.05.005

**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.

Copyright © 2020 Li, Huang, Chen, Wang, Li, Xie, Dai and Zou. 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.



# SPVec: A Word2vec-Inspired Feature Representation Method for Drug-Target Interaction Prediction

Yu-Fang Zhang<sup>1</sup>, Xiangeng Wang<sup>1</sup>, Aman Chandra Kaushik<sup>1,2</sup>, Yanyi Chu<sup>1</sup>, Xiaoqi Shan<sup>1</sup>, Ming-Zhu Zhao<sup>3</sup>, Qin Xu<sup>1\*</sup> and Dong-Qing Wei<sup>1,4\*</sup>

<sup>1</sup> State Key Laboratory of Microbial Metabolism, and SJTU-Yale Joint Center for Biostatistics and Data Science, School of Life Sciences and Biotechnology, and Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China, <sup>2</sup> Wuxi School of Medicine, Jiangnan University, Wuxi, China, <sup>3</sup> Instrumental Analysis Center, Shanghai Jiao Tong University, Shanghai, China, <sup>4</sup> Peng Cheng Laboratory, Shenzhen, China

## OPEN ACCESS

### Edited by:

Zunnan Huang,  
Guangdong Medical University, China

### Reviewed by:

Francesco Ortuso,  
University of Catanzaro, Italy  
Ling Wang,  
South China University of  
Technology, China

### \*Correspondence:

Qin Xu  
xuqin523@sjtu.edu.cn  
Dong-Qing Wei  
dqwei@sjtu.edu.cn

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 11 October 2019

**Accepted:** 12 December 2019

**Published:** 10 January 2020

### Citation:

Zhang Y-F, Wang X, Kaushik AC,  
Chu Y, Shan X, Zhao M-Z, Xu Q and  
Wei D-Q (2020) SPVec: A  
Word2vec-Inspired Feature  
Representation Method for  
Drug-Target Interaction Prediction.  
Front. Chem. 7:895.  
doi: 10.3389/fchem.2019.00895

Drug discovery is an academical and commercial process of global importance. Accurate identification of drug-target interactions (DTIs) can significantly facilitate the drug discovery process. Compared to the costly, labor-intensive and time-consuming experimental methods, machine learning (ML) plays an ever-increasingly important role in effective, efficient and high-throughput identification of DTIs. However, upstream feature extraction methods require tremendous human resources and expert insights, which limits the application of ML approaches. Inspired by the unsupervised representation learning methods like Word2vec, we here proposed SPVec, a novel way to automatically represent raw data such as SMILES strings and protein sequences into continuous, information-rich and lower-dimensional vectors, so as to avoid the sparseness and bit collisions from the clumsily manually extracted features. Visualization of SPVec nicely illustrated that the similar compounds or proteins occupy similar vector space, which indicated that SPVec not only encodes compound substructures or protein sequences efficiently, but also implicitly reveals some important biophysical and biochemical patterns. Compared with manually-designed features like MACCS fingerprints and amino acid composition (AAC), SPVec showed better performance with several state-of-art machine learning classifiers such as Gradient Boosting Decision Tree, Random Forest and Deep Neural Network on BindingDB. The performance and robustness of SPVec were also confirmed on independent test sets obtained from DrugBank database. Also, based on the whole DrugBank dataset, we predicted the possibilities of all unlabeled DTIs, where two of the top five predicted novel DTIs were supported by external evidences. These results indicated that SPVec can provide an effective and efficient way to discover reliable DTIs, which would be beneficial for drug reprofiling.

**Keywords:** drug-target interaction, representation learning, Word2vec, machine learning, feature embedding

## INTRODUCTION

Drug discovery is an issue of global importance, both academically and commercially. Generally, drugs have interactions with specific molecular targets, which are known as drug-target interactions (DTIs). Accurate identification of DTIs can significantly facilitate the processes of drug discovery. Thus, modern drug development calls for more effective and efficient techniques to identify true DTIs from the vast libraries of chemical compounds and protein targets. Numerous efforts have been poured into predictions of DTIs. However, it is still challenging to identify new drugs and their corresponding targets because of the limited knowledge about complex relationships between chemical space and proteomics space. Since *in vivo* and *in vitro* testings are rather costly and time-consuming (Kuruvilla et al., 2002; Haggarty et al., 2003; Valentin et al., 2018), scientists' focus moves more than ever to *in silico* techniques predict potential drug-target associations on a large scale, in which machine learning (ML) is one of the most attractive approaches.

Various machine learning methods have been developed in the last decades, in which the most widely used models are binary classifiers like Random Forest (RF) (Ho, 1998), Support Vector Machine (SVM) (Cortes and Vapnik, 1995), Deep Neural Network (DNN) (Liu et al., 2017), Gradient Boosting Decision Tree (GBDT) (Friedman, 2001), and so on. The performance of machine learning methods relies heavily on data representation (or features). Therefore, the design of data preprocessing and data transformation is of great concern to ensure that the data representation can support efficient machine learning algorithms. Numeric methods have been proposed to excavate drug and target features from their chemical structures and genomic sequences, respectively, such as fingerprints (Morgan, 1965; Ewing et al., 2006) and other molecular descriptors (Van Aalten et al., 1996; Hong et al., 2008) for drugs, amino acid composition (AAC) (Nakashima and Nishikawa, 1994) and physico-chemical properties (Cai et al., 2002) of target proteins, and so on. For example, Nascimento et al. (2016) used "normalized Smith-Waterman, mismatch and spectrum kernels" for the target protein sequences and "the spectrum, Lambda-k, Marginalized, MinMax, and Tanimoto kernels" for the drug's chemical structure to predict DTIs. In the work by Nanni et al. (2014), the drugs were represented by FP2 fingerprints and the representations on the targets were based on autocovariance, entropy, discrete wavelet, and substitution, and so on. The representation of the drug-target pairs was done by concatenating the target descriptors with the FP2 fingerprints of the drug. In the works by He et al. (2010), multiple chemical functional groups for drug-related features and pseudo AAC for protein-related features were extracted to describe drug-target pairs. Chen et al. (2012) integrated protein-protein similarity network, drug-drug similarity network, and known drug-target interaction networks into a heterogeneous network, and then implemented the random walk algorithm on this heterogeneous network for the prediction of DTIs. Rayhan et al. (2017) exploited their algorithms using both structural and evolutionary information to generate informative features. Based on these traditional features, the performance of

machine learning algorithms for predictions of DTIs have been gradually improved to a quite high level. However, these feature extraction methods require tremendous manpower and expert insights, and the effectiveness of these features also requires tremendous computations to be proved. Tedious processes of "feature engineering" have to be done before these features can be fed into downstream ML models. In order to facilitate the application of machine learning technologies, it is necessary to make them less dependent on feature engineering.

Representation learning (RL) (Bengio et al., 2013) is a way to introduce artificial intelligence (AI) and prior knowledge to automatically learn continuous, information-rich and lower-dimensional vectors from raw data that can be easily and directly used in ML models. An RL algorithm attempts to discover the latent features that describe the structure of a dataset under certain (either explicit or implicit) assumptions. Nowadays, RL has shown an influential role in effectively extracting features and solving the problem in computer vision, pattern recognition and natural language processing (NLP) (Mikolov et al., 2013a; Sharif Razavian et al., 2014). RL aims to automatically learn the representations (or features) from raw data that can be effectively utilized by downstream machine learning models to improve the performance of the model. Word2vec (Mikolov et al., 2013b) is one of the most popular RL methods, making NLP problems easier to tackle. Inspired by the distributed hypotheses that words found in similar environments usually have similar meanings, the Word2vec model predicts the center word based on its neighbor words in the window of a given size. This method simultaneously learn several language concepts such as Collobert and Weston (2008): (1) the meaning of the word; (2) how words are combined to form a concept (i.e., grammar); (3) how a concept relates to the task. Word2vec effectively removes word-meaning extraction subtasks by providing pre-trained word embeddings for learning algorithms. The word representations computed using Word2vec are very interesting because the learned vectors explicitly encode many linguistic regularities and patterns. Somewhat surprisingly, many of these patterns can be represented as linear translations. For example, the vector of "Paris" minus the vector of "France" plus the vector of "Italy" is very close to "Rome." In addition to its original utility as a word-embedding method, some of its ideas are effective in sequential data of non-language tasks (Jaeger et al., 2018; Zhang et al., 2019).

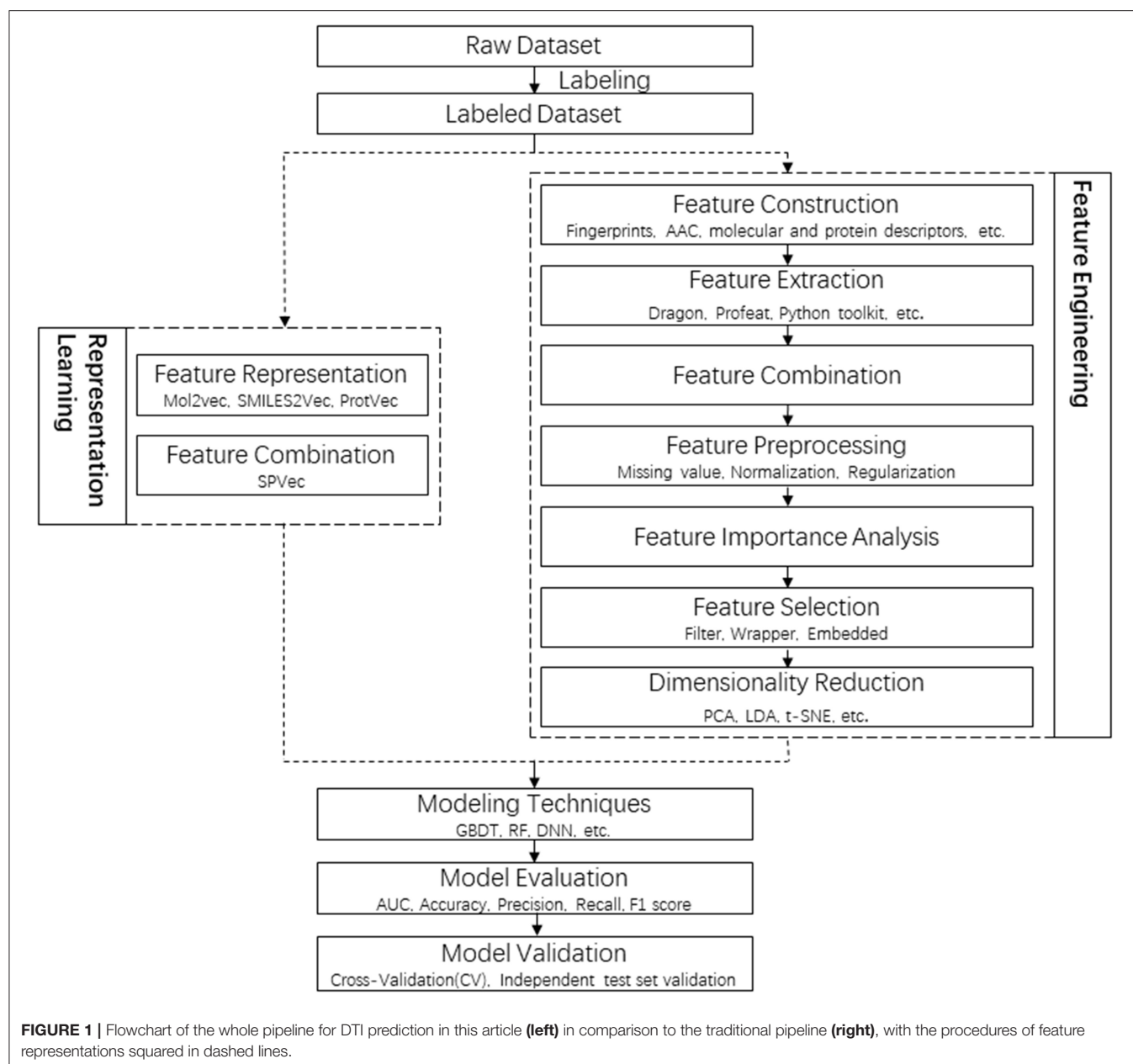
Recently, RL brought several breakthroughs in compound space and protein space. Convolutional neural networks were successfully applied on molecular graphs (Kearnes et al., 2016; Coley et al., 2017) and depictions of molecules (Goh et al., 2017b). Latent semantic structure indexing (LaSSI) (Schneider et al., 2017) techniques were adopted to compute chemical similarity from molecular descriptors. Word2vec (Asgari and Mofrad, 2015) has been adapted to protein sequences (ProtVec) for classification of protein families and predication of disordered proteins. Wan and Zeng (2016) used term frequency-inverse document frequency (tf-idf) to learn compound representations from Morgan fingerprints. While substructures of a molecule are hashed to a binary fingerprint (possibly sparse) in the case of the fingerprints, the Mol2vec approach, proposed by Jaeger et al. (2018), forms a vector with continuous and dense values.



The SMILES2Vec (Goh et al., 2017a) a model introduces a direct conversion of chemical structures from SMILES (Simplified Molecular-Input Line-Entry System) strings into vectors. These works show that RL technologies represented by Word2vec can automatic learn low-dimensional features from compound and protein feature space and achieve excellent performances, suggesting its advantages in both efficiency and effectiveness.

In this study, new SPVec vectors were constructed via the combination of SMILES2Vec and ProtVec to represent specific DTIs, where the drug representation was simplified by using SMILES directly. The whole pipeline of DTI prediction in this article is shown in **Figure 1**, in comparison with a tradition pipeline. Not like RL who can atomically learn lower-dimensional features without human resources and expert

insights, traditional feature engineering usually contains a lot of steps, including feature extraction, feature selection and dimensionality reduction, while every step need professional knowledge and extra time. For example, Fingerprint-based features are sparse and high-dimensional, thus dimensional reduction is necessary. Feature importance analysis and feature selection might be indispensable for mixed features, such as physicochemical properties, structural and evolutionary information and interaction information. It only takes several hours for feature presentation by SPVec training on a modern quad-core CPU, while dozens of days are required for traditional feature engineering. To evaluate the performance of SPVec, the constructed vectors was fed into several state-of-art machine learning classifiers such as GBDT, RF and DNN on BindingDB



(Gilson et al., 2016). The performance and robustness of SPVec were also confirmed by an external validation using DrugBank database. Also, we predicted the possibilities of all unlabeled DTIs in DrugBank database (Law et al., 2014), where two of the top five predicted novel DTIs were supported by external evidences. The results indicated that SPVec can discover reliable DTIs, which could be beneficial for drug reprofiling.

## METHOD

### Datasets

BindingDB is a public, web-accessible database of measured binding affinities, focusing chiefly on the interactions of target proteins with small, drug-like molecules, was utilized to evaluate the performance of SPVec. The whole BindingDB claims to contain 1,756,093 binding data for 7,371 protein targets and 780,240 small molecules (updated on 2019-05-01). Considering the validity of the features represented, inorganic compounds and protein targets with sequence identity > 75% were removed. In addition, considering the druggability, we excluded interactions with IC50 value missing or > 300 nM. Finally we got 36,014 small molecular drugs and 2,099 targets from BindingDB, which may generate over 75 million DTI pairs. Among them, 83,676 pairs are known as positive DTIs, and the rest are undetermined and treated as unlabeled data, from which 83,676 drug-target pairs were randomly selected as a negative dataset.

To further validate our model, we also collected data of DTIs from DrugBank. The data of drugs, targets and their interactions were separated by the date April 20, 2016, with those before it regarded as old while those after it regarded as new. In this way, we constructed five positive datasets as shown in **Table 1**: (1) dataset\_1 consists of all old drugs, old targets and their old interaction pairs; (2) dataset\_2 consists of all old drugs, old targets and their new interaction pairs; (3) dataset\_3 consists of all new drugs, old targets and their interaction pairs; (4) dataset\_4 consists of all old drugs, new targets and their interaction pairs; (5) dataset\_5 consists of all new drugs, new targets and their interaction pairs. The largest dataset\_1 with all old data was used for model training, while the other four datasets with new data were used to validate the robustness of the models. The generation of corresponding negative datasets of these five datasets are same as that from Binding DB, except that the unlabeled data pool of dataset\_2 is the rest of positive interactions of dataset\_1 and dataset\_2 (6068×3839-14534-3348).

### Feature Representations

SPVec is a Word2vec-inspired technique to represent latent features of small compounds and target proteins. Word2vec refers to the method that for any word  $w$  in dictionary  $\mathcal{D}$ , specify a fixed length of the real value vector  $V(w) \in \mathbb{R}^m$ , where  $V(w)$  is called the word vector of  $w$  and  $m$  is the length of the word vector. All of these vectors form a word vector space, and each vector can be regarded as a point in the space. The lexical or semantic similarity between them can be judged by the “distance” between the points.

In particular, we mainly used the Skip-gram model implemented with the Negative-sampling (NEG) method

to train the Word2vec-like models. The classical Skip-gram model consists of three layers: the input layer, the projection layer, and the output layer. Take a sample  $(w, \text{Context}(w))$  for example, assuming that  $\text{Context}(w)$  consists of  $c$  words before and after  $w$ , then a brief description of these three layers is as follows: the input layer is the word vector  $V(w) \in \mathbb{R}^m$  of the current sample; the projection layer is identity projection, which means projecting  $V(w)$  to  $V(w)$ ; the output layer is a binary Huffman tree, which takes every word appearing in the corpus as the leaf node and frequency of the word as weight. In the revised Skip-gram model here, the negative samples were generated by relatively simple random NEG method instead of Huffman trees, so as to improve training speed and improve the quality of the resulting word vectors. Given that a negative sample subset  $\text{NEG}(w) \neq \emptyset$  for  $w$  and  $\forall \tilde{w} \in \mathcal{D}$ , we define  $L^w(\tilde{w})$  as the label of word  $\tilde{w}$ , where the label of a positive sample is 1, and that of a negative sample is 0. For a given sample  $(w, \text{Context}(w))$ , we want to maximize the following function:

$$g(w) = \prod_{\tilde{w} \in \text{Context}(w)} \prod_{w \in \{u\} \cup \text{NEG}^{\tilde{w}}(w)} p(u|\tilde{w}) \quad (1)$$

where

$$p(u|\tilde{w}) = \left[ \sigma(V(\tilde{w})^T \theta^u) \right]^{L^w(u)} \times \left[ 1 - \sigma(V(\tilde{w})^T \theta^u) \right]^{1-L^w(u)}, \quad (2)$$

here  $\text{NEG}^{\tilde{w}}(w)$  is a generated subset of negative samples when processing words  $\tilde{w}$ . For a given corpus  $\mathcal{C}$ , the final objective function is:

$$\mathcal{L} = \log G = \log \prod_{w \in \mathcal{C}} g(w). \quad (3)$$

Maximizing this objective function can be performed using the stochastic gradient descent technique.

The same principles in the work by Wan and Zeng (2016) were followed to choose the hyperparameters of Skip-gram. That is, the embedding dimension was set as  $d = 100$ , the context window size was set as  $c = 12$ , and the number of negative examples was set as  $k = 15$ . Using this revised Skip-gram model, SMILES2Vec and ProtVec models were trained for feature representations of drug compounds and target proteins, respectively, and combined into SPVec to represent their interactions. Different from the works by Jaeger et al. (2018), we directly use SMILES of drug molecules rather than Morgan fingerprints as “sentences” to learn the representations, as SMILES strings are more like “sentences” and don’t need additional calculations. At the same time, the protocol of Asgari and Mofrad (2015) was followed in training of ProtVec here, where protein sequences were regarded as “sentences” and every three non-overlapping amino acids were regarded as a “word.”

To benchmark our SPVec approach against classical feature extraction approaches, we also extracted manually-designed features of chemical structures and protein sequences. For the ligands, we adopted the MACCS fingerprint (Corey and Wipke,

1969), one of the most widely used “structural fingerprints” based on pre-defined chemical substructures and finally got 166-dimensional compound feature vectors. At the same time, we considered the 20-dimensional AAC as protein descriptors, which were computed via PROFEAT (Zhang et al., 2016), a web server for computing commonly used protein features from their amino acid sequences.

The SPVec features compose a 100-dimensional space in such a way that similar objects are modeled into nearby points. To explore biochemical implications from SPVec features, the feature vectors of small molecular drugs and protein targets in DrugBank are projected from this 100-dimensional space into a 3D or 2D space for easier visualization using t-Distributed Stochastic Neighbor Embedding (t-SNE) (Der Maaten and Hinton, 2008), which is a non-linear dimensionality reduction technique for visualization of high dimensional data in a low-dimensional space, generally in two or three dimensions.

## Machine Learning Models

The feature embeddings learned by SPVec model were then fed into various machine learning models to predict the likelihood of their interactions. The performance in DTIs prediction by three state-of-art machine learning methods RF, GBDT, and DNN was used to evaluate the utility of SPVec embeddings. RF is an ensemble method that combines the probabilistic predictions of a number of decision tree-based classifiers to improve the generalization ability over a single estimator. GBDT is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion as other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function. DNN is a supervised learning algorithm which could learn non-linear models. It has one or more non-linear hidden layers between the input and output. For each hidden layer, different numbers of hidden neurons can be assigned. Each hidden neuron yields a weighted linear summation of the values from the previous layer, and the non-linear activation function is followed. The weights are learned through backpropagation algorithm or variations upon it. All these models were implemented by Python v3.6 and scikit-learn library (Pedregosa et al., 2011). All the datasets and source codes, as well as a Python module for user-friendly application of the SPVec method are available at <https://github.com/dqwei-lab/SPVec>.

## RESULTS AND DISCUSSION

### Biochemical Implications of SPVec Features

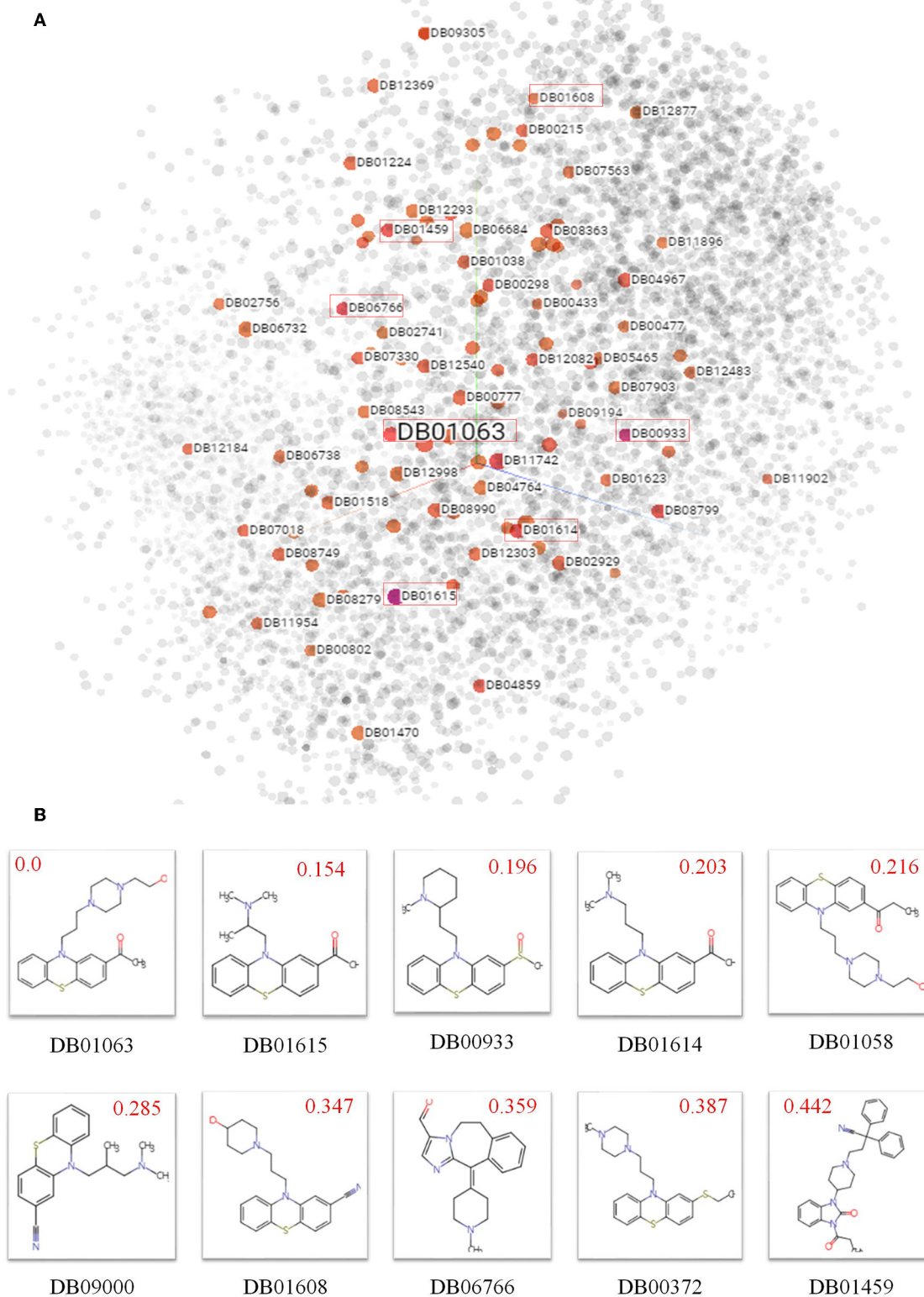
To explore biochemical implications from SPVec features, small molecular drugs in DrugBank are represented into vectors by SMILES2Vec and projected from a 100-dimensional space to a 3D space by t-SNE, shown in **Figure 2A**, where each point represents a small drug molecule. Since the SMILES2Vec vectors are sums of substructure vectors, they may implicitly capture substructure importance via the vector weight, thus

the drugs closer to each other may have more structural and functional similarities. For example, the boxed points stand for part of the top 10 chemicals (because some are masked by other points) similar to Acetophenazine (Drug ID: DB01063), an antipsychotic drug of moderate-potency used in the treatment of disorganized, psychotic thinking and false perceptions. **Figure 2B** shows the molecular structures of these top 10 similar chemicals, most of which have the phenothiazine substructure (a Sulfur atom and a Nitrogen atom connected with two benzene rings) with neuroleptic and anti-histamine properties. Some of them share the same target. For example, Acepromazine (DrugBank ID: DB01614), Thiethylperazine (DrugBank ID: DB00372) and Acetophenazine have two common targets, D(2) dopamine receptor and D(1A) dopamine receptor. Periciazine (DrugBank ID: DB01608) and Acetophenazine have two common targets, D(1A) dopamine receptor and Androgen receptor). Oppositely, Ceftibuten (DrugBank ID: DB01459), which is relatively far from Acetophenazine, has no Phenothiazine substructure. And in terms of functionalities, Ceftibuten is typically used to treat acute bacterial exacerbations of chronic bronchitis (ABECB), acute bacterial otitis media, pharyngitis, and tonsillitis, which is different from Acepromazine either. Obviously, molecules with similar functional groups are close in the generated SMILES2Vec vector space.

Although the ProtVec is only trained based on the primary sequences of proteins, it shows some important biochemical and biophysical implications (Asgari and Mofrad, 2015). In order to study these features, we visualized the distribution of ProtVec vectors by mass, volume, polarity, hydrophobicity. In **Figure 3**, each point represents a protein, with a color according to its scale in each property. The distribution of these points turns out that proteins with similar biochemical and biophysical properties tend to be closer. This observation indicates that not only encodes protein sequences effectively and efficiently, the ProtVec also implicitly reveals some important biophysical and biochemical patterns of the protein, while AAC only contains information of protein compositions (Nakashima and Nishikawa, 1994).

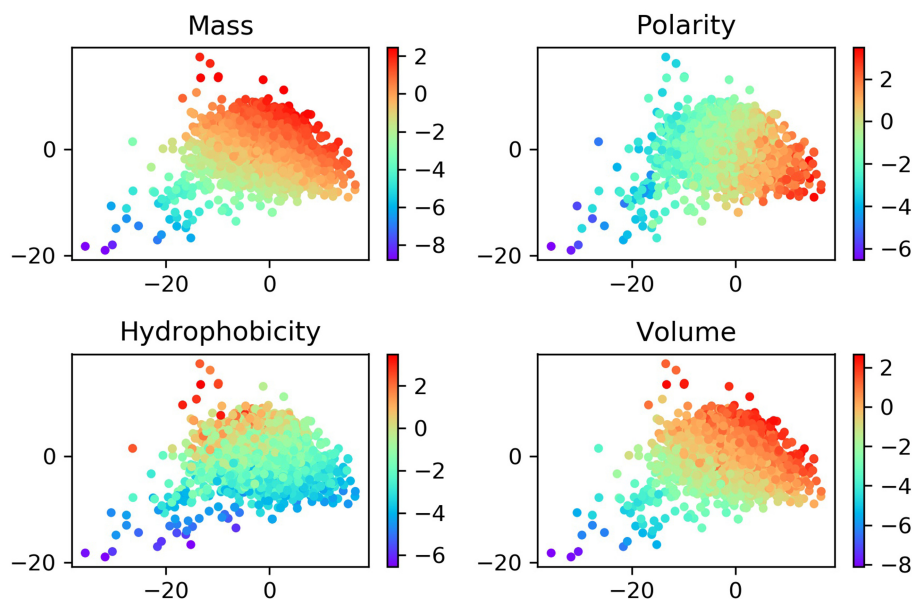
### Performance of SPVec in Comparison With Traditional Feature Representation Methods

We tested the performance of DTIs predictions for SPVec combined with three state-of-art ML classifiers (GBDT, RF, and DNN) using BindingDB, and compared with the combination of MACCS and AAC features as a baseline. To validate the performance of SMILES2Vec and ProtVec independently, we also constructed another two feature combinations, that is, MACCS-ProtVec and SMILES2Vec-AAC. A summary of classification performances of these four feature combinations on the BindingDB dataset are shown in **Table 2**, with their ROC curves shown in **Figure S1**. It is obvious that DTIs predictions based on SPVec vectors are relatively improved than those on the classical feature combination (i.e., MACCS-AAC) when using any of the ML classifiers. For predictions by the



**FIGURE 2 |** Biochemical implications from SMILES2Vec features. **(A)** Visualizations of the SMILES2Vec vector space of drugs in DrugBank using t-SNE. **(B)** The top 10 drugs most similar to Acetophenazine (DrugBank ID: DB01063) according to their SMILES2Vec vectors. Red values show their cosine distances with Acetophenazine. The smaller the value, the more similar in the chemical structures.





**FIGURE 3** | Normalized distributions of biochemical and biophysical properties in a 2D space projected by t-SNE from the 100-dimensional ProtVec protein-space. In these plots, each point represents a protein, and the colors indicate the scale for each property.

GBDT, RF, and DNN classifiers, the AUCs using SMILES2Vec-ProtVec are 13.35, 15.67, and 11.66% higher than MACCS-AAC, respectively. When only molecules are characterized via SMILES2Vec, the AUCs of SMILES2Vec-ProtVec are about 8.86%, 11.57%, and 9.09% higher than SMILES2Vec-AAC. And when molecules are characterized via MACCS, the AUCs of MACCS-ProtVec were about 8.91, 9.42, and 6.85% higher than MACCS-AAC. Therefore, in DTIs predictions single feature representations by ProtVec or SMILES2Vec also partly improve the classification performances. It is also reasonable to expect their individual performances in other tasks related to only drugs or proteins, such as compound property predictions and protein classifications. **Table 2** also indicates that features represented by SPVec are quite reliable with different ML models. Based on the datasets from BindingDB, the GBDT, RF, and DNN models resulted in no important difference for classification tasks of DTIs, and all achieved similarly higher AUC score, accuracy, precision, recall, and F1-score.

The performance of SPVec based DTIs predictions was also compared with earlier results using different popular classical features or modeling methods, as summarized in **Table 3**. Compared with these classical features, the features represented by SPVec are much lower in dimensions, which masterly avoid the “Curse of Dimensionality,” and enable ML models to achieve better performances. Especially when some kinds of features are hard to obtain, such as the 3D molecular and protein descriptors, the advantages of SPVec is more evident. It’s worth to note that You et al. (2019) and Yu et al. (2012) used DrugBank database with different versions (released on 14 Nov. 2017 and 1 June 2011, respectively), while the datasets for the other predictions (Ezzat et al., 2016, 2017) were from the version 4.3 of DrugBank database (released on 17 Nov. 2015) in which there are 12,674

**TABLE 1** | Number of entries of the five different datasets obtained from DrugBank dataset.

Datasets	Dataset_1	Dataset_2	Dataset_3	Dataset_4	Dataset_5
Drug	6,068	6,068	537	6,068	537
Target	3,839	3,839	3,839	160	160
Interactions	15,434	3,348	1,735	264	37

drug-target interactions between 5,877 drugs and their 3,348 protein interaction partners in total. However, as shown in **Table S1**, the performances of SPVec did not change a lot using the different versions of database. AUC of DNN, GBDT and RF only increased by 0.0315, 0.0039, and 0.0088, respectively. Therefore, the better performance of SPVec compared with earlier results in **Table 3** is still guaranteed.

## Evaluation of the Robustness of SPVec

In order to test the robustness of SPVec in DTIs predictions, especially in the newly found interactions, five datasets were constructed from DrugBank, as described in section Datasets. We used dataset\_1 as the training set to learn features and construct the ML models and then tested their performances on the datasets with new data. The classification performances on dataset\_1 via  $10 \times 5$ -fold cross-validation and performances on independent test sets like dataset\_1, dataset\_2, and dataset\_3 using GBDT, RF and DNN are summarized in **Table 4** with their ROC curves shown in **Figure 4**. As in **Table 4**, the ML approaches equipped with the SPVec features got quite high AUC on the training set, which is similar to the results on BindingDB. Although DNN architecture was outperformed by

the tree-based methods GBDT and RF in both cases, we would like to note the possibility that further fine-tuning might a little bit improve the prediction performance of the SPVec-DNN combination. **Table 4** shows that SPVec performed satisfactorily on the test sets, suggesting acceptable generalization capacity and competitive performance of SPVec for the prediction of novel DTIs in drug repositioning or drug rediscovery, which is also suggested by the ROC curves in **Figure 4**. Among the four test sets, all three classifiers achieve highest AUC on dataset\_2 to predict new interactions between old drugs and old targets, while the prediction results on the interactions with new drugs or new targets are much worse, which is extraordinary obvious in the ROC curves of dataset\_5 in **Figure 4**. A possible explanation is that the newly found drugs

or targets are not studied adequately and many potential DTIs between them have not been identified yet. Thus, the reduced accuracy of the data impairs the accuracy of the models. It is also worth noting that the negative sample was constructed by randomly selection from the unlabeled data, where the portion of unidentified potential positive DTI pairs may be even higher in smaller datasets. At last, the distributions of the new DTIs in the vector space of the test sets may be deviated from that of the training set, and impair the robustness of the models.

Particularly, the SPVec-GBDT method achieves the best performance among these three classifiers in DTIs predictions on the four test sets, with the AUCs as 0.8945, 0.7502, 0.7356, and 0.68, respectively. Although GBDT and RF showed similar results on the first four datasets, GBDT outperformed RF on dataset\_5. This indicates that the SPVec-GBDT method may have better generalization capacity to achieve more robust prediction results, even for new drug-target pairs with limited or no interactions information, which may suggest that the SPVec-GBDT prediction model is possibly highly pertinent to the prediction of novel DTIs in drug repositioning.

### Prediction and Validation on Unidentified DTIs

Although the SPVec vectors with continuous values show competitive performance in the task of DTIs predictions, a dominant issue in the prediction of DTIs is that only confirmed positive interactions are deposited in the databases while those unlabeled interactions are unknown to be really positive or negative. For example, the newly found interactions in Drugbank might be thought negative in dataset\_1. To further evaluate the validity of SPVec predictions on DTIs, the possibilities of all unlabeled DTIs in DrugBank dataset were evaluated using SPVec-GBDT and the top five ranked interactions were

**TABLE 2 |** Results of classification performance of four feature combinations using three classifiers on BindingDB via 10 × 5-fold cross-validation, with the highest scores highlighted in the bold font.

Feature combinations	Model	AUC	Accuracy	Precision	Recall	F1-score
SPVec (SMILES2Vec-ProtVec)	GBDT	0.9923	<b>0.9680</b>	0.9695	<b>0.9667</b>	<b>0.9681</b>
	RF	<b>0.9927</b>	0.9675	<b>0.9808</b>	0.9540	0.9672
	DNN	0.9617	0.9332	0.9287	0.9248	0.9197
SMILES2Vec-AAC	GBDT	<b>0.9037</b>	<b>0.8272</b>	<b>0.8563</b>	<b>0.7873</b>	<b>0.8204</b>
	RF	0.8770	0.7974	0.8657	0.7050	0.7772
	DNN	0.8708	0.8124	0.7993	0.7879	0.7126
MACCS-ProtVec	GBDT	<b>0.9479</b>	<b>0.8810</b>	<b>0.8908</b>	<b>0.8690</b>	<b>0.8798</b>
	RF	0.9302	0.8542	0.8712	0.8322	0.8512
	DNN	0.9136	0.8034	0.8025	0.8097	0.8074
MACCS-AAC	GBDT	<b>0.8588</b>	<b>0.7811</b>	<b>0.8077</b>	<b>0.7392</b>	<b>0.7719</b>
	RF	0.8360	0.7468	0.8366	0.6150	0.7089
	DNN	0.8451	0.7832	0.7884	0.7726	0.7724

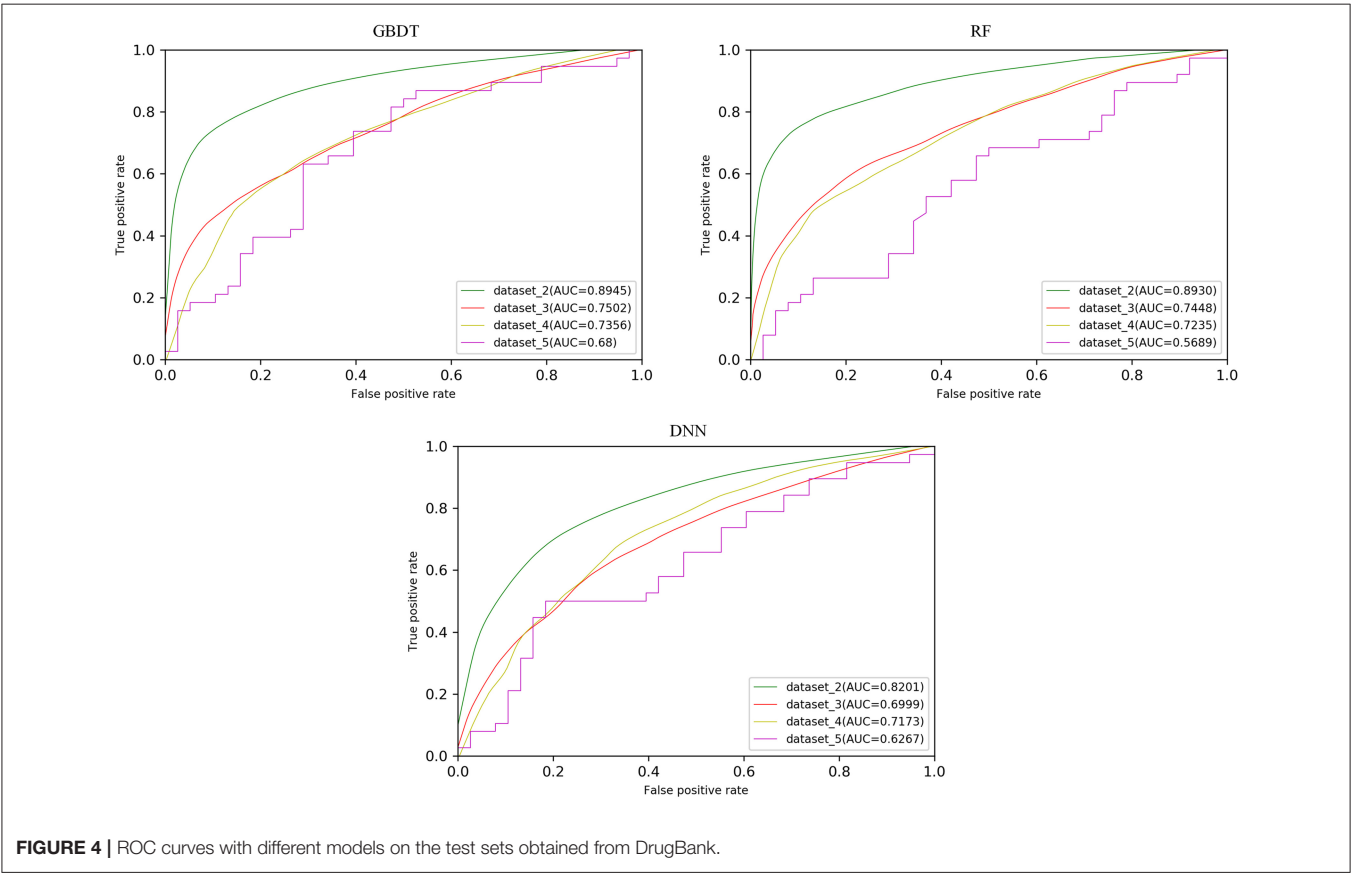
**TABLE 3 |** AUCs of SPVec and other models on DTI predictions using DrugBank.

Drug features	Drug dim.	Protein features	Protein dim.	ML method	AUC	References
Drug structure information	2,216	AAC, DC <sup>a</sup> and TC <sup>b</sup>	11,943	DNN	0.81	You et al., 2019 <sup>c</sup>
Constitutional, topological and molecular descriptors, 2D autocorrelations, topological charge indices, eigenvalue-based indices	1,664	AAC; DC <sup>a</sup> ; autocorrelation; Composition, Transition, Distribution descriptors; Quasi-sequence-order	1,080	RF	0.8950	Yu et al., 2012 <sup>c</sup>
Constitutional, topological and geometrical descriptors	193	AAC; DC <sup>a</sup> ; autocorrelation; composition, transition and distribution; quasi-sequence-order; amphiphilic pseudo-amino acid composition and total amino acid properties	1,260	DT RF	0.760 0.855	Ezzat et al., 2016
PubChem fingerprints indicating presence or absence of 881 known chemical substructures	881	Fingerprints of 876 different protein domains that are obtained from the Pfam database	876	EnsemDT	0.882	Ezzat et al., 2017
SMILES2Vec	100	ProtVec	100	RF	0.855	This work
				GBDT	0.9467	
				DNN	0.8637	

<sup>a</sup>DC, dipeptide composition; <sup>b</sup>TC, tripeptide composition; <sup>c</sup> These models are trained on different versions of DrugBank, whose AUCs are only as references.

**TABLE 4 |** Results of classification performance using three classifiers on datasets obtained from DrugBank, with the highest scores highlighted in the bold font.

Dataset	Model	AUC	Accuracy	Precision	Recall	F1-score
<b>Training set</b>		<b>10 × 5-fold cross-validation</b>				
Dataset_1	GBDT	0.9506	<b>0.9323</b>	<b>0.9456</b>	0.9367	<b>0.9343</b>
	RF	<b>0.9557</b>	0.9234	0.9378	<b>0.9369</b>	0.9337
	DNN	0.8952	0.8732	0.8345	0.8437	0.8654
<b>Test sets</b>		<b>Independent validation</b>				
Dataset_2	GBDT	<b>0.8945</b>	0.8628	<b>0.8747</b>	<b>0.8696</b>	<b>0.8637</b>
	RF	0.8930	<b>0.8753</b>	0.8645	0.8467	0.8555
	DNN	0.8201	0.8026	0.8138	0.8199	0.8144
Dataset_3	GBDT	<b>0.7502</b>	<b>0.7389</b>	<b>0.7340</b>	<b>0.7245</b>	<b>0.7333</b>
	RF	0.7448	0.7299	0.7198	0.7243	0.7230
	DNN	0.6999	0.6922	0.6825	0.6798	0.6832
Dataset_4	GBDT	<b>0.7356</b>	<b>0.7223</b>	<b>0.7167</b>	<b>0.7177</b>	<b>0.7201</b>
	RF	0.7235	0.7034	0.7108	0.7078	0.71
	DNN	0.7173	0.6899	0.6884	0.6896	0.6866
Dataset_5	GBDT	<b>0.68</b>	<b>0.6703</b>	<b>0.6679</b>	<b>0.6664</b>	<b>0.6688</b>
	RF	0.5689	0.5605	0.5398	0.5321	0.5411
	DNN	0.6267	0.6098	0.607	0.6122	0.6114



tested by external supporting evidences from several reference databases like PubChem (Wang et al., 2009), KEGG (Kanehisa and Goto, 2000), ChEMBL (Gaulton et al., 2017) and biomedical literatures. As a result, two of the top five predicted DTIs were confirmed by existing evidences (Table 5). The tyrosine-protein kinase Yes (Target ID: P07947, also known as Yes1) has been implicated as a potential therapeutic target in lots of cancers including breast cancers, melanomas, and rhabdomyosarcomas.

**TABLE 5 |** Top five novel DTIs predicted by SPVec-GBDT.

Drug ID	Target ID	Drug name	Target name	Validation source
DB11805	P07947	Saracatinib	The tyrosine-protein kinase Yes	Patel et al., 2013
DB09282	P42262	Molsidomine	Glutamate receptor 2	None
DB05524	Q99640	Pelitinib	Membrane-associated tyrosine and threonine-specific cdc2-inhibitory kinase	<a href="https://pubchem.ncbi.nlm.nih.gov/compound/6445562">https://pubchem.ncbi.nlm.nih.gov/compound/6445562</a>
DB03017	Q16620	Lauric acid	BDNF/NT-3 growth factors receptor	None
DB13165	P11362	Ripasudil	Fibroblast growth factor receptor 1	None

Saracatinib (Drug ID: DB11805) was identified by Patel et al. (2013) as a potent Yes1 kinase inhibitor with the IC<sub>50</sub> as low as 6.2nM. Our results also predicted the interaction between Pelitinib (Drug ID: DB05524) and membrane-associated tyrosine and threonine-specific cdc2-inhibitory kinase (Target ID: P42262) which was confirmed by PubChem database. Pelitinib (EKB-569) is a potent, low molecular weight, selective, and irreversible inhibitor of epidermal growth factor receptor (EGFR) in development as an anticancer agent, while membrane-associated tyrosine and threonine-specific cdc2-inhibitory kinase is the kinase domain of human myt1. These results demonstrate that the SPVec-DTIs model has highly useful pertinence for the prediction of novel DTIs.

## CONCLUSION

Combining SMILES2Vec and ProtVec, SPVec could transfer SMILES strings of drug compounds and protein sequences into information-rich and lower-dimensional vectors automatically. Visualization of SPVec vectors nicely illustrates that the derived vectors from similar structures locate closely in the vector space, suggesting that they may implicitly reveals some important biophysical and biochemical patterns. Based on BindingDB and DrugBank database, SPVec vectors were fed into several state-of-art machine learning methods like GBDT, RF and DNN to train DTIs prediction models. The results using BindingDB have shown that the proposed models can achieve better

prediction performance than manually extracted features like the combination of MACCS and AAC. Also, the results tested on DrugBank datasets indicated that our approach, especially SPVec-GBDT, can discover reliable DTIs in newly found drugs and targets, which might be beneficial for drug re-profiling. At last, all the unlabeled DTIs in DrugBank database was repredicted by the SPVec-GBDT model, and two of the top five predicted novel DTIs were confirmed by external evidences from other databases or biomedical literatures. In addition, SPVec vectors also have the advantages of automatic learning and lower dimensionality, which may significantly speed up training and reduces memory requirements, making it a highly potential method of feature representation for DTI predictions.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the <https://github.com/dqwei-lab/SPVec>.

## AUTHOR CONTRIBUTIONS

Y-FZ, QX, and D-QW made the conception and designed the study. Y-FZ and XW collected and organized the database. Y-FZ, AK, and YC performed the statistical analysis. QX and Y-FZ wrote the manuscript. XS contributed to part of the first draft of the manuscript. M-ZZ contributed to part of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

## FUNDING

This work was supported by the grants from the National Natural Science Foundation of China (Contract nos. 31770772, 61832019, and 61503244), the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, and Joint Research Funds for Translational Medicine at Shanghai Jiao Tong University (ZH2018ZDA06).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2019.00895/full#supplementary-material>

## REFERENCES

- Asgari, E., and Mofrad, M. R. (2015). Continuous distributed representation of biological sequences for deep proteomics and genomics. *PLoS ONE* 10:e0141287. doi: 10.1371/journal.pone.0141287
- Bengio, Y., Courville, A. C., and Vincent, P. (2013). Representation learning: a review and new perspectives. *IEEE T. Pattern Anal.* 35, 1798–1828. doi: 10.1109/TPAMI.2013.50
- Cai, Y., Liu, X., Xu, X., and Chou, K. (2002). Support vector machines for prediction of protein subcellular location by incorporating quasi-sequence-order effect. *J. Cell. Biochem.* 84, 343–348. doi: 10.1002/jcb.10030
- Chen, X., Liu, M., and Yan, G. (2012). Drug–target interaction prediction by random walk on the heterogeneous network. *Mol. Biosyst.* 8, 1970–1978. doi: 10.1039/c2mb00002d
- Coley, C. W., Barzilay, R., Green, W. H., Jaakkola, T. S., and Jensen, K. F. (2017). Convolutional embedding of attributed molecular graphs for physical property prediction. *J. Chem. Inf. Model.* 57:1757. doi: 10.1021/acs.jcim.6b00601
- Collobert, R., and Weston, J. (2008). A unified architecture for natural language processing: deep neural networks with multitask learning. *ACM* 8, 160–167. doi: 10.1145/1390156.1390177
- Corey, E. J., and Wipke, W. T. (1969). Computer-assisted design of complex organic syntheses. *Science* 166, 178–192. doi: 10.1126/science.166.3902.178



- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Der Maaten, L. V., and Hinton, G. E. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Ewing, T., Baber, J. C., and Feher, M. (2006). Novel 2D fingerprints for ligand-based virtual screening. *J. Chem. Inf. Model.* 46, 2423–2431. doi: 10.1021/ci060155b
- Ezzat, A., Wu, M., Li, X. L., and Kwok, C. K. (2016). Drug-target interaction prediction via class imbalance-aware ensemble learning. *BMC Bioinf.* 17, 267–276. doi: 10.1186/s12859-016-1377-y
- Ezzat, A., Wu, M., Li, X. L., and Kwok, C. K. (2017). Drug-target interaction prediction using ensemble learning and dimensionality reduction. *Methods* 129, 81–88. doi: 10.1016/j.ymeth.2017.05.016
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232. doi: 10.1214/aos/1013203451
- Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., et al. (2017). The ChEMBL database in 2017. *Nucleic Acids Res.* 45, D945–D954. doi: 10.1093/nar/gkw1074
- Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. (2016). BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* 44, 1045–1053. doi: 10.1093/nar/gkv1072
- Goh, G. B., Hodas, N. O., Siegel, C., and Vishnu, A. (2017a). Smiles2vec: an interpretable general-purpose deep neural network for predicting chemical properties. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1712.02034> (accessed December 6, 2017).
- Goh, G. B., Siegel, C., Vishnu, A., Hodas, N. O., and Baker, N. (2017b). Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1706.06689> (accessed June 20, 2017).
- Haggarty, S. J., Koeller, K. M., Wong, J. C., Butcher, R. A., and Schreiber, S. L. (2003). Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays. *Chem. Biol.* 10, 383–396. doi: 10.1016/S1074-5521(03)00095-4
- He, Z., Zhang, J., Shi, X., Hu, L., Kong, X., Cai, Y., et al. (2010). Predicting drug-target interaction networks based on functional groups and biological features. *PLoS ONE* 5:e9603. doi: 10.1371/journal.pone.0009603
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE T. Pattern Anal.* 20, 832–844. doi: 10.1109/34.709601
- Hong, H., Xie, Q., Ge, W., Qian, F., Fang, H., Shi, L., et al. (2008). Mold(2), molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *J. Chem. Inf. Model.* 48, 1337–1344. doi: 10.1021/ci800038f
- Jaeger, S., Fulle, S., and Turk, S. (2018). Mol2vec: unsupervised machine learning approach with chemical intuition. *J. Chem. Inf. Model.* 58, 27–35. doi: 10.1021/acs.jcim.7b00616
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi: 10.1093/nar/28.1.27
- Kearnes, S., McCloskey, K., Berndl, M., Pande, V., and Riley, P. (2016). Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aid. Mol. Des.* 30, 1–14. doi: 10.1007/s10822-016-9938-8
- Kuruvilla, F. G., Shamji, A. F., Sternson, S. M., Hergenrother, P. J., and Schreiber, S. L. (2002). Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. *Nature* 416, 653–657. doi: 10.1038/416653a
- Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A., Liu, Y., et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42, 1091–1097. doi: 10.1093/nar/gkt1068
- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing* 234, 11–26. doi: 10.1016/j.neucom.2016.12.038
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1706.06689> (accessed June 20, 2019).
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *ACM* 13, 3111–3119.
- Morgan, H. L. (1965). The generation of a unique machine description for chemical structures-A technique developed at chemical abstracts service. *J. Chem. Doc.* 5, 107–113. doi: 10.1021/c160017a018
- Nakashima, H., and Nishikawa, K. (1994). Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. *J. Mol. Biol.* 238, 54–61. doi: 10.1006/jmbi.1994.1267
- Nanni, L., Lumini, A., and Brahnam, S. (2014). A set of descriptors for identifying the protein–drug interaction in cellular networking. *J. Theor. Biol.* 359, 120–128. doi: 10.1016/j.jtbi.2014.06.008
- Nascimento, A. C. A., Prudêncio, R. B. C., and Costa, I. G. (2016). A multiple kernel learning algorithm for drug-target interaction prediction. *BMC Bioinf.* 17:46. doi: 10.1186/s12859-016-0890-3
- Patel, P. R., Sun, H., Li, S. Q., Shen, M., Khan, J., Thomas, C. J., et al. (2013). Identification of potent yes1 kinase inhibitors using a library screening approach. *Bioorg. Med. Chem. Lett.* 23, 4398–4403. doi: 10.1016/j.bmcl.2013.05.072
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.1524/auto.2011.0951
- Rayhan, F., Ahmed, S., Shatabda, S., Farid, D. M., Mousavian, Z., Dehngani, A., et al. (2017). iDTI-ESBoost: identification of drug target interaction using evolutionary and structural features with boosting. *Sci. Rep.* 7:17731. doi: 10.1038/s41598-017-18025-2
- Schneider, N., Fechner, N., Landrum, G. A., and Stiefl, N. (2017). Chemical topic modeling: exploring molecular data sets using a common text-mining approach. *J. Chem. Inf. Model.* 57, 1816–1831. doi: 10.1021/acs.jcim.7b00249
- Sharif Razavian, A., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. *IEEE* 5, 512–519. doi: 10.1109/CVPRW.2014.131
- Valentin, J., Guillon, J., Jenkinson, S., Kadambi, V. J., Ravikumar, P., Roberts, S., et al. (2018). *In vitro* secondary pharmacological profiling: an IQ-drusafe industry survey on current practices. *J. Pharmacol. Tox. Met.* 93, 7–14. doi: 10.1016/j.vascn.2018.07.001
- Van Aalten, D. M. F., Bywater, R. P., Findlay, J. B. C., Hendlich, M., Hooft, R. W. W., and Vriend, G. (1996). PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput. Aid. Mol. Des.* 10, 255–262. doi: 10.1007/BF00355047
- Wan, F., and Zeng, J. (2016). Deep learning with feature embedding for compound-protein interaction prediction. *bioRxiv [Preprint]*. Available online at: <https://www.biorxiv.org/content/10.1101/086033v1> (accessed November 07, 2016).
- Wang, Y., Xiao, J., Suzek, T. O., Jian, Z., 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
- You, J., McLeod, R. D., and Hu, P. (2019). Predicting drug-target interaction network using deep learning model. *Comput. Biol. Chem.* 80, 90–101. doi: 10.1016/j.compbiolchem.2019.03.016
- Yu, H., Chen, J., Xu, X., Li, Y., Zhao, H., Fang, Y., et al. (2012). A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. *PLoS ONE* 7:e37608. doi: 10.1371/journal.pone.0037608
- Zhang, H., Liao, L., Cai, Y., Hu, Y., and Wang, H. (2019). IVS2vec: a tool of inverse virtual screening based on word2vec and deep learning techniques. *Methods* 66, 57–65. doi: 10.1016/j.ymeth.2019.03.012
- Zhang, P., Tao, L., Zeng, X., Qin, C., Chen, S., Zhu, F., et al. (2016). A protein network descriptor server and its use in studying protein, disease, metabolic and drug targeted networks. *Brief. Bioinform.* 18, 1057–1070. doi: 10.1093/bib/bbw071

**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.

Copyright © 2020 Zhang, Wang, Kaushik, Chu, Shan, Zhao, Xu and Wei. 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.



# The Use of Computational Approaches in the Discovery and Mechanism Study of Opioid Analgesics

Bangyi Zhao<sup>1†</sup>, Wei Li<sup>1†</sup>, Lijie Sun<sup>2</sup> and Wei Fu<sup>1\*</sup>

<sup>1</sup> Department of Medicinal Chemistry, School of Pharmacy, Fudan University, Shanghai, China, <sup>2</sup> Shijiazhuang No. 4 Pharmaceutical Co., Ltd., Shijiazhuang Economic and Technological Development Zone, Shijiazhuang, China

## OPEN ACCESS

### Edited by:

Zunnan Huang,  
Guangdong Medical University, China

### Reviewed by:

Renyu Liu,  
University of Pennsylvania,  
United States  
Mingyue Zheng,  
Shanghai Institute of Materia Medica  
(CAS), China  
Chaoyang Xue,  
Rutgers University, Newark,  
United States

### \*Correspondence:

Wei Fu  
wfu@fudan.edu.cn

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

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 22 February 2020

Accepted: 31 March 2020

Published: 15 May 2020

### Citation:

Zhao B, Li W, Sun L and Fu W (2020)  
The Use of Computational  
Approaches in the Discovery and  
Mechanism Study of Opioid  
Analgesics. *Front. Chem.* 8:335.  
doi: 10.3389/fchem.2020.00335

Opioid receptors that belong to class A G protein-coupled receptors (GPCRs) are vital in pain control. In the past few years, published high-resolution crystal structures of opioid receptor laid a solid basis for both experimental and computational studies. Computer-aided drug design (CADD) has been established as a powerful tool for discovering novel lead compounds and for understanding activation mechanism of target receptors. Herein, we reviewed the computational-guided studies on opioid receptors for the discovery of new analgesics, the structural basis of receptor subtype selectivity, agonist interaction mechanism, and biased signaling mechanism.

**Keywords:** opioid receptor, analgesics, computer-aided drug design, molecular dynamics, agonist mechanism, G-protein-biased activation

## INTRODUCTION

G protein-coupled receptors (GPCRs), also known as hepta-helical receptors, are characterized as  $\alpha$ -helix domains which span the cell membrane seven times (King et al., 2003). As a large family of membrane proteins, GPCRs are targeted by about 40% of all modern drugs (Overington et al., 2006). Opioid receptors are a subfamily of GPCRs and are responsible for powerful analgesic effects. They are mainly divided into four subtypes,  $\mu$ ,  $\delta$ ,  $\kappa$ , and nociceptin/orphanin FQ (N/OFQ or NOP). Lately, Claff et al. (2019) reported the first active crystal structure of the  $\delta$  opioid receptor. This will enhance the future structure-based development of opioid agonists along with some previously revealed crystal structures. Inspired by the progress in the determination of opioid receptor, here we summarize the recent development in the use of computational approaches in the discovery and mechanism study of opioid analgesics.

When an agonist binds to the opioid receptor, it will activate G proteins, inhibit adenylate cyclase, and promote the extrusion of potassium ions. These events can finally lead to the hyperpolarization of cell membranes (Snyder, 1977; Simon and Hiller, 1978). These events also block or minimize the transmission of painful stimuli and raise the pain threshold. For centuries, opioid analgesics have been used to manage acute and chronic pain with the level from moderate to severe. These analgesics, like morphine, fentanyl, etc., mainly target the  $\mu$  opioid receptor. These drugs have strong analgesic effects that cannot be replaced by other analgesic drugs, and therefore,

become the first-line treatment of cancer pain, surgical pain, and other kinds of pain. However, usage of these drugs has brought about obvious side effects such as constipation, drug tolerance, addiction, and respiratory depression. These unwanted side effects put limits on their clinical use. Also reported are some other analgesics in clinical use with relatively less addiction. These include tramadol, buprenorphine, butorphanol, and dezocine. The analgesics with less addiction are mostly targeted to multiple opioid receptors. For example, butorphanol is a partial agonist of  $\mu$  and  $\kappa$  receptors, while dezocine is recently recognized as an agonist of  $\mu$  opioid receptor and an inhibitor of norepinephrine transporter ( $\mu$ OR-NRI) (Wang et al., 2017).

However, there is strong clinical need for new analgesics with less side effects. As inspired, researchers are generating novel ideas, of which are continuously being evaluated. These include the design of  $\kappa$  or  $\delta$  opioid selective agonists, new bifunctional  $\mu$  receptor agonists ( $\mu/\delta$ ,  $\mu$ OR-NRI, NOP/ $\mu$ OR) and G protein-biased ligand. The  $\kappa$  opioid agonists can offer effective analgesia without causing typical opioid-related adverse effects, which means selective  $\kappa$  opioid can serve agonists as analgesics (Assana et al., 2014; Grechko et al., 2016; Zaitseva et al., 2018). Similarly, the  $\delta$  opioid receptor ( $\delta$ OR) has also been identified as a potent target for the discovery of novel analgesic drugs, with less side effects and better analgesia in animal models (Desmeules et al., 1993; Fraser et al., 2000; Brandt et al., 2001; Mika et al., 2001). After studying the interaction between  $\mu$  receptor and other targets, researchers raised the concept of bifunctional  $\mu$  receptor agonists. They found that the  $\delta$  receptor agonists can enhance the analgesic effect of the  $\mu$  receptor agonists, while the  $\delta$  receptor antagonists can significantly improve or even completely block the side effects of addiction, tolerance, and respiratory inhibition caused by the  $\mu$  receptor agonists (Abdelhamid et al., 1991; Porreca et al., 1992). They also found that the activation of NOP can reduce the reinforcing effects by the way of decreasing dopamine levels (Liu et al., 2001). Additional study of opioid receptor function has demonstrated that G protein conjugation pathway of  $\mu$  opioid receptor mainly induces analgesic effect, while the  $\beta$ -arrestin pathway can lead to side effects such as respiratory depression, nausea, and constipation (Violin et al., 2014). These new concepts need to be tested with extensive follow-up studies, and computational study is an appropriate and efficient test.

The concept of computer-aided drug design (CADD) firstly appeared in 1981 (Van Drie, 2007) and immediately was used extensively in the drug discovery. Since then, CADD has intensively improved the efficiency of traditional high-throughput screening (HTS) method, and it has significantly lowered the cost while retaining the same level of lead compound discovery (Sliwoski et al., 2014). Nowadays, CADD methods have been widely used in the modern drug discovery and mechanism studies on all drug-targeted proteins. In this review, we will focus on the applications of computational approaches in these topics: (1) discovering opioid analgesics; (2) the molecular mechanisms of opioid subtype selectivity; (3) the mechanism of the opioid receptor activation, and specifically, the G protein-biased activation.

## COMPUTATIONAL APPROACHES IN THE DISCOVERY OF OPIOID ANALGESICS

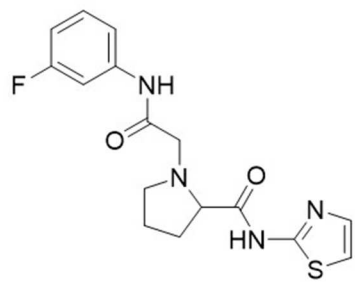
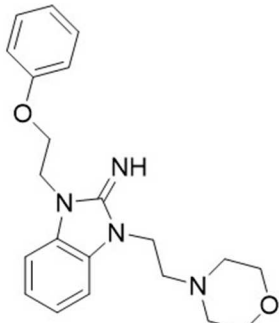
### Structure-Based Virtual Screening

Structure-based virtual screening (SBVS) starts from the three-dimensional (3D) structure of a target receptor to design potential active compounds (Cavasotto and Orry, 2007) through the molecular docking operations and reasonable scoring for a library of compounds. Due to various experimental difficulties in determining the structures of GPCR, the number of determined GPCR structures is still small, but keeps growing. As such, homology modeling offers a relatively reasonable strategy to build the model structure of the target receptor based on its amino acids sequence and an available homologous structure (as a template) of a related protein.

As a main tool, molecular docking commonly applied used in SBVS to evaluate the interactions between ligands and a particular receptor in order to rank the binding affinities of these ligands. The ligands are docked into the active site of a receptor through conformational search and a pre-built scoring function during the virtual screening. Researchers have developed many conformational search algorithms, including molecular dynamics (MD) simulation, systematical methods, Monte Carlo (MC) search, and genetic algorithms (GAs). Scoring functions include the molecular mechanics-based scoring functions, the empirical scoring functions, the knowledge-based scoring functions, and the consensus scoring functions (Sliwoski et al., 2014). For a particular receptor protein, both a feasible search algorithm and an accurate scoring function have to be carefully tested and selected.

Luckily, many efforts have been made in determining the crystal structures of opioid receptors. In 2012, studies revealed the inactive crystal structures of all four opioid receptor subtypes, i.e., the  $\mu$  receptor (Manglik et al., 2012), the  $\delta$  receptor (Granier et al., 2012), the  $\kappa$  receptor (Wu et al., 2012), and N/OFQ (Thompson et al., 2012) receptor. Later in 2015, the crystal structure of active  $\mu$  (O'Connor et al., 2015) and  $\kappa$  (Huang et al., 2015) opioid receptors were also resolved. In 2019,  $\delta$ OR's active crystal structure has been finally obtained (Claff et al., 2019). These crystal structures can be directly used for SBVS. Aiming at discovering potent non-addictive analgesics, Negri et al. (2013) virtually screened through docking 4,554,059 compounds of the ZINC database to inactive crystal structure of  $\kappa$  opioid receptor and found a selective novel agonist, MCKK-17S (Figure 1, top). In 2014, Daga et al. (2014) firstly applied SBVS to find novel hits for the nociception opioid receptor. In that study, they built an active state homology model of NOP based on the antagonist-bound crystal structure and refined it by the enrichment analysis. Shape-based approaches were also applied to improve the hit rate of the screening, and a new chemical scaffold was finally discovered. Among these newly discovered compounds, AT4 showed the best receptor affinity ( $K_i = 1.42 \pm 0.6 \mu\text{M}$ ) (Figure 1, bottom).

Although a number of crystal structures of opioid receptor have been reported, it is still a challenge to figure out the actual active conformation of the receptor when docking it with different ligands. It has been demonstrated that the affinities

	$K_i$ ( $\mu$ M)		
	$\kappa$ OR	$\delta$ OR	$\mu$ OR
 <b>MCKK-17</b>	17R: $2900 \pm 960$ 17S: $120 \pm 38$	$290 \pm 320$ >10000	$510 \pm 640$ >10000
 <b>AT-4</b>	$K_i$ ( $\mu$ M) NOP  $1.42 \pm 0.6$		

**FIGURE 1** | Structure of MCKK-17 and AT-4 found by structure-based virtual screening (SBVS) and their  $K_i$  to respective receptors.

predicted for the  $\mu$  opioid receptor by docking failed to significantly correlate with the reported experimental values (Cui et al., 2013). This indicates that the receptor may go through conformational transitions upon ligand binding. Thus, the results of SBVS based on docking may be some kind of misleading.

## Ligand-Based Virtual Screening

Ligand-based virtual screening (LBVS) method compares known active compounds of the target receptor with compounds in a database like SPECS and ZINC and select ligands with better activity. Compared to SBVS, the LBVS method is usually faster and does not depend on the information about the structure of the target receptor (Lemmen and Lengauer, 2000). Based on the complexity of ligand structure information that is used, LBVS can be divided into three classes, one-dimensional (1D), two-dimensional (2D), and 3D. Among them, 3D methods have better performance since it takes the conformational flexibility of compounds into consideration. However, 3D methods are challenging as the bioactive conformation of a ligand frequently does not match the lowest-energy conformation in many LBVS cases (Leach and Gillet, 2007). They can be further divided into five classes as atomic distance-based, surface-based, Gaussian function-based, field-based, and pharmacophore-based methods (Shin et al., 2015).

As early as in the year 2008, Zhang et al. (2009) have generated two pharmacophore models for the  $\kappa$ -agonists using

Catalyst/HypoGen and Phase respectively. These models could predict the structure-activity relationship and help to develop some new compounds. In the paper by Zhang et al., they also evaluated and compared the two models. These two models shared one hydrogen-bond receptor and one positive ionizable function and differ only in their definition of the hydrophobic point and aromatic ring features.

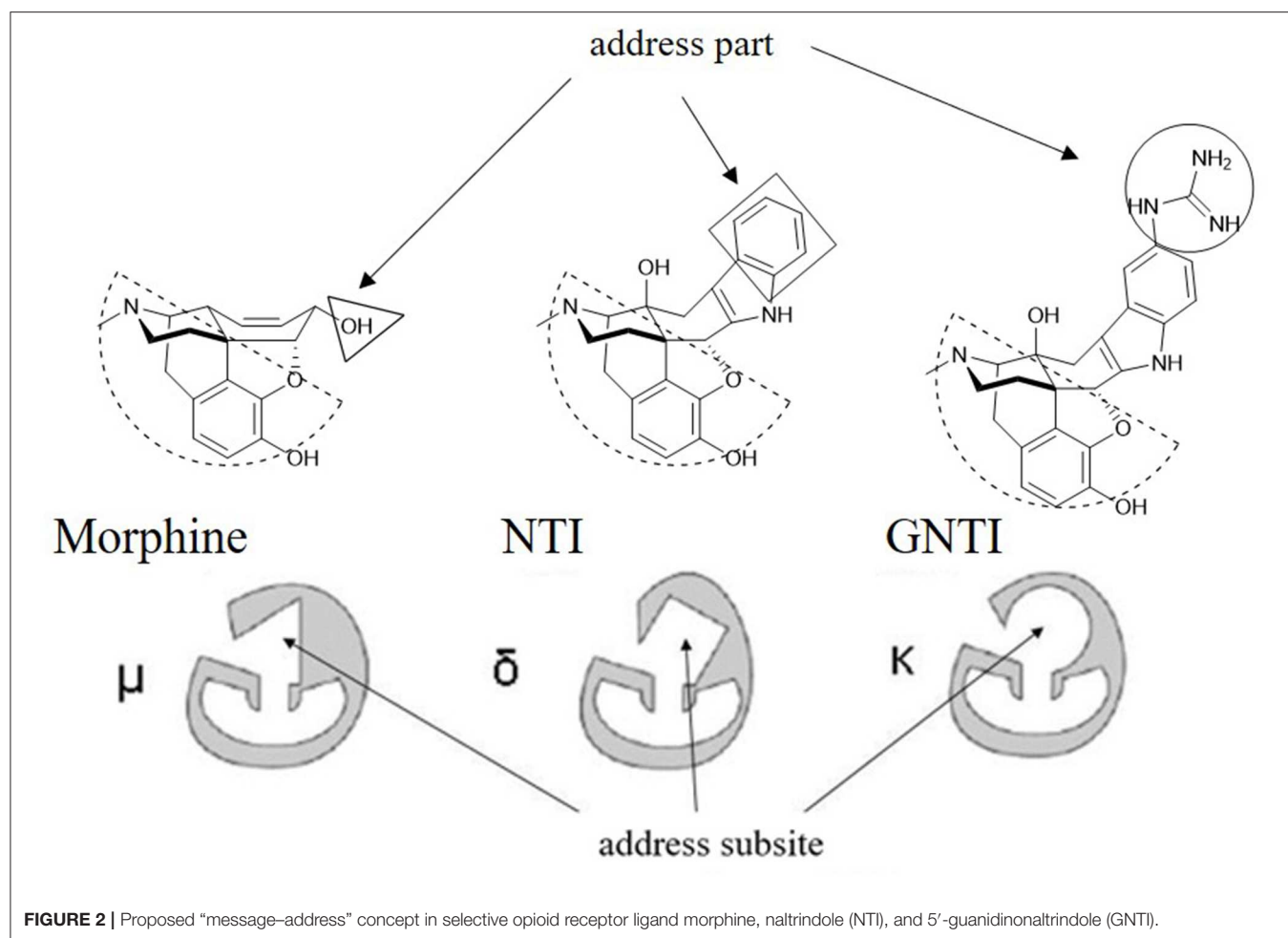
So far, the obvious weakness of LBVS method is its ability to identify the bioactive conformation of a ligand. As it also depends on the structure of the discovered active compounds, LBVS method cannot bring about very novel and diversified active structures.

## COMPUTATIONAL APPROACHES IN MECHANISM STUDY OF OPIOID RECEPTORS

### Molecular Docking to Illustrate Opioid Receptor Selectivity

As reported, the  $\kappa$  and  $\delta$  opioid receptors have been recognized as potent targets to develop analgesic drugs with less side effects. It has been found that some drugs in the clinic showed less side effects and are proved to interact with more than one subtype of opioid receptors. Compounds as butorphanol and buprenorphine are typical examples. To





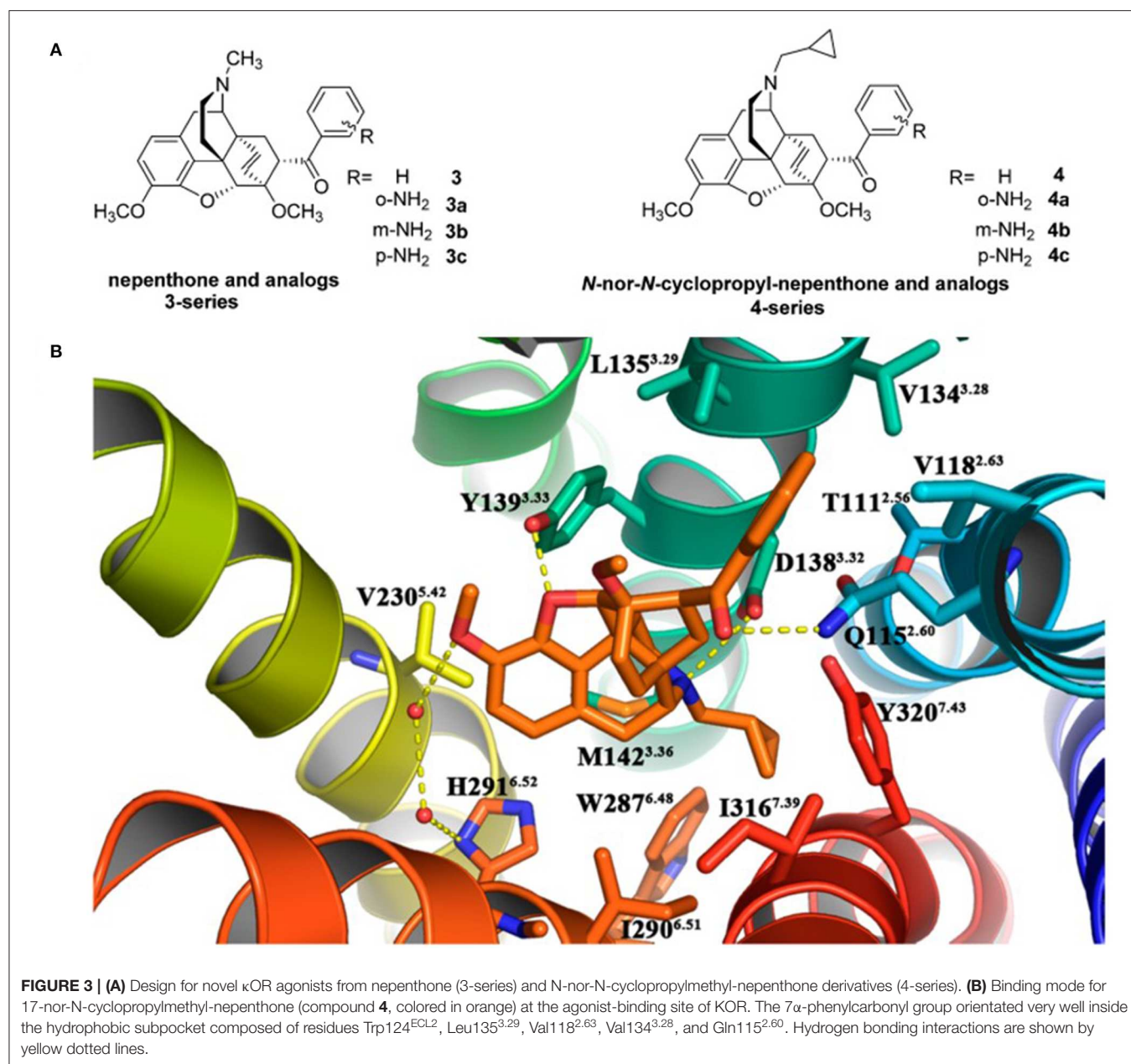
better understand the molecular mechanism of such multi-target ligands and subtype selectivity, researchers proposed the classic “message–address” concept to explore the binding mode between the opioid receptors and their ligands (Schwyzer, 1977). That is, the “message” part of the ligand is responsible for receptor recognition and affinity, while the “address” part of the ligand determines the selectivity and/or potency (Figure 2). It was successfully applied in the discovery of  $\delta$ OR antagonist naltrindole (Portoghese et al., 1990). The “message” and “address” parts interact with different residues at the active site of opioid receptor. Different subtypes have similar “message” subsites but totally different “address” subsites. This concept was widely used not only to explain the selectivity of active compounds but also to design better selective molecules through computer-aided molecular docking and/or MD simulations.

### The $\kappa$ Opioid Receptor

Recently, Li et al. (2017) designed and synthesized two new series of nepenthone derivatives, among which compound **4** displayed the highest affinity ( $K_i = 0.4 \pm 0.1$  nM) and the highest selectivity ( $\mu/\kappa = 339$ ,  $\delta/\kappa = 2,034$ ) toward  $\kappa$ OR

(Figure 3A). Molecular docking operations were carried out to explore how compound **4** binds with the  $\kappa$ OR (Figure 3B). The active  $\kappa$ OR structure was built based on the crystal structure of the active murine  $\mu$ OR. Compound **4** interacts with conserved residues D<sup>3.32</sup>, I<sup>6.51</sup>, M<sup>3.36</sup>, Y<sup>7.43</sup>, W<sup>6.48</sup>, and Y<sup>3.33</sup> at the active site of  $\kappa$ OR. The carbonyl compound **4** formed a hydrogen bond with residue Q<sup>2.60</sup>, which induces the 7 $\alpha$ -phenylcarbonyl group insert into the hydrophobic subpocket formed by residues W<sup>ECL2</sup>, L<sup>3.29</sup>, V<sup>2.63</sup>, V<sup>3.28</sup>, and Q<sup>2.60</sup> of the receptor.

Later on, Xiao et al. (2019) discovered a benzylamine derivative (compound **4**, SLL-039) as a highly potent and selective  $\kappa$  opioid agonist ( $\kappa$ ,  $K_i = 0.47$  nM,  $\kappa/\mu = 682$ ,  $\kappa/\delta = 283$ ) (Figure 4A). Also, molecular docking was used to explain the selectivity (Figure 4B). Molecular docking revealed three possible reasons for its high selectivity and activity. Firstly, the benzamide group could form a hydrogen bond with the Cys210<sup>EL2</sup> of the receptor, which was not available in the  $\mu$  opioid receptor. Secondly, the benzamide phenyl group of SLL-039 fit into a tight hydrophobic subpocket composed of Val118<sup>2.63</sup>, Trp124<sup>EL1</sup>, and Glu209<sup>EL2</sup> in the  $\kappa$ OR receptor, which is regarded as a novel “address” subsite. Finally, the relative structural rigidity



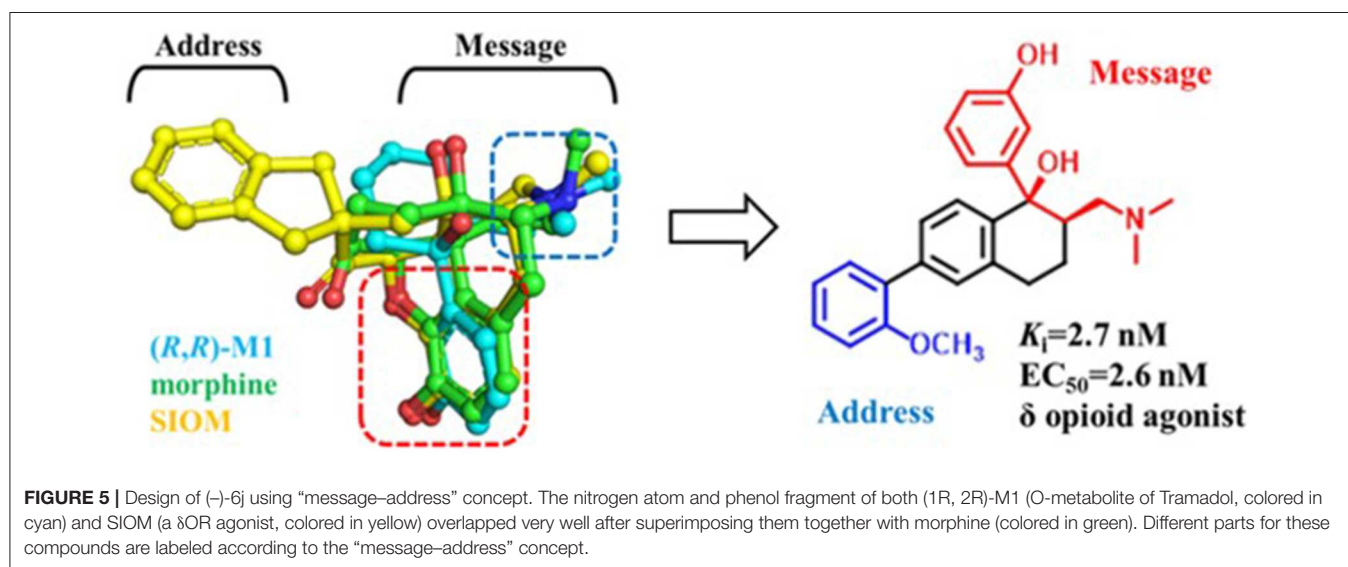
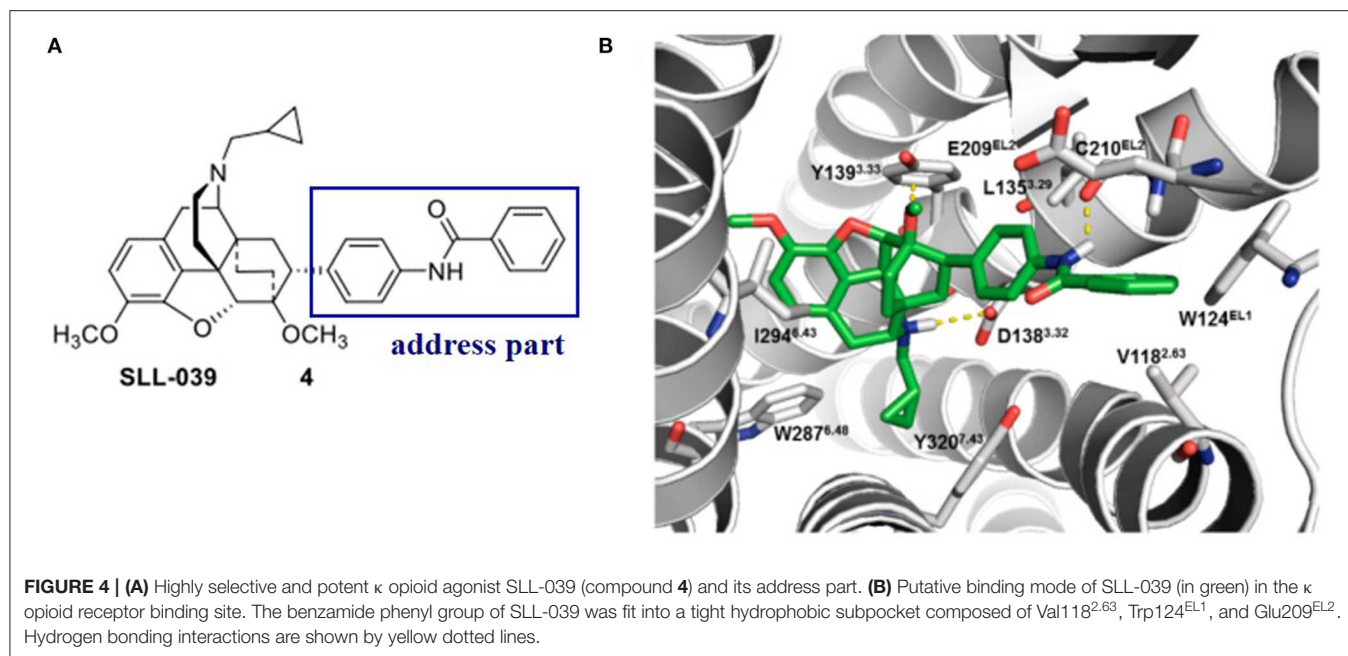
of benzamide carbonyl linker could produce conformational constraints to the structure and help it fit into the subpocket of  $\kappa$ OR but is not able to fit the corresponding sites in other receptor subtypes.

In these studies, molecular docking operations offered a visual modeling of the interaction between ligands and receptors and thus played a key role in exploring the molecular mechanism of the subtype selectivity of these newly discovered compounds toward  $\kappa$ OR.

### The $\delta$ Opioid Receptor

Similarly,  $\delta$ OR also has an “address” subsite which is different from that of  $\kappa$ OR. By revisiting the “message–address” concept,

Shen et al. (2016) designed a series of highly selective agonists of  $\delta$ OR from Tramadol. Among these new compounds, the compound (–)-6j displayed the best activity ( $EC_{50} = 2.6$  nM) and  $\delta$ OR selectivity (more than 1,000-fold) (Shen et al., 2016) (Figure 5). Through molecular docking operations, it was revealed that the subsite of the “message” part in (–)-6j, i.e., the protonated nitrogen atom formed a strong salt bridge with residue D<sup>3.32</sup> of  $\delta$ OR, while the phenolic group formed a strong hydrogen bond with the neighboring water molecule at the active site of the receptor. The “address” part of (–)-6j, the 6-ortho-methoxyphenyl, stretched into the hydrophobic pocket composed of residues L<sup>7.35</sup>, V<sup>6.55</sup>, and W<sup>6.58</sup> of  $\delta$ OR and formed cation– $\pi$  interaction with K<sup>5.39</sup> (Figure 6). The binding mode



of (-)-6j is consistent with the case of the  $\delta$ OR agonist SIOM (Portoghese et al., 1993). Overall, this study has proved that the “message–address” concept, together with the computational docking operations, can be applied successfully to design novel active compounds with better subtype selectivity.

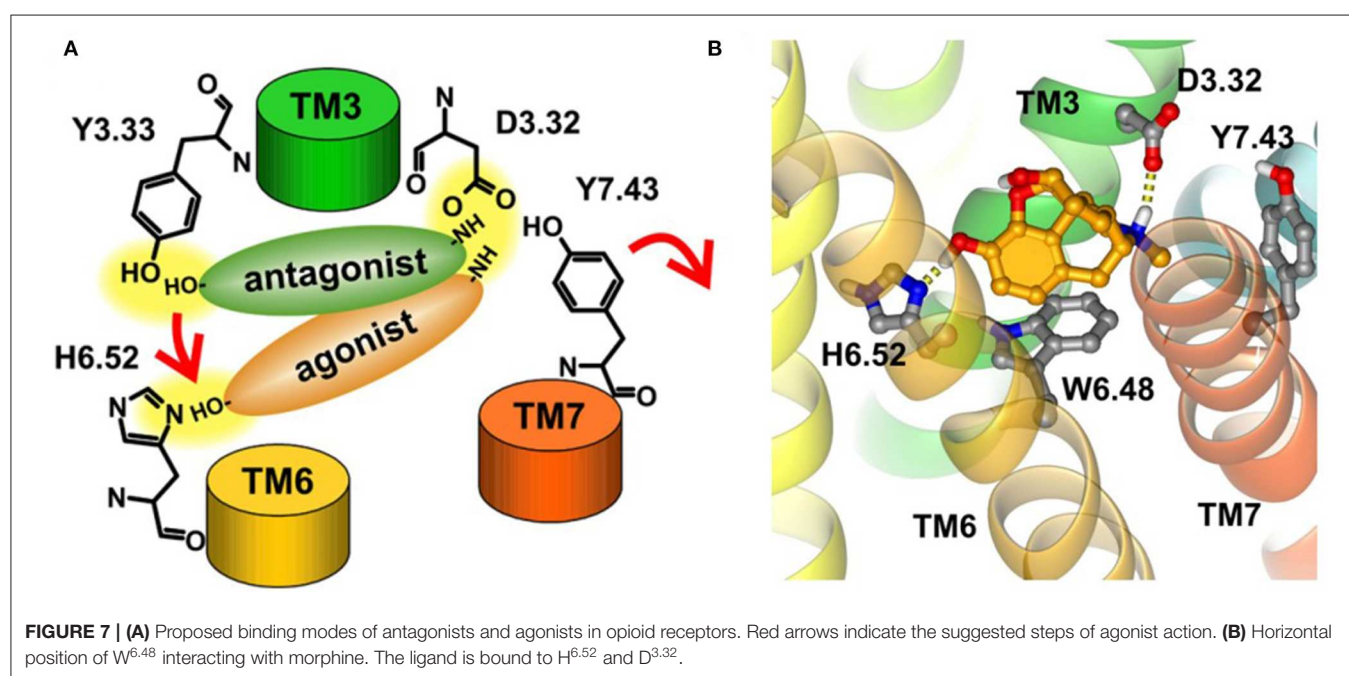
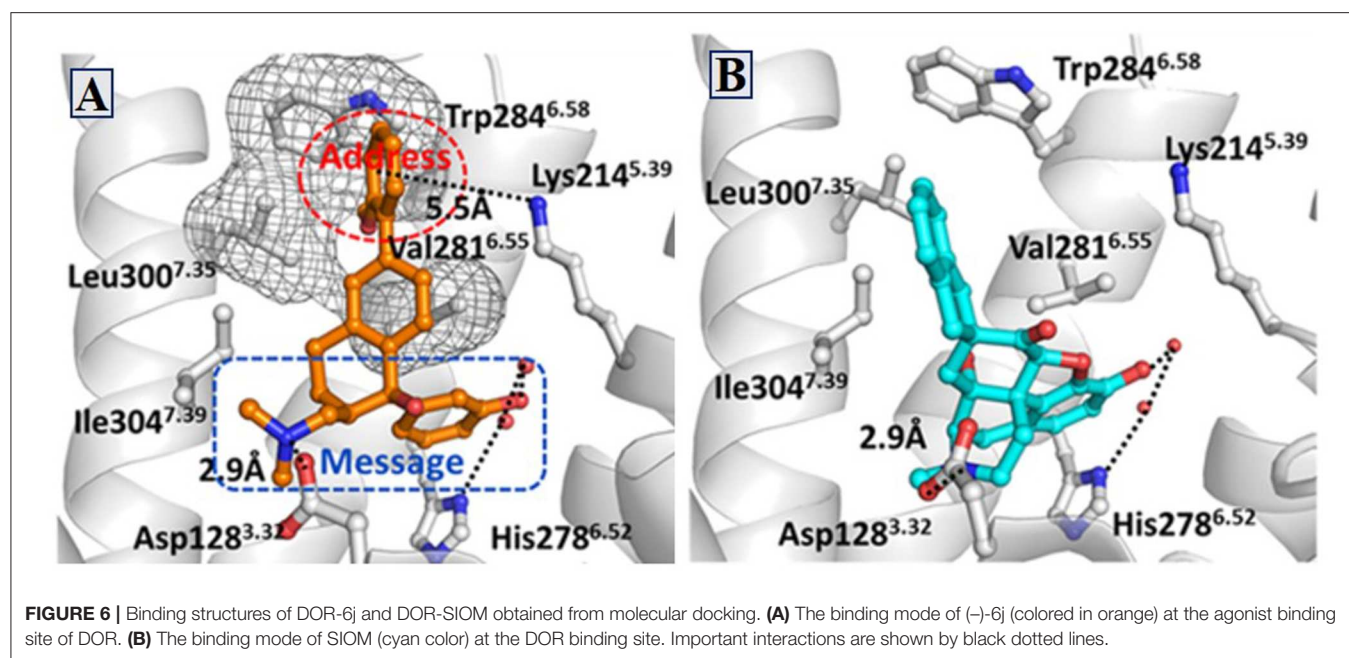
## Molecular Dynamic Simulations to Study the Activation Mechanism

GPCRs form an important nexus between the extracellular stimuli and intracellular second messengers. Their conformations change frequently between ground state (R) and active state (R\*) all the time. After binding with the agonist or antagonist, either the receptor's active state or its inactive

state will be stabilized, and the ratio between different states will be altered. However, how an agonist induces the receptor to become active state is still not very clear. Although a number of GPCR structures are now available in the protein data bank (PDB), it is still not easy to illuminate the molecular nature of the conformational changes.

With the advancement of computing algorithms an powerful resources, computational studies on GPCRs are going deeper and more visible in the research community. As one of the powerful tools, the MD simulations have been used to figure out more puzzles about the GPCR and their interactions with downstream proteins at atomic level. By using MD techniques, researches can simulate the conformational evolvement of the receptor





from a certain initial state on the basis of Newton mechanics and conduct the structural and functional studies. Up to now, many published studies have focused on the questions about the activation mechanism of opioid receptors (Kolinski and Filipek, 2008; Huang et al., 2015; Cheng et al., 2016; An et al., 2019). Among these studies, researchers are able to explore the biased activation mechanism toward a specific G protein (Schneider et al., 2016; Cheng et al., 2017).

### The Activation Mechanism of Opioid Receptor

Early in 2008, Kolinski and Filipek performed MD studies and analyses on the ligand–receptor complexes about two agonists (morphine and N-methyl-morphine), a selective antagonist  $\beta$ -funaltrexamine ( $\beta$ -FNA), and a non-selective antagonist naltrexone (NTX) through the trajectories of 20 ns MD simulations on the  $\mu$  opioid receptor. In this work, they revealed the very first steps of receptor activation. For all the analyzed



ligands, the protonated nitrogen atom was bound to D<sup>3.32</sup> of the receptor, but their phenolic (C3) OH group interacted with a different binding anchor site. It seemed that antagonists tended to bind to Y<sup>3.33</sup> residue, whereas agonists prefer H<sup>6.52</sup> residue of the receptor. This indicated that the agonists could penetrate deeper into the active site of the receptor structure and induce a series of conformational transitions. During this movement, the hydrogen bond between D<sup>3.32</sup> and Y<sup>7.43</sup> inside the receptor was broken (**Figure 7A**), and these two residues turn to other neighbor partners, respectively. These findings suggested that the D<sup>3.32</sup>-Y<sup>7.43</sup> interaction should be the important stabilizing factor for the inactive state of opioid receptors. In addition, the study also found that W<sup>6.48</sup> acted as a rotamer toggle switch for the receptor activation. During the MD simulations, it was found that the agonists changed the rotamer of W<sup>6.48</sup> to a horizontal position (perpendicular to TM6), while the antagonists maintained the initial vertical position of W<sup>6.48</sup> (**Figure 7B**). However, the previously reported TM3–TM6 links inside the receptor were firmly maintained in the binding of both the agonists and antagonists. The researchers also indicated that the rotamer configurations of C<sup>6.47</sup>, W<sup>6.48</sup>, and F<sup>6.52</sup> of the receptor were coupled with each other, modulating the bend angle of TM6 around the kink at P<sup>6.50</sup>. In 2015, Huang et al. (2015) indicated in their study the conformational movement of the transmembrane helix. At the cytoplasmic surface of the  $\mu$ OR, they observed large outward movement of TM6 (10 Å) and a slight inward shift of TM5 and TM7.

Meanwhile, Cheng et al. (2016) discovered that residues of E6.58 and I6.55 played a pivotal role in the activation of  $\kappa$ OR. Fluctuation of I6.55 triggered the conformational change, immediately induced the motion of E6.58, and finally led to the rearrangement of each transmembrane helix. To better investigate the conformation dynamics, in 2019, An et al. (2019) carried out long-time Gaussian accelerated MD (GaMD) simulations on  $\kappa$ OR binding with agonist MP1104 and antagonist JDTic. They found that the free of  $\kappa$ OR was the most stable, while the active  $\kappa$ OR-apo would gradually change into inactive state. When  $\kappa$ OR binds with agonists, some crucial motifs (DYNNM and CWXP) inside the receptor will be stabilized, increasing the ratio of active state in conformation equilibria. Antagonist binding with the receptor could not shake the inactive conformation, and these crucial motifs of the receptor were maintained well, although there was a stable intermediate (I) state between the active state and inactive state of the receptor (An et al., 2019).

### The Mechanism of G Protein-Biased Activation

It is interesting to explore the conformational differences of the same receptor toward the binding of different G proteins. That is, the G protein-biased activation. To investigate the mechanism of G protein-biased activation, Cheng et al. constructed five  $\mu$ OR systems: the G protein-biased agonists TRV130 and BU72, antagonists  $\beta$ -FNA and NTX, and the free receptor (Cheng et al., 2017). According to their study, W<sup>6.48</sup> and Y<sup>7.43</sup> of  $\mu$ OR were proposed as a paired activation switch. These residues may play an important role in the binding of  $\beta$ -arrestin, thus regulating G protein signal transduction. As the first representative  $\mu$ OR

ligand with G protein bias, TRV130 was found to directly interact with Y<sup>7.43</sup> and make it closer to W<sup>6.48</sup>. Such interaction would stabilize the rotamer of W<sup>6.48</sup> side chain, mostly at  $-70^\circ$  and downregulate the binding of  $\beta$ -arrestin. Also, W<sup>7.35</sup> was observed to be stabilized by a hydrophobic interaction. All these findings have been validated by experimental mutation studies (Hothersall et al., 2017; Sun et al., 2017).

Unlike traditional molecular docking, MD simulations can take good care of the conformational changes of the receptor. The only constraint of MD is the huge demand of computing resources. For a particular opioid receptor, including the G protein into the simulating system will definitely better simulate the whole process of ligand binding and receptor activation. However, it will also add about 50,000 atoms into the system, making the system very huge. Thus, such computing effort is very possibly undergoing in research labs, but still in the air of such publications. Also, it has been indicated that the timescale of activation for a GPCR receptor could be at the scale of microsecond or seconds (Dror et al., 2011), which will hopefully be revealed soon to the research community.

## PERSPECTIVES

Scientists have been making many research efforts to discover analgesics with less side effects, although the way looks not short. As experimental techniques become more matured, more and more crystal structures of opioid receptors and other GPCRs are quickly coming out. The availability of these new crystal structures in PDB at high resolutions is stimulating new rounds of virtual screening and activation mechanism study. Also, as the timescale of current MD simulations is increasing rapidly, some important conformational changes missed in the previous explorations may be captured and interpreted more reasonably for the opioid receptors in the near future. In addition, a number of new techniques, like enhanced sampling of MD simulations, have been developed and effectively applied to the studies on GPCRs. These techniques include targeted MD (Schlitter et al., 1994), steered MD (Isralewitz et al., 2001), and accelerated MD (Hamelberg et al., 2004; Miao et al., 2015). It is reasonable to expect that the molecular mechanisms of opioid receptors will be investigated and demonstrated more completely in the next couple of years.

## AUTHOR CONTRIBUTIONS

BZ and WL wrote the manuscript. LS and WF revised the manuscript. All authors were involved in the preparation of the manuscript and approved the final version.

## FUNDING

This work was financially supported by research grants from the National Natural Science Foundation of China (No. 81773635) and the Shanghai Science and Technology Development Funds (14431900500).

## ACKNOWLEDGMENTS

We are very grateful to Dr. Xiaoqin Huang from the Center of Research Computing and the Center for Theoretical Biological

Physics of Rice University in the USA for many useful discussions on the selection of review sections and the modification of this manuscript.

## REFERENCES

- Abdelhamid, E. E., Sultana, M., Portoghese, P. S., and Takemori, A. E. (1991). Selective blockage of delta opioid receptors prevents the development of morphine tolerance and dependence in mice. *J. Pharmacol. Exp. Ther.* 258, 299–303. doi: 10.1016/0160-5402(91)90033-2
- An, X., Bai, Q., Bing, Z., Zhou, S., Shi, D., Liu, H., et al. (2019). How does agonist and antagonist binding lead to different conformational ensemble equilibria of the  $\kappa$ -opioid receptor: insight from long-time gaussian accelerated molecular dynamics simulation. *ACS Chem. Neurosci.* 10, 1575–1584. doi: 10.1021/acscchemneuro.8b00535
- Assana, C., Feller, C., Fogel, M., Frelka, A., Gottfried, S., Jaber, R., et al. (2014). Kappa opioid receptor-based treatments for pain. *FASEB J.* 28:LB88.
- Brandt, M. R., Furness, M. S., Mello, N. K., Rice, K. C., and Negus, S. S. (2001). Antinociceptive effects of delta-opioid agonists in rhesus monkeys: effects on chemically induced thermal hypersensitivity. *J. Pharmacol. Exp. Ther.* 296, 939–946. doi: 10.1016/S1056-8719(01)00154-X
- Cavasotto, C. N., and Orry, A. J. (2007). Ligand docking and structure-based virtual screening in drug discovery. *Curr. Top. Med. Chem.* 7, 1006–1014. doi: 10.2174/156802607780906753
- Cheng, J., Cheng, T., Li, W., Liu, G., Zhu, W., and Tang, Y. (2017). Computational insights into the G-protein-biased activation and inactivation mechanisms of the  $\mu$  opioid receptor. *Acta Pharmacol. Sin.* 39, 154–164. doi: 10.1038/aps.2017.158
- Cheng, J., Sun, X., Li, W., Liu, G., Tu, Y., and Tang, Y. (2016). Molecular switches of the kappa opioid receptor triggered by 6'-GNTI and 5'-GNTI. *Sci. Rep.* 6:18913. doi: 10.1038/srep18913
- Claff, T., Yu, J., Blais, V., Patel, N., Martin, C., Wu, L., et al. (2019). Elucidating the active  $\delta$ -opioid receptor crystal structure with peptide and small-molecule agonists. *Sci. Adv.* 5:eaax9115. doi: 10.1126/sciadv.aax9115
- Cui, X., Yeliseev, A., and Liu, R. (2013). Ligand interaction, binding site and G protein activation of the mu opioid receptor. *Eur. J. Pharmacol.* 702, 309–315. doi: 10.1016/j.ejphar.2013.01.060
- Daga, P. R., Polgar, W. E., and Zaveri, N. T. (2014). Structure-based virtual screening of the nociceptin receptor: hybrid docking and shape-based approaches for improved hit identification. *J. Chem. Inform. Model.* 54, 2732–2743. doi: 10.1021/ci500291a
- Desmeules, J. A., Kayser, V., and Guilbaud, G. (1993). Selective opioid receptor agonists modulate mechanical allodynia in an animal-model of neuropathic pain. *Pain* 53, 277–285. doi: 10.1016/0304-3959(93)90224-D
- Dror, R. O., Arlow, P., Maragakis, T. J., Mildorf, A. C., Pan, H., Xu, D. W., et al. (2011). Activation mechanism of the  $\beta_2$ -adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18684–18689. doi: 10.1073/pnas.1110499108
- Fraser, G. L., Gaudreau, G. A., Clarke, P. B. S., Menard, D. P., and Perkins, M. N. (2000). Antihyperalgesic effects of delta opioid agonists in a rat model of chronic inflammation. *Br. J. Pharmacol.* 129, 1668–1672. doi: 10.1038/sj.bjp.0703248
- Granier, S., Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Weis, W. I., et al. (2012). Structure of the delta-opioid receptor bound to naltrindole. *Nature* 485, 400–404. doi: 10.1038/nature11111
- Grechko, O. Y., Spasov, A. A., and Shtareva, D. M. (2016). Opioid kappa receptors as a molecular target for the creation of a new generation of analgesic drugs. *Pharm. Chem. J.* 50, 1–9. doi: 10.1007/s11094-016-1388-z
- Hamelberg, D., Mongan, J., and McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120, 11919–11929. doi: 10.1063/1.1755656
- Hothersall, J. D., Torella, R., Humphreys, S., Hooley, M., Brown, A., McMurray, G., et al. (2017). Residues W320 and Y328 within the binding site of the muopioid receptor influence opiate ligand bias. *Neuropharmacology* 118, 46–58. doi: 10.1016/j.neuropharm.2017.03.007
- Huang, W., Manglik, A., Venkatakrishnan, A. J., Laeremans, T., Feinberg, E. N., Sanborn, S. L., et al. (2015). Structural insights into  $\mu$ -opioid receptor activation. *Nature* 524, 315–321. doi: 10.1038/nature14886
- Israelowitz, B., Gao, M., and Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* 11, 224–230. doi: 10.1016/S0959-440X(00)00194-9
- King, N., Hittinger, C. T., and Carroll, S. B. (2003). Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* 301, 361–363. doi: 10.1126/science.1083853
- Kolinski, M., and Filipek, S. (2008). Molecular dynamics of  $\mu$  opioid receptor complexes with agonists and antagonists. *Open Struct. Biol. J.* 2, 8–20. doi: 10.2174/1874199100802010008
- Leach, A. R., and Gillet, V. J. (2007). *An Introduction to Chemoinformatics, Revised Edn.* Dordrecht: Springer, p. 27. doi: 10.1007/978-1-4020-6291-9
- Lemmen, C., and Lengauer, T. (2000). Computational methods for the structural alignment of molecules. *J. Comput. Aided Mol. Des.* 14, 215–232. doi: 10.1023/A:1008194019144
- Li, W., Long, J., Qian, Y., Long, Y., Xu, X., Wang, Y., et al. (2017). The pharmacological heterogeneity of nepenthone analogs in conferring highly selective and potent  $\kappa$ -opioid agonistic activities. *ACS Chem. Neurosci.* 8, 766–776. doi: 10.1021/acscchemneuro.6b00321
- Liu, Z., Wang, Y., Zhang, J., Ding, J., Guo, L., Cui, D., et al. (2001). Orphanin FQ: an endogenous antagonist of rat brain dopamine transporter. *Neuroreport* 12, 699–702. doi: 10.1097/00001756-200103260-00017
- Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., et al. (2012). Crystal structure of the mu-opioid receptor bound to a morphinan antagonist. *Nature* 485, 321–326. doi: 10.1038/nature10954
- Miao, Y., Feher, V. A., and McCammon, J. A. (2015). Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J. Chem. Theor. Comput.* 11, 3584–3595. doi: 10.1021/acs.jctc.5b00436
- Mika, J., Przewlocki, R., and Przewlocka, B. (2001). The role of delta-opioid receptor subtypes in neuropathic pain. *Eur. J. Pharmacol.* 415, 31–37. doi: 10.1016/S0014-2999(01)00814-7
- Negri, A., Rives, M. L., Caspers, M. J., Prisinzano, T. E., Javitch, J. A., and Filizola, M. (2013). Discovery of a novel selective kappa-opioid receptor agonist using crystal structure-based virtual screening. *J. Chem. Inform. Model.* 53, 521–526. doi: 10.1021/ci400019t
- O'Connor, C., White, K. L., Doncescu, N., Didenko, T., Roth, B. L., Czaplicki, G., et al. (2015). NMR structure and dynamics of the agonist dynorphin peptide bound to the human kappa opioid receptor. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11852–11857. doi: 10.1073/pnas.1510117112
- Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006). How many drug targets are there. *Nat. Rev. Drug. Discov.* 5, 993–996. doi: 10.1038/nrd2199
- Porreca, F., Takemori, A. E., Sultana, M., Portoghese, P. S., Bowen, W. D., and Mosberg, H. I. (1992). Modulation of mu-mediated antinociception in the mouse involves opioid delta-2 receptors. *J. Pharmacol. Exp. Ther.* 263, 147–152. doi: 10.1007/BF01061471
- Portoghese, P. S., Moe, S. T., and Takemori, A. E. (1993). A selective delta-1 opioid receptor agonist derived from oxymorphone - Evidence for separate recognition sites for delta-1 opioid receptor agonists and antagonists. *J. Med. Chem.* 36, 2572–2574. doi: 10.1021/jm00069a017
- Portoghese, P. S., Sultana, M., and Takemori, A. E. (1990). Design of peptidomimetic delta-opioid receptor antagonists using the message address concept. *J. Med. Chem.* 33, 1714–1720. doi: 10.1021/jm00168a028
- Schlitter, J., Engels, M., and Krüger, P. (1994). Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graph.* 12, 84–89. doi: 10.1016/0263-7855(94)80072-3

- Schneider, S., Provasi, D., and Filizola, M. (2016). How oliceridine (TRV-130) binds and stabilizes a  $\mu$ -opioid receptor conformational state that selectively triggers G protein signaling pathways. *Biochemistry* 55, 6456–6466. doi: 10.1021/acs.biochem.6b00948
- Schwyzler, R. (1977). ACTH: a short introductory review. *Ann. N.Y. Acad. Sci.* 297, 3–26. doi: 10.1111/j.1749-6632.1977.tb41843.x
- Shen, Q., Qian, Y., Huang, X., Xu, X., Li, W., Liu, J., et al. (2016). Discovery of potent and selective agonists of  $\delta$  opioid receptor by revisiting the “message-address” concept. *ACS Med. Chem. Lett.* 7, 391–396. doi: 10.1021/acsmedchemlett.5b00423
- Shin, W. H., Zhu, X., Bures, M. G., and Kihara, D. (2015). Three-dimensional compound comparison methods and their application in drug discovery. *Molecules* 20, 12841–12862. doi: 10.3390/molecules200712841
- Simon, E. J., and Hiller, J. M. (1978). The opiate receptors. *Annu. Rev. Pharmacol. Toxicol.* 18, 371–394. doi: 10.1007/bf00965627
- Sliwoski, G., Kothiwale, J., and Meiler, E. W. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi: 10.1016/j.vascn.2010.02.005
- Snyder, S. H. (1977). Opiate receptors in the brain. *N. Engl. J. Med.* 296, 266–271. doi: 10.1056/NEJM197702032960511
- Sun, X., Laroche, G., Wang, X., Agren, H., Bowman, G. R., Giguere, P. M., et al. (2017). Propagation of the allosteric modulation induced by sodium in the delta-opioid receptor. *Chemistry* 23, 4615–4624. doi: 10.1002/chem.201605575
- Thompson, A. A., Liu, W., Chun, E., Katritch, V., Wu, H., Vardy, E., et al. (2012). Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485, 395–399. doi: 10.1038/nature11085
- Van Drie, J. H. (2007). Computer-aided drug design: the next 20 years. *J. Comput. Aided Mol. Des.* 21, 591–601. doi: 10.1007/s10822-007-9142-y
- Violin, J. D., Crombie, A. L., Soergel, D. G., and Lark, M. W. (2014). Biased ligands at G-protein-coupled receptors: promise and progress. *Trends Pharmacol. Sci.* 35, 308–316. doi: 10.1016/j.tips.2014.04.007
- Wang, Y., Mao, X., Li, T., Gong, N., and Zhang, M. (2017). Dezocine exhibits antihypersensitivity activities in neuropathy through spinal  $\mu$ -opioid receptor activation and norepinephrine reuptake inhibition. *Sci. Rep.* 7:43137. doi: 10.1038/srep43137
- Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G. W., Vardy, E., et al. (2012). Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* 485, 327–332. doi: 10.1038/nature10939
- Xiao, L., Wang, Y., Zhang, M., Wu, W., Kong, L., Ma, Y., et al. (2019). Discovery of a highly selective and potent  $\kappa$  opioid receptor agonist from N-Cyclopropylmethyl-7 $\alpha$ -phenyl-6,14-endoethanotetrahydronorthebaines with Reduced Central Nervous System (CNS) Side Effects Navigated by the Message-Address Concept. *J. Med. Chem.* 62, 11054–11070. doi: 10.1021/acs.jmedchem.9b00857
- Zaitseva, N. I., Galan, S. E., and Pavlova, L. A. (2018). Prospects of a search for kappa-opioid receptor agonists with analgesic activity (review). *Pharm. Chem. J.* 51, 843–851. doi: 10.1007/s11094-018-1703-y
- Zhang, J., Liu, G., and Tang, Y. (2009). Chemical function-based pharmacophore generation of selective  $\kappa$ -opioid receptor agonists by catalyst and phase. *J. Mol. Model.* 15, 1027–1041. doi: 10.1007/s00894-008-0418-5

**Conflict of Interest:** LS was employed by Shijiazhuang No. 4 Pharmaceutical Co., Ltd.

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

Copyright © 2020 Zhao, Li, Sun and Fu. 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.



# Molecular Determinants Elucidate the Selectivity in Absciscic Acid Receptor and HAB1 Protein Interactions

Jing-Fang Yang<sup>3,4</sup>, Chun-Yan Yin<sup>2</sup>, Di Wang<sup>3,4</sup>, Chen-Yang Jia<sup>3,4</sup>, Ge-Fei Hao<sup>1\*</sup> and Guang-Fu Yang<sup>3,4,5\*</sup>

<sup>1</sup> State Key Laboratory Breeding Base of Green Pesticide and Agricultural Bioengineering, Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Research and Development Center for Fine Chemicals, Guizhou University, Guiyang, China, <sup>2</sup> School of Life Science, Wuchang University of Technology, Wuhan, China, <sup>3</sup> Key Laboratory of Pesticide & Chemical Biology, Ministry of Education, College of Chemistry, Central China Normal University, Wuhan, China, <sup>4</sup> International Joint Research Center for Intelligent Biosensor Technology and Health, Central China Normal University, Wuhan, China, <sup>5</sup> Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China

## OPEN ACCESS

### Edited by:

Xiao Jun Yao,  
Macau University of Science and  
Technology, Macau

### Reviewed by:

Tingjun Hou,  
Zhejiang University, China  
Qing-Chuan Zheng,  
Jilin University, China

### \*Correspondence:

Ge-Fei Hao  
gefei\_hao@foxmail.com  
Guang-Fu Yang  
gfyang@mail.ccnu.edu.cn

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 26 February 2020

Accepted: 23 April 2020

Published: 04 June 2020

### Citation:

Yang J-F, Yin C-Y, Wang D, Jia C-Y,  
Hao G-F and Yang G-F (2020)  
Molecular Determinants Elucidate the  
Selectivity in Absciscic Acid Receptor  
and HAB1 Protein Interactions.  
Front. Chem. 8:425.  
doi: 10.3389/fchem.2020.00425

The abscisic acid (ABA), as a pivotal plant hormone, plays a key role in controlling the life cycle and adapting to the environmental stresses. The receptors of ABA are the Pyrabactin resistance/Pyrabactin resistance-like/regulatory component of ABA receptors (PYR/PYL/RCAR, PYLs for simplicity), which regulate the protein phosphatase 2Cs (PP2Cs) in the signal pathway. As an important ABA-mimicking ligand, Pyrabactin shows the activation function to parts of members of PYLs, such as PYR1 and PYL1. Due to the antagonism of Pyrabactin to PYL2, it was used as a probe to discover a part of ABA receptors. Since then, many researchers have been trying to find out the determinants of the selective regulation of PYLs and PP2Cs interaction. However, the roles of residues on the selective regulation of PYR1/PYL2 and PP2Cs interaction induced by Pyrabactin are still ambiguous. This research investigated the selective activation mechanism of Pyrabactin through the sequence alignment, molecular docking, molecular dynamics simulation, and binding free energy calculation. Furthermore, the electrostatic and hydrophobic interaction differences induced by Pyrabactin and agonists were compared. The results indicate that Leu137/Val114, Ser85/Ser89, and Gly86/Gly90 from the pocket and gate of PYR1/PYL2 are the vital residues for the selective activation of Pyrabactin. Meanwhile, the electrostatic interaction between PP2Cs and PYLs complexed with agonists was improved. This mechanism provides strong support for the design of selective agonists and antagonists.

**Keywords:** abscisic acid, PYR/PYL/RCARs, pyrabactin, selectivity, molecular dynamics

**Abbreviations:** ABA, abscisic acid; PYR/PYL/RCARs, the Pyrabactin resistance/Pyrabactin resistance-like/regulatory component of ABA receptors; PP2Cs, protein phosphatase 2Cs; SnRK2s, SNF1-related protein kinase 2s; ELE, electrostatic; MD, molecular dynamics; MM/GBSA, the molecular mechanics/generalized born surface area; RMSD, the atomic root mean square deviations; STD, the standard deviation; VDW, van der Waals.



## INTRODUCTION

Drought stress is a severe abiotic stress, which may lower the crop yield of the world (Vishwakarma et al., 2017; Zhang et al., 2018). The phytohormone abscisic acid (ABA), as an antistress regulator, has aroused widespread concern (Park et al., 2009). Indeed, it not only regulates many aspects of plant growth and development but also responds to environmental stresses (e.g., drought, salinity, cold, and UV radiation) (Verma et al., 2016). Generally speaking, ABA-deficient plants show defects in stomatal regulation, seed dormancy, and germination (Finkelstein, 2013). Due to the essentiality of ABA in the plants, the discoveries of ABA receptor protein and the core signaling complexes may transmit cues for the understanding of subsequent molecular events and plant antistress phenotype (Sah et al., 2016).

The bona fide ABA receptor and the regulation mechanism were discovered and understood in 2009 (Ma et al., 2009; Park et al., 2009; Pennisi, 2009). The receptors of ABA are Pyrabactin resistance/Pyrabactin resistance-like/regulatory component of ABA receptors (PYR/PYL/RCARs, simplified as PYLs). There are 14 members in *Arabidopsis thaliana*, consisting of PYR1, PYL1–PYL13 (Yang et al., 2019). Besides, PYLs contain a binding pocket with a loop as gate, and it is closed in response to ABA (Melcher et al., 2009), which will create a binding surface outside the gate loop for protein phosphatase 2Cs (PP2Cs), including ABI1, ABI2, HAB1, HAB2, and PP2CA. Then, SNF1-related protein kinase 2s (SnRK2s) are activated to phosphorylate downstream effectors (Yamaguchi-Shinozaki and Shinozaki, 2006; Melcher et al., 2009). As a gate-lock mechanism, it offers a key clue for the discovery of ABA mimic molecule and plays a pivotal role in revealing some specific regulation pathways.

ABA mimic molecules, which function as agonists or antagonists of PYLs and have the potential application in agriculture, may deepen the knowledge of the signaling and promote the study of the ABA receptor. Hence, there is a wide interest in the discovery of ABA mimic molecules. Pyrabactin, an early synthetic ABA mimic, functions as an agonist of PYR1 and PYL1 but as an antagonist of PYL2 (Park et al., 2009; Yuan et al., 2010). It is sufficient to activate ABA responses in seeds but yields a minimal response in vegetative tissues and cannot make plants survive from drought tolerance (Okamoto et al., 2013). Another selective agonist toward a part of PYLs (PYR1, PYL1–PYL3, and PYL5), Quinabactin (also known as AM1), is a promising agrochemical to elicit stomatal closure and enhance crop drought resistance (Cao et al., 2013; Okamoto et al., 2013). After structure optimization, the fluoro-substitution compound (AMF4) was synthesized. AMF4 has a long-lasting effect to promote the stomatal closure, induce the expression of stress-responsive genes, and activate the same PYLs as AM1 (Cao et al., 2017). Overall, Pyrabactin is the only selective agonist of PYR1 and PYL2, and it is rotated for 90° in PYL2 relative to that in PYR1, which is not found for AM1 and AMF4. Because of the significance of Pyrabactin for the selective activation of PYLs, a lot of research have focused on its distinct selectivity to PYLs. It is believed that some important amino acids, such as Val67 and Val114 of PYL2, are important for the selectivity (Peterson

et al., 2010; Yuan et al., 2010). However, the dynamics roles and the energy determinants of the residues on the PYLs–PP2Cs interface remain unclear, which are crucial for understanding the selectivity and performing precise ABA mimic molecule design.

Therefore, in this study, the complex structures of HAB1 and PYR1/PYL2 with ABA, AM1, AMF4, and Pyrabactin binding were used to perform molecular simulation and reveal the selective activation determinants of PYLs. Additionally, the sequence alignment, structures comparison, molecular dynamics (MD) simulation, and binding free energy and decomposition were performed based on public tools and our protocols (Hao et al., 2016; Yang et al., 2020). Beyond that, the interactions between PYR1/PYL2 and HAB1 were analyzed and decomposed. Finally, it is found that the electrostatic (ELE) and hydrophobic interactions between PYR1/PYL2 and HAB1 induced by agonists are conserved. At the same time, pocket residues Leu137/Val114 and gate residues Ser85/89 and Gly86/90 of PYR1/PYL2 are the key residues for selectivity of Pyrabactin. More importantly, this finding provides guidance for the design of agonists and antagonists of PYLs.

## METHOD

### Sequences Alignment

There are already a large number of tools available to analyze sequences and structures, such as ClustalW (Thompson et al., 1994), Muscle (Edgar, 2004), T-coffee (Di Tommaso et al., 2011), and Dialign (Al Ait et al., 2013). Discovery Studio is a popular commercial software to perform combined analyses. Therefore, 14 *Arabidopsis thaliana* PYLs sequences downloaded from the National Center for Biotechnology information (NCBI) were aligned using an automatic tool in Discovery Studio 2.5 (BIOVIA Discovery Studio, 2009).

### Molecular Docking and Structures Preparation

The crystal structures of PYR1 (ABA)-HAB1 (PDB ID: 3QN1) (Dupeux et al., 2011), PYL1 (Pyrabactin)-ABI1 (PDB ID: 3NMN) (Melcher et al., 2010), PYL2 (ABA)-HAB1 (PDB ID: 3KB3) (Melcher et al., 2009), PYL2 (AM1)-HAB1 (PDB ID: 4LA7) (Okamoto et al., 2013), and PYL2 (AMF4)-HAB1 (PDB ID: 5VSR) (Cao et al., 2017) downloaded from the Protein Data Bank (PDB) database were superimposed and analyzed by Pymol software (Berman et al., 2000; DeLano, 2002), some of which were used for molecular docking and MD.

AutoDock 4.2 was applied to molecular docking (Morris et al., 2009). In this study, AM1, AMF4, and Pyrabactin were docked into the crystal structure of PYR1-HAB1 (PDB ID: 3QN1) (Yuan et al., 2010) to generate initial structures for MD. The Lamarckian genetic algorithm (LGA) was applied for the conformational search of the ligand (Morris et al., 1998). The grid size was set as 40 × 40 × 40, and the grid space was set to the default value of 0.375 Å. A total of 256 runs were launched for each compound. Most of the parameters for the docking calculation were set to the default values recommended by the software.

To reveal the conformational selective mechanism of Pyrabactin to PYR1 and PYL2, the structures of PYR1

(Pyrabactin)-HAB1 from docking and crystal structure PYL2 (Pyrabactin) (PDB ID: 3NS2) were superimposed. Then, the Pyrabactin was extracted from two complexes and saved with the other ABA receptor to get the complexes PYR1 (Pyrabactin) and PYL2 (Pyrabactin) with the initial Pyrabactin structure from PYL2 (Pyrabactin) and PYR1 (Pyrabactin)-HAB1, respectively. Meanwhile, HAB1 was saved with the PYL2 (Pyrabactin) to build up the initial structure of PYL2 (Pyrabactin)-HAB1 for later MD.

## MD Simulation and Trajectory Analysis

MD simulations were performed using the AMBER 16 software package with the ff14SB force field (Case et al., 2016). The ligand electrostatic potentials were computed at the HF/6-31G\* level in the Gaussian 03 program (Frisch et al., 2004). The RESP fitting technique in AMBER was used to determine the partial charges (Wang et al., 2000). The force-field parameters for the ligands were generated with the general AMBER force field (gaff) by the Antechamber program (Case et al., 2005). Each complex was immersed in a cubic box of the TIP3P water model with an 8.0-Å minimum solute-wall distance. Na<sup>+</sup> or Cl<sup>-</sup> ions were added to neutralize each complex system.

The complex systems were optimized before the simulation as follows. First, the movement was allowed only for hydrogen atoms. Next, the side chains were relaxed. Finally, all atoms were permitted to move freely. In each stage, energy minimization was executed by the steepest descent method for the first 1,000 steps and the conjugate gradient method for the subsequent 2,000 steps. After that, the systems were set up to obtain stable MD trajectories. Complex systems were gradually heated from 10 to 300 K in 200 ps, and more than 500 ps equilibrating calculation was executed at 1 atm and 300 K with applying periodic boundary conditions in the NPT ensemble to avoid edge effects. The 8 ns MD simulation of each system was performed. The snapshots extracted at every picosecond of the stable interval from the last 6 ns production MD trajectory using the CPPTRAJ module of AMBER were used for structural and energetic analysis. Meanwhile, the snapshots from the last 6 ns MD simulation processed were used to detect hydrogen bonds. The systems and timescales for all molecular dynamics simulations are listed in the Table S1.

## Binding Energy and Decomposition Calculation

For each snapshot, the binding energy ( $\Delta H$ ) of the protein (ligand)-protein complex was calculated by the molecular mechanics/generalized born surface area (MM/GBSA) method as in the following equation (Genheden and Ryde, 2015).

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S \quad (1)$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{ele}} + \Delta E_{\text{vdw}} + \Delta E_{\text{int}} \quad (2)$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{GB}} + \Delta G_{\text{np}} \quad (3)$$

The binding free energy ( $\Delta G_{\text{bind}}$ ) equals the changes in the molecular mechanics component in gas phase ( $\Delta E_{\text{MM}}$ ), solvation free energy ( $\Delta G_{\text{solv}}$ ), and entropic contribution ( $-T\Delta S$ ). The molecular mechanics free energy ( $\Delta E_{\text{MM}}$ ) is further split into

electrostatic ( $\Delta E_{\text{ele}}$ ), van der Waals ( $\Delta E_{\text{vdw}}$ ), and bond, angle, and dihedral ( $\Delta E_{\text{int}}$ ) energies. The solvation free energy ( $\Delta G_{\text{solv}}$ ) can be divided into electrostatic solvation free energy ( $\Delta G_{\text{GB}}$ ) and a nonpolar solvation free energy ( $\Delta G_{\text{np}}$ ). The  $\Delta G_{\text{GB}}$  to the solvation energy is computed with a GB module of the AMBER suite.

The decomposition analysis was also performed by mm\_pbsa module of AMBER. The detailed procedure was described by Gohlke et al. (2003).

## RESULTS AND DISCUSSION

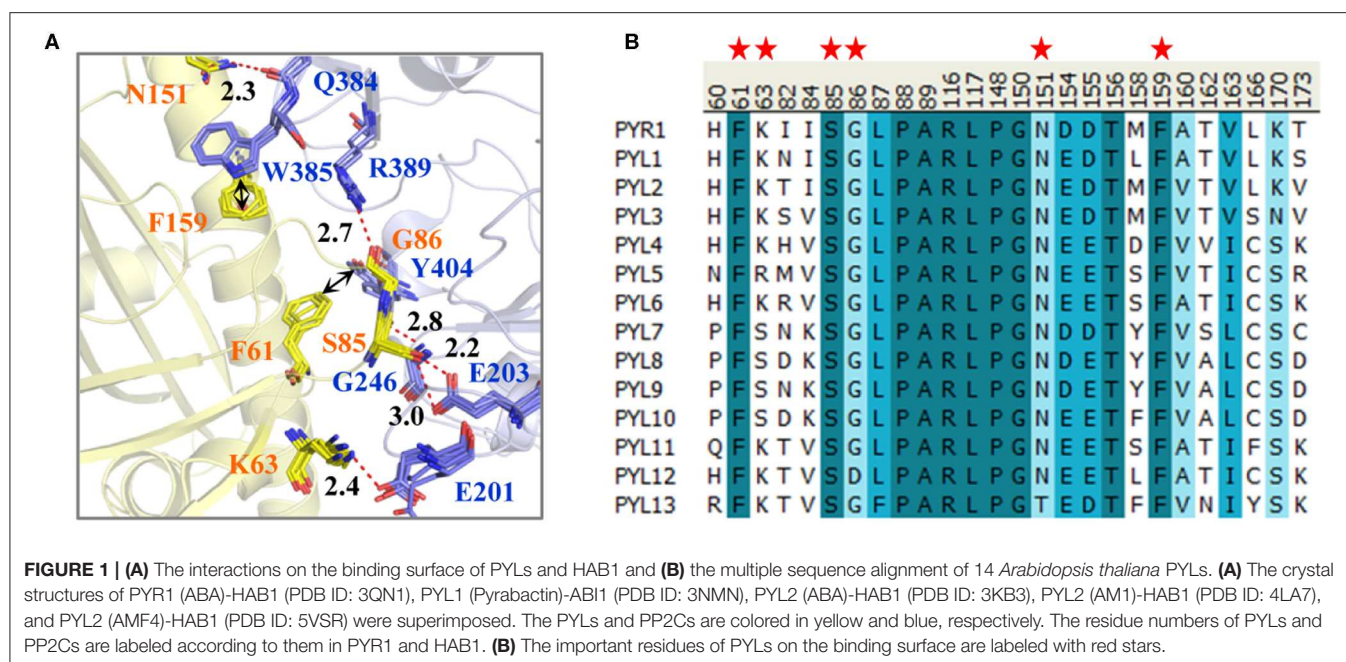
### Sequence and Structure Comparison of PP2Cs Binding Domain of PYLs

To compare the binding surfaces of different PYLs and PP2Cs, we collected the PYLs-PP2Cs complex structures, including PYR1 (ABA)-HAB1 (PDB ID: 3QN1) (Dupeux et al., 2011), PYL1 (Pyrabactin)-ABI1 (PDB ID: 3NMN) (Melcher et al., 2010), PYL2 (ABA)-HAB1 (PDB ID: 3KB3) (Melcher et al., 2009), PYL2 (AM1)-HAB1 (PDB ID: 4LA7) (Okamoto et al., 2013), and PYL2 (AMF4)-HAB1 (PDB ID: 5VSR) (Cao et al., 2017). Subsequently, the collected structures were superimposed and compared (Figure 1A). Take PYR1-HAB1 as an example; Ser85 of PYR1 forms hydrogen bonds with Gly246 and Glu203 of HAB1. Meanwhile, Lys63, Gly86, and Asn151 of PYR1 form ELE interaction with Glu201, Arg389, and Gln384 of HAB1, respectively. In addition to hydrogen bonds, there is a  $T-\pi$  interaction between Phe61 of PYR1 and Tyr404 of HAB1 as well as a  $\pi-\pi$  interaction between Phe159 of PYR1 and Trp385 of HAB1. These interactions almost exist in all the complexes, except for the hydrogen bond between Lys90 of PYL1 (corresponding to Lys63 of PYR1) and Glu140 of ABI1 (Glu201 of HAB1). However, the long-range electrostatic interaction between them still exists. In a word, the interactions on the binding surface are very conservative.

For further exploring the conservativeness of *A. thaliana* PYLs and PP2Cs, the multiple sequence alignment of 14 *A. thaliana* PYL sequences was performed (Figure 1B). It could be found that the residues on the position of Phe61, Ser85, and Phe159 of PYR1 are highly conservative. Gly86 and Asn151 of PYR1 are replaced by Asp and Thr in PYL12 and PYL13, respectively. Nonetheless, the side chain of Gly86 does not influence the interaction between PYLs and PP2Cs shown by the binding mode. The least conservative site is the position of Lys63 of PYR1, which is Lys in PYR1, PYL1-PYL4, PYL6, and PYL11-PYL13, but Ser in PYL7-PYL10. Through comparison, we found that the binding surfaces of *A. thaliana* PYLs are in high conservation, especially for PYR1 and PYL1-PYL3.

### Comparison of the Binding Models of Pyrabactin in PYR1 and PYL2

It has been found that the interactions on the PYLR1/PYL2-PP2Cs binding surface are conservative. How does the Pyrabactin induce the selectivity? The X-ray crystal structures show that there were two absolutely different conformations for Pyrabactin in PYR1/PYL1 and PYL2 (Yuan et al., 2010). Apparently, PYL2



binding with Pyrabactin is insensitive to PP2Cs. However, PP2C is inhibited by PYR1 (binding with Pyrabactin) effectively. In other words, the influence of the conformation of Pyrabactin is significant. Therefore, four complexes, constructed based on the crystal structures PYR1 (PDB ID: 3QN1) and PYL2 (PDB ID: 3NS2) as well as the initial Pyrabactin structure from PYR1 (Conf1) and PYL2 (PDB ID: 3NS2, Conf2), were used to perform MD simulations in order to compare the binding models of Pyrabactin in atomic level (Table S1).

To verify the equilibration of the systems, the atomic root mean square deviations (RMSD) were calculated, and the convergences of energies were analyzed (Knapp et al., 2011; Dawson and Gygi, 2018). As displayed in the RMSD plots, the RMSDs of all the systems reach a certain value (Figure S1). It seems that there was no big conformational change of PYL2 complexed with Pyrabactin in two starting conformations (Figure S1). However, the RMSD values of the backbone of PYR1 and the heavy atoms of Pyrabactin with the initial structure in the PYL2 were much higher than those of the other three systems, indicating that the conformations of PYR1 and Pyrabactin were changed in this system (Figure S1B). In terms of the energy, the standard deviations (STDs) of binding free energy of these systems were low ranging from 1.01 to 1.61 kcal mol<sup>-1</sup> (Table S2). All these results suggest that the systems had already reached equilibrations, and these trajectories may be used for further analysis.

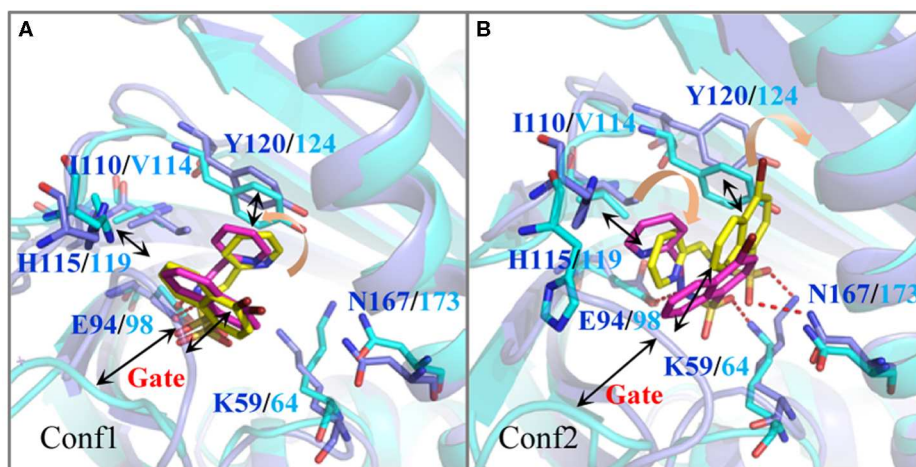
For further revealing the impact of the initial structure of Pyrabactin on PYR1 and PYL2, the binding free energies and binding modes of these systems were compared. As we all know, MM/GBSA is a powerful tool in drug design to rank the binding affinities for systems without metals (Wang et al., 2019). Therefore, it was applied in this study. To be specific, the binding free energy of Pyrabactin in Conf1 (−20.08 kcal mol<sup>-1</sup>) was

**TABLE 1 |** The binding free energy (kcal/mol) of Pyrabactin in Conf1 and Conf2 and PYLs (PYR1 and PYL2).

	$\Delta E_{ele}$	$\Delta E_{vdw}$	$\Delta E_{MM}$	$\Delta G_{solv}$	$\Delta G_{cal}$
PYR1-Pyrabactin (Conf1)	−16.18	−39.75	−55.93	35.86	−20.08
PYR1-Pyrabactin (Conf2)	−20.12	−39.51	−59.63	41.51	−18.12
PYL2-Pyrabactin (Conf1)	−15.00	−36.12	−51.12	44.30	−6.82
PYL2-Pyrabactin (Conf2)	−23.57	−37.09	−60.66	47.34	−13.33

lower than that of Pyrabactin in Conf2 (−18.12 kcal mol<sup>-1</sup>) with PYR1. Meanwhile, the binding affinity of Pyrabactin in Conf2 (−13.33 kcal mol<sup>-1</sup>) is higher than the other one (−6.82 kcal mol<sup>-1</sup>, Table 1) with PYL2. Therefore, the Conf1 and Conf2 were favored conformations in the pockets of PYR1 and PYL2, respectively. Beyond that, the binding free energies of Pyrabactin and PYR1 were lower than those of Pyrabactin and PYL2 (Figure 2). Besides, the closed gate improved the  $\Delta E_{vdw}$  and reduced the influence of  $\Delta G_{solv}$ . For Conf1, it did not undergo a big conformation change in PYL2 after the MD. There was a direct hydrogen bond between the sulfonamide of Pyrabactin and Glu94/98 of PYR1/PYL2. Furthermore, the naphthalene and pyridine of Pyrabactin formed the  $T-\pi$  and  $\pi-\pi$  interaction with His115/119 and Tyr120/124, respectively (Figure 2A). The only difference was that the pyridine of Pyrabactin moved near to the Val114 of PYL2 because of the short chain of Val, which induced the naphthalene to move always from the gate. As indicated by the result, it was hard for Pyrabactin in Conf1 to induce the closure of the gate of PYL2. For the Pyrabactin in Conf2, there was a deflection in PYR1. The longer side chain of Ile110 of PYR1 conflicted with the pyridine of Pyrabactin, which was pushed near to the Asn167 and Tyr120 (Figure 2B).





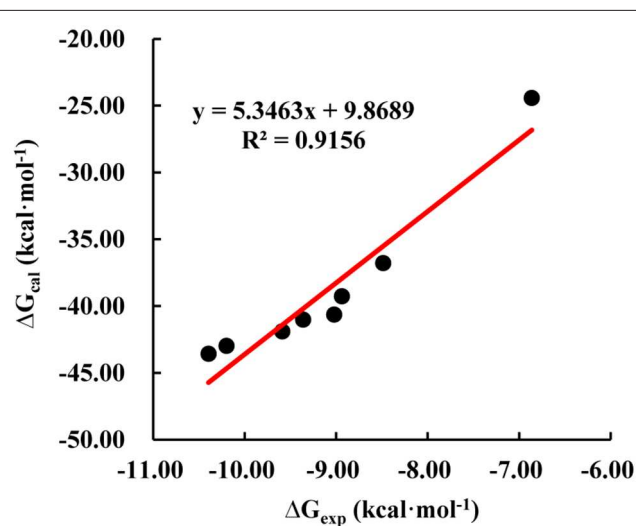
**FIGURE 2 |** The comparison of the binding modes of Pyrabactin in PYR1 and PYL2 after MD. The starting structures of Pyrabactin in the PYR1 (Conf1, **A**) and PYL2 (Conf2, **B**) complexes were used. The PYR1 and PYL2 are colored in blue and cyan, and the corresponding Pyrabactin are colored in yellow and magentas.

This deflection made the protein outward and improved the penalty of  $\Delta G_{\text{solv}}$ , inducing the low binding affinity of Pyrabactin in Conf2 with PYR1. The hydrogen bond with Lys59/64 of PYR1/PYL2 remained, while the hydrogen bond with Glu98 of PYL2 was replaced by Asn167 of PYR1. These hydrogen bonds improved the  $\Delta E_{\text{ele}}$  of this conformation both in PYR1 and PYL2. Meanwhile, the van der Waals (VDW) interaction between the naphthalene and pyridine of Pyrabactin and His115/119 and Tyr120/124 of PYR1/PYL2 stabilized the binding modes (**Figure 1B**). Apparently, the Pyrabactin in Conf2 disordered the structure of PYR1 and broke the active conformation of PYR1. At the same time, Pyrabactin was far from the closed gate in this binding mode, indicating that it is hard to induce the active conformation of PYL2.

Therefore, the different residue Ile110/Val114 is the determinant for the conformational selectivity of Pyrabactin in PYR1 and PYL2, which is consistent with the previous experimental data. The V114I mutant of PYL2 is able to inhibit the phosphatase activity of ABI1 in the presence of Pyrabactin just like PYR1 (Peterson et al., 2010; Yuan et al., 2010). This single residue alteration influences the Pyrabactin conformation in the pockets of PYLs, directly determining the state of the gate and the function of PYLs.

## Selective Activation Mechanism of PYR1 and PYL2

In order to obtain dynamics conformation, we applied MD simulation on eight systems, including HAB1 complexed with PYR1 and PYL2 binding with ABA, AM1, AMF4, and Pyrabactin, respectively. Moreover, RMSD value per picosecond and binding energy per nanosecond in the last 6 ns were calculated to explore the dynamic stability of eight systems. In this process, the RMSD values of the backbone of PYR1/PYL2-HAB1 and the heavy atoms of ligands were lower than 2.5 and 0.5 Å (**Figure S2**). With regard to the binding energy, all the STDs were

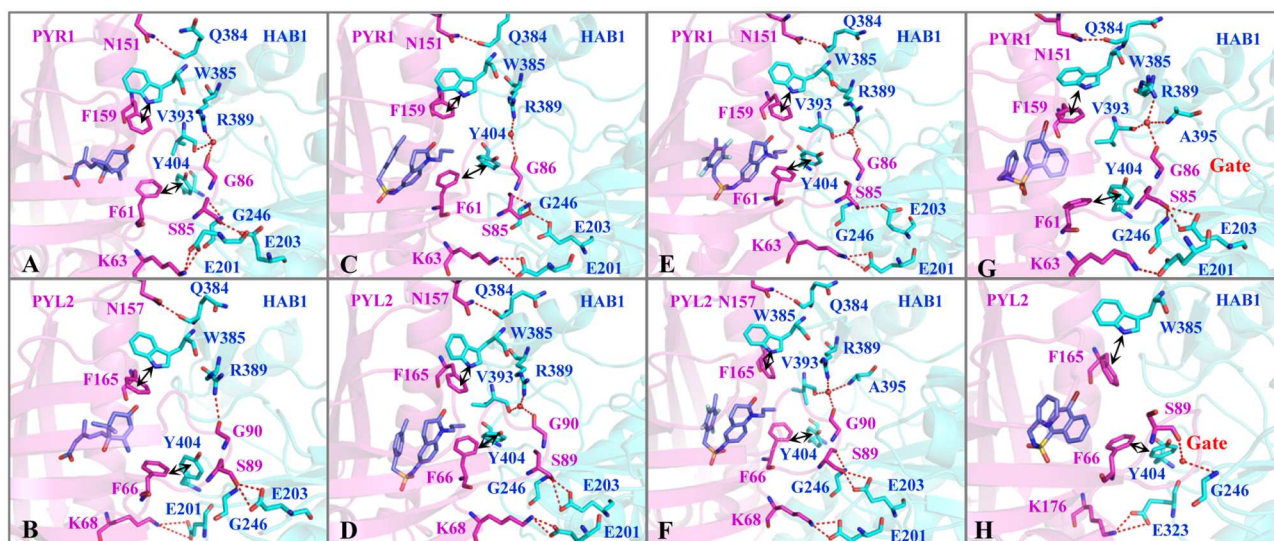


**FIGURE 3 |** The correlation of calculated ( $\Delta G_{\text{cal}}$ ) and experimental ( $\Delta G_{\text{exp}}$ ) binding free energy.

lower than  $2.21 \text{ kcal mol}^{-1}$  (**Table S3**). These results revealed that all the systems reached the equilibrium stage. Additionally, the linear relationship between the calculated and experimental binding energy was fitted to further validate the result (**Figure 3**). The calculated data ( $\Delta G_{\text{cal}}$ ,  $-43.58$  to  $-24.43 \text{ kcal mol}^{-1}$ ) is consistent with the experimental data ( $\Delta G_{\text{exp}}$ ,  $-10.40$  to  $-6.86 \text{ kcal mol}^{-1}$ ) with high correlation coefficient  $R^2$  (0.92), suggesting that the trajectories from molecular dynamics were reliable.

The binding modes were further analyzed to study the interactions on the binding surface of HAB1 and PYR1/PYL2. As for agonists, the conservative ELE interactions were kept during MD, such as the hydrogen bond of Lys63/68-Glu201, Ser85/89-Gly246, Ser85/89-Glu203, and Asn151/157-Gln384 between





**FIGURE 4 |** The interactions on the binding surface of PYLs and HAB1 after MD. The important residues of (A,C,E,G) PYR1 and (B,D,F,H) PYL2 are colored in magentas. The residues of HAB1 are colored in cyan. The ligands (A,B) ABA, (C,D) AM1, (E,F) AMF4, and (G,H) Pyrabactin are colored in blue.

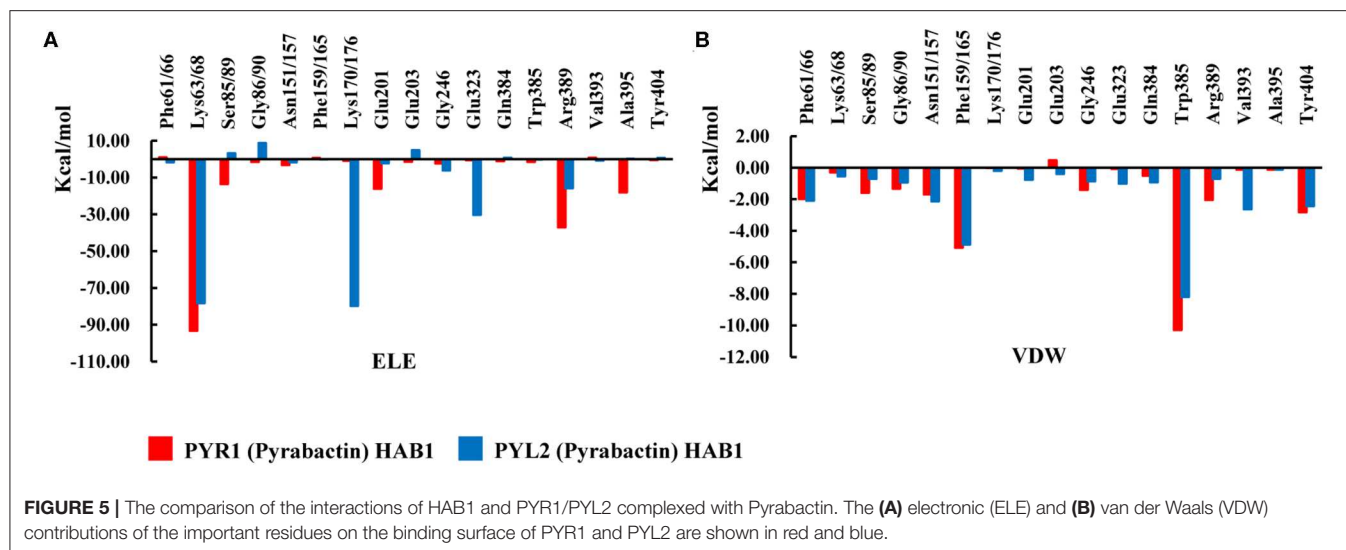
**TABLE 2 |** The binding free energy (kcal/mol) of HAB1 and PYLs (PYR1 and PYL2) complexed with ligands from calculation and experiment.

	$\Delta E_{ele}$	$\Delta E_{vdw}$	$\Delta E_{MM}$	$\Delta G_{solv}$	$\Delta G_{cal}$	IC <sub>50</sub> (nM)	$\Delta G_{exp}$
PYR1-ABA	-276.76	-92.95	-369.71	330.44	-39.27	307.0	-8.9
PYL2-ABA	-266.80	-91.13	-357.94	316.92	-41.02	151.0	-9.4
PYR1-AM1	-279.65	-91.12	-370.78	328.87	-41.90	103.0	-9.6
PYL2-AM1	-252.25	-89.87	-342.12	301.48	-40.64	267.0	-9.0
PYR1-AMF4	-252.33	-90.69	-343.02	300.04	-42.98	119.1	-10.2
PYL2-AMF4	-267.41	-86.34	-353.75	310.17	-43.58	85.7	-10.4
PYR1-Pyrabactin	-257.12	-89.83	-346.95	310.16	-36.79	656.0	-8.5
PYL2-Pyrabactin	-197.74	-79.48	-277.22	252.80	-24.43	>10,000	-6.9

PYR1/PYL2 and HAB1. Meanwhile, Gly86/90 of PYR1/PYL2 contacted with Arg389 of HAB1 through hydrogen bond or water bridge (Table S4). Furthermore, the  $T-\pi$  interaction between Phe61/66 of PYR1/PYL2 and Tyr404 of HAB1 as well as the  $\pi-\pi$  interaction between Phe159/165 of PYR1/PYL2 and Trp385 of HAB1 promoted the complexes formation (Figures 4A–G). Due to the conservative interactions on the binding surface, there is no absolute difference between ELE ( $-279.65$  to  $-252.25$  kcal mol<sup>-1</sup>) and VDW ( $-92.95$  to  $-86.34$  kcal mol<sup>-1</sup>) contribution in the systems binding with agonists (Table 2). However, the ELE contribution on the binding surface of PYL2 (Pyrabactin) and HAB1 dropped to  $-197.74$  kcal mol<sup>-1</sup>, which induced their low binding affinity directly (Table 2). Based on the binding mode and the hydrogen bond monitoring result, this research found that the conserved hydrogen bonds were broken, such as the hydrogen bond of Lys68–Glu201, Gly86–Arg389, Ser89–Glu203, and Asn151–Gln384 between PYL2 and HAB1. Nonetheless, the Ser89 of PYL2 contacted with Gly246 of HAB1 through “water bridge,” in which a new hydrogen bond was formed between

Lys176 of PYL2 and Glu323 of HAB1. On the other hand, it seems that the hydrophobic interaction between PYL2 (Pyrabactin) and HAB1 was no better than that of PYR1 (Pyrabactin)–HAB1 (Figure 4H). The  $\pi-\pi$  interaction between Phe165 of PYL2 and Trp385 of HAB1 was not stable as other systems, and the hydrophobic interactions between the gate of PYL2 and HAB1 were reduced.

To reveal protein–protein interactions influenced by Pyrabactin from the energy aspect, the energy decompositions of amino acid residues on the binding surface of HAB1 and PYR1/PYL2 complexed with Pyrabactin were performed. In comparison to PYR1, the ELE contributions of Lys68, Ser89, Gly90 in PYL2 and Glu201, Glu203, and Arg389 in HAB1, induced by the loss of hydrogen bonds or water bridges, were reduced by 15.04, 16.61, 10.18, 13.81, 6.18, and 21.17 kcal mol<sup>-1</sup>, respectively (Figure 5A). On the other hand, the ELE contributions of Lys176 and Glu323 of PYL2 and HAB1 were improved from  $-0.82$  and  $-0.56$  kcal mol<sup>-1</sup> to  $-79.77$  and  $-30.31$  kcal mol<sup>-1</sup> because of the new hydrogen bond formed.



The VDW contributions on the binding surface were also influenced: the contribution of Trp385 in HAB1, which formed  $\pi$ - $\pi$  interaction with Phe165 of PYL2, was reduced by  $-2.09$  kcal mol $^{-1}$ ; the contributions of Ser89 and Gly90 on the gate were also lower than those with PYR1. The decomposition analysis of the binding energy was in accordance with the protein-protein interaction analysis results. From above, the distributions of Ser89 and Gly90 of PYL2 binding with Pyrabactin to ELE and VDW were both reduced significantly. These differences derived from the conformation of the gate determined by the Pyrabactin.

In short, the closed gate formed a better binding surface of HAB1 and formed more interactions with the downstream proteins (Figures 4G,H). This result is consistent with the gate-latch-lock mechanism underlying ABA signaling. When ABA binds to PYR1, the Pro88 of PYR1 on the gate moves toward the pocket to close the gate, whereas the Ser85 on the gate is flipped outward the cavity to contact with PP2Cs (Zhang et al., 2015). Meanwhile, Ser89 of PYL2 was reported as an important residue to form the tertiary complex (Yin et al., 2009). Furthermore, the mutations in PYL2 residues involving the formation of the gate compromise its ability to activate the reporter in response to ABA (Melcher et al., 2009). Therefore, gate-closed-induced ligands will be candidates for the ABA mimicking.

## CONCLUSION

In this study, we studied the selective activation mechanism of PYLs through the sequence alignment, molecular dynamics simulation, and binding free energy calculation methods. Even though the residues on the binding surface of PYR1 and PYL2 are conserved, the gate conformations of PYR1 and PYL2 induced by Pyrabactin are different. The Val114 in the pocket of PYL2 leads to a rotated binding model of Pyrabactin, which leads to an opened gate. This reduces the binding free energy of

PYL2 and HAB1. The energy contribution changes of Lys63/68, Ser85/89, and Gly86/90 on the binding surface elucidate the selectivity of PYR1 and PYL2 complexed with Pyrabactin to HAB1. Therefore, the gate conformation influences the functions of PYLs directly. The results elucidate molecular determinants of the selectivity of PYLs and HAB1 interactions, which may provide new ideas for further agrochemical design and drought tolerance research.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## AUTHOR CONTRIBUTIONS

G-FY and G-FH conceived of the research plan. J-FY, C-YY, and DW performed the project. G-FY, G-FH, J-FY, and C-YJ wrote the manuscript.

## FUNDING

This work was supported in part by the National Key R&D Program (2017YFD0200501), the National Natural Science Foundation of China (21772059, 91853127, and 31960548), and the Science and Technology Project of Guizhou Province (No. [2017]1028). Program of Introducing Talents of Discipline to Universities of China (111 Program, D20023).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.00425/full#supplementary-material>

## REFERENCES

- Al Ait, L., Yamak, Z., and Morgenstern, B. (2013). DIALIGN at GOBICS—multiple sequence alignment using various sources of external information. *Nucleic Acids Res.* 41, W3–W7. doi: 10.1093/nar/gkt283
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235
- BIOVIA Discovery Studio (2009). *Discovery Studio 2.5*. San Diego, CA: Accelrys Inc.
- Cao, M., Liu, X., Zhang, Y., Xue, X., Zhou, X. E., Melcher, K., et al. (2013). An ABA-mimicking ligand that reduces water loss and promotes drought resistance in plants. *Cell. Res.* 23, 1043–1054. doi: 10.1038/cr.2013.95
- Cao, M. J., Zhang, Y. L., Liu, X., Huang, H., Zhou, X. E., Wang, W. L., et al. (2017). Combining chemical and genetic approaches to increase drought resistance in plants. *Nat. Commun.* 8:1183. doi: 10.1038/s41467-017-01239-3
- Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham, T. E. III, Darden, T. A., Duke, R. E., et al. (2016). *AMBER 2016*. San Francisco, CA: University of California.
- Case, D. A., Cheatham, T. E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M. Jr., Onufriev, A., et al. (2005). The amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688. doi: 10.1002/jcc.20290
- Dawson, W., and Gygi, F. (2018). Equilibration and analysis of first-principles molecular dynamics simulations of water. *J. Chem. Phys.* 148:124501. doi: 10.1063/1.5018116
- DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. San Carlos, CA: Delano Scientific.
- Di Tommaso, P., Moretti, S., Xenarios, I., Orobittg, M., Montanyola, A., Chang, J. M., et al. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 39, W13–W17. doi: 10.1093/nar/gkr245
- Dupeux, F., Antoni, R., Betz, K., Santiago, J., Gonzalez-Guzman, M., Rodriguez, L., et al. (2011). Modulation of abscisic acid signaling *in vivo* by an engineered receptor-insensitive protein phosphatase type 2C allele. *Plant Physiol.* 156, 106–116. doi: 10.1104/pp.110.170894
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Finkelstein, R. (2013). Abscisic acid synthesis and response. *Arabidopsis Book* 11:e0166. doi: 10.1199/tab.0166
- Frisch, M. J. T., Trucks, G., Schlegel, H. B., Scuseri, G. E. W., Robb, M. A., Cheeseman, J. R., et al. (2004). *Gaussian 03, Revision B.03*. Wallingford, CT: Gaussian, Inc.
- Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Dis.* 10, 449–461. doi: 10.1517/17460441.2015.1032936
- Gohlke, H., Kiel, C., and Case, D. A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* 330, 891–913. doi: 10.1016/S.0022-2836(03)00610-7
- Hao, G. F., Jiang, W., Ye, Y. N., Wu, F. X., Zhu, X. L., Guo, F. B., et al. (2016). ACFIS: a web server for fragment-based drug discovery. *Nucleic Acids Res.* 44, W550–W556. doi: 10.1093/nar/gkw393
- Knapp, B., Frantal, S., Cibena, M., Schreiner, W., and Bauer, P. (2011). Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? *J. Comput. Biol.* 18, 997–1005. doi: 10.1089/cmb.2010.0237
- Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A., et al. (2009). Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* 324, 1064–1068. doi: 10.1126/science.1172408
- Melcher, K., Ng, L. M., Zhou, X. E., Soon, F. F., Xu, Y., Suino-Powell, K. M., et al. (2009). A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. *Nature* 462, 602–608. doi: 10.1038/nature08613
- Melcher, K., Xu, Y., Ng, L. M., Zhou, X. E., Soon, F. F., Chinnusamy, V., et al. (2010). Identification and mechanism of ABA receptor antagonism. *Nat. Struct. Mol. Biol.* 17, 1102–1108. doi: 10.1038/nsmb.1887
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., et al. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662. doi: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256
- Okamoto, M., Peterson, F. C., Defries, A., Park, S. Y., Endo, A., Nambara, E., et al. (2013). Activation of dimeric ABA receptors elicits guard cell closure, ABA-regulated gene expression, and drought tolerance. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12132–12137. doi: 10.1073/pnas.1305919110
- Park, S.-Y., Fung, P., Nishimura, N., Jensen, D. R., Fujii, H., Zhao, Y., et al. (2009). Absciscic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* 324, 1068–1071. doi: 10.1126/science.1173041
- Pennisi, E. (2009). Stressed out over a stress hormone. *Science* 324:1012. doi: 10.1126/science.324\_1012
- Peterson, F. C., Burgie, E. S., Park, S. Y., Jensen, D. R., Weiner, J. J., Bingman, C. A., et al. (2010). Structural basis for selective activation of ABA receptors. *Nat. Struct. Mol. Biol.* 17, 1109–1113. doi: 10.1038/nsm.1898
- Sah, S. K., Reddy, K. R., and Li, J. (2016). Absciscic acid and abiotic stress tolerance in crop plants. *Front. Plant. Sci.* 7:571. doi: 10.3389/fpls.2016.00571
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.2.4673
- Verma, V., Ravindran, P., and Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* 16, 1–10. doi: 10.1186/s12870-016-0771-y
- Vishwakarma, K., Upadhyay, N., Kumar, N., Yadav, G., Singh, J., Mishra, R. K., et al. (2017). Absciscic acid signaling and abiotic stress tolerance in plants: a review on current knowledge and future prospects. *Front. Plant. Sci.* 8:161. doi: 10.3389/fpls.2017.00161
- Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem. Rev.* 119, 9478–9508. doi: 10.1021/acs.chemrev.9b00055
- Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21, 1049–1074. doi: 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F
- Yamaguchi-Shinozaki, K., and Shinozaki, K. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* 57, 781–803. doi: 10.1146/annurev.arplant.57.032905.105444
- Yang, J.-F., Chen, M. X., Zhang, J.-H., Hao, G.-F., and Yang, G.-F. (2019). Genome-wide phylogenetic and structural analysis reveals the molecular evolution of the ABA receptor gene family. *J. Exp. Bot.* 71, 1322–1336. doi: 10.1093/jxb/erz511
- Yang, J. F., Wang, F., Chen, Y. Z., Hao, G.-F., and Yang, G. F. (2020). LARMD: integration of bioinformatic resources to profile ligand-driven protein dynamics with a case on the activation of estrogen receptor. *Brief. Bioinformatics* doi: 10.1093/bib/bbz141. [Epub ahead of print].
- Yin, P., Fan, H., Hao, Q., Yuan, X., Wu, D., Pang, Y., et al. (2009). Structural insights into the mechanism of abscisic acid signaling by PYL proteins. *Nat. Struct. Mol. Biol.* 16, 1230–1236. doi: 10.1038/nsmb.1730

- Yuan, X., Yin, P., Hao, Q., Yan, C., Wang, J., and Yan, N. (2010). Single amino acid alteration between valine and isoleucine determines the distinct pyrabactin selectivity by PYL1 and PYL2. *J. Biol. Chem.* 285, 28953–28958. doi: 10.1074/jbc.M110.160192
- Zhang, H., Li, Y., and Zhu, J. K. (2018). Developing naturally stress-resistant crops for a sustainable agriculture. *Nat. Plants* 4, 989–996. doi: 10.1038/s41477-018-0309-4
- Zhang, X. L., Jiang, L., Xin, Q., Liu, Y., Tan, J. X., and Chen, Z. Z. (2015). Structural basis and functions of abscisic acid receptors PYLs. *Front. Plant. Sci.* 6:88. doi: 10.3389/fpls.2015.00088

**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.

Copyright © 2020 Yang, Yin, Wang, Jia, Hao and Yang. 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.





# Diterpenoids and Triterpenoids From Frankincense Are Excellent Anti-psoriatic Agents: An *in silico* Approach

Sobia Ahsan Halim<sup>1</sup>, Ajmal Khan<sup>1</sup>, Rene Csuk<sup>2</sup>, Ahmed Al-Rawahi<sup>1</sup> and Ahmed Al-Harrasi<sup>1\*</sup>

<sup>1</sup> Natural and Medical Sciences Research Center, University of Nizwa, Nizwa, Oman, <sup>2</sup> Organic Chemistry, Martin-Luther University Halle-Wittenberg, Halle (Saale), Germany

## OPEN ACCESS

### Edited by:

Ruo-Xu Gu,  
Max Planck Institute for Biophysical  
Chemistry, Germany

### Reviewed by:

Beibei Wang,  
University of Electronic Science and  
Technology of China, China  
Yong Wang,  
University of Copenhagen, Denmark

### \*Correspondence:

Ahmed Al-Harrasi  
aharrasi@unizwa.edu.om

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 25 February 2020

**Accepted:** 11 May 2020

**Published:** 25 June 2020

### Citation:

Halim SA, Khan A, Csuk R,  
Al-Rawahi A and Al-Harrasi A (2020)  
Diterpenoids and Triterpenoids From  
Frankincense Are Excellent  
Anti-psoriatic Agents: An *in silico*  
Approach. *Front. Chem.* 8:486.  
doi: 10.3389/fchem.2020.00486

Psoriasis is a chronic autoimmune disease that affects 2–3% of the global population and requires an effective treatment. Frankincense has been long known for its potent anti-inflammatory activities. In this study, a structural bioinformatics approach was used to evaluate the efficacy of individual active components of frankincense, macrocyclic diterpenoid derivatives (**1–27**), and boswellic acids (**28–46**) in the treatment of psoriasis. Initially, major druggable targets of psoriasis were identified. Subsequently, structure-based screening was employed by using three different docking algorithms and scoring functions (MOE, AutoDock Vina, and MVD) for the target fishing of compounds against 18 possible targets of psoriasis. Janus Kinase 1, 2, 3 (JAK 1/2/3), eNOS, iNOS, interleukin-17 (IL-17), and Tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) were identified as the preferred molecular targets for these compounds. This computational analysis reflects that frankincense diterpenoids and triterpenoids can serve as excellent anti-psoriatic agents by targeting major cytokines (TNF- $\alpha$ , IL-17, IL-13, IL-23, and IL-36 $\gamma$ ), exacerbated in psoriasis, and inflammatory pathways particularly JAK1/2/3, eNOS, iNOS, MAPK2, and IFN $\gamma$ . The results were compared with the reported experimental findings which correlates well with our *in-silico* verdicts.

**Keywords:** psoriasis, frankincense, diterpenoid, boswellic acids, *in silico* target fishing, molecular docking

## INTRODUCTION

Psoriasis is a chronic and most inexplicable autoimmune skin disease that affect 2–3% of the population worldwide (Baliwag et al., 2015). It is characterized by increased propagation of the epidermis that can multiply up to 10 times faster than normal in psoriasis with dilation of dermal capillaries. As underlying cells reach the skin's surface and die, their sheer volume creates red plaques covered with white scales that cause itchy and scaly skin, swelling, pain, and disfiguring skin lesions (MacDonald and Burden, 2007). Psoriasis can occur at any age and equally in men and women. The disease is more common in adults than children. The estimates in children vary between 0.7% (Augustin et al., 2010) in Europe to almost none in Asia (Bø et al., 2008; Chen et al., 2008). The variation in the prevalence of psoriasis has also been linked to geographical locations, as it is less common in countries closer to the Equator (Egypt, Sri-Lanka, Taiwan) as opposed to countries that are further away (Europe, Australia, and North America) (Parisi et al., 2013).

Findings have consistently reported an increasing trend of the prevalence of psoriasis (Icen et al., 2009; Egeberg et al., 2017; Eder et al., 2019).

Several factors are involved in causing psoriasis, such as bacterial infection, genetic/environmental factors, and autoimmune disorders. Psoriasis is associated with several comorbidities, including cardiovascular disease (CVD), lymphoma, and extensive depression (Ni and Chiu, 2014; Takeshita et al., 2017). It happens through chronic interactions between hyper-proliferative keratinocytes and infiltrating, activated immune cells. The immune system plays a critical role in the pathogenesis of psoriasis. T cells (particularly Th1 and Th17) are heavily present in psoriatic lesions. Moreover, TNF $\alpha$  and iNOS producing inflammatory dendritic cells, massively infiltrate psoriatic skin, and polarize T cells to Th1 and Th17 fates. Additionally, psoriatic skin is infiltrated by macrophages, innate immune cells, and an increased number of endothelial cells, which exacerbate the pathogenesis of psoriasis. The genetics behind psoriasis is complex and multifactorial. *PSORS* (psoriasis-susceptibility) loci harbor several genes that are involved in psoriasis; for example, *HLA-Cw6*, *ERAP1*, *ERAP2*, and *MICA* are involved in antigen presentation. Furthermore, several other genes span an array of functions, i.e., T-cell development and polarization (*RUNX1*, *RUNX3*, *STAT3*, *TAGAP*, *IL4*, and *IL13*), development of innate immunity (*CARD14*, *c-REL*, *TRAF3IP2*, *DDX58*, and *IFIH1*), the IL-23 axis (*IL12Bp40*, *IL23Ap19*, *IL23R*, *JAK2*, and *TYK2*), and negative regulators of immune responses (*TNIP1*, *TNFAIP3*, *NFKBIA*, *ZC3H12C*, *IL36RN*, and *SOC31*) (Al Robaee, 2010; Harden et al., 2015; Woo et al., 2017).

Currently, there are three different types of treatment used to reduce the inflammation and skin irritation/itching, including topical treatments, light therapy, and systemic medications (Winterfield et al., 2005; Gisondi et al., 2017; Golbari et al., 2018). Topical treatments serve as first-line therapies that include use of topical corticosteroids, vitamin D analogs, anthralin, retinoids, and calcineurin inhibitors, and they are frequently prescribed to treat mild to moderate psoriasis. The overuse of corticosteroids causes thinning of the skin. Vitamin D analogs (Calcipotriene, Calcitriol) and anthralin reduce skin cell growth, remove scales, and make skin smoother. These analogs treat mild to moderate psoriasis along with other treatments; however, they promote skin irritation. Similarly, topical retinoids may decrease inflammation, but cause skin irritation and increase sensitivity to sunlight. Moreover, oral retinoids cause risk of birth defects and are not recommended for pregnant and breast-feeding women. Calcineurin inhibitors, particularly tacrolimus and pimecrolimus, also reduce inflammation and plaque accumulation; however, an increased risk of skin cancer and lymphoma is associated with these inhibitors, and they are therefore not recommended for long-term or continuous use (Choi et al., 2017).

Second- or third-line therapies, including phototherapy and systemic therapies, are given to patients with severe psoriasis or treatment-resistant disease. In phototherapy, natural or artificial ultraviolet (UV) light is used to treat mild psoriasis by exposing skin to controlled amounts of natural sunlight, artificial

ultraviolet A (UVA), or UV B (UVB) light, either alone or in combination with medications. Exposure to UV in sunlight or artificial light slows skin cell turnover and reduces scaling and inflammation. Daily exposure to small amounts of sunlight may improve psoriasis, but intense sun exposure can worsen symptoms and cause skin damage. Controlled doses of UVB light improves the symptoms of mild to moderate psoriasis; however, it may cause short-term side effects like redness, itching, and dry skin. UVA light penetrates deeper into the skin than UVB, improves skin, and is often used to treat severe psoriasis. Despite this, it can cause short-term side effects, such as nausea, headaches, burning, and itching, or long-term side effects, such as dry/wrinkled skin, freckles, increased sun sensitivity, and increased risk of skin cancer and melanoma (Pardasani et al., 2000).

Patients with severe psoriasis are treated with systemic treatment (including retinoids, methotrexate, and cyclosporine), which are associated with severe side effects. Retinoids may cause lip inflammation and hair loss. Methotrexate helps psoriasis by decreasing the production of skin cells and suppressing inflammation; however, it may also cause stomach upset, loss of appetite, and fatigue. With long-term use, methotrexate can cause liver damage and decreased production of red and white blood cells and platelets. Cyclosporine suppresses the immune system and is similar to methotrexate in effectiveness but can only be taken short term because it may increase risk of infection, cancer, kidney problems, and high blood pressure at high doses or long-term therapy (Hoffman et al., 2016).

The presence of cytokines, dendritic cells, and T lymphocytes in psoriatic lesions has encouraged the development of biologic therapies for psoriasis (Schadler et al., 2019). These therapies include monoclonal antibodies (mAB) against tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) (infliximab, adalimumab, golimumab), interleukin (IL)-12 and IL-23 (ustekinumab), IL-17A (secukinumab and ixekizumab), and inhibitors of TNF- $\alpha$  (etanercept) and phosphodiesterase 4 (apremilast). These drugs are usually used to treat psoriatic patients who have failed to respond to traditional therapy or are associated with psoriatic arthritis. However, these drugs have strong effects on the immune system and may permit life-threatening infections. In particular, people taking these treatments must be screened for tuberculosis. All these treatments are only used to manage the disease at each time it surfaces (Tollefson et al., 2018). Therefore, new and safer chemical agents are urgently required for the effective treatment of psoriasis.

Frankincense is known for its superior anti-inflammatory potential (Hussain et al., 2017; Al-Harrasi et al., 2018a). The active constituents of frankincense, including incensole and several boswellic acid derivatives, suppress the expression of tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), interleukin-1 $\beta$  (IL-1 $\beta$ ), and nuclear factor- $\kappa$ B (NF- $\kappa$ B) (Moussaieff et al., 2007). Due to our deep interest in exploring medicinal properties of frankincense, a computational pipeline was created to investigate the anti-psoriatic potential of the active components of frankincense. In this study, computational target fishing was applied for the identification of potential druggable molecular targets of psoriasis and the binding potential of cembrenoid

diterpenoids and triterpenoids found in several species of *Boswellia* was scrutinized by *in silico* reverse molecular docking. The computational analyses reveal the promising binding potential of these compounds with the proteins associated with psoriatic pathways.

## MATERIALS AND METHODS

The computational experiments were performed on a Windows 10 workstation with Intel® Core™ i7-7700HQ CPU@2.80GHz processor and 12 GB RAM. For docking, MOE (Molecular Operating Environment), MVD (Molegro Virtual Docker) and ADT Vina (AutoDock Tools Vina) were used. Protein–ligand interactions were visualized on Chimera software (Pettersen et al., 2004).

### Identification of Druggable Proteins in Psoriasis

The major biological pathways of psoriasis were deduced by a literature survey (Rácz and Prens, 2009; Bejarano and Valdecantos, 2013; Baliwag et al., 2015; Hugh and Weinberg, 2018; Yadav et al., 2019), and this revealed that Tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), interleukin-1 $\alpha$  (IL-1 $\alpha$ ), IL-1 $\beta$ , IL-13, IL-12/23, IL-17, IL-22, IL-36 $\gamma$ , Interferon- $\gamma$  (IFN- $\gamma$ ), Nuclear Factor- $\kappa$ B (NF- $\kappa$ B), endothelial nitric oxide synthase (eNOS), inducible NOS (iNOS), Peroxisome proliferator-activated receptor gamma (PPAR- $\gamma$ ), MAP Kinase-Activated Protein Kinase 2 (MAPK2), Janus Kinase 1 (JAK1), JAK2, JAK3, and the Signal transducer and activator of transcription 3 (STAT3) play important roles in the pathogenesis of psoriasis. Moreover, the druggable macromolecules were also confirmed by Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<https://www.kegg.jp/>), which showed that the NF- $\kappa$ B, IL-17, and IL-36 pathways are particularly involved in psoriasis (KEGG ID: H01656). Thus, the aforementioned 18 proteins were selected as potential drug targets in our docking studies. The three-dimensional (3D-) coordinates of selected targets were retrieved from the RCSB Protein Data Bank (PDB, <https://www.rcsb.org/>) with good resolution. The binding site of each protein was elucidated by visual analysis by UCSF chimera (Pettersen et al., 2004) (<https://www.cgl.ucsf.edu/chimera/>), a literature review, and the PDBsum database (Laskowski et al., 1997) (<http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum>). For docking, each protein was treated individually. The standard protonation state of each protein was set according to neutral pH, and partial charges were applied based on the MMFF94x force field by MOE v2014.09. All the heteroatoms were deleted from protein structures. By careful analysis, only those water molecules were retained in the protein structures which are involved in protein–ligand bridging, while the rest of the water molecules were removed. The PDB codes of selected targets, their resolution, and their binding residues are tabulated in **Table 1**. The positive controls were added in the dataset as a reference ligand to test the screening accuracy of docking programs. For IL-17, IL-36 $\gamma$ , PPAR- $\gamma$ , MAPK2, JAK1/2/3, TNF- $\alpha$ , iNOS, and eNOS, the co-crystallized ligands were chosen as a positive

control. The known inhibitors of IFN- $\gamma$ , IL-22, IL-12/23, IL-1 $\alpha$ , IL-1 $\beta$ , NF- $\kappa$ B, STAT3, and IL-13 were selected from <https://www.medchemexpress.com/>.

### Selection and Preparation of Ligands for Docking

A set of 46 macrocyclic diterpenoids and triterpenes from different species of *Boswellia* was chosen from our *in-house* compound's library for reverse docking analysis. The structures of the selected compounds are shown in **Table 2**. The 2D-structure of each ligand was drawn on ChemDraw (<https://www.perkinelmer.com/category/chemdraw>) and converted into a 3D-format by Molecular Operating Environment (MOE 2014.09) (Florence et al., 2014) (<https://www.chemcomp.com/>). Hydrogen atoms were added, partial charges were applied (based on MMFF94x force field) on each structure, and the structures were minimized using the MMFF94x force field (eps = r, Cutoff until the RMS gradient of 0.1 kcal/mol/Å was achieved). The protonation state of each compound was assigned based on the neutral pH. All the compounds were imported into MOE database to be used in docking.

### Reverse Molecular Docking by MOE

A reverse docking approach was applied to search the appropriate druggable binders of macrocyclic diterpenoids and triterpenes by using MOE docking suit (Lee et al., 2016; Xu et al., 2018). After the preparation of protein and ligand files, docking was performed by using the Triangle Matcher docking algorithm and London dG scoring function. Induced Fit protocol (implemented in MOE) was applied during docking. The ligand binding residues were selected to define the active/ligand binding site (**Table 1**). By default, 30 docked conformations of each ligand were retained after docking. At the end of docking, the interactions of each ligand with the binding residues were visualized by Protein–ligand Interaction Fingerprints (PLIF) setup of MOE and Chimera. Later, conformational sampling was performed, and the best docked pose of each ligand was selected based on the docking score, rank, and binding interactions with the target. PLIF calculates hydrogen bonding, protein–ligand solvent bridging, ionic attraction, surface contact, metal ligation, arene attraction, and protein–ligand solvent interactions between protein and ligand. The 2D interactions of ligands with their targets were visualized by the MOE protein–ligand interaction tool. Subsequently, Chimera was used to depict the 3D view of protein–ligand interactions.

### Molegro Virtual Docker

For each docking, protein and ligand structures were imported into the MVD (version 2019.7.0.0) workspace in “pdb” and “mol2” format, respectively (Thomsen and Christensen, 2006). The ligands were prepared by MOE by applying charges and explicit hydrogen, and the valences, bond orders, and protons were thus correct in the ligands. The ligand binding site in the protein was defined by cavity detection algorithm of MVD. Thirty docking runs were performed for each ligand with a maximum iteration of 1,500 and maximum population size of 50, which resulted in 30 docked poses of each ligand after docking.

**TABLE 1** | Drug targets used in docking experiments.

S. no.	Molecular targets	PDB ID	Resolution (Å)	Positive control	Binding residues
1	TNF- $\alpha$	2AZ5	2.1	<sup>†</sup> <b>6,7-dimethyl-3-[(methyl[2-[methyl[1-[3-(trifluoromethyl)phenyl]-1H-indol-3-yl)methyl]amino]ethyl]amino)methyl]-4h-chromen-4-one</b>	Leu57(A/B), Tyr59(A/B), Ser60(A/B), Gln61(A), Tyr119(A/B), Leu120(A/B), Gly121(A/B), Gly122(A), Tyr151(A/B).
2	IL-1 $\alpha$	5UC6	2.1	*IX207-887, RP54745	Arg16, Lys60, Asp64, Asp65, Ala66, Lys67, Ile68, Trp113, Ile118
3	IL-1 $\beta$	1ITB	2.5	*Byakandelicol, Diacerein	Arg11, Ser13, Gln14, Gln15, Met20, Ser21, Gly22, Lys27, Leu29, His30, Leu31, Gln32, Gly33, Gln34, Asp35, Met36, Gln38, Gln126, Ala127, Glu128, Asn129, Met130, Pro131, Thr147, Gln149
4	IL-13	3LB6	3.05	*Suplatast (Tosilate)	Arg11, Ile14, Glu15, Leu101, Lys104, Lys105, Phe107, Arg108, Met33, Trp35, Asp87, Thr88, Lys89, Ile90, Glu91
5	IL-17	5HI5	1.8	<sup>†</sup> <b>(4S,20R)-7-chloro-N-methyl-4-[[[1-methyl-1H-pyrazol-5-yl]carbonyl]amino]-3,18-dioxo-2,19-diazatetracyclo[20.2.2.1~6,10~.1~11,15~]octacos-1(24),6(28),7,9,11(27),12,14,22,25-nonaene-20-carboxamide</b>	Asn36, Pro37, Trp62, Pro63, Val65, Ile66, Trp67, Gln94, Glu95, Ile96, Leu97, Val98, Leu99, Leu112
6	IL-22	3DLQ	1.9	*GSK2981278	Phe47, Gln49, Thr53, Ser64, Asp67, Thr70, Asp71, Arg73, Lys162, Gly165, Glu166, Asp168, Arg175
7	IL-12/23	5MZV	2.8	*Apilimod, Isomucronulatol, Tyrphostin A1	Leu56, Arg57, Glu58, Trp156, Leu160, Lys164
8	IL-36 $\gamma$	6P9E	2.0	<sup>†</sup> <b>(2S)-2-[[4-(3-amino-4-methylphenyl)-6-methylpyrimidin-2-yl]oxy]-3-methoxy-3,3-diphenylpropanoic acid</b>	Arg121, Lys123, Val58, Leu130, Leu165, Ile27
9	IFN- $\gamma$	1FG9	2.9	*AX-024, Pralnacasan	Gln1, Glu9, Lys12, Gly18, Ser20, Ala23, Asp24, Asn25, Gly26, Lys108, His111, Glu112, Gln115, Ala118
10	NF- $\kappa$ B	1A3Q	2.1	*(-)-DHMEQ, Daxanabinol	Monomer 1: Arg52, Arg54, Tyr55, Gly56, Cys57, Glu58, His62, Lys143, Lys221, Ser222, Lys252, Gln284 Monomer II: Gly50, Tyr55, Cys57, Lys143, Ser220, Asn227, Lys252, Lys255, Val281, Gln284, Wat1-13
11	eNOS	4D1P	1.731	<sup>†</sup> <b>6-(((3S, 5R)-5-(((6-amino-4-methylpyridin-2-yl)methoxy)methyl)pyrrolidin-3-yl)oxy)methyl)-4-methylpyridin-2-amine</b>	Val336, Trp356, Glu361, Phe105, Phe353, Trp447, Trp74, Wat2049, 2282, 2374, 2283, 2372, 2370, 2270, 2279, 2373, 2349 Heme binding: Trp178, Arg183, Val185, Gly186, Cys184, Tyr475, Ser226, Phe353, Ser354, Trp356, Met358, Glu361, Trp447, Phe473, Tyr475
12	iNOS	3E7G	2.2	<sup>†</sup> <b>Ethyl 4-[(4-methylpyridin-2-yl)amino]piperidine-1-carboxylate</b>	Ligand binding site: Tyr347, Glu377, Gln263, Arg266, Pro350, Val352, Phe369, Gly371, Trp372, Tyr373, Asp382, Arg388 Heme binding: Cys200, Tyr491, Trp194, Ala197, Gly202, Ser242, Phe369, Asn370, Trp372, Glu377, Trp463, Tyr489, Tyr491, Wat4015, 4048, 4023, 4116, 4121, 4123, 4072
13	PPAR- $\gamma$	1NYX	2.65	<sup>†</sup> <b>(2s)-2-ethoxy-3-{4-[2-(10h-phenoxazin-10-yl)ethoxy]phenyl}propanoic acid</b>	His323, His449, Tyr473, Ile281, Phe282, Cys285, Gln286, Arg288, Ser289, Ile326, Leu330, Leu240, Ile341, Ser342, Leu453, Leu469, Wat14
14	MAPK2	3KC3	2.9	<sup>†</sup> <b>N4-(7-(benzofuran-2-yl)-1H-indazol-5-yl)pyrimidine-2,4-diamine</b>	Lys93, Leu141, Thr206, Leu70, Gly73, Val78, Ala91, His108, Met138, Glu139, Cys140, Leu141, Asp142, Leu193, Asp207
15	JAK 1	6N7B	1.81	<sup>†</sup> <b>N-[3-(5-chloro-2-methoxyphenyl)-1-methyl-1H-pyrazol-4-yl]-1H-pyrazolo[4,3-c]pyridine-7-carboxamide</b>	Leu881, Gly882, Glu883, Gly884, Gly887, Lys888, Val889, Ala906, Met956, Glu957, Phe958, Leu959, Gly962, Glu966, Arg1007, Asn1008, Leu1010, Gly1020, Asp1021, Wat1303, 1351, 1397
16	JAK 2	4BBE	1.9	<sup>†</sup> <b>N-[4-[2-[(4-morpholin-4-ylphenyl)amino]pyrimidin-4-yl]phenyl]ethanamide</b>	Lys882, Leu932, Leu855, PHE 860, VAL 863, LYS 882, MET 929, GLU 930, Tyr931, Leu932, Gly935, Leu983, Asp994, Gln853
17	JAK 3	6AAK	2.67	<sup>†</sup> <b>4-[[[(1S,3R)-5-oxidanyl-2-adamantyl]amino]-1H-pyrrolo[2,3-b]pyridine-5-carboxamide</b>	Leu905, Glu903, Leu828, Ala853, Val884, Met902, Tyr904, Leu905, Gly908, Arg953, Asn954, Leu956, Wat1302, 1309

(Continued)



TABLE 1 | Continued

S. no.	Molecular targets	PDB ID	Resolution (Å)	Positive control	Binding residues
18	STAT3	6QHD	2.85	*STAT3-IN-3, STAT2-IN-1	Arg382(A/B),Ser465(A/B), Asn466(A/B),His332(A/B), Lys340(A/B), Ile467(A/B), Arg417A, Gln469A, Wat1101C, 1102C, 1103C, 1101D, 1103D, 815B, 818A, 821B, 808A

TNF- $\alpha$ , Tumor necrosis factor- $\alpha$ ; IL-1 $\alpha$ , Interleukin-1 $\alpha$ ; IL-1 $\beta$ , Interleukin-1 $\beta$ ; IL-13, Interleukin-13; IL-17, Interleukin-17; IL-22, Interleukin-22; IL-23, Interleukin-23; IL-36 $\gamma$ , Interleukin-36 gamma; IFN- $\gamma$ , Interferon-Gamma; NF- $\kappa$ B, nuclear factor kappa-light-chain-enhancer of activated B cells; eNOS, endothelial Nitric oxide synthase; iNOS, inducible Nitric oxide synthase; PPAR- $\gamma$ , Peroxisome proliferator-activated receptors- $\gamma$ ; MAPK, Mitogen activated protein Kinase; JAK, Janus Kinase; STAT3, Signal transducer and activator of transcription 3.

<sup>†</sup> These ligands were taken from Protein databank ([www.rcsb.org](http://www.rcsb.org)).

<sup>\*</sup> These ligands were taken from MedChemExpress (<https://www.medchemexpress.com/>).

Later, the compounds were ranked on the basis of MolDock and re-rank scores.

### AutoDock Vina

AutoDock Vina (version 1.1.2) was used in this study (Trott and Olson, 2010). AutoDock Vina requires the receptor and the ligand files in PDBQT [Protein Data Bank, Partial Charge (Q) and Atom Type (T)] format. We used the same PDB files of the receptors and the mol2 files of the ligands that were used in MOE, and we converted the structures to pdbqt format using PyRx Virtual Screening tool [<https://pyrx.sourceforge.io/>]. The docking search space was defined around 3 Å of selected residues or ligand (where available) with a grid box size of 25 × 25 × 25 Å. The number of binding modes and exhaustiveness of search was set to 30 and eight, respectively. Later, each of the conformation was visualized on UCSF Chimera, and the best docked orientation was selected based on docking score and binding interactions with the protein.

### Analysis Measures and Conformational Sampling

To select the most appropriate inhibitor against selected target, a consensus approach was used. The docked library was sorted on the basis of docking scores, and the suitable inhibitor for a particular target was selected when it is ranked among the top 10 ligands of the screened dataset by two out of three programs. The most optimal binding mode of each compound was chosen by conformational sampling. All the docked poses of each compound (generated by each program) were compared, and the binding mode that was analogous in all the docking methods was declared as the probable binding mode.

## RESULTS AND DISCUSSION

### Re-docking

The robustness of the docking methods was scrutinized by re-docking experiments. For this purpose, 10 (IL-17, IL-36 $\gamma$ , PPAR- $\gamma$ , MAPK2, JAK1/2/3, TNF- $\alpha$ , iNOS, and eNOS) out of 18 selected targets were chosen due to the presence of co-crystallized ligands (inhibitors) in their PDB structures. All the programs successfully re-produced the X-ray confirmed orientation of each ligand with root mean square deviation (RMSD) of  $\leq 2.5$  Å. Additionally, the screening reliability was justified by the

ranking position of known inhibitors of all the target, and it was observed that all the positive controls were ranked at the top of the screened library. The results indicate that the used docking methods are reliable in predicting the binding orientation and in the selection of known inhibitors. The re-docking results are summarized in supporting information Table S1.

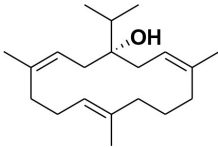
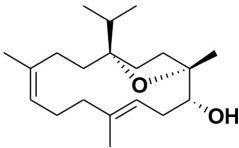
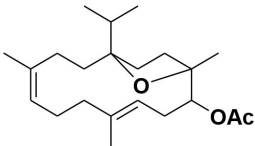
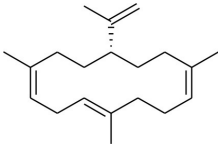
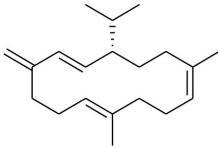
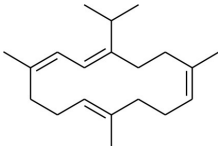
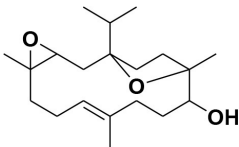
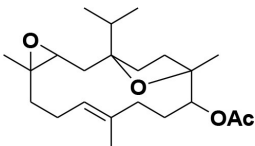
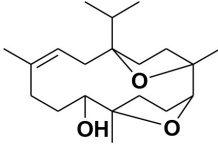
### Molecular Targeting of Cembrane Diterpenes and Boswellia Triterpenes

In the current study, a reverse docking approach was employed to predict the binding potential of selected cembrenoid diterpenoids and *Boswellia* triterpenoids with the 18 drug targets. For this purpose, three different docking algorithms and scoring functions were used. Those proteins, which were collectively ranked with higher score by all the docking methods, were considered as “best” targets. Based on the docking score, JAK1/2/3, eNOS, and iNOS were proposed as the best targets for all the compound by MOE and ADT vina, while MVD only picked 17/46 and 12/46 compounds as excellent inhibitors for JAK1 and eNOS, respectively. Similarly, ADT Vina ranked JAK1 and eNOS among the top ranked targets for all the compounds, while JAK2, JAK3, and iNOS were retrieved as good binders for 42/46, 40/46, and 39/46 compounds, respectively. MVD demonstrated that 40/46, and 37/46 compounds can efficiently target JAK2 and 3, respectively. While >39, 26, and 50% of the screened library may target JAK1, eNOS, and iNOS, respectively. Thus, JAK1/2/3, eNOS and iNOS were considered as the most appropriate druggable candidates for the selected di- and tri- terpenoids.

Among the selected cytokines, IL-17 was retrieved as an interesting target since MVD showed that all the compounds may neatly fit at the binding pocket of IL-17, while MOE and ADT Vina suggested that 42/46 and 40/46 compounds can effectively target IL-17, respectively. According to MOE and ADT Vina, 45/46 compounds possess high to moderate binding affinities for TNF- $\alpha$ , while MVD showed that >63% of the screened database holds good inhibitory potential against TNF- $\alpha$ . Therefore, IL-17 and TNF- $\alpha$  were also considered as good targets for these compounds.

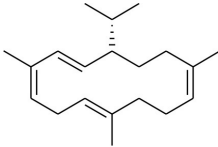
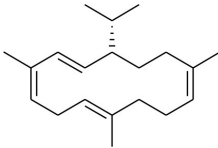
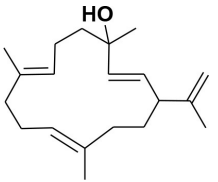
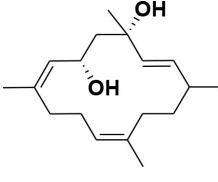
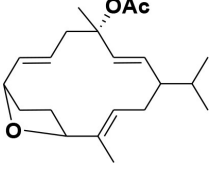
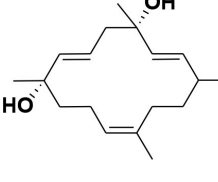
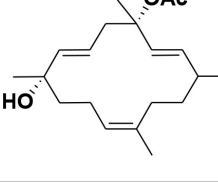
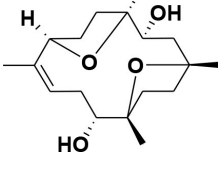
Based on the *in-silico* outcomes, JAK1/2/3, eNOS, iNOS, IL-17, and TNF- $\alpha$  are the most appropriate targets for the purified chemical constituents of frankincense. Moreover, the consensus

**TABLE 2** | Chemical structures of compounds 1–46.

Compounds	Name	Source	Structure	References
1	Serratol	<i>Boswellia serrata</i>		Schmidt et al., 2011
2	Incensole	<i>Boswellia carterii</i>		Corsano and Nicoletti, 1967
3	Incensole acetate	<i>Boswellia carterii</i>		Obermann, 1977; Moussaieff et al., 2007
4	Cembrene A	<i>Boswellia carterii</i>		Shmidt et al., 1970; Patil et al., 1973
5	Isocembrene	<i>Boswellia carterii</i>		Wahab et al., 1987
6	Cembrene C	<i>Boswellia carterii</i>		Barakat and Rullkötter, 1993
7	Incensole oxide	<i>Boswellia carterii</i>		Nicoletti and Forcellese, 1968
8	Incensole oxide acetate	<i>Boswellia carterii</i>		Hamm et al., 2005
9	Isoincensole oxide	<i>Boswellia carterii</i>		Yamago et al., 1998

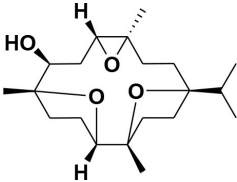
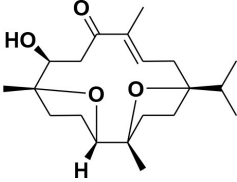
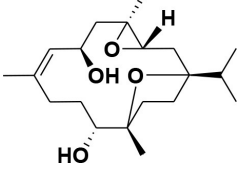
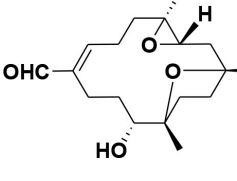
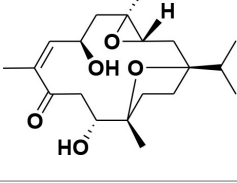
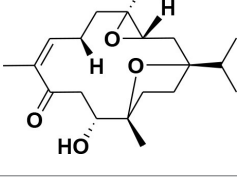
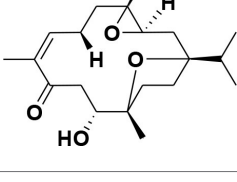
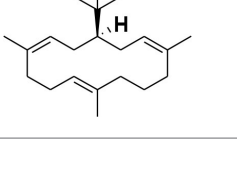
(Continued)

TABLE 2 | Continued

Compounds	Name	Source	Structure	References
10	Cembrene	<i>Boswellia carterii</i>		Basar et al., 2001
11	Isocembrene	<i>Boswellia carterii</i>		Basar et al., 2001
12	Thunbergol	<i>Boswellia carterii</i>		Mikhaeil et al., 2003
13	Duva-4,8,13-triene-1,3- $\alpha$ -diol	<i>Boswellia carterii</i>		Mikhaeil et al., 2003
14	Duva-3,9,13-triene-1- $\alpha$ -hydroxy-5,8-oxide-1-acetate	<i>Boswellia carterii</i>		Mikhaeil et al., 2003
15	Duva-3,9,13-triene-1,5- $\alpha$ -diol-1-acetate	<i>Boswellia carterii</i>		Mikhaeil et al., 2003
16	Duva-3,9,13-triene-1,5- $\alpha$ -diol	<i>Boswellia carterii</i>		Mikhaeil et al., 2003
17	Boscartin A	<i>Boswellia carterii</i>		Ren et al., 2015

(Continued)

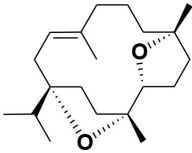
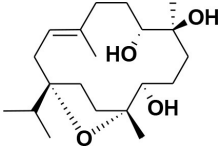
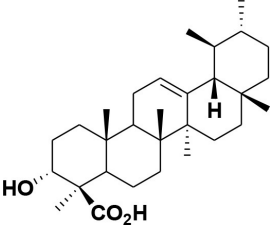
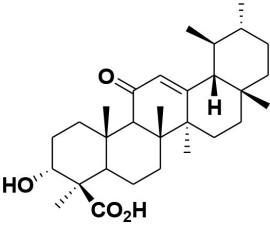
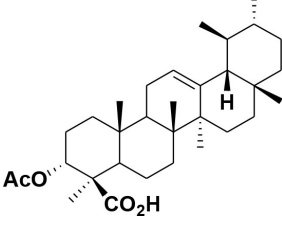
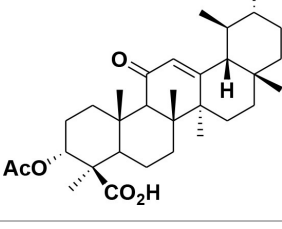
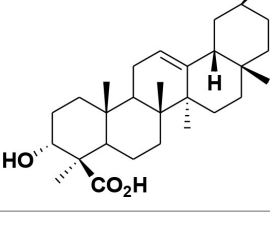
TABLE 2 | Continued

Compounds	Name	Source	Structure	References
18	Boscartin B	<i>Boswellia carterii</i>		Ren et al., 2015
19	Boscartin C	<i>Boswellia carterii</i>		Ren et al., 2015
20	Boscartin D	<i>Boswellia carterii</i>		Ren et al., 2015
21	Boscartin E	<i>Boswellia carterii</i>		Ren et al., 2015
22	Boscartin F	<i>Boswellia 2carterii</i>		Ren et al., 2015
23	Boscartin G	<i>Boswellia carterii</i>		Ren et al., 2015
24	Boscartin H	<i>Boswellia carterii</i>		Ren et al., 2015
25	Iso-serratol	<i>Boswellia carterii</i>		Basar, 2005

(Continued)

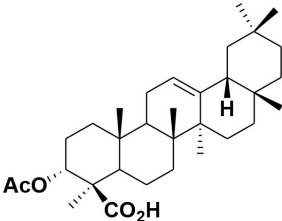
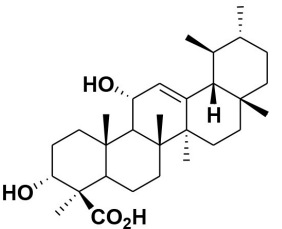
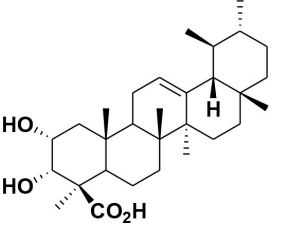
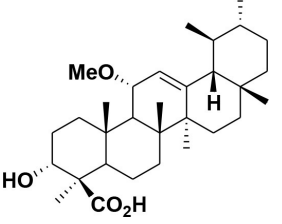
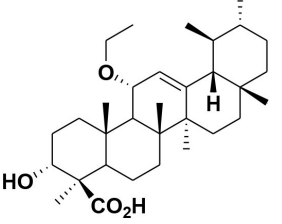
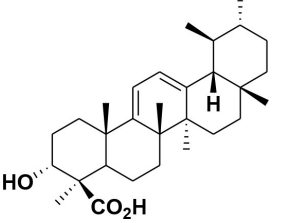


TABLE 2 | Continued

Compounds	Name	Source	Structure	References
26	Incensfuran	<i>Boswellia papyrifera</i>		Rehman et al., 2017
27	Iso-incensolol	<i>Boswellia carterii</i>		Li et al., 2010
28	$\beta$ -Boswellic acid	<i>Boswellia sacra</i>		Pardhy, 1978; Al-Harrasi et al., 2018a
29	11-keto- $\beta$ -Boswellic acid	<i>Boswellia sacra</i>		Pardhy, 1978; Al-Harrasi et al., 2018a
30	3- $\alpha$ -acetyl- $\beta$ -Boswellic acid	<i>Boswellia sacra</i>		Pardhy, 1978; Al-Harrasi et al., 2018a
31	3- $\alpha$ -acetyl-11-keto- $\beta$ -Boswellic acid	<i>Boswellia sacra</i>		Pardhy, 1978; Al-Harrasi et al., 2018a
32	$\alpha$ -Boswellic acid	<i>Boswellia carterii</i>		Akihisa et al., 2006

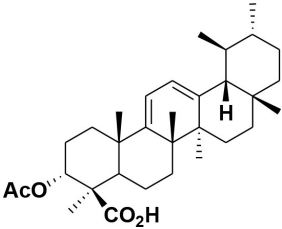
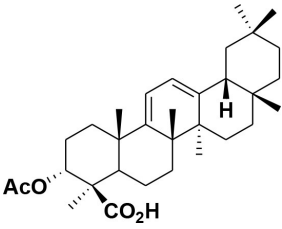
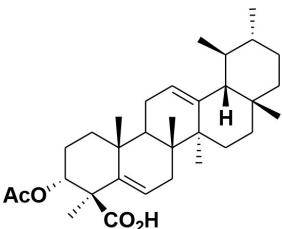
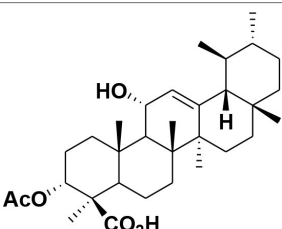
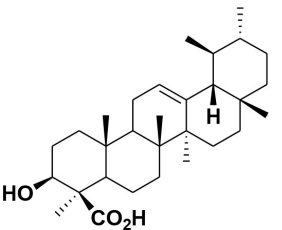
(Continued)

TABLE 2 | Continued

Compounds	Name	Source	Structure	References
33	3- $\alpha$ -acetyl- $\alpha$ -Boswellic acid	<i>Boswellia carterii</i>		Akihisa et al., 2006
34	3,11-dihydroxy-12-ene-24-oic acids	<i>Boswellia sacra</i>		Al-Harrasi et al., 2018b
35	2a,3a-dihydroxy-urs-12-ene-24-oic acid	<i>Boswellia serrata</i>		Mahajan et al., 1995
36	11- $\alpha$ -methoxy- $\beta$ -boswellic acid	<i>Boswellia carterii</i>		Ota and Houghton, 2008
37	11- $\alpha$ -ethoxy- $\beta$ -boswellic acid	<i>Boswellia carterii</i>		Al-Harrasi et al., 2013b
38	9,11-dehydro- $\beta$ -boswellic acid	<i>Boswellia sacra</i>		Ali et al., 2014

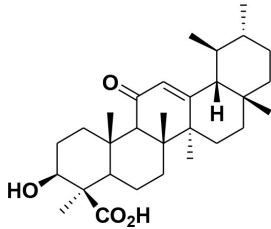
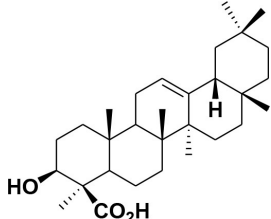
(Continued)

TABLE 2 | Continued

Compounds	Name	Source	Structure	References
39	3-O-acetyl-9,11-dehydro- $\beta$ - boswellic acid	<i>Boswellia sacra</i>		Al-Harrasi et al., 2013a
40	9,11-dehydro- $\alpha$ - boswellic acid	<i>Boswellia serrata</i>		Büchele et al., 2003
41	3-O-acetyl-9,11-dehydro- $\alpha$ - boswellic acid	<i>Boswellia serrata</i>		Büchele et al., 2003
42	3a-acetoxyurs-5:12-dien-24-oic acid	<i>Boswellia sacra</i>		Ali et al., 2014
43	3-acetyl-11-hydroxy-12-ene-24-oic acids	Common Boswellia		Corsano and Iavarone, 1964
44	3-epi- $\beta$ -Boswellic acid	Synthetic		Kumar et al., 2012

(Continued)

TABLE 2 | Continued

Compounds	Name	Source	Structure	References
45	3-epi-11-keto- $\beta$ -Boswellic acid	Synthetic		Kumar et al., 2012
46	3-epi- $\alpha$ -Boswellic acid	Synthetic		Al-Harrasi et al., 2018b

results of all the docking methods depicted that MAPK2, PPAR- $\gamma$ , IL-13, IL-23, IL-36 $\gamma$ , and IFN $\gamma$  can serve as possible binding proteins for several di- and tri- terpenes. Meanwhile, IL-1 $\alpha$ , IL-1 $\beta$ , NF- $\kappa$ B, IL-22, and STAT3 were identified as the least potential targets for these cembrenoid derivatives. The best inhibitors for each target are given in supporting information **Table S2** and the docking scores are tabulated in **Tables S3–S5**. The predicted binding potential of each compound with the selected target is presented as a heatmap in **Figure 1**.

## Top Predicted Targets

### JAK1/2/3

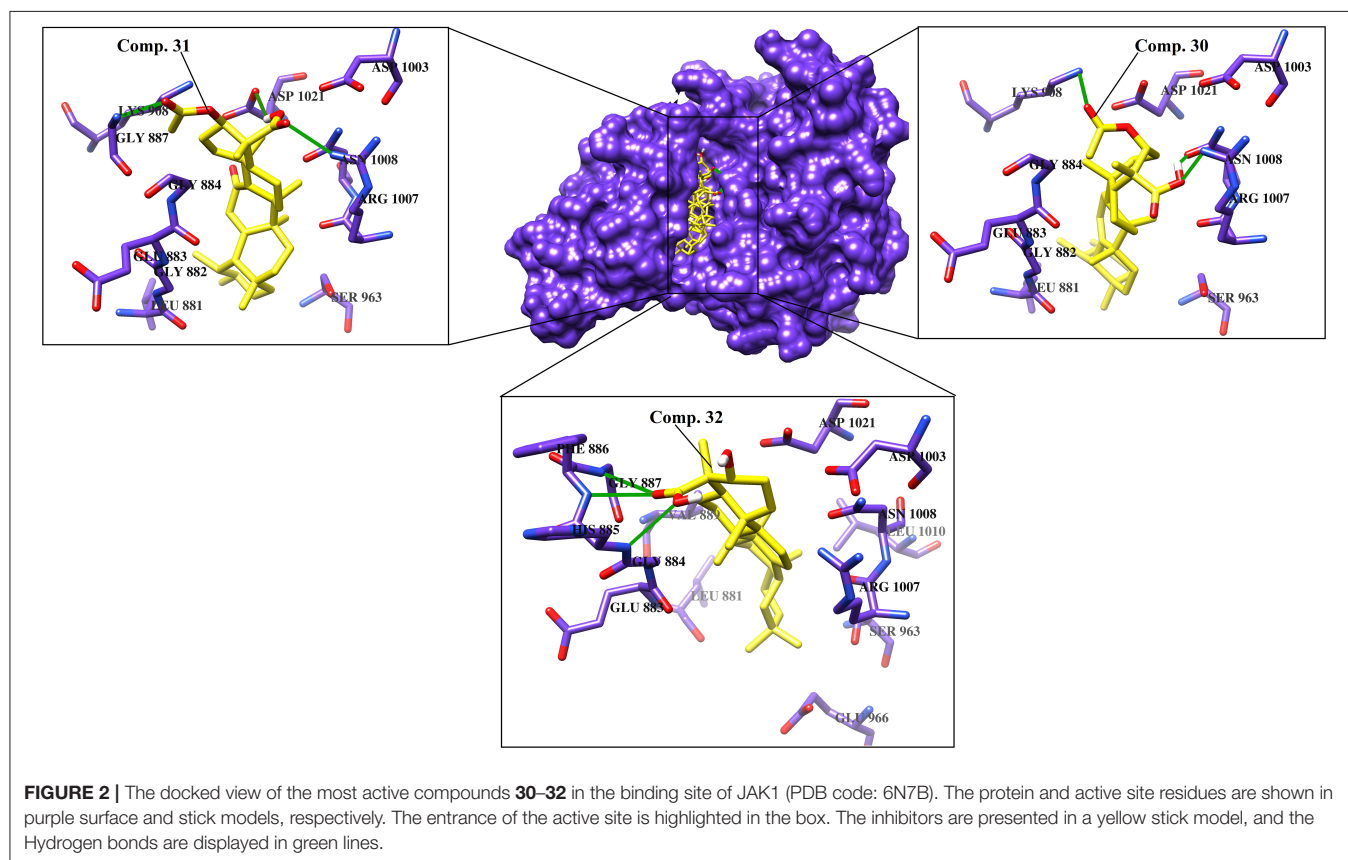
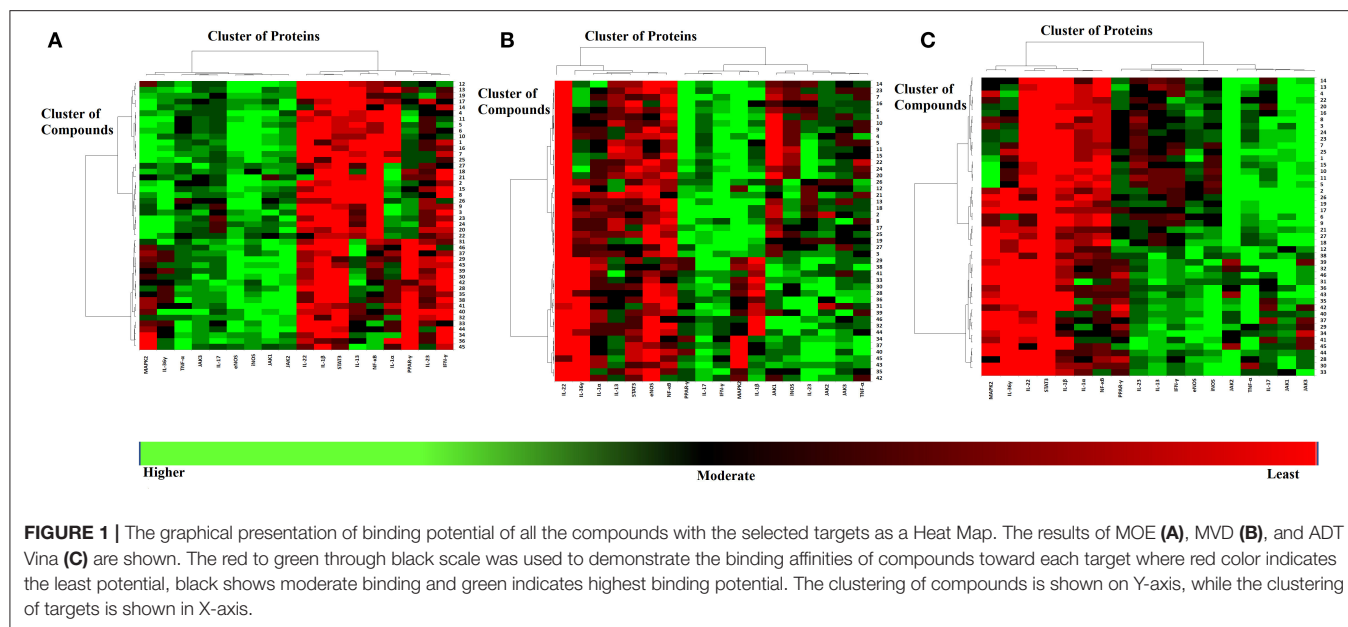
Various cytokines are regulated by the JAK-STAT pathway, particularly TH17 signaling, which is important in the pathogenesis of psoriasis. JAK inhibitors are efficacious against psoriasis, alopecia areata, and atopic dermatitis (Ciechanowicz et al., 2019). When docked at the active site of JAK1, all the di- and tri- terpenoids were predicted as the potent inhibitors by MOE and ADT vina; however, MVD ranked 17/46 compounds (**12**, **28–34**, **36–38**, **40–41**, and **43–46**) as the best binders of JAK1. Among them, four molecules (**6**, **30–32**) were mutually ranked by all the docking methods among the top 10 ligands; however, interaction analysis suggest that only **30–32** can serve as potent inhibitor of JAK1. The binding modes of **30–32** suggests that these compounds neatly fit at the active site of JAK1 where Lys908, Arg1007, Asn1008, Phe886, His885, and Gly887 particularly stabilizes these compounds through H-bonding. The carboxylic group of 3- $\alpha$ -acetyl- $\beta$ -Boswellic acid (**30**) mediates bidentate interactions with the side chains of Asn1008 and Arg1007, while the acetyl moiety interacts with the side chain of Lys908. Similarly, carboxylic group of 3- $\alpha$ -acetyl-11-keto- $\beta$ -Boswellic acid (**31**) interacts with the side chains of Asn1008 and Asp1021, whereas the acetyl group mediates H-bonding with the amino group of Gly887. However, the carboxylic group of  $\alpha$ -Boswellic acid (**32**) mediates H-bonding

with the amino groups of His885, Phe886, and Gly887. These pentacyclic triterpene boswellic acids possesses promising anti-inflammatory properties (Al-Harrasi et al., 2018a). The docked conformations of **30–32** are presented in **Figure 2**.

JAK2 and JAK3 were also identified as superior targets for each molecule by MOE, while MVD and ADT vina demonstrated that 40/46 compounds can effectively target JAK2. The consensus examination depicted that five cembrane diterpenes (**3**, **14**, **16**, **20**, **23**,) and three boswellic acid derivatives (**28**, **37**, and **43**) are the most prominent inhibitor of JAK2. The binding mode of incensole acetate (**3**) showed that the acetate moiety of the compound mediates H-bonding with the side chain of Lys882, which is homologous to Lys908 and Lys855 in JAK1 and JAK3, respectively. Lys882 also mediate H-bonding with the acetate groups of **14** and **16**. The docked view of **20** reflects that this compound has a chance to bind with several residues (Leu932, Tyr931, and Arg980) due to the presence of several polar -OH groups in its structure. The -OH and the carbonyl groups of **23** interact with the side chains of Lys882 and Asp994, respectively. This indicates that Lys882 is a crucial residues in the stabilization of diterpenoids at the active site of JAK2. The carboxylic groups of triterpenoid **28** and **43** were found to be H-bonded with the side chain of Arg980, while the carboxylic moiety of **37** interacts with the side chain of Asn981 via H-bonding. Arg980 and Asn981 of JAK2 are homologous to Arg1007 and Asn1008 of JAK1, respectively. This suggest that these Arg and Asn residues play important roles in the binding of boswellic acids in the active site of JAK1/2.

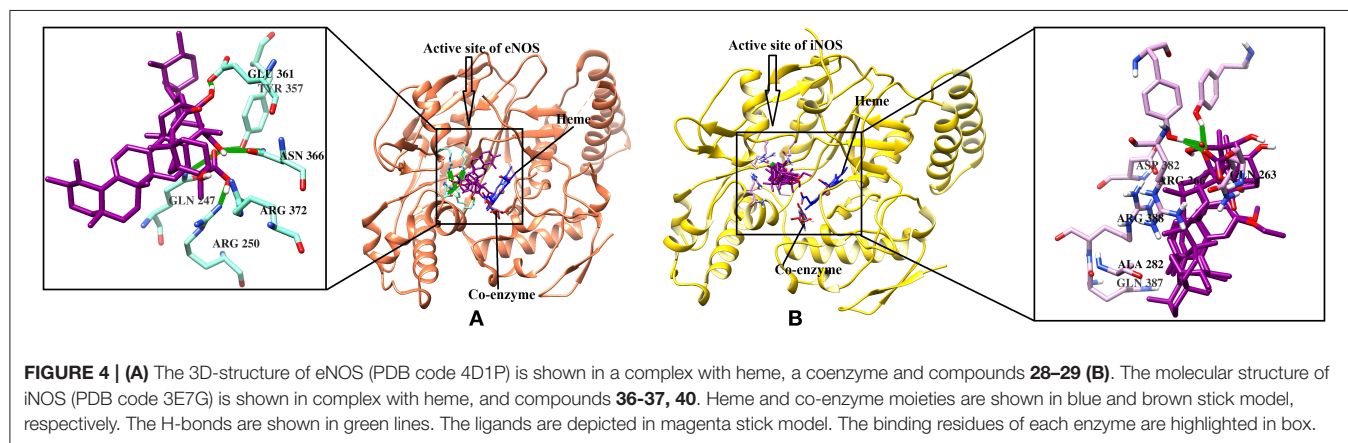
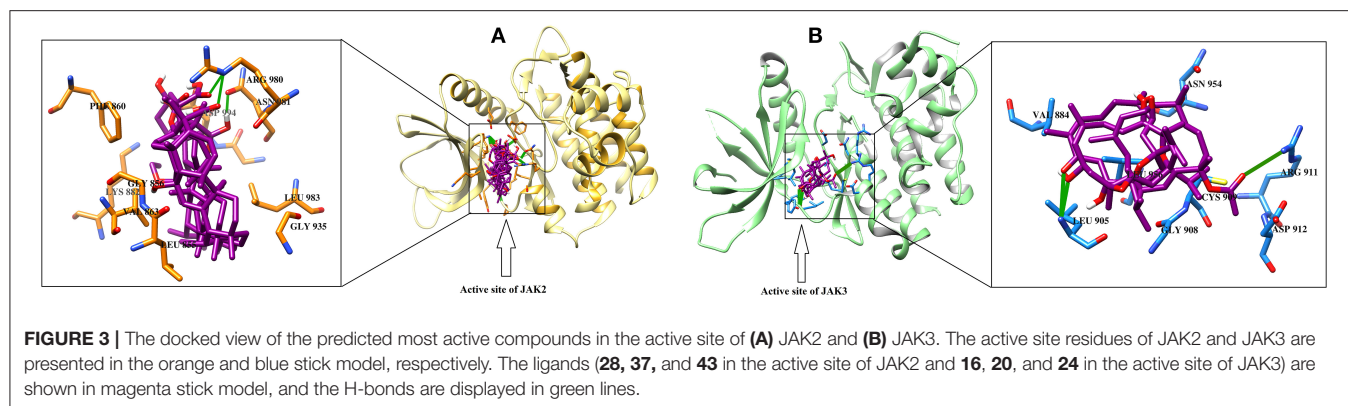
For JAK3, only 9/46 and 6/46 compounds showed the least inhibitory potential in MVD and ADT vina docking, respectively, while rest of the compounds exhibited higher binding potential. Meanwhile, compounds **16**, **20**, **24**, and **36** were communally ranked at the top of the screened dataset by all three scoring functions. It was observed that compounds **16** and **20** can inhibit both the enzymes. 3- $\alpha$ -acetyl-11-keto- $\beta$ -boswellic acid (**31**) is





known to inhibit the activation of STAT3 by inhibiting the phosphorylation of JAK2 and Src. For the dephosphorylation of STAT3, compound **31** induces Src homology region 2 domain-containing phosphatase 1 (SHP-1) (Kunnumakkara et al., 2009). Our docking results are consistent with these experimental

findings. Compound **31** was identified as a good inhibitor of JAK1/2/3 by all the scoring functions. The binding modes of selected hits (**16**, **20**, **24**, and **36**) showed that the methyl acetate of **16** binds with Arg911 and Cys909, while the -OH group of **20** and the carbonyl group of **24** interacts with the amino group



of Leu905. The binding mode of 11- $\alpha$ -methoxy- $\beta$ -boswellic acid (**31**) depicted that this compound did not produce significant binding interactions within the active site of JAK3. The binding modes of best predicted inhibitors in the active site of JAK2/3 are shown in **Figure 3**.

### eNOS and iNOS

eNOS and iNOS are involved in immune response, and they are important enzymes in the pathogenesis of psoriasis (Mittal et al., 2009; Rácz and Prens, 2009). The post docking analysis showed that all the compounds bind at the entrance of the active site of eNOS and iNOS. All the derivatives were ranked as good inhibitors of eNOS and iNOS by MOE. Similarly, ADT vina also showed that all the cembrenoids possess greater inhibitory affinities for eNOS; however, except for **1**, **2**, **5**, **7**, **10**, **11**, and **15**, the rest of the compounds exhibits excellent binding tendency for iNOS. Despite this, the MolDock score filtered only few compounds as good binders of eNOS (**Table S2**) and 50% of the docked library (**2**, **3**, **8**, **12**, **13**, **18**, **21**, **26**, **29**, **31–34**, **36–38**, and **40–46**) for iNOS. Among all, compounds **15–16**, **28–29**, and **37–38** as well as compounds **8**, **21**, **36–37**, and **40** were of utmost importance for eNOS and iNOS, respectively. It is shown that compound **37** targets both eNOS and iNOS effectively, suggesting it could be the most important inhibitor of eNOS and iNOS. The binding modes of the most active compound **28–29** and **36–37** and **40** in the active site of eNOS and iNOS, respectively, are shown in **Figure 4**.

### IL-17

IL-17 and TNF- $\alpha$  emerged as good targets for most of the compounds by all the docking methods. An IL-17 dimer binds with IL-17RA and forms a protein-protein interface, which is also blocked by recently identified cyclo-/linear- peptide inhibitors (Blauvelt and Chiricozzi, 2018; Bremilla et al., 2018). Based on the CA criterion, compounds **16**, **18**, **28**, **29**, and **32** were retrieved as the most promising inhibitors for IL-17. The mode of interactions reveals that **18**, **28**, **29**, and **32** are well accommodated at the antagonist binding site of IL-2 where Gln94, Glu95, and Leu97 play important roles in the stabilization of these compounds (supporting information **Figure S1**).

### TNF- $\alpha$

TNF- $\alpha$  plays a crucial role in the exacerbation of inflammation in psoriasis. The management of psoriasis is done by inhibiting the production of TNF- $\alpha$  by several inhibitors like adalimumab, certolizumab pegol, etanercept, golimumab, and infliximab. MOE and ADT vina ranked TNF- $\alpha$  as a moderate binder for all the compounds except compound **26** and **30**, respectively. While MVD predicted that 63% of compounds can exhibit good-to-moderate inhibition of TNF- $\alpha$ . According to CA, compounds **20**, **36**, **38**, **40**, **43**, and **45** were regarded as the most effective binders of TNF- $\alpha$ . mediated favorable interactions with the Ser60, Leu120, Gly121, and Tyr151 and displayed excellent to moderate inhibition of TNF- $\alpha$  *in silico*. The interactions of ligands with TNF- $\alpha$  are shown in **Figure S2**.





activation of NF- $\kappa$ B by inhibiting TAK/TAB-mediated I $\kappa$ B kinase (IKK) activation loop phosphorylation (Berrrrpohl et al., 2007; Moussaieff et al., 2007). It is also reported that **3** did not suppress I $\kappa$ B $\alpha$  phosphorylation in co-stimulated T cells, indicating that the kinase inhibition is neither direct nor does it affect all NF- $\kappa$ B activation pathways. Moreover, acetyl-11-keto- $\beta$ -boswellic acid (**31**) and acetyl- $\alpha$ -boswellic acid (**33**) inhibits NF- $\kappa$ B signaling by specifically inhibiting I $\kappa$ B $\alpha$  kinase (IKK), which is pivotal for the degradation of the NF- $\kappa$ B inhibitor I $\kappa$ B, as well as the phosphorylation of p65, two steps essential for NF- $\kappa$ B activation and the subsequent cytokine expression. Syrovets et al. (2005) proved that acetyl-boswellic acids inhibits LPS-mediated TNF- $\alpha$  induction in monocytes by direct interaction with I $\kappa$ B kinase (Syrovets et al., 2005). Our docking results also indicate that these compounds have less of a potential to directly bind with NF- $\kappa$ B.

IL-17 is a pro-inflammatory cytokine produced by Th-17 cell in response to their stimulation by IL-23. IL-17 interacts with its receptor to mediate proinflammatory and allergic responses, and, as a result, production of several cytokines (IL-1 $\beta$ , TGF- $\beta$ , TNF- $\alpha$ , IL-6, GM-CSF, and G-CSF), chemokines (GRO- $\alpha$ , IL-8, and MCP-1), and prostaglandins (PGE2) occurs via macrophages, keratinocytes, fibroblasts, epithelial, and endothelial cells. Activation of IL-17 signaling is observed in the pathogenesis of various autoimmune disorders, including rheumatoid arthritis, asthma, lupus, allograft rejection, anti-tumor immunity, psoriasis, and multiple sclerosis, and IL-17 inhibitors are being investigated as possible treatments for these disorders. The monoclonal antibodies (mAb) secukinumab and ixekizumab inhibit IL-17 while brodalumab blocks the IL-17 receptor (IL-17RA). By interfering with IL-17 signaling, these antibodies interrupt the inflammatory cycle of psoriasis. The mAb against IL-23 (ustekinumab) is also used to treat psoriasis by reducing IL-17, and, thus, the IL-23/IL-17 pathway plays a major role in psoriasis. Recently, Stürner et al. (2014) revealed that compound **29** (11-keto- $\beta$ -Boswellic acid, KBA) reduces the differentiation of human CD4 $^{+}$  T cells to Th17 cells while slightly increasing Th2- and Treg-cell differentiation. Furthermore, KBA reduces the IL-1 $\beta$ -triggered IL-17A release of memory Th17 cells. KBA may affect IL-1 $\beta$  signaling by preventing IL-1R-associated kinase 1 phosphorylation and subsequently decreasing STAT3 phosphorylation at Ser727, which is required for Th17-cell differentiation. The effects of KBA on Th17 differentiation and IL-17A release make the compound a good candidate for potential treatment of Th17-driven diseases. Boswellic acids reduce Th17 differentiation via blockade of IL-1 $\beta$ -mediated IRAK1 signaling (Stürner et al., 2014). Our docking results also confirm that compound **29** (KBA) possesses excellent binding affinity for IL-17.

The monoclonal antibody (ustekinumab) inhibits IL-12/IL-23, which are both associated with psoriatic inflammation and psoriatic arthritis. Moreover, mAb against IL-23 (tildrakizumab, Risankizumab, and guselkumab) also reduce psoriatic symptoms and slow disease progression. IL-23 is involved in the proliferation of CD4 $^{+}$  Th17 cells to produce IL-17 and along with IL-17, associated with several autoimmune and inflammatory disorders such as psoriasis, psoriatic/rheumatoid arthritis, multiple sclerosis, Crohn's disease, uveitis, and

inflammatory bowel disease (Braegelmann et al., 2016; Chan et al., 2018; Fotiadou et al., 2018). Boscartin F (**22**), 3- $\alpha$ -acetyl-11-keto- $\beta$ -Boswellic acid (AKBA, **31**), 3-O-acetyl-9,11-dehydro- $\beta$ -boswellic acid (**39**), 3-epi- $\beta$ -Boswellic acid (**44**), and 3-epi-11-keto- $\beta$ -Boswellic acid (**45**) were identified as excellent inhibitors of IL-23 in this study. Previously, the effect of **31** (3- $\alpha$ -acetyl-11-keto- $\beta$ -Boswellic acid) was studied on activated dendritic cells in psoriasis-like mouse model that revealed that this compound improved the psoriasis-like skin lesions, reduced the thickness of epidermis, ameliorated the infiltration of CD3 $^{+}$  and CD11c $^{+}$  cells in skin lesions, decreased the activation of local dendritic cells, inhibited the expression and secretion of IL-12 and IL-23, inhibited the maturation and differentiation of DCs to promote T-cell differentiation, and inhibited the activation of TLR7/8 and IRF signaling pathways. Thus, AKBA might have an anti-inflammatory effect on psoriasis by inhibiting the maturation and activation of DCs via the TLR8 and IRF signaling pathways (Wang et al., 2018). Our docking results also suggests that AKBA tends to inhibit IL-23.

The cytokine IL-1 $\alpha$  is produced by macrophages, neutrophils, and epithelial and endothelial cells, while IL-1 $\beta$  is produced by activated macrophages and mediates inflammatory response after binding with their receptor. Both the cytokines are excessively produced in psoriatic lesions (Tsai and Tsai, 2017). Both the cytokines were depicted as moderate to least binding proteins for this set of compounds. IL-22 is produced by T cells and is involved in the modulation of tissue responses during inflammation and found significantly higher in psoriatic patients (Fujita, 2013; Hofny et al., 2017; Voglis et al., 2018). IL-22 was also depicted as a least possible target for di- and tri-terpene derivatives.

PPAR- $\gamma$  regulates lipid and glucose metabolism. Recently, the pathophysiological role of PPAR- $\gamma$  was reported in the inflammatory immune response and identified as a major drug target for the treatment of psoriasis, benign epidermal tumors, and atopic dermatitis (Sertznig and Reichrath, 2011). Liu et al. (2013) described that AKBA (**31**) downregulates the expression of PPAR- $\gamma$ 2 and loss of phenotypic markers of mature human adipocytes and mobilizes lipolysis (Liu et al., 2013). In the current study, **31** was recognized as an excellent inhibitor of PPAR- $\gamma$  by MVD.

In psoriatic patients, STAT3 is activated in lesional keratinocytes, which can lead to the development of regenerative epidermal phenotype observed in psoriasis (Sano et al., 2005). Incensole (**2**) is a potent inhibitor of STAT3 (Pollastro et al., 2016). It is reported that **29** (KBA) inhibits the phosphorylation of STAT3 (Stürner et al., 2014). In this study, both the compounds were identified as moderate inhibitors of STAT3 by MVD. Our *in-silico* experiment correlates well with the reported experimental findings. The docking studies are used in the discovery of novel inhibitors against several therapeutic targets (Halim et al., 2013). However, docking must be combined with more sophisticated computational techniques like molecular dynamic simulation to enhance the efficiency of virtual screening. This work needs to be further explored by molecular dynamic simulation to study the dynamic behavior of protein upon protein-ligand complex formation.



## CONCLUSION

The anti-psoriatic potential of several *Boswellia* diterpenoids and triterpenoids was explored by computational analysis. The compounds have displayed selective docking to human anti-inflammatory (JAK1/2/3, eNOS, iNOS, and TNF- $\alpha$ ) and anti-psoriatic (IL-17) molecular targets. Moreover, several other proteins (MAPK2, PPAR- $\gamma$ , IL-13, IL-23, IL-36 $\gamma$ , and IFN $\gamma$ ) were retrieved as leading biological targets for these compounds. Thus, these compounds have reflected excellent anti-psoriatic and anti-inflammatory potency *in silico*. The results indicate that *Boswellia* di- and tri-terpenoids can serve as promising chemical scaffolds for the development and improvement of inhibitors to treat psoriasis.

## DATA AVAILABILITY STATEMENT

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

## AUTHOR CONTRIBUTIONS

AA-H, RC, and AA-R conceived and designed the study. SH performed all computational studies and analyzed the data. SH

and AK wrote the manuscript with inputs and comments from all co-authors. All authors have read and approved the final version of the manuscript.

## FUNDING

This project was supported by grant from The Oman Research Council (TRC) through the funded project (BFP/RGP/HSS/19/198).

## ACKNOWLEDGMENTS

The authors would like to thank the University of Nizwa for the generous support of this project. We thank technical staff for assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.00486/full#supplementary-material>

## REFERENCES

- Abdallah, M. A., Abdel-Hamid, M. F., Kotb, A. M., and Mabrouk, E. A. (2009). Serum interferon- $\gamma$  is a psoriasis severity and prognostic marker. *Cutis* 84, 163–168.
- Akihisa, T., Tabata, K., Banno, N., Tokuda, H., Nishihara, R., Nakamura, Y., et al. (2006). Cancer chemopreventive effects and cytotoxic activities of the triterpene acids from the resin of *Boswellia carteri*. *Biol. Pharm. Bull.* 29, 1976–1979. doi: 10.1248/bpb.29.1976
- Al Robaee, A. A. (2010). Molecular genetics of Psoriasis (Principles, technology, gene location, genetic polymorphism and gene expression). *Int. J. Health Sci.* 4:103.
- Al-Harrasi, A., Ali, L., Rehman, N. U., Hussain, H., Hussain, J., Al-Rawahi, A., et al. (2013a). Nine triterpenes from *Boswellia sacra* Flückiger and their chemotaxonomic importance. *Biochem. Syst. Ecol.* 51, 113–116. doi: 10.1016/j.bse.2013.08.026
- Al-Harrasi, A., Ali, L., Ur Rehman, N., Hussain, J., Hussain, H., Al-Rawahi, A., et al. (2013b). 11 $\alpha$ -Ethoxy- $\beta$ -boswellic acid and nizwanone, a new boswellic acid derivative and a new triterpene, respectively, from *Boswellia sacra*. *Chem. Biodivers.* 10, 1501–1506. doi: 10.1002/cbdv.201200231
- Al-Harrasi, A., Hussain, H., Csuk, R., and Khan, H. Y. (2018a). *Chemistry and Bioactivity of Boswellic Acids and Other Terpenoids of the Genus Boswellia*. London: Elsevier. doi: 10.1007/978-3-030-16725-7
- Al-Harrasi, A., Rehman, N. U., Khan, A. L., Al-Broumi, M., Al-Amri, I., Hussain, J., et al. (2018b). Chemical, molecular and structural studies of *Boswellia* species:  $\beta$ -Boswellic Aldehyde and 3-epi-11 $\beta$ -Dihydroxy BA as precursors in biosynthesis of boswellic acids. *PLoS ONE* 13:e0198666. doi: 10.1371/journal.pone.0198666
- Ali, L., Hussain, J., Al-Rawahi, A., and Al-Harrasi, A. (2014). Two new and four known triterpenoids from *Boswellia sacra* flückiger. *Rec. Nat. Prod.* 8:407.
- Augustin, M., Glaeske, G., Radtke, M., Christophers, E., Reich, K., and Schäfer, I. (2010). Epidemiology and comorbidity of psoriasis in children. *Br. J. Dermatol.* 162, 633–636. doi: 10.1111/j.1365-2133.2009.09593.x
- Baliwag, J., Barnes, D. H., and Johnston, A. (2015). Cytokines in psoriasis. *Cytokine* 73, 342–350. doi: 10.1016/j.cyto.2014.12.014
- Barakat, A. O., and Rullkötter, J. (1993). Gas chromatographic/mass spectrometric analysis of cembrenoid diterpenes in kerogen from a lacustrine sediment. *Org. Mass Spectrom.* 28, 157–162. doi: 10.1002/oms.1210280306
- Basar, S. (2005). *Phytochemical investigations on Boswellia species: Comparative studies on the essential oils, pyrolysates and boswellic acids of Boswellia carterii Birdw., Boswellia serrata Roxb., Boswellia frereana Birdw., Boswellia neglecta S. Moore and Boswellia rivae Engl (Doctoral dissertation, Verlag nicht ermittelbar)*.
- Basar, S., Koch, A., and König, W. A. (2001). A verticillane-type diterpene from *Boswellia carterii* essential oil. *Flavour Fragr. J.* 16, 315–318. doi: 10.1002/ffj.992
- Bejarano, J. J. R., and Valdecantos, W. C. (2013). Psoriasis as autoinflammatory disease. *Dermatol. Clin.* 31, 445–460. doi: 10.1016/j.det.2013.04.009
- Berrrrpohl, D., You, Z., Lo, E. H., Kim, H.-H., and Whalen, M. J. (2007). TNF alpha and Fas mediate tissue damage and functional outcome after traumatic brain injury in mice. *J. Cereb. Blood Flow Metab.* 27, 1806–1818. doi: 10.1038/sj.jcbfm.9600487
- Blauvelt, A., and Chiricozzi, A. (2018). The immunologic role of IL-17 in psoriasis and psoriatic arthritis pathogenesis. *Clin. Rev. Allergy Immunol.* 55, 379–390. doi: 10.1007/s12016-018-8702-3
- Bø, K., Thoresen, M., and Dalgard, F. (2008). Smokers report more psoriasis, but not atopic dermatitis or hand eczema: results from a Norwegian population survey among adults. *Dermatology* 216, 40–45. doi: 10.1159/000109357
- Braegelmann, J., D'Erme, A. M., Akmal, S., Maier, J., Braegelmann, C., and Wenzel, J. (2016). Interleukin-36 $\gamma$  (IL-1F9) identifies psoriasis among patients with erythroderma. *Acta Dermato Venereol.* 96, 386–387. doi: 10.2340/00015555-2265
- Bremilla, N. C., Senra, L., and Boehncke, W.-H. (2018). The IL-17 family of cytokines in psoriasis: IL-17A and beyond. *Front. Immunol.* 9:1682. doi: 10.3389/fimmu.2018.01682
- Büchle, B., Zugmaier, W., and Simmet, T. (2003). Analysis of pentacyclic triterpenic acids from frankincense gum resins and related phytopharmaceuticals by high-performance liquid chromatography. Identification of lupeolic acid, a novel pentacyclic triterpene. *J. Chromatogr. B* 791, 21–30. doi: 10.1016/S1570-0232(03)00160-0
- Chan, T. C., Hawkes, J. E., and Krueger, J. G. (2018). Interleukin 23 in the skin: role in psoriasis pathogenesis and selective interleukin 23 blockade as treatment. *Ther. Adv. Chronic Dis.* 9, 111–119. doi: 10.1177/2040622318759282

- Chen, G.-Y., Cheng, Y.-W., Wang, C.-Y., Hsu, T.-J., Hsu, M. M.-L., Yang, P.-T., et al. (2008). Prevalence of skin diseases among schoolchildren in Magong, Penghu, Taiwan: a community-based clinical survey. *J. Formosan Med. Assoc.* 107, 21–29. doi: 10.1016/S0929-6646(08)60004-2
- Choi, J. W., Kim, B. R., and Youn, S. W. (2017). Adherence to topical therapies for the treatment of psoriasis: surveys of physicians and patients. *Ann. Dermatol.* 29, 559–564. doi: 10.5021/ad.2017.29.5.559
- Ciechanowicz, P., Rakowska, A., Sikora, M., and Rudnicka, L. (2019). JAK-inhibitors in dermatology: current evidence and future applications. *J. Dermatol. Treat.* 30, 648–658. doi: 10.1080/09546634.2018.1546043
- Corsano, S., and Iavarone, C. (1964). Isolamento dall'incenso dell'acido 3-acetil-11-ossi- $\beta$ -boswellico. *Gazz. Chim. Ital.* 94, 328–339.
- Corsano, S., and Nicoletti, R. (1967). The structure of incensole. *Tetrahedron* 23, 1977–1984. doi: 10.1016/S0040-4020(01)82601-6
- D'Erme, A. M., Wilschmann-Theis, D., Wagenpfeil, J., Hölzel, M., Ferring-Schmitt, S., Sternberg, S., et al. (2015). IL-36 $\gamma$  (IL-1F9) is a biomarker for psoriasis skin lesions. *J. Invest. Dermatol.* 135, 1025–1032. doi: 10.1038/jid.2014.532
- Eder, L., Widdifield, J., Rosen, C. F., Cook, R., Lee, K. A., Alhusayen, R., et al. (2019). Trends in the prevalence and incidence of psoriasis and psoriatic arthritis in Ontario, Canada: a population-based study. *Arthritis Care Res.* 71, 1084–1091. doi: 10.1002/acr.23743
- Egeberg, A., Skov, L., Gislasen, G. H., Thyssen, J. P., and Mallbris, L. (2017). Incidence and prevalence of psoriasis in Denmark. *Acta Derm. Venereol.* 97, 808–812. doi: 10.2340/00015555-2672
- Florence, N. T., Benoit, M. Z., Jonas, K., Alexandra, T., Désiré, D. D. P., Pierre, K., et al. (2014). Antidiabetic and antioxidant effects of *Annona muricata* (Annonaceae), aqueous extract on streptozotocin-induced diabetic rats. *J. Ethnopharmacol.* 151, 784–790. doi: 10.1016/j.jep.2013.09.021
- Fotiadou, C., Lazaridou, E., Sotiriou, E., and Ioannides, D. (2018). Targeting IL-23 in psoriasis: current perspectives. *Psoriasis Targets Ther.* 8:1. doi: 10.2147/PTT.S98893
- Fujita, H. (2013). The role of IL-22 and Th22 cells in human skin diseases. *J. Dermatol. Sci.* 72, 3–8. doi: 10.1016/j.jdermsci.2013.04.028
- Gisoni, P., Del Giglio, M., and Girolomoni, G. (2017). Treatment approaches to moderate to severe psoriasis. *Int. J. Mol. Sci.* 18:2427. doi: 10.3390/ijms18112427
- Golbari, N. M., Porter, M. L., and Kimball, A. B. (2018). Current guidelines for psoriasis treatment: a work in progress. *Cutis* 101, 10–12.
- Goldminz, A., Au, S., Kim, N., Gottlieb, A., and Lizzul, P. (2013). NF- $\kappa$ B: an essential transcription factor in psoriasis. *J. Dermatol. Sci.* 69, 89–94. doi: 10.1016/j.jdermsci.2012.11.002
- Halim, S. A., Abdalla, O. M., Mesaik, M. A., Wadood, A., and Kontoyianni, M. (2013). Identification of novel Interleukin-2 inhibitors through computational approaches. *Mol. Divers.* 17, 345–355. doi: 10.1007/s11030-013-9431-4
- Hamm, S., Bleton, J., Connan, J., and Tchaplal, A. (2005). A chemical investigation by headspace SPME and GC-MS of volatile and semi-volatile terpenes in various oil samples. *Phytochemistry* 66, 1499–1514. doi: 10.1016/j.phytochem.2005.04.025
- Harden, J. L., Krueger, J. G., and Bowcock, A. M. (2015). The immunogenetics of psoriasis: a comprehensive review. *J. Autoimmun.* 64, 66–73. doi: 10.1016/j.jaut.2015.07.008
- Hoffman, M. B., Hill, D., and Feldman, S. R. (2016). Current challenges and emerging drug delivery strategies for the treatment of psoriasis. *Expert Opin. Drug Deliv.* 13, 1461–1473. doi: 10.1080/17425247.2016.1188801
- Hofny, E. R., Morsy, H., Hasaball, A., and Twisy, H. (2017). Serum level of interleukin-22 in patients with psoriasis and its correlation with disease severity. *J. Curr. Med. Res. Pract.* 2:133. doi: 10.4103/JCMRP.JCMRP\_27\_16
- Hugh, J. M., and Weinberg, J. M. (2018). Update on the pathophysiology of psoriasis. *Cutis* 102, 6–12.
- Hussain, H., Al-Harrasi, A., Csuk, R., Shamraiz, U., Green, I. R., Ahmed, I., et al. (2017). Therapeutic potential of boswellic acids: a patent review (1990–2015). *Expert Opin. Ther. Pat.* 27, 81–90. doi: 10.1080/13543776.2017.1235156
- Icen, M., Crowson, C. S., McEvoy, M. T., Dann, F. J., Gabriel, S. E., and Kremers, H. M. (2009). Trends in incidence of adult-onset psoriasis over three decades: a population-based study. *J. Am. Acad. Dermatol.* 60, 394–401. doi: 10.1016/j.jaad.2008.10.062
- Johansen, C., Funding, A. T., Otkjaer, K., Kragballe, K., Jensen, U. B., Madsen, M., et al. (2006). Protein expression of TNF- $\alpha$  in psoriatic skin is regulated at a posttranscriptional level by MAPK-activated protein kinase 2. *J. Immunol.* 176, 1431–1438. doi: 10.4049/jimmunol.176.3.1431
- Kumar, A., Shah, B. A., Singh, S., Hamid, A., Singh, S. K., Sethi, V. K., et al. (2012). Acyl derivatives of boswellic acids as inhibitors of NF- $\kappa$ B and STATs. *Bioorg. Med. Chem. Lett.* 22, 431–435. doi: 10.1016/j.bmcl.2011.10.112
- Kunnumakkara, A. B., Nair, A. S., Sung, B., Pandey, M. K., and Aggarwal, B. B. (2009). Boswellic acid blocks STAT3 Signaling, proliferation, and survival of multiple myeloma via the protein tyrosine phosphatase SHP-1. *Mol. Cancer Res.* 7:118. doi: 10.1158/1541-7786.MCR-08-0154
- Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., and Thornton, J. M. (1997). PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* 22, 488–490. doi: 10.1016/S0968-0004(97)01140-7
- Lee, A., Lee, K., and Kim, D. (2016). Using reverse docking for target identification and its applications for drug discovery. *Expert Opin. Drug Discov.* 11, 707–715. doi: 10.1080/17460441.2016.1190706
- Li, F., Xu, K., Yuan, S., Yan, D., Liu, R., Tan, J., et al. (2010). Macrocyclic diterpenes from *Boswellia carterii* Birdwood (frankincense). *Chin. J. Org. Chem.* 30, 107–111.
- Liu, J.-J., Toy, W. C., Liu, S., Cheng, A., Lim, B. K., Subramaniam, T., et al. (2013). Acetyl-keto- $\beta$ -boswellic acid induces lipolysis in mature adipocytes. *Biochem. Biophys. Res. Commun.* 431, 192–196. doi: 10.1016/j.bbrc.2012.12.136
- Lowes, M. A., Suárez-Fariñas, M., and Krueger, J. G. (2014). Immunology of psoriasis. *Annu. Rev. Immunol.* 32, 227–255. doi: 10.1146/annurev-immunol-032713-120225
- MacDonald, A., and Burden, A. (2007). Psoriasis: advances in pathophysiology and management. *Postgrad. Med. J.* 83, 690–697. doi: 10.1136/pgmj.2007.061473
- Mahajan, B., Taneja, S., Sethi, V., and Dhar, K. (1995). Two triperpenoids from *Boswellia serrata* gum resin. *Phytochemistry* 39, 453–455. doi: 10.1016/0031-9422(95)99386-3
- Mikhael, B. R., Maatooq, G. T., Badria, F. A., and Amer, M. M. (2003). Chemistry and immunomodulatory activity of frankincense oil. *Z. Naturforschung C* 58, 230–238. doi: 10.1515/znc-2003-3-416
- Mittal, R., Malhotra, S., Pandhi, P., Kaur, I., and Dogra, S. (2009). Efficacy and safety of combination Acitretin and Pioglitazone therapy in patients with moderate to severe chronic plaque-type psoriasis: a randomized, double-blind, placebo-controlled clinical trial. *Arch. Dermatol.* 145, 387–393. doi: 10.1001/archdermatol.2009.5
- Moorchong, N., Kulaar, J. S., Chatterjee, M., Vasudevan, B., Tripathi, T., and Dutta, V. (2014). Role of NF- $\kappa$ B in the pathogenesis of psoriasis elucidated by its staining in skin biopsy specimens. *Int. J. Dermatol.* 53, 570–574. doi: 10.1111/ijd.12050
- Moussaeff, A., Shein, N. A. A., Tsenter, J., Grigoriadis, S., Simeonidou, C., Alexandrovich, A. G., et al. (2008). Incensole acetate: a novel neuroprotective agent isolated from *Boswellia carterii*. *J. Cereb. Blood Flow Metab.* 28, 1341–1352. doi: 10.1038/jcbfm.2008.28
- Moussaeff, A., Shohami, E., Kashman, Y., Frider, E., Schmitz, M. L., Renner, F., et al. (2007). Incensole acetate, a novel anti-inflammatory compound isolated from *Boswellia* resin, inhibits nuclear factor- $\kappa$ B activation. *Mol. Pharmacol.* 72, 1657–1664. doi: 10.1124/mol.107.038810
- Ni, C., and Chiu, M. W. (2014). Psoriasis and comorbidities: links and risks. *Clin. Cosmet. Investig. Dermatol.* 7:119. doi: 10.2147/CCID.S44843
- Nicoletti, R., and Forcellese, M. (1968). The structure of incensole-oxide. *Tetrahedron* 24, 6519–6525. doi: 10.1016/S0040-4020(01)96864-4
- Obermann, H. (1977). Differences in the chemistry and odor of incense resins. *Dragoco Rep.* 11, 260–265.
- Ota, M., and Houghton, P. J. (2008). Boswellic acids with acetylcholinesterase inhibitory properties from frankincense. *Nat. Prod. Commun.* 3, 21–26. doi: 10.1177/1934578X0800300105
- Pardasani, A. G., Feldman, S., and Clark, A. R. (2000). Treatment of psoriasis: an algorithm-based approach for primary care physicians. *Am. Fam. Physician* 61, 725–733.
- Pardhy, R. S. (1978). b-Boswellic acid, acetyl-b-boswellic acid, acetyl-11-keto-b-boswellic acid and 11-keto-b-boswellic acid, four pentacyclic triterpene acids from the resin of *Boswellia serrata* roxb. *Ind. J. Chem. Sect. B* 16, 176–178.
- Parisi, R., Symmons, D. P., Griffiths, C. E., and Ashcroft, D. M. (2013). Global epidemiology of psoriasis: a systematic review of incidence and prevalence. *J. Invest. Dermatol.* 133, 377–385. doi: 10.1038/jid.2012.339

- Patil, V., Nayak, U., and Dev, S. (1973). Chemistry of ayurvedic crude drugs—II: Guggulu (resin from *Commiphora mukul*)-2: Diterpenoid constituents. *Tetrahedron* 29, 341–348. doi: 10.1016/S0040-4020(01)93299-5
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084
- Pollastro, F., Golin, S., Chianese, G., Putra, M. Y., Schiano Moriello, A., De Petrocellis, L., et al. (2016). Neuroactive and anti-inflammatory frankincense cembranes: a structure–activity study. *J. Nat. Prod.* 79, 1762–1768. doi: 10.1021/acs.jnatprod.6b00141
- Rácz, E., and Prens, E. P. (2009). Molecular pathophysiology of psoriasis and molecular targets of antipsoriatic therapy. *Expert Rev. Mol. Med.* 11:e38. doi: 10.1017/S146239940900129X
- Rehman, N. U., Hussain, H., Al-Shidhani, S., Avula, S. K., Abbas, G., Anwar, M. U., et al. (2017). Incensuran: isolation, X-ray crystal structure and absolute configuration by means of chiroptical studies in solution and solid state. *RSC Adv.* 7, 42357–42362. doi: 10.1039/C7RA07351H
- Ren, J., Wang, Y.-G., Wang, A.-G., Wu, L.-Q., Zhang, H.-J., Wang, W.-J., et al. (2015). Cembranoids from the gum resin of *Boswellia carterii* as potential antiulcerative colitis agents. *J. Nat. Prod.* 78, 2322–2331. doi: 10.1021/acs.jnatprod.5b00104
- Sano, S., Chan, K. S., Carbajal, S., Clifford, J., Peavey, M., Kiguchi, K., et al. (2005). Stat3 links activated keratinocytes and immunocytes required for development of psoriasis in a novel transgenic mouse model. *Nat. Med.* 11, 43–49. doi: 10.1038/nm1162
- Schadler, E. D., Ortel, B., and Mehlis, S. L. (2019). Biologics for the primary care physician: review and treatment of psoriasis. *Dis. Month* 65, 51–90. doi: 10.1016/j.disamonth.2018.06.001
- Schmidt, T. J., Kaiser, M., and Brun, R. (2011). Complete structural assignment of serranol, a cembrane-type diterpene from *Boswellia serrata*, and evaluation of its antiprotozoal activity. *Planta Med.* 77, 849–850. doi: 10.1055/s-0030-1250612
- Sertznig, P., and Reichrath, J. (2011). Peroxisome proliferator-activated receptors (PPARs) in dermatology: challenge and promise. *Dermato Endocrinol.* 3, 130–135. doi: 10.4161/derm.15025
- Shmidt, E., Kashtanova, N., and Pentegova, V. (1970). Neocembrene—a new diterpene hydrocarbon from *Picea obovata* and *Pinus koraiensis*. *Chem. Nat. Compd.* 6, 705–707. doi: 10.1007/BF00565337
- Stürner, K. H., Verse, N., Yousef, S., Martin, R., and Sospedra, M. (2014). Boswellic acids reduce T h17 differentiation via blockade of IL-1 $\beta$ -mediated IRAK 1 signaling. *Eur. J. Immunol.* 44, 1200–1212. doi: 10.1002/eji.201343629
- Syrovets, T., Büchele, B., Krauss, C., Laumonnier, Y., and Simmet, T. (2005). Acetyl-boswellic acids inhibit lipopolysaccharide-mediated TNF- $\alpha$  induction in monocytes by direct interaction with I $\kappa$ B kinases. *J. Immunol.* 174, 498–506. doi: 10.4049/jimmunol.174.1.498
- Takeshita, J., Grewal, S., Langan, S. M., Mehta, N. N., Ogdie, A., Van Voorhees, A. S., et al. (2017). Psoriasis and comorbid diseases: epidemiology. *J. Am. Acad. Dermatol.* 76, 377–390. doi: 10.1016/j.jaad.2016.07.064
- Thomsen, R., and Christensen, M. H. (2006). MolDock: a new technique for high-accuracy molecular docking. *J. Med. Chem.* 49, 3315–3321. doi: 10.1021/jm051197e
- Tollefson, M. M., Van Houten, H. K., Asante, D., Yao, X., and Kremers, H. M. (2018). Association of psoriasis with comorbidity development in children with psoriasis. *JAMA Dermatol.* 154, 286–292. doi: 10.1001/jamadermatol.2017.5417
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334
- Tsai, Y.-C., and Tsai, T.-F. (2017). Anti-interleukin and interleukin therapies for psoriasis: current evidence and clinical usefulness. *Ther. Adv. Musculoskelet. Dis.* 9, 277–294. doi: 10.1177/1759720X17735756
- Voglis, S., Moos, S., Kloos, L., Wanke, F., Zayoud, M., Pelczar, P., et al. (2018). Regulation of IL-22BP in psoriasis. *Sci. Rep.* 8, 1–8. doi: 10.1038/s41598-018-23510-3
- Wahab, S. A., Aboutabl, E., El-Zalabani, S., Fouad, H., De Pooter, H., and El-Fallaha, B. (1987). The essential oil of olibanum. *Planta Med.* 53, 382–384. doi: 10.1055/s-2006-962745
- Wang, M.-X., Zhao, J.-X., Meng, Y.-J., Di, T., Xu, X., Xie, X.-J., et al. (2018). Acetyl-11-keto- $\beta$ -boswellic acid inhibits the secretion of cytokines by dendritic cells via the TLR7/8 pathway in an imiquimod-induced psoriasis mouse model and *in vitro*. *Life Sci.* 207, 90–104. doi: 10.1016/j.lfs.2018.05.044
- Wang, Y., Weng, H., Song, J. F., Deng, Y. H., Li, S., and Liu, H. B. (2017). Activation of the HMGB1-TLR4-NF- $\kappa$ B pathway may occur in patients with atopic eczema. *Mol. Med. Rep.* 16, 2714–2720. doi: 10.3892/mmr.2017.6942
- Winterfield, L., Menter, A., Gordon, K., and Gottlieb, A. (2005). Psoriasis treatment: current and emerging directed therapies. *Ann Rheum Dis.* 64, ii87–ii90. doi: 10.1136/ard.2004.032276
- Woo, Y. R., Cho, D. H., and Park, H. J. (2017). Molecular mechanisms and management of a cutaneous inflammatory disorder: psoriasis. *Int. J. Mol. Sci.* 18, 2684. doi: 10.3390/ijms18122684
- Xu, X., Huang, M., and Zou, X. (2018). Docking-based inverse virtual screening: methods, applications, and challenges. *Biophys. Reports.* 4, 1–16. doi: 10.1007/s41048-017-0045-8
- Yadav, M., Sardana, I., Sharma, A., Sharma, N., Nagpal, K., and Malik, P. (2019). Emerging pathophysiological targets of psoriasis for future therapeutic strategies. *Infect. Disord Drug Targets.* doi: 10.2174/1871526519666190617162701. [Epub ahead of print].
- Yamago, S., Furukawa, M., Azuma, A., and Yoshida, J.-I. (1998). Synthesis of optically active dendritic binaphthols and their metal complexes for asymmetric catalysis. *Tetrahedron Lett.* 39, 3783–3786. doi: 10.1016/S0040-4039(98)00616-9

**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.

Copyright © 2020 Halim, Khan, Csuk, Al-Rawahi and Al-Harrasi. 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 Approaches in Preclinical Studies on Drug Discovery and Development

Fengxu Wu<sup>1,2†</sup>, Yuquan Zhou<sup>1,3†</sup>, Langhui Li<sup>1,4†</sup>, Xianhuan Shen<sup>1,4</sup>, Ganying Chen<sup>1,3</sup>, Xiaoqing Wang<sup>1,4</sup>, Xianyang Liang<sup>1,3</sup>, Mengyuan Tan<sup>1,4</sup> and Zunnan Huang<sup>1,4,5\*</sup>

<sup>1</sup> Key Laboratory of Big Data Mining and Precision Drug Design of Guangdong Medical University, Research Platform Service Management Center, Dongguan, China, <sup>2</sup> Key Laboratory of Pesticide & Chemical Biology, Ministry of Education, College of Chemistry, Central China Normal University, Wuhan, China, <sup>3</sup> The Second School of Clinical Medicine, Guangdong Medical University, Dongguan, China, <sup>4</sup> Key Laboratory for Research and Development of Natural Drugs of Guangdong Province, School of Pharmacy, Guangdong Medical University, Dongguan, China, <sup>5</sup> Marine Biomedical Research Institute of Guangdong Zhanjiang, Zhanjiang, China

## OPEN ACCESS

### Edited by:

Simone Brogi,  
University of Pisa, Italy

### Reviewed by:

Hongbin Yang,  
University of Cambridge,  
United Kingdom  
Bruno Villoutreix,  
Institut National de la Santé et de la  
Recherche Médicale  
(INSERM), France

### \*Correspondence:

Zunnan Huang  
zn\_huang@yahoo.com

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

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 29 March 2020

Accepted: 14 July 2020

Published: 11 September 2020

### Citation:

Wu F, Zhou Y, Li L, Shen X, Chen G,  
Wang X, Liang X, Tan M and Huang Z  
(2020) Computational Approaches in  
Preclinical Studies on Drug Discovery  
and Development.  
Front. Chem. 8:726.  
doi: 10.3389/fchem.2020.00726

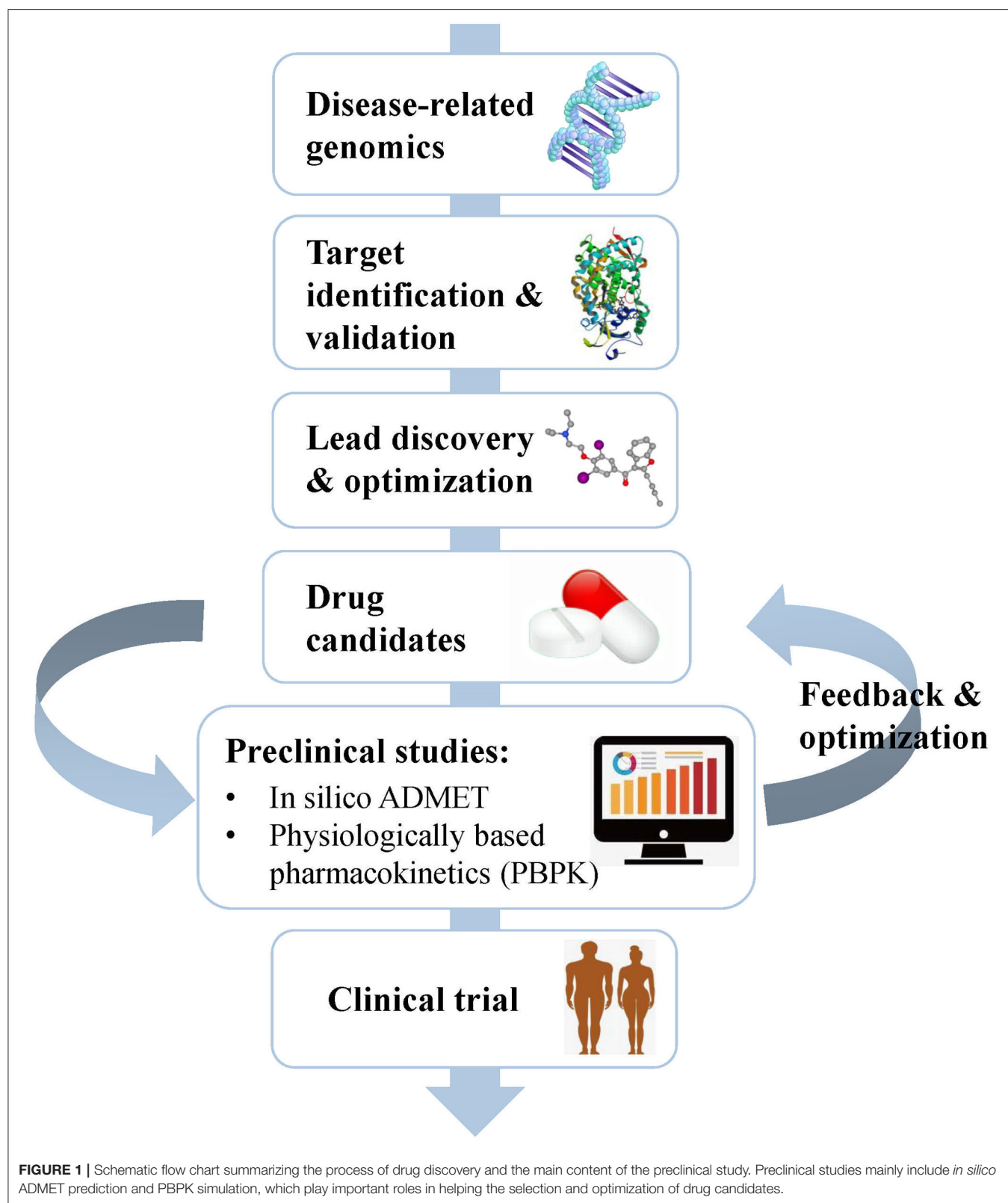
Because undesirable pharmacokinetics and toxicity are significant reasons for the failure of drug development in the costly late stage, it has been widely recognized that drug ADMET properties should be considered as early as possible to reduce failure rates in the clinical phase of drug discovery. Concurrently, drug recalls have become increasingly common in recent years, prompting pharmaceutical companies to increase attention toward the safety evaluation of preclinical drugs. *In vitro* and *in vivo* drug evaluation techniques are currently more mature in preclinical applications, but these technologies are costly. In recent years, with the rapid development of computer science, *in silico* technology has been widely used to evaluate the relevant properties of drugs in the preclinical stage and has produced many software programs and *in silico* models, further promoting the study of ADMET *in vitro*. In this review, we first introduce the two ADMET prediction categories (molecular modeling and data modeling). Then, we perform a systematic classification and description of the databases and software commonly used for ADMET prediction. We focus on some widely studied ADMT properties as well as PBPK simulation, and we list some applications that are related to the prediction categories and web tools. Finally, we discuss challenges and limitations in the preclinical area and propose some suggestions and prospects for the future.

**Keywords:** drug discovery, pre-clinical studies, ADMET, pharmacokinetics, PBPK modeling

## INTRODUCTION

Drug development is a complicated, risky, and time-consuming process that can be divided into several stages, including disease-related genomics, target identification and validation, lead discovery and optimization, preclinical studies, and clinical trials (Tang et al., 2006) (**Figure 1**). During early drug discovery, the activities and specificities of candidate drugs are usually assessed at an early stage, and pharmacokinetics and toxicities are evaluated at a relatively late stage (Selick et al., 2002). However, the undesirable efficacy and safety, mainly caused by absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics, resulted in the failure of many candidate drugs in the final stage (Caldwell et al., 2009). Cook et al. (2014) comprehensively





reviewed the results of AstraZeneca's small-molecule drug projects from 2005 to 2010 based on a longitudinal study. They found that unacceptable safety and toxicity were the most

important reasons for the failure of more than half of all project closures. As with the development of drug discovery, it was realized that it is important to filter and optimize the ADMET

properties for drugs at an early stage, which has been accepted and widely used to reduce the attrition rate in drug research and development. A “fail early, fail cheap” strategy has been employed by many pharmaceutical companies (Yu and Adedoyin, 2003). Pharmacokinetics and toxicity assessments of preclinical drugs are of great value in reducing the failure rate of new chemical entities (NCEs) in clinical trials (Kola and Landis, 2004; Yang et al., 2018b; Ferreira and Andricopulo, 2019). In recent years, *in vitro* and *in vivo* ADMET prediction methods have been widely used, but it is impractical to perform complex and expensive ADMET experiments on a large number of compounds (Cheng et al., 2013; Patel C. N. et al., 2020). Thus, an *in silico* strategy to predict ADMET properties has become very attractive as a cost-saving and high-throughput alternative to experimental measurement methods.

With the rapid development of computer technologies, the high-throughput screening of compounds, application of combinatorial chemistry, and ability of compound synthesis have increased dramatically. The early demands for ADMET data on lead compounds have also significantly increased, and methods for evaluating ADMET *in vitro* are gradually increasing. Many *in silico* methods have been successfully applied to the *in vitro* prediction of ADMET, and *in silico* models have also been developed to replace *in vivo* models for the prediction of pharmacokinetics, toxicity, and other parameters (Zhu et al., 2011; Wang et al., 2015; Alqahtani, 2017). *In silico* ADMET prediction has progressed with the continuous development of cheminformatics and has entered the era of big data (Ferreira and Andricopulo, 2019). Two *in silico* approach categories can be used for ADMET prediction: molecular modeling and data modeling. Molecular modeling is based on the three-dimensional structures of proteins. It includes multiple methods such as molecular docking, molecular dynamics (MD) simulation, and quantum mechanics (QM) calculation (Bowen and Guener, 2013; Cheng et al., 2013; Silva-Junior et al., 2017). Data modeling includes quantitative structure–activity relationship (QSAR) (Cumming et al., 2013) and physiologically-based pharmacokinetic (PBPK) modeling (Fan and de Lannoy, 2014). Due to the increase in number of properties that need to be predicted, a series of ADMET software programs capable of comprehensive property prediction have been developed. The development from *in silico* approaches to ADMET software has undergone a long process of predicting property parameters from less to more at early to late timepoints (Figure 2). This review first provides a detailed introduction to the two *in silico* approaches of ADMET prediction. Then, we summarize the widely used databases and software related to ADMET prediction. Finally, we analyze the problems and challenges faced by computer model prediction methods as well as the tools, and we propose some of our own prospects for future development in this area.

## IN SILICO APPROACHES

### Molecular Modeling

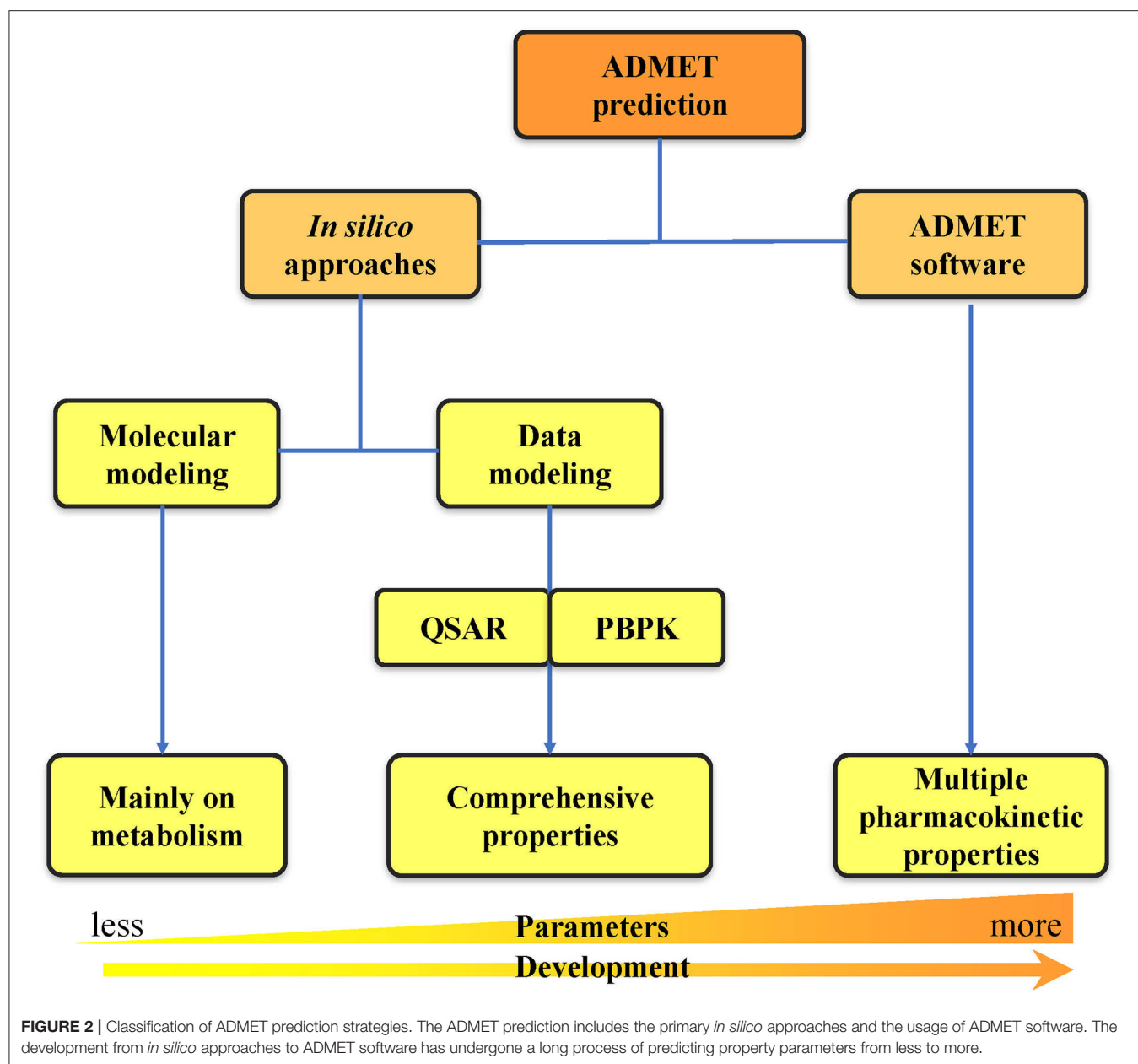
Molecular modeling, based on the three-dimensional structures of proteins, is an important category in predicting ADMET

properties and includes methods such as pharmacophore modeling, molecular docking, MD simulations, and QM calculations (Figure 3). As more and more three-dimensional structures of ADMET proteins become available, molecular modeling can complement or even surpass QSAR studies (Moroy et al., 2012). Applying molecular modeling to perform ADMET prediction is a challenge because the ADMET proteins usually have flexible and large binding cavities. Many promising results of molecular modeling in predicting compound metabolism have been reported. The methods in these cases can be generally divided into ligand-based and structure-based and help not only to analyze metabolic properties but also to further optimize compound toxicity, bioavailability, and other parameters (Lin et al., 2003).

### Ligand-Based Methods

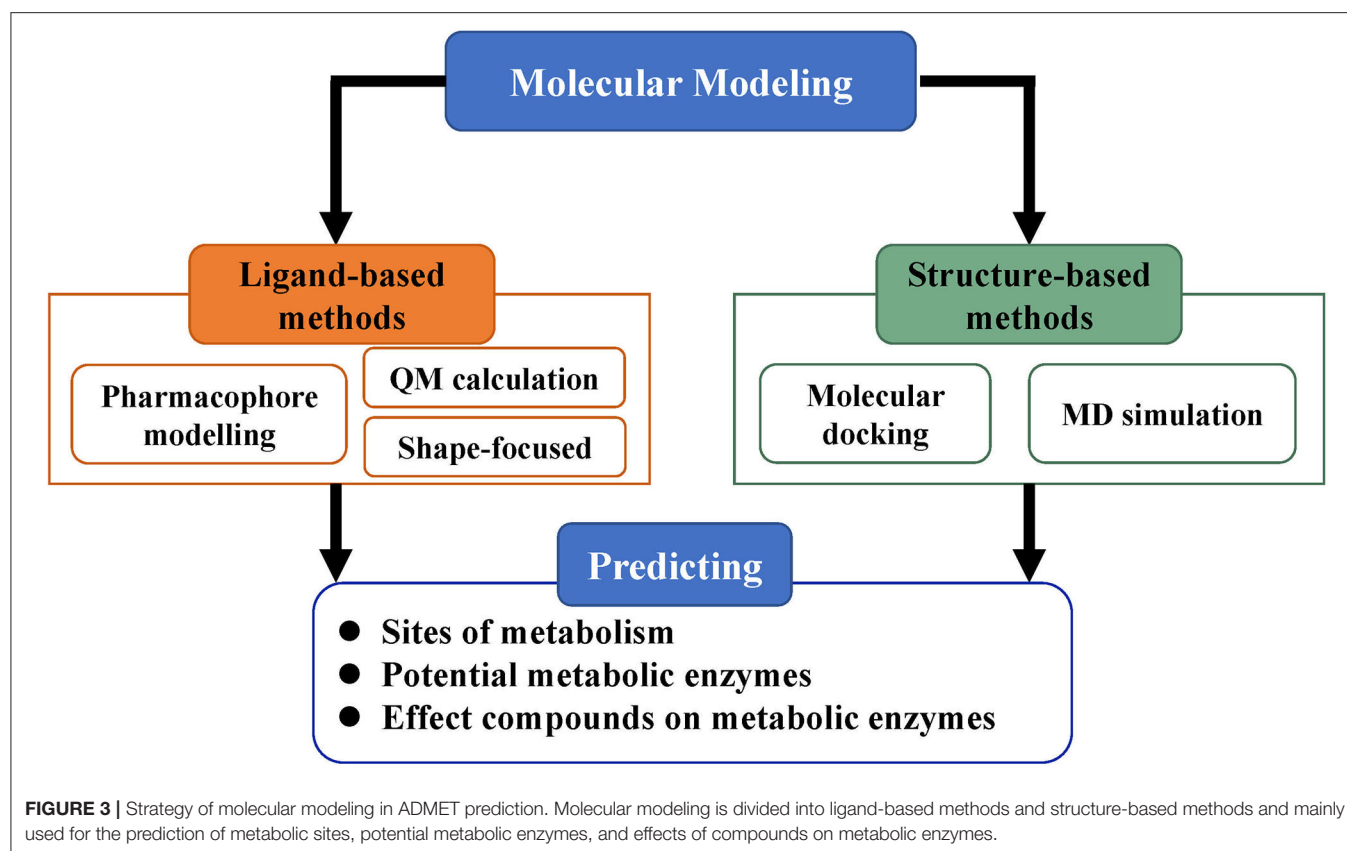
Ligand-based methods derive information on proteins' active sites based on the shapes, electronic properties, and conformations of inhibitors, substrates or metabolites; this information depends on the assumption that the metabolic properties of compounds are entirely the result of their chemical structures and characteristics (de Groot et al., 2004; Andrade et al., 2014). In this category, pharmacophore modeling is one of the most widely used methods. The interactions between ligands and receptors can be predicted by constructing a pharmacophore model to cover the structures or properties of ligands in three-dimensional space and then to simulate the spatial and chemical properties of binding sites (de Groot, 2006). Therefore, the availability of ligand data is essential to the construction of pharmacophore models. In recent years, there have been many cases of using pharmacophore models to screen promising compounds with outstanding ADMET properties (Nandekar et al., 2016; El-Zahabi et al., 2019; Mohan et al., 2020; Patel D. B. et al., 2020; Rawat and Verma, 2020). For example, Nandekar et al. (2016) generated and validated a pharmacophore model to screen anticancer compounds acting via cytochrome P450 1A1 (CYP1A1). Nine compounds that have preferred pharmacophore characteristics and are capable of generating reactive metabolites were finally selected for further study. Rawat and Verma (2020) developed a pharmacophore model to discover new dual target inhibitors of *Plasmodium falciparum* dihydroorotate dehydrogenase (PfDHODH) and cytochrome bc1 complex (PfCytbc1) to treat malaria. The molecule MMV007571, which has been validated as an efficient multi-target inhibitor, was used to extract features from the binding information for the model construction. The model was used to screen a library including more than 40,000 molecules. After a series of experiments, two compounds were developed with the desired properties in binding potential and pharmacokinetic characters.

The shape-focused method is another category of ligand-based methods. This method is based on the fact that the shapes of a ligand and the binding site of its receptor should be complementary. Thus, molecules that have a comparable shape should be able to bind to the same receptor. It is more possible for a ligand to bind with the same target if it has greater similarity with the reference molecule (Putta and Beroza, 2007). This method requires only one reference



molecule (shape) to perform a screening. However, more models should be constructed to cover more diverse chemical space if different shapes are available (Perez-Nueno and Ritchie, 2011), particularly for highly flexible proteins. Some studies using the shape-focused method have been reported in recent years (Reddy et al., 2013; Chen et al., 2015; Kumar et al., 2015; Prabhu and Singh, 2019). For example, Chen et al. (2015) presented a shape-based virtual screening to find new cores for the design of acetylcholinesterase (AChE) inhibitors. The shape of commercial inhibitor tacrine was used to search for new potential inhibitors. Two hit compounds were finally identified with good ADMET properties and better activity than tacrine.

With the improvement of computer hardware performance, the time-consuming QM calculation in ADMET prediction has become possible and popular. The QM calculation can be used to evaluate the bond break, which is a step required for metabolic transformation (Andrade et al., 2014). Moreover, this calculation uses an accurate means of describing electrons in atoms and molecules (Modi, 2003). Hence, QM calculation is very helpful in ADMET prediction. Extensively increasing studies involving the application of QM methods have been conducted to describe ADMET properties of new compounds (Li et al., 2012; Taxak et al., 2013; Kavitha et al., 2015; Sasahara et al., 2015; Evangelista et al., 2016; Mondal et al., 2017). The *ab initio* (Hartree-Fock), semiempirical (AM1 and PM3), and density functional theory



(DFT) approaches are most commonly used in these studies (Silva-Junior et al., 2017). For example, Mondal et al. (2017) used the DFT method to study the absorption profile and antimicrobial activity for five sulfonamide Schiff bases. The geometries compared well with the experimental value. Sasahara et al. (2015) used the DFT method to evaluate the metabolic selectivity of antipsychotic thioridazine by CYP450 2D6. This study revealed the importance of the substrate orientation in the reaction center of this enzyme for the metabolic reaction.

All ligand-based methods need to address the problem of the uncertainty of metabolic enzyme binding sites. If reliable structural data for a metabolic enzyme are lacking, the properties of ligand binding to the enzyme can only be speculative, and minor modifications to a ligand may cause a significant decrease in ligand-protein affinity (Kirchmair et al., 2012). Therefore, it is difficult to predict ADMET using ligand-based methods without reliable protein structure data. However, the use of ligand-based methods in metabolism prediction can easily eliminate inappropriate compounds and reduce the number of compounds that fail during the synthetic evaluation cycle and more expensive late stages.

### Structure-Based Methods

Structure-based methods can be used not only to predict the ADMET properties of compounds, but also to study specific interactions between small molecules and ADMET proteins. In general, they focus on obtaining binding modes from the

static structures of protein-ligand complexes, regardless of time-dependent conformational fluctuations (Cheng T. et al., 2012). For structure-based methods, changes in receptor conformation must be considered mainly because of the possible interaction between the existing large and flexible binding cavities with diverse ligands. By performing MD calculation to simulate the dynamic changes of spatial shape, we may obtain an adequate conformation sampling to search a stable and reliable binding mode for ADMET prediction. In recent years, structure-based methods have been widely used, for instance, to predict the binding patterns of substrates (Macalino et al., 2015), conformational changes in enzymes (Cheng T. et al., 2012), and their catalytic effects on physiological systems (Cui and Karplus, 2003), to evaluate substrate affinity, instability and metabolic pathways (Sun and Scott, 2010), and to assess the relationship between metabolism and carcinogenicity (Fratev and Benfenati, 2008).

The molecular modeling strategy makes important contributions to the rationalization of metabolic reactions of compounds, allowing the simulation of binding modes between drugs and macromolecules in the ADMET process at atomic or molecular levels. With the rapid development of structural elucidation and pharmacokinetic calculation techniques, structure-based methods are becoming increasingly predictive and accurate. However, the molecular modeling strategy is still limited by drawbacks such as a requirement to accurately analyze the structural flexibility of proteins (Kazmi et al., 2019).



Additionally, high-resolution experimental structural data of the target will be more conducive to our accurate prediction of the drug metabolic fate (Doss et al., 2014). Surely, combining improved structure-based and ligand-based methods can create synergistic effects in metabolic prediction, enabling more comprehensive descriptions of metabolic reactions (Issa et al., 2017; Kar and Leszczynski, 2017).

## Data Modeling

There are two widely used data modeling methods to predict ADMET-related properties, QSAR, and PBPK modeling (Cheng et al., 2013). ADMET analysis and prediction in QSAR mainly depends on many molecular descriptors, including topological, geometrical physicochemical, or electronic descriptors. Many properties, such as blood-brain barrier (BBB), clinical adverse effects, percent protein binding (%PPB), lipophilicity (logP), preclinical toxicological endpoints, and metabolism of pharmaceutical substances, can be predicted using the QSAR method. The PBPK modeling always predicts parameters concerning the dose size and dose frequency, such as the volume of distribution at steady-state ( $V_{ss}$ ), total drug clearance (CL), and fraction of dose that reaches the portal vein (Fabs), because most drugs are taken orally.

## QSAR

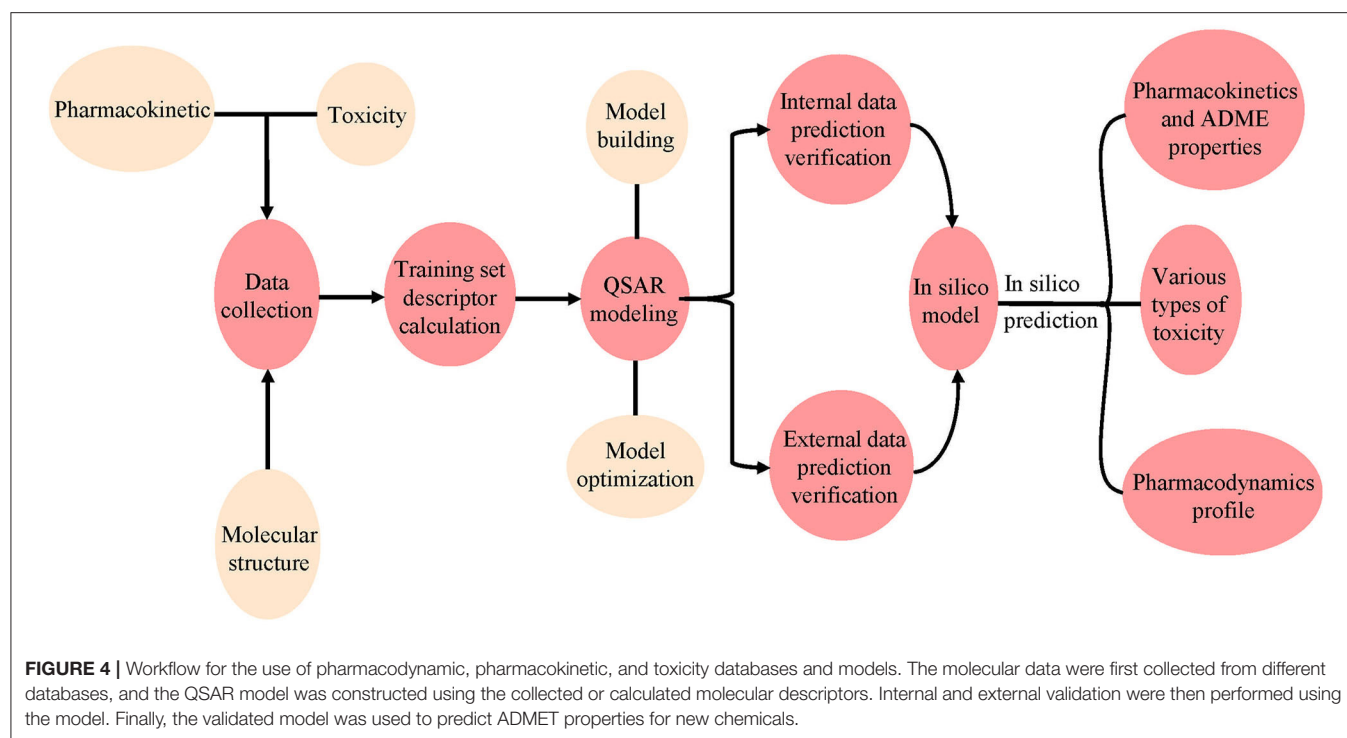
QSAR, which employs mathematical models to describe relationships between molecular structures and their biological activities, has been used in pharmaceutical chemistry since the 1960s (Hansch, 1981). The classical QSAR developed by Hansch was used to predict ADMET (Hansch, 1981). It is mainly based on the hypothesis that similar molecules exhibit similar properties (Patterson et al., 1996). Thus, two prediction methods are primarily considered: (1) prediction based on molecular similarity (pharmacophore- and molecular fragment-based methods) and (2) prediction based on property similarity (log P, log D, and others) (Yongye and Medina-Franco, 2013). In these two methods, the accuracy of prediction depends on the attribute characteristics in the applicability domain contained in the training set. Thus, when using a model constructed with a specific training set, prediction should be performed using compounds that have a similar structure space to those in the training set to improve the prediction accuracy of the QSAR model, since compounds with similar distribution in chemical space are more likely to exhibit similar biological activities (Huang and Fan, 2011).

QSAR is a method for using various biochemical and physical data to construct models. In QSAR studies, compounds can be mathematically codified as molecular descriptors, and the relationship between molecular descriptors and defined properties is constructed by statistical methods, after which a generated model is used to predict the corresponding properties of new compounds (Michielan and Moro, 2010). This method first transforms molecular structure into molecular descriptors, which are then used to establish prediction models by using statistical approaches or machine learning techniques such as support vector machine (SVM) and K Nearest Neighbor (kNN) (Wang S. et al., 2016; Wu et al., 2019; Yang et al., 2019;

Fu et al., 2020). For example, Schyman et al. (2017) used the variable nearest neighbor (vNN) method to develop 15 ADMET prediction models and to use them to quickly assess some potential drug candidates, including toxicity, microsomal stability, mutagenicity, and likelihood of causing drug-induced liver injury. Belekar et al. (2015) developed a computational model to identify compounds as breast cancer resistance protein (BCRP) inhibitors or not by using various machine learning approaches like SVM, k-NN, and the artificial neural network (ANN). The prediction accuracy of all three approaches was over 85%. Finally, internal and external cross-validation were performed to confirm the reliability of the model before it is used on new predictions to find molecules outside the training set (Figure 4).

QSAR modeling uses a large number of descriptors that allow lookups, enable structure/response associations, and help with similarity and substructure searches (Khan, 2010). Most of the available descriptors can be divided into three categories: (1) two-dimensional molecular topology information; (2) three-dimensional molecular structure; and (3) physicochemical and electronic descriptors, which are commonly used to predict ADMET-related properties (Danishuddin and Khan, 2016; Tabeshpour et al., 2018; Zhang et al., 2018). Jiang et al. (2020) developed a series of QSAR models by using 379 molecular descriptors to discriminate BCRP inhibitors. The descriptors characterized the physicochemical, two-dimensional substructures, and drug-like properties of the studied compounds. Lapins et al. (2018) constructed a QSAR model to predict the lipophilicity of compounds by using a signature molecular descriptor, which is related to the molecular two-dimensional topology information from 1.6 million compounds. Xu et al. (2017) developed three deep learning-based QSAR models to evaluate the acute oral toxicity (AOT) of compounds. The atom and bond information extracted from over 2,000 two-dimensional molecule structures were used as descriptors to construct the models. The best model achieved an external prediction accuracy over 94%, which is more efficient than traditional fingerprints or descriptors. Bujak et al. (2015) predicted the permeability of the BBB of chemical compounds using molecular energy-related descriptors in combination with the well-known lipophilicity descriptors. The data indicate that the QSAR model has important information value, and these descriptors may have supportive value in predicting the blood brain distribution (Bujak et al., 2015). Therefore, accurate prediction of ADMET parameters mainly depends on the selection of a suitable modeling method, molecular descriptors of specific ADMET endpoints, and large experimental data sets related to these endpoints. Only in this way can the ADMET properties of the candidate compounds be predicted precisely.

At present, many tools used for ADMET prediction have been developed based on QSAR methods. These tools utilize different descriptors to define the collected data, and then the mathematical model fitted from the training set is used to predict the properties. We listed three widely used QSAR-based ADMET prediction tools and related studies herein. (1) The Danish QSAR Database (<http://qsar.food.dtu.dk/>) collected estimates from over 200 QSAR models from free and commercial



platforms, including descriptors like ecotoxicity, environmental fate, physicochemical properties, and ADMET. Trivedi et al. (2019, 2020) used the online tool Danish QSAR database to determine the ADMET properties for potential hits for Dengue fever and H1N1 flu, respectively. Multicase acute aquatic toxicity, carcinogenicity, arylhydroxylase activity, lethal body burden, bioconcentration, mutagenicity, biodegradation, environmental partitioning, and general properties are included in the ADMET properties. Finally, 12 compounds were identified as potential leads against dengue fever and 18 compounds against H1N1 flu. (2) The OCED Toolbox (<https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm>), a package for toxicity prediction, was also developed based on QSAR. Han et al. (2019) used OECD QSAR Toolbox 4.1 to predict the genotoxicity for ceftazidime (CAZ) and its impurities to improve quality control of drugs. (3) ADMET Predictor™ (<https://www.simulations-plus.com/software/admetpredictor/>) is another tool utilizing QSAR to predict ADMET parameters of compounds. Takac et al. (2019) used ADMET Predictor™ to investigate the potential impact and safety profile with respect to the environment and health for 25 selected entactogen molecules. The chemical structure (including 1D and 2D) information was used as the input for ADMET Predictor™. Lipophilicity parameters, volume of distribution, jejunal permeability, solubility, and logarithm of the brain/blood partition coefficient were predicted in this case. Alarn and Khan (2019) used the ADMET Predictor™ to predict pharmacokinetics, pharmacodynamics, and toxicity parameters of flavone analogs to reveal their anticancer activity. Different physicochemical properties were calculated as descriptors to build the model, and then numerous properties were predicted, such as solubility, lipophilicity, permeability,

absorption, bioavailability, BBB, transporters, plasma-protein binding, and volume of distribution.

Although the use of QSAR models has made considerable progress in ADMET prediction, these models cannot yet be used to replace *in vitro* or *in vivo* studies for all endpoints. The QSAR method is always limited by its model expansion capability, and large experimental data are always needed for model construction. The narrow data distribution may induce over fitting and lead to inaccurate prediction results. For example, Verheyen et al. (2017) estimated the QSAR models used for the prediction of eye and skin irritation/corrosion in Derek Nexus, Toxtree and Case Ultra. They found that the prediction results were unsatisfactory because of the narrow application range and low accuracy. Thus, validation and documentation for a constructed model is important prior to use.

### PBPK Modeling

Most traditional models for predicting drug pharmacokinetics are empirical models. With a deeper understanding of the pharmacokinetic mechanism of drugs, PBPK models have been developed to predict PK properties (De Buck and Mackie, 2007). The PBPK model is an arithmetical model that combines drug data (e.g., drug concentration and clearance rate) and species physiology parameters to replicate the PK profile of a drug in plasma and tissues, aiming to describe *in vivo* drug pharmacokinetics that are related to tissue volume, administration routes, blood flow, biotransformation pathways, and interactions with tissues or organs in the body (Espie et al., 2009; Jones et al., 2015). The origin of the PBPK models can be traced back to Teorell's work in 1935. Teorell introduced a multicompartment model to simulate pharmacokinetics, which

organically combined physiology and biology for the first time (Teorell, 1935; Zhao et al., 2011). Teorell's work has since attracted serious attention to the PBPK model.

In recent years, PBPK modeling has been substantially improved, making it more widely applicable for the research and development of drugs (Edginton et al., 2008; Rowland et al., 2011; Zhuang and Lu, 2016). Moreover, the increase in preclinical data, especially *in vitro* data, has promoted the development of PBPK models and simulations in drug discovery (Zhuang and Lu, 2016). PBPK modeling describes the physical and biological disposition of each compartment by dividing organisms into individual organs, and the most common processes are related to blood transportation and penetration, distribution between blood and organ tissue, and metabolic excretion, among others (Schmitt and Willmann, 2004). Since PBPK integrates large amounts of drug-specific data, parameters, and species physiology (systematic data), there are two kinds of parameters in PBPK models, which use the concentration-time curves of all organs and blood as output information (Nestorov, 2003). The first type consists of physiological parameters, such as tissue volume, blood flow, and cardiac output. Recently, due to the extensive application of *in vitro-in vivo* extrapolation (IVIVE) in PBPK, many researchers have predicted the disposal of drugs *in vivo* through *in vitro* metabolism and transport data, indicating that metabolic enzymes and transporter expression data have become important physiological parameters (Rostami-Hodjegan, 2012). The second type consists of drug-related parameters, such as the plasma ratio, organ/blood partition coefficient, and permeability (Nestorov, 2003). Recently, PBPK models have been widely constructed to predict drug-related parameters (Chow et al., 2016; Pathak et al., 2019; Song et al., 2020). For example, Chow et al. used a physiologically based model to predict drug solubility and effective permeability (Chow et al., 2016) to examine the potential impact of excipients on oral drug absorption.

## DATABASES

In the past 10 years, with rapid development, a number of related databases storing pharmacokinetic parameters have emerged. We collected some of the most commonly used databases and classified them as ADMET-related databases and auxiliary databases. A brief introduction to these databases, including website links, data scales, and descriptions, is provided in **Tables 1, 2**. For the ADMET-related databases, users can submit information on the compounds they want to query through the corresponding modules. Then, shape screening or pharmacophore screening will be performed to obtain additional targets or bioactivity information on similar ligands that match the query molecule. The ADMET-related properties can also be obtained from the query result. The auxiliary databases mainly focus on providing structural information about compounds. Although some ADMET-related information is provided in the search results, it is not complete, and not every compound is associated with such information.

## ADMET-Related Databases

At present, many *in silico* models are used to predict ADMET, but massive amounts of data are needed to build them. The quality and quantity of the data are closely related to the accuracy of model prediction, so reliable experimental data are the key to successful prediction (Dearden, 2007; Alqahtani, 2017). Currently, there are some databases that can help ADMET prediction, such as the ADME Database (Shang et al., 2017), SuperToxic (Schmidt et al., 2009), PKKB (Cao et al., 2012), and DSSTox (Williams et al., 2017). By using these databases, users can obtain helpful data sets for use in external algorithms to generate prediction models. The databases can also be used directly to perform prediction through search functions, such as the similarity search or prediction. In addition, these models can be updated as new experimental data are added to the database.

The ADME Database (<https://www.fujitsu.com/jp/group/kyushu/en/solutions/industry/lifescience/admedatabase/>), developed by Zagreb University and Fujitsu in 2004, is a commercial database that specializes in pharmacokinetics information. It provides comprehensive data on drug-metabolizing enzymes and drug transporters that are specific to humans. The data have been widely used in drug research and development, such as ADME prediction and drug-drug interactions. Users can search for classification, metabolic reactions, and kinetics-related information about compounds by structure or substructure. However, the database currently limits large-scale downloads of user data, as well as public dissemination of some models.

SuperToxic collects toxins from different sources (animals, plants, synthetic, etc.), compiles ~60 000 compounds with their structures, and integrates some chemical properties and commercial availability information (Schmidt et al., 2009). These compounds are classified based on their toxicity, which derives from more than 2 million measurements. The values can be used to study the relationship between chemical structures and functions of toxins for evaluating the risk of their use. Users can easily query the structure and toxicity information of all compounds with corresponding properties through a structure search, name search, or property search. SuperToxic also allows users to browse the data by choosing an alphabetic character or numbers to present all entries starting with the selection (Schmidt et al., 2009). The available CASRN or NSC numbers in the database can also be recorded. The toxicity information retrieved from the database includes the dosage, type of test (toxicity measurement, such as LD50), and cell lines or organisms that determine the toxicity (Schmidt et al., 2009). The database also integrates software packages that are widely used in modern composite database construction, such as Marvin Sketch (molecule drawing), JMol (visual inspection), and MyChem/OpenBabel (property calculation). Furthermore, SuperToxic was connected to the Protein Data Bank, UniProt, and KEGG databases to identify potential targets in biochemical pathways to search for compounds (Schmidt et al., 2009).

The EPA Distributed Structure-Searchable Toxicity (DSSTox) database, which provides a series of documented, standardized and complete structure annotated toxicity information files,

**TABLE 1** | Some widely used ADMET-related databases.

Database name	Availability	Description	URL
ACD/Percepta	Commercial	Used to calculate PhysChem properties, ADME properties, and toxicity endpoints	<a href="https://www.acdlabs.com/products/percepta/index.php">https://www.acdlabs.com/products/percepta/index.php</a>
ADME Database	Commercial	A commercial database for studying drug interactions and ADME, updated quarterly	<a href="https://www.fujitsu.com/jp/group/kyushu/en/solutions/industry/lifescience/admedatabase/">https://www.fujitsu.com/jp/group/kyushu/en/solutions/industry/lifescience/admedatabase/</a>
AurSCOPE ADME	Commercial	A complete annotated, structured knowledge base can be used to design predictive models and identify potential drug interactions	<a href="http://www.aureus-sciences.com/">http://www.aureus-sciences.com/</a>
Knowitall	Commercial	The world's largest mass spectrometry library, can be used to quickly and accurately identify the spectrum	<a href="http://www.bio-rad.com/">http://www.bio-rad.com/</a>
ToxBank	Commercial	A cross-cluster project can be used for comprehensive data analysis of toxicology and alternative detection of repeated dose toxicity tests	<a href="http://toxbank.net/">http://toxbank.net/</a>
ACToR	Free	Exploring and visualizing complex computational toxicology information	<a href="https://actor.epa.gov/actor/home.xhtml">https://actor.epa.gov/actor/home.xhtml</a>
BRENDA	Free	Database of enzyme and enzyme-ligand information	<a href="http://www.brenda-enzymes.org/">http://www.brenda-enzymes.org/</a>
CEBS	Free	A toxicological resource that can be used to model, predict analyze and assess effects of time and dose on responses to experimental conditions	<a href="https://manticore.niehs.nih.gov/cebssearch">https://manticore.niehs.nih.gov/cebssearch</a>
ChemTunes	Free	A unique cheminformatics platform and expert QC'ed database that facilitate and support the safety and risk assessment process for chemical substances	<a href="https://www.mn-am.com/products/chemtunes">https://www.mn-am.com/products/chemtunes</a>
CTD	Free	A public resource storing scientific data about relationships between genes, chemicals, and human diseases	<a href="http://ctdbase.org/">http://ctdbase.org/</a>
DIDB	Free	Allows researchers to perform the assessment of human PK-based drug interactions and drug safety	<a href="https://www.druginteractioninfo.org/">https://www.druginteractioninfo.org/</a>
DSSTox	Free	A public resource of chemical structure information with existing toxicity data for supporting better predictive toxicology	<a href="https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database">https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database</a>
eChemPortal	Free	Chemical substance search, chemical property data search and GHS search are available	<a href="http://www.oecd.org/chemicalsafety/risk-assessment/echemportal/globalportaltoinformationchemicalsubstances.htm">http://www.oecd.org/chemicalsafety/risk-assessment/echemportal/globalportaltoinformationchemicalsubstances.htm</a>
IDA2PM	Free	An integrated database for data access transformation and analysis visualization and prognostic modeling	<a href="http://idaapm.helsinki.fi/">http://idaapm.helsinki.fi/</a>
Liceptor Database	Free	A ligand database containing 2D structures, related molecular descriptors and bioactivity data consisting of assays, functions and therapeutic hints	<a href="http://www.evolvus.com/">http://www.evolvus.com/</a>
Metabolism and Transport Database	Free	A database related to small molecule transport metabolism that can be used for computational analysis and modeling	<a href="http://www-metabase.ch.cam.ac.uk/">http://www-metabase.ch.cam.ac.uk/</a>
MetaCyc	Free	A database of metabolic pathways that can predict the metabolic pathways of sequenced genomes	<a href="https://metacyc.org/">https://metacyc.org/</a>
NTP	Free	NTP studied at any chemical with the potential to impact health	<a href="https://sandbox.ntp.niehs.nih.gov/neurotox/">https://sandbox.ntp.niehs.nih.gov/neurotox/</a>
Pharmacokinetics Knowledge Base (PKKB)	Free	A free database that collects ADMET data and can be used for ADMET modeling	<a href="http://cadd.zju.edu.cn/pkbb/">http://cadd.zju.edu.cn/pkbb/</a>
Repdose	Free	A toxicity database storing <i>in vitro</i> and <i>in vivo</i> data and a prediction system for the safety and risk evaluation	<a href="https://repdose.item.fraunhofer.de/about_repdose.html">https://repdose.item.fraunhofer.de/about_repdose.html</a>
RTECS	Free	Contains additional information related to the chemical industry and occupational safety and health that can be used to assess workers' exposure to chemicals	<a href="https://www.atsdr.cdc.gov/substances/index.asp">https://www.atsdr.cdc.gov/substances/index.asp</a>
SuperToxic	Free	Can perform similarity screening, risk assessment, and link to other databases	<a href="http://bioinformatics.charite.de/supertoxic/">http://bioinformatics.charite.de/supertoxic/</a>
T3DB	Free	Describe the relationship between toxins and targets and the mechanism of toxic action, which can be used for prediction of toxicity, prediction of toxic targets, etc.	<a href="http://www.t3db.ca/">http://www.t3db.ca/</a>
TOXNET	Free	It is composed of a set of databases dealing with environmental health of toxicological hazardous chemicals and related fields	<a href="https://toxnet.nlm.nih.gov/">https://toxnet.nlm.nih.gov/</a>
ToxRefDB	Free	Store data from <i>in vivo</i> animal toxicity tests and provide toxicity endpoints for predictive modeling	<a href="https://catalog.data.gov/dataset/toxcast-toxrefdb">https://catalog.data.gov/dataset/toxcast-toxrefdb</a>



**TABLE 2 |** Some auxiliary databases for ADMET prediction.

Database Name	Availability	Description	Scale	URL
PubChem	Free	A public database storing biological properties of small molecules	Contains biological test results for more than 700,000 compounds	<a href="http://pubchem.ncbi.nlm.nih.gov/">http://pubchem.ncbi.nlm.nih.gov/</a>
DrugBank	Free	Storing information about drugs and related targets	Contains 13,490 drug entries	<a href="https://www.drugbank.ca/">https://www.drugbank.ca/</a>
STITCH	Free	A search tool and resource for interactions of chemicals and proteins	Interaction data for more than 68,000 different chemicals	<a href="http://stitch.embl.de/">http://stitch.embl.de/</a>
ChEMBLdb	Free	An open large bioactive database that can display metabolite pathways and link to metabolite metabolizing enzymes and information document source data	Containing more than 1.6 million individual compound structures denoted in the database, with 14 million activity values from more than 1.2 million assays	<a href="https://www.ebi.ac.uk/chembl/db/">https://www.ebi.ac.uk/chembl/db/</a>
BindingDB	Free	A web-based resource containing experimental binding affinities, focusing mainly on the interactions between potential drug-targets and drug-like molecules	1,794,819 binding data, for 7,438 protein targets and 796,104 small molecules	<a href="http://www.bindingdb.org/bind/index.jsp">http://www.bindingdb.org/bind/index.jsp</a>
ChemProt	Free	A resource to perform <i>in silico</i> estimation of small molecules with the integration of molecular, cellular, and disease-related proteins complexes	over 1.7 million compounds with 7.8 million bioactivity measurements for 19,504 proteins	<a href="http://potentia.cbs.dtu.dk/ChemProt/">http://potentia.cbs.dtu.dk/ChemProt/</a>
SIDER	Free	Contains information about marketed drugs and their adverse reactions, which can be used to quickly track trace the origin of an extracted side effect	5,868 side effect and 1,430 drugs and 139756 drug-SE pairs	<a href="http://sideeffects.embl.de/drugs/">http://sideeffects.embl.de/drugs/</a>
MetaADEDB	Free	A comprehensive database of adverse drug reactions (adr), which is used to predict clinical adr and the source of adverse drug side effects of personalized drugs, and to predict drug interactions with targets and drug action patterns	3,060 chemicals (including more than 1,300 FDA approved and experimental drugs) and 13,256 ADEs	<a href="http://lmmd.ecust.edu.cn/online_services/metaadedb/">http://lmmd.ecust.edu.cn/online_services/metaadedb/</a>
TTD	Free	Provides data about known and studied therapeutic macromolecules, the targeted disease, pathway information, and the associated drugs directed at each of these targets	2,589 targets and 31,614 drugs	<a href="http://bidd.nus.edu.sg/group/cjttd/">http://bidd.nus.edu.sg/group/cjttd/</a>
KEGG	Free	A reference repository that is widely used to integrate and interpret large-scale datasets obtained by genome sequencing and other high-throughput experimental technologies	4 categories (systems, genomic, chemical and health information) from 18 databases	<a href="https://www.kegg.jp/">https://www.kegg.jp/</a>

can be very useful for SAR model development (Richard and Williams, 2002; Williams et al., 2017). To allow wider use of the database, DSSTox was designed to use a structure data file (SDF), a public and industry-standard import/export file format storing chemical structures and property information that can be used as input for any chemical relational database (CRD) application or converted to data tables. It is one of the best-curated public datasets available at present, and the data stored in it are regarded as a standard reference for publicly available structural toxicity-based data.

Except for the databases introduced above, some newly constructed ADMET-related databases should also be of concern, for example, Comparative Toxicogenomics Database (CTD) (Davis et al., 2019), The Toxicity Reference Database (ToxRefDB) (Watford et al., 2019), and The Chemical Effects in Biological Systems database (CEBS) (Lea et al., 2017). CTD is a powerful and public database designed to enhance understanding of how environmental exposures influence human health. It provides

data on chemical–gene/protein interactions, chemical–disease and gene–disease relationships that are combined with pathway and function data to help develop hypotheses about the inherent mechanisms of diseases affected by the environment (Davis et al., 2019). ToxRefDB collects data from *in vivo* animal toxicity tests and provides toxicity endpoints for predictive modeling. Approximately 28,000 datasets from nearly 400 endpoints have been generated and stored. The recent update of ToxRefDB has added connections to other resources and significantly enhanced the utility of predicting toxicology (Watford et al., 2019). CEBS offers a toxicology resource that compiles individual and concise animal data from 11,000 test articles and more than 8,000 studies encompassing all available National Toxicology Program (NTP) carcinogenicity, genetic toxicity, and short-term toxicity studies. The high-quality data in CEBS is very useful for constructing a more accurate model for toxicity prediction (Lea et al., 2017). We may infer from the recent constructed databases that the quantity and quality of the data

determine the quality of the model and will be our focus in the future.

## Auxiliary Databases

In addition to ADMET-related databases, databases of biological activity, pathways, and side effects are important for ADMET prediction. Most of these databases are free and open to visitation, such as DrugBank (Wishart et al., 2018), PubChem (Kim et al., 2019), and ChEMBL (Gaulton et al., 2017). Although they are rarely used to predict ADMET-related properties directly, they can provide structural information to build models or be queried for information about compounds. Users can also download the predicted compound structure and use it as an input file for other software.

DrugBank, a comprehensive database, integrated thousands of well-studied drugs and drug targets with their physical, chemical, biological, and pharmaceutical data (Wishart et al., 2006, 2008). DrugBank 4.0 was further expanded to contain data on ADMET and other kinds of QSAR information (Law et al., 2013). DrugBank 5.0, the latest version, has further updated this information (Wishart et al., 2018). Users can use chemical shifts or mass-to-charge ratio ( $m/z$ ) lists to search DrugBank's spectral library for approximate or exact matches. DrugBank also systematically classifies compounds into different types based on structural features and structural similarities and allows users to query it by using a simple text (Law et al., 2013).

Pubchem, a public database of small molecules with their biological properties, consists of three interconnected parts: (1) Compound, storing over 102 million unique chemical structures provided by various depositors; (2) Substance, containing more than 251 million records including complexes, extracts, mixtures, and non-characteristics; and (3) BioAssay, containing more than 1,067,000 bioassays, providing composite adjacent structures, substructures, similar structures, bioactivity data, and other search functions (Kim et al., 2019).

ChEMBL is an open data database containing two-dimensional structures, calculated properties (molecular weight, lipophilicity, etc.) and abstract biological activities (pharmacology and ADMET data) for numerous drug-like bioactive compounds (Gaulton et al., 2017). It is composed of three different datasets that were originally developed by Inpharmatica, including StARlite, CandiStore, and DrugStore (Overington, 2009). The data in ChEMBL were extracted from the scientific literature and designed to meet the needs of users to intelligently cluster relevant information and integrate data across therapeutic studies and areas.

## SOFTWARE

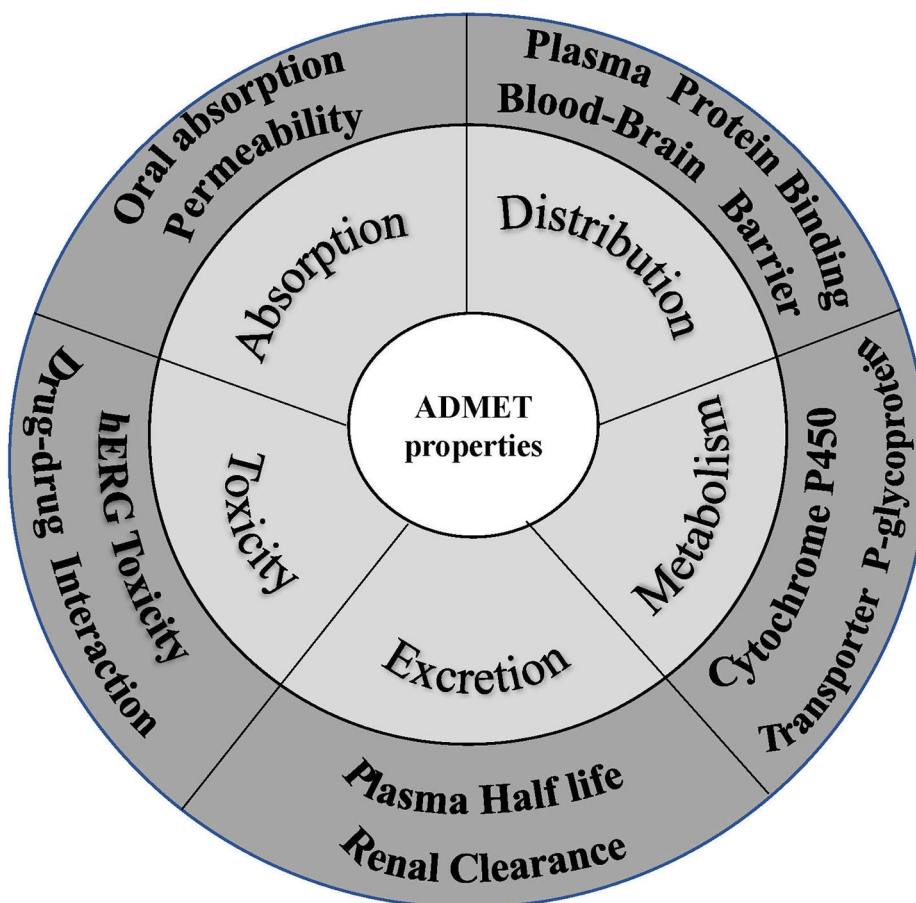
Favorable ADMET characteristics are important as early requirements for drug candidates to reduce late failure and cost. However, many ADMET properties are highly dependent on each other, so they need to be optimized simultaneously in preclinical studies on drug discovery and development. Nevertheless, concurrent optimization of multiparameter ADMET is the most difficult and least attractive stage. As a result, early prediction of ADMET involved only some simple

properties, such as  $\log P$ ,  $\log D$ , and  $\log S$ . With increasing experimental data, an increasing number of *in silico* models were developed to predict more complex ADMET parameters, such as the human intestinal absorption rate, oral bioavailability, blood-brain barrier permeation rate, Caco-2 permeability, human intestinal absorption, drug interactions, P-glycoprotein, plasma protein binding rate, CYP metabolic enzymes, and kidney clearance (Pires et al., 2015; Dong et al., 2018) (**Figure 5**). People have also attempted to integrate these models to predict ADMET parameters concurrently, and many studies have described these *in silico* models and their predicted properties (Dickins and van de Waterbeemd, 2004; Wang et al., 2015). In addition, software integrating these models to predict ADMET parameters concurrently has been developed. **Tables 3, 4** list some of these software packages (free and commercial) with their functions. We also compared five commonly used software in **Table 5** with each other to visualize their detailed functions.

SwissADME is a hybrid web server that was developed by the Swiss Institute of Bioinformatics (Daina et al., 2017). It supports diverse input formats and can predict and analyze the ADME properties of numerous compounds in batches submitted from all over the world. This software outputs different types of physicochemical properties of drugs: water solubility, lipophilicity, physicochemical properties, druglikeness, pharmacokinetics, and medicinal chemistry, which can be directly exported and saved as a data file in CSV (comma-separated values) format and read by programs such as WordPad and Excel (Daina et al., 2017). In addition, it supplies a bioavailability radar map to quickly and intuitively evaluate the druglikeness of small molecules, facilitating its use for non-experts without professional knowledge (Daina et al., 2017). The server uses a variety of rules to evaluate the same property and provides the evaluation criteria and basis for most of the predicted data. However, in the prediction of whether the compound is a substrate or an inhibitor of the CYP enzyme, only a propensity is output, rather than a probability output similar to that of admetSAR. SwissADME also supplies a link for the one-click submission of the queried molecules to other servers in the Swiss series for further analysis (Daina et al., 2017).

Although *in vivo* toxicology is still the gold standard for identifying drug side effects, it is now believed that this method will not help reduce the large consumption rate in late clinical development (Merlot, 2010). Many computational tools have been developed to predict drug toxicity, helping to decrease the attrition rate of molecular compounds in drug discovery and reduce drug development time and cost. In recent years, the predictive power of these toxicology prediction systems has tremendously improved, covering more complex toxicological endpoints, such as hepatotoxicity, teratogenicity, nephrotoxicity, and carcinogenicity (Muster et al., 2008; Yang et al., 2018b). Currently, many commercial and free web-based toxicity predictors are available, such as Lazar (Maunz et al., 2013) and Toxtree (Patlewicz et al., 2008; Bhatarai et al., 2016).

Lazar, developed by *in silico* toxicology GMBH, is a tool based on OpenTox (an integrated interface for an interoperable prognostic toxicology framework) to predict toxicological endpoints such as carcinogenicity, reproductive toxicity, and



**FIGURE 5 |** Crucial properties of ADMET.

long-term toxicity (Hardy et al., 2010). It uses data mining algorithms to predict the toxicity of new compounds based on the experimental training data. Data sets with chemical structure and biological activity can be used as training data. Thus, Lazar can serve as a universal predictive algorithm for any biological endpoint if sufficient experimental data are available, so users no longer need to consider chemical, biological, or toxicological expertise but derive *in silico* models from statistical standards (Maunz et al., 2013). Users need only to input the structure of the compound, and Lazar will search the database for a series of similar compounds and corresponding experimental data, which will be used to construct a local QSAR model. The prediction results using the model will be displayed in a graphical interface, which provides structural features and compounds similar to the query compounds and toxicity properties for each fragment (Helma, 2006; Maunz et al., 2013).

Toxtree is a free software program that was commissioned by the European Chemicals Agency (ECB), the joint research center of the European Commission (Bhatia et al., 2015). It was originally designed to enable effective development of the Cramer decision tree. The latest version of Toxtree included additional projects, such as corrosion rules, BfR/SICRET skin irritation,

and the Verhaar scheme, with a total of 14 functional modules (Bhatia et al., 2015). It includes physiochemical exclusion rules and structural alert inclusion rules, which are used to categorize compounds (Bhattacharai et al., 2016). Unlike Lazar, it has no training set. Its prediction is based on structural filters, so there is no applicability domain. It handles molecular structure information by using a decision tree model for risk assessment (Bhattacharai et al., 2016). Users can access it at <http://toxtree.sourceforge.net/> to predict the toxicity of structures of interest.

ADMETlab is a platform for systematic ADMET estimation based on a comprehensive collection of ADMET databases (Dong et al., 2018). The platform includes four functional modules, which are used for drug similarity assessment (based on Lipinski's rule of five and the Druglikeness model), ADMET endpoint prediction, system evaluation, and database/similarity search (Dong et al., 2018). Among them, "ADMET prediction" is the main module used; the other three are auxiliary modules. Users can query one or more compounds with the platform by using the SMILES, uploading an SDF format file, or drawing the chemical structure using the embedded JME editor. After the compound is uploaded, the platform will use multiple pharmacokinetics models built by the different integrated data

**TABLE 3 |** Some free ADMET software and related properties in their prediction.

Software	URL	LogS	LogP	LogD	Sol	TPSA	HIA	Caco-2	PPB	BBB	V <sub>d</sub>	Met	CL	Tox	P-gp	per	pKa	MDCK
ACD/I-lab	<a href="http://www.acdlabs.com/home/">http://www.acdlabs.com/home/</a>	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓		✓	
ADMETlab	<a href="http://admet.scbdd.com/">http://admet.scbdd.com/</a>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
admetSAR	<a href="http://lmd.ecust.edu.cn/admetSar1/">http://lmd.ecust.edu.cn/admetSar1/</a>				✓		✓	✓		✓		✓		✓	✓			
FAF-Drug4	<a href="http://fafdrugs4.mti.univ-paris-diderot.fr">http://fafdrugs4.mti.univ-paris-diderot.fr</a>		✓			✓												
Lazar	<a href="https://www.in-silico.de">https://www.in-silico.de</a>													✓				
OCHEM	<a href="http://ochem.eu">http://ochem.eu</a>	✓	✓		✓		✓			✓		✓		✓		✓		
OECD Toolbox	<a href="http://toolbox.oasis-lmc.org/">http://toolbox.oasis-lmc.org/</a>											✓		✓				
OSIRIS property explorer	<a href="http://www.organic-chemistry.org/prog/peo/">http://www.organic-chemistry.org/prog/peo/</a>	✓	✓		✓	✓								✓				
pkCSM	<a href="https://smarctyp.sund.ku.dk">https://smarctyp.sund.ku.dk</a>				✓		✓	✓		✓	✓	✓	✓	✓	✓	✓		
SMARTCyp	<a href="https://nodepit.com/node/org.lhasalimited.knime.metabolism.encapsulated.smartcyp.SMARTCypNodeFactory">https://nodepit.com/node/org.lhasalimited.knime.metabolism. encapsulated.smartcyp.SMARTCypNodeFactory</a>											✓						
SwissADME	<a href="http://www.swissadme.ch">http://www.swissadme.ch</a>	✓	✓		✓	✓	✓			✓		✓			✓	✓		
ToxCreate	<a href="https://github.com/opentox/toxcreate">https://github.com/opentox/toxcreate</a>						✓							✓				
ToxTree	<a href="http://toxtree.sourceforge.net/#carousel0">http://toxtree.sourceforge.net/#carousel0</a>											✓		✓				
VCCLAB (ALOGPS 2.1)	<a href="http://www.vcclab.org">http://www.vcclab.org</a>	✓	✓	✓	✓													✓
vNN-ADMET	<a href="https://vnnadmet.bhsai.org/vnnadmet/login.xhtml">https://vnnadmet.bhsai.org/vnnadmet/login.xhtml</a>									✓		✓		✓	✓			

logS, aqueous solubility; LogP, octanol-water partition coefficient; LogD, octanol-water distribution coefficient; Sol, solubility; TPSA, topological polar surface area; HIA, human intestinal absorption; PPB, plasma protein binding; BBB, blood-brain barrier; V<sub>d</sub>, volume of distribution; Met, metabolism; CL, clearance; Tox, toxicity; P-gp, P glycoprotein; Per, Permeability; pKa, acidity coefficient; MDCK, madin Darby canine kidney cell line.

**TABLE 4 |** Some commercial ADMET software and related properties in their prediction.

Software	URL	LogS	LogP	LogD	Sol	TPSA	HIA	Caco-2	PPB	BBB	V <sub>d</sub>	Met	CL	Tox	P-gp	Per	pKa	MDCK
ACD/Percepta Platform	<a href="https://www.acdlabs.com/products/percepta/">https://www.acdlabs.com/products/percepta/</a>	✓	✓	✓	✓	✓				✓	✓	✓		✓	✓		✓	
ADMEWORKS	<a href="https://www.fujitsu.com/jp/group/kyushu/en/solutions/industry/lifescience/admeworks/">https://www.fujitsu.com/jp/group/kyushu/en/solutions/industry/ lifescience/admeworks/</a>	✓	✓		✓		✓			✓		✓		✓	✓			
CompuDrug's Pallas System	<a href="http://www.compudrug.com/pallas_system">http://www.compudrug.com/pallas_system</a>		✓	✓		✓						✓		✓			✓	
Derek Nexus	<a href="https://www.lhasalimited.org/products/derek-nexus.htm">https://www.lhasalimited.org/products/derek-nexus.htm</a>													✓				
MCASE, CASE, CASETOX	<a href="https://www.multicase.com/">https://www.multicase.com/</a>													✓				
MetaSite	<a href="http://www.moldiscovery.com/soft_metasite.php">http://www.moldiscovery.com/soft_metasite.php</a>											✓						
PASS	<a href="http://genexplain.com/pass/">http://genexplain.com/pass/</a>											✓		✓				
Schrodinger QikProp	<a href="https://www.schrodinger.com/qikprop/">https://www.schrodinger.com/qikprop/</a>	✓	✓	✓	✓			✓		✓				✓				✓
Simulations Plus ADMET Predictor	<a href="https://www.simulations-plus.com/software/admetpredictor/">https://www.simulations-plus.com/software/admetpredictor/</a>		✓	✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
StarDrop	<a href="https://www.optibrium.com/stardrop/stardrop-p450-models.php">https://www.optibrium.com/stardrop/stardrop-p450-models.php</a>		✓	✓	✓	✓	✓		✓	✓				✓	✓			
TIMES	<a href="http://oasis-lmc.org/products/software/times.aspx">http://oasis-lmc.org/products/software/times.aspx</a>													✓				
TOPKAT	<a href="http://www.moldiscovery.com/software/vsplus/">http://www.moldiscovery.com/software/vsplus/</a>													✓				
VolSurf+	<a href="http://www.moldiscovery.com/software/vsplus/">http://www.moldiscovery.com/software/vsplus/</a>		✓	✓	✓					✓	✓	✓					✓	

logS, aqueous solubility; LogP, octanol-water partition coefficient; LogD, octanol-water distribution coefficient; Sol, solubility; TPSA, topological polar surface area; HIA, human intestinal absorption; PPB, plasma protein binding; BBB, blood-brain barrier; V<sub>d</sub>, volume of distribution; Met, metabolism; CL, clearance; Tox, toxicity; P-gp, P glycoprotein; Per, Permeability; pKa, acidity coefficient; MDCK, madin Darby canine kidney cell line.



sets to make extensive predictions of the ADMET properties. Model prediction results are output in an interactive data table containing predicted values and structures. The software allows batch prediction, and users can apply the “drug similarity assessment” module to filter out compounds that are unlikely to be lead compounds or drugs, achieving the purpose of preliminary screening (Dong et al., 2018). Users must select one model to acquire results for one or multiple molecules, which is proper for screening compounds at specific endpoints, and the results will provide reasonable ADMET recommendations for each endpoint. Therefore, users can perform rapid screening of ADMET properties based on these independent specific prediction models and even further deliberately optimize the chemical structures of compounds, making them more likely to become drugs (Dong et al., 2018). Considering the very large amount of collected data and numerous constructed QSPR models, ADMETlab is currently one of the most comprehensive platforms used in ADMET prediction.

AdmetSAR is a free and comprehensive tool for ADMET property prediction (Cheng F. et al., 2012). The ADMET-related property data stored in AdmetSAR were collected from published literature. AdmetSAR also includes a searchable tool called ADMET-Simulator, which combines predictive and superior QSAR models in a toolbox based on chemical informatics and can predict ~50 ADMET endpoints. AdmetSAR enables users to easily search for ADMET properties by querying CASRN, the common name, or the structure (Yang et al., 2018a). The new version of admetSAR (version 2.0) mainly focuses on *in silico* prediction of chemical ADMET properties (Yang et al., 2018a). More than 40 predictive models trained by state-of-the-art machine learning methods were implemented in admetSAR. Four functions were developed: (1) customizable ADMET risk filters, (2) QSAR-based ADMET property prediction, (3) toxicity prediction for environmental chemicals, and (4) environmental hazard assessment. ADMETopt (Yang et al., 2018c) is a new module added in version 2.0 for lead compound optimization according to the predicted ADMET properties.

Except for ADMET software, an increasing number of PBPK software have been developed to perform systematics of the drug process in human body. The development of PBPK software has further promoted the use of PBPK modeling methods (Bouzon and Walther, 2008; Edginton et al., 2008; Perdaems et al., 2010). PBPK software was used to build and use the PBPK models, which can be useful for the estimation of pharmacokinetic parameters during the drug development process. At present, PBPK software can be roughly divided into two categories, the user customized software and user-friendly software. A brief introduction to the two types of software is provided in Table 6, including the company/institute and website links. In Table 7, we compare the functions of the commonly used software packages for both categories, such as WinNonlin (<https://www.certara.com/>) and GastroPlus (<https://www.simulations-plus.com/>).

It was initially used in engineering and mathematics of the user customized software because the essence of PBPK modeling is mathematical modeling (Bouzon et al., 2012). Hence, the application of these software programs in PBPK modeling is a natural transformation. The user customized software requires

users to write their own programs to build the PK model at the beginning stage of development. This procedure requires users to have proficient programming skills as well as expertise in the field. Recently, specific PK or PBPK modules and equation libraries, as well as visual graphical interfaces, have been added to some of these software programs (Bouzon et al., 2012). By using these software programs, users can quickly generate standard PBPK models by following existing templates that already contain standard codes and equations, greatly facilitating user operation.

The user-friendly type, customized for PBPK modeling, has a graphical interface. It requires no modeling language and programming, so it is relatively simple to operate. Originally, some such software programs were specifically modeled for predicting a specific property of the ADME process, such as absorption (GastroPlus, Tian et al., 2011) or metabolism (SimCYP, Jamei et al., 2009). These software programs gradually evolved into complex PBPK modeling tools for the entire body. Recently, the function of updated versions has become increasingly sophisticated. Now, they not only model their specific areas but also simulate the whole-body pharmacokinetic process, which is absorption, metabolism, and excretion, etc. (Li M. et al., 2017; Byun et al., 2020). These software programs can perform various tasks, such as simulation, parameter evaluation, and sensitivity analysis, simply by inputting specific drug parameters and choosing certain model options (Bouzon et al., 2012).

## APPLICATIONS

The *in silico* applications of predicting ADMET profiles in 2016–2018 were collected by searching PubMed. We analyzed the search results and briefly introduce how software programs and methods predict the properties of ADMETs.

### Molecular Modeling

Most applications of molecular modeling focus on predicting the strength of the interaction between a molecule and a metabolic enzyme or transporter. For instance, Niu et al. (2016) performed docking studies of flavokawain A (FKA) and its target CYP450. FKA shows obvious inhibition of different CYP isoforms, and subsequent inhibition experiments showed that CYP3A2 was the primary isoform contributing to the metabolism of FKA. Gong et al. (2018) performed a molecular docking experiment to study the binding mode between sauchinone and CYPs. The results showed the interactions of sauchinone in the active site of CYP2B6, 2C19, 2E1, and 3A4. In addition to the above examples, the details of 22 representative studies are included in Table 8.

### QSAR

Prediction of the pharmacokinetic properties of compounds using QSAR relies mainly on traditional models or software developed based on constructed data sets. Table 9 lists seven typical applications of the QSAR method. For example, Khan et al. (2016) utilized the QSAR model in ACD/I-lab to determine multiple ADMET properties (such as logS, logP, logD, BBB) for 6 compounds targeting heat-shock protein 90 (Hsp90).

**TABLE 5** | Comparison of five commonly used ADMET software programs.

Tools	Availability	Batch computation	Endpoints	Database	Druglikeness rules	Druglikeness model	Systematic evaluation	Medicinal chemistry friendliness	Physico-chemistry properties	Similarity	QSAR model	Algorithms	Training sets	Pattern recognition
SwissADME	Free	✓	Number, 19 Contents: B, A, D, M	×	✓	×	✓	✓	✓	✓	×	×	×	×
ADMETlab	Free	✓	Number: 31 Contents: B, A, D, M, E, T	✓ (288,967 entries; 5 similarity searching strategies)	✓ (5 rules)	✓	✓	×	✓	✓	✓	✓	✓	×
admetSAR 2.0	Free	×	Number: 47 Contents: B, A, D, M, E, T	✓ (210,000 entries)	×	×	✓	×	✓	✓	✓	✓	✓	×
Lazar	Free	×	Number: 11 Contents: T: Acute toxicity; BBB; Carcinogenicity, LOAEL, Maximum Recommended Daily Dose, Mutagenicity	×	×	×	×	×	×	✓	✓	✓	✓	✓
ToxTree	Free	×	Number: 6 Contents: M, T	×	×	×	×	×	×	✓	×	×	×	✓

*B*, basic physicochemical property; *A*, absorption; *D*, distribution; *M*, metabolism; *E*, excretion; *T*, toxicity.

Then, six compounds were designed according to BBB and antiangiogenic properties. One molecule (compound 6) was observed to inhibit Hsp90 with a predicted efficiency of BBB permeation of 0.55 kcal/mol in comparison to the experimental

value of 0.625 kcal/mol. Ajay Kumar et al. (2018) performed 3D-QSAR studies to filter compounds based on ADME properties by using Schrodinger. Fifty hit compounds targeted to transforming growth factor- $\beta$  (TGF- $\beta$ ) type I were screened based on predicted ADME properties (such as BBB, logS, and Lipinski's rule of five). Seven molecules were finally selected as the lead compound for subsequent research.

TABLE 6 | Two types of several PBPK software programs.

Software	Company/Institute	URL
USER CUSTOMIZED SOFTWARE		
Kinetica	Thermo Fisher Scientific Inc.	<a href="http://kineticadownload.com/Kinetica5.1-SP1/Default.asp">http://kineticadownload.com/Kinetica5.1-SP1/Default.asp</a>
MATLAB-simulink	The MathWorks Inc.	<a href="https://www.mathworks.com/">https://www.mathworks.com/</a>
NONMEM	ICON	<a href="https://www.iconplc.com/innovation/nonmem/">https://www.iconplc.com/innovation/nonmem/</a>
SAAM II	Washington University	<a href="https://tegvirginia.com/software/saam-ii/">https://tegvirginia.com/software/saam-ii/</a>
WinNonlin	Certara USA Inc.	<a href="https://www.certara.com/software/phoenix-winnonlin/">https://www.certara.com/software/phoenix-winnonlin/</a>
USER-FRIENDLY SOFTWARE		
Cloe® PK	Cyprotex	<a href="https://www.cyprotex.com/insilico/">https://www.cyprotex.com/insilico/</a>
GastroPlus	Simulations Plus Inc.	<a href="https://www.simulations-plus.com/">https://www.simulations-plus.com/</a>
Medici-PK	Computing in Technology	<a href="http://www.cit-wulkow.de/">http://www.cit-wulkow.de/</a>
PK—Sim	Bayer Technology Services	<a href="http://www.systems-biology.com/sb/">http://www.systems-biology.com/sb/</a>
SimCYP	Certara USA Inc.	<a href="https://www.certara.com/software/simcyp-pbpbk/">https://www.certara.com/software/simcyp-pbpbk/</a>

PBPK

PBPK modeling has been consistently performed to predict pharmacokinetics with the help of some widely used software programs (Table 7). Seventy-four applications using three software packages are listed in Table 10. WinNonlin is one of the most widely used software programs in the prediction of pharmacokinetics. For example, Gestrich et al. (2018) used WinNonlin (v6.4) to analyze compartmental and non-compartmental gentamicin plasma concentrations vs. time. The peak drug concentrations and AUCs in young adults and older alpacas were compared, and both were significantly lower in young adults than in geriatric alpacas. The increased drug exposure and decreased clearance in geriatric alpacas created a greater risk of ADRs and/or therapeutic failure. Another software, NONMEM, is also widely used to predict the impact of drugs on the target population. It is the “gold standard” software package for analysis of population PK/PD data. For example, Polepally et al. (2018) used non-linear mixed-effects modeling in NONMEM (version 7.3) to analyze concentration-time data to estimate the effect of age on the pharmacokinetic parameters of lamotrigine (LTG). By comparing the pharmacokinetic

TABLE 7 | Comparison of commonly used PBPK software programs for two categories.

User customized software	Latest version	Modeling language	Auxiliary tool	Operation method	Model
WinNonlin	WinNonlin v6.4	R	Phoenix® NLME™	Built-in options or a combination of graphics and text commands	Non-compartmental analysis (NCA)
NONMEM	NONMEM v7.3	FORTTRAN	Wings for NONMEM, priana	Fully based on text	Two-compartment or three-compartment model
User-friendly software	Company	Species	Routes of administration	Features	Application
GastroPlus	Simulations Plus Inc.	Human, rat, dog, mouse, monkey, user defined	i.v., p.o., ocular, pulmonary, lingual, sublingual, buccal	ACAT model PEAR function	Ten modules for PBPK modeling, built-in multi-person physiological treatment model, IIVC; single simulation, batch simulation, parameter sensitivity analysis (PSA), population simulation
SimCYP	Certara USA Inc.	Human, rat, mouse, dog	i.v., p.o., pulmonary, skin	ADAM model, Extensive data base with physiological information	PBPK modeling function, IIVE; DDI simulation and population differential prediction; time course of simulated metabolites
PK-Sim	Bayer Technology Services	Human, rat, mouse, dog, monkey, minipig	i.v., p.o., subcutaneous, dermal, user defined	WB-PBPK model, various PBPK calculation methods	High-flexibility-all model parameters accessible for specific investigations; Simulating physiological variability in reactions

**TABLE 8 |** Applications of molecular modeling in predicting metabolic properties from 2016 to 2018.

Year	Compound	Metabolizing enzyme	Tool	References
2016	4 Steroid derivatives	CYP1B1 (3PM0)	Gold-5.4	Poirier et al., 2016
2016	8 omeprazole-based analogs	CYP2C19 (4GQS)	GOLD 5.2	Li et al., 2016
2016	Flavokawain A	CYP2D6 (3QM4)	SYBYL-X2.0	Niu et al., 2016
2016	Ketoconazole, Resveratrol, MAR, DAR and TAR	CYP3A4 (2V0M)	Discovery Studio 4.0	Basheer et al., 2016
2016	Naringenin, 6',7'-dihydrox-ybergamottin	CYP1A1 (4I8V)	Autodock	Santes-Palacios et al., 2016
2016	Pyridine, Piperidine and Azol scaffolds	UBE2D4	Discovery Studio 3.5	Ramatenki et al., 2016
2016	Rab38 inhibitors	Rab38	Discovery Studio 4.0	Abdelmonsef et al., 2016
2017	1'-S-1'-acetoxychavicol acetate	CYP1A2 (2HIF), CYP2D6 (3QM4), CYP3A4 (4D6Z)	AutoDock Vina version 1.1.2	Haque et al., 2017
2017	12 estrone (E1), 17 $\beta$ -estradiol (E2) derivatives	CYP1B1(3PM0)	GOLD 5.4	Dutour et al., 2017
2017	Progesterone (PGS)	CYP3A4 (1W0F)	Schrödinger Suite, 2012	Du H. et al., 2017
2017	Quinoxaline, Diazepine, Piperazine	UBE2NL enzyme	LigPrep version 5.6, Schrödinger	Ramatenki et al., 2017
2017	Resveratrol, Nitrostilbene, Dimethoxy-nitrostilbene, Ketoconazole	CYP3A4(2V0M)	Discovery Studio 4.0 (CDOCKER), Schrodinger Suite 2016 platform (Glide docking)	Basheer et al., 2017
2017	Sulfonyl hydrazones	MAO-A and B	AutoDock Vina	Abid et al., 2017
2017	Wilfortrine, Wilforine, Wilfordine, Euonymine, Wilforgine	CYP3A4 (1W0F)	Discovery Studio (CDOCKER)	Wang L. et al., 2017
2017	XIAP	Caspase-3	AutoDock Vina	Prokop et al., 2017
2017	$\alpha$ -Naphthoflavone (ANF) 7-ethoxyresorufin (7ER)	CYP1A2 (2H14)	GOLD 5.2.2	Watanabe et al., 2017
2018	15 Vinca derivatives	CYP3A4 (3NXU), CYP3A5	GOLD 5.2	Saba and Seal, 2018
2018	3 Compounds	CYP3A4 (4NY4)	Schrödinger Release 201702	Vaz et al., 2018
2018	Bavachin <sup>a</sup> , Neobavaisoflavone <sup>b</sup> , Corylifol A <sup>c</sup>	CYP1A2 (2H14) <sup>a,b,c</sup> , CYP2C9 (1R9O) <sup>a,b,c</sup> , CYP2C19 (4GQS) <sup>a,b,c</sup> , CYP2D6 (3TGB) <sup>a,b,c</sup> , CYP3A4 (1W0F) <sup>a,b,c</sup> , CYP2E1 (3T3Z) <sup>a,b,c</sup>	Discovery Studio 4.1 (CDOCKER)	Wang L. et al., 2018
2018	FAK and Triazinic inhibitors	FAK	Schrödinger 9.0	Cheng P. et al., 2018
2018	Metconazole (MEZ) isomers	CYP3A4 (2V0M)	AutoDock Vina	Zhuang S. et al., 2018
2018	Paracetamol <sup>a</sup> , Pilocarpine <sup>b</sup>	CYP2E1 (3T3Z) <sup>a,b</sup> , CYP3A4 <sup>a</sup>	SYBYL-X 1.3	Wang Y. et al., 2018
2018	Sauchinone	CYP3A4 (3UA1), CYP2B6 (3IBD), CYP2C19 (4GQS), CYP2E1 (3GPH)	AutoDock 4.2.6	Gong et al., 2018
2018	Sulfaphenazole, Chondroitin disaccharide $\Delta$ di-4S (C4S), Glucosamine 3-sulfate, Glucosamine 6-sulfate, Diacerein, Rhein	CYP2C9 (1R9O)	Discovery Studio 4.0 (CDOCKER)	Tan et al., 2018

*Superscript values denote that the metabolizing enzymes in the second column correspond to the query molecules in the first column, respectively.*

characteristics of young adult and elderly epilepsy patients, it was concluded that the bioavailability of LTG was not affected by age (Polepally et al., 2018). However, LTG CL in the elderly was 27.2% lower than in young epilepsy patients. These findings are very useful for clinicians to offer optimal epilepsy care and support to elderly patients starting low-dose treatment (Polepally et al., 2018).

GastroPlus and SimCYP are the mainstream PBPK emulation software programs. In recent years, applications have shown

a growing trend in the use of these software programs along with the improvement of software functions. For example, Ye et al. (2018) constructed an absorption model using GastroPlus to predict the potential effects of different gastric pH levels on the pharmacokinetics of lanabecestat and found that changes in gastric pH had a minimal influence on clinical exposure to lanabecestat. They also compared the bioavailability of two tablet formulations and an oral solution. The results showed that the 90% confidence intervals for geometric mean



**TABLE 9 |** Applications of QSAR in predicting ADMET properties from 2016 to 2018.

Year	Tool/Method	Compound	Properties	References
2016	ACD/I-lab	Angiogenic inhibitor for brain tumor	logS, logP, logD, logBB, hERG inhibition, HBA, HBD, MW	Khan et al., 2016
2016	QikProp 4.6	N-pyridyl, Pyrimidine benzamides	logS, logBB, MDCK, logKP, metab, CNS, loghERG, HOA	Malik et al., 2016
2017	ACD ChemSketch	Flavonoids	MW, IC50, Index of refraction, Surface tension, Density, Polarity, logP	Das et al., 2017
2017	CoMFA model	CITCO, $\alpha$ -naphtholphthalein, diethylstilbestrol, TPP, phenytoin, (R)-ethotoin, (S)-ethotoin	Drug-drug interactions, logP	Kato et al., 2017
2017	Discovery Studio v3.5	Novel dibenzofuran derivatives	PSA, Solubility, HIA, Cytochrome P450 2D6, BBB, PPB, Hepatotoxicity	Ma et al., 2017
2017	Self-organizing molecular field analysis (SOMFA)	Curcumin analogs	MW, IC50, logP	Verma and Thareja, 2017
2018	CoMFA and CoMSIA	Scopoletin Phenolic Ether Derivatives	LC50, TPSA	Luo et al., 2018
2018	MLR, SYBYLX v1.3, CoMFA and CoMSIA	Amyloid $\beta$ aggregation inhibitors	IC50	Aswathy et al., 2018
2018	Quantum Mechanics/ Molecular Mechanics (QM/MM)	DCHA inhibitors	IC50	Kollar and Freceer, 2018
2018	QikProp 4.6	TGF-type I inhibitors	IC50, BBB, logS	Ajay Kumar et al., 2018

LogP, octanol-water partition coefficient; LogD, octanol-water distribution coefficient; logS, aqueous solubility; logBB, blood-brain barrier; hERG, human ether-a-go-go related gene block; HOA, human oral absorption; PPB, plasma protein binding; MW, Molecular weight; LC50, median lethal concentration; TPSA, topological polar surface area; PSA, polar surface area; IC50, half maximal inhibitory concentration; HIA, human intestinal absorption; MDCK, madin Darby canine kidney cell line.

ratios were within standard bioequivalence boundaries for all other pharmacokinetic parameters, indicating that both tablet formulations were located in the accepted bioequivalence criteria compared with the oral solution (Ye et al., 2018). Boland et al. (2018) used SIMCYP to generate a dose-concentration model by using data from different genders, ages, and oral morphine formulations. The model was then validated against clinical pharmacokinetics data and used to calculate the association of the morphine dose with the plasma concentration. Finally, the analysis showed that older age, female sex, modified-release formulation, and inferior renal function were related to higher plasma concentrations (Boland et al., 2018). This result can help clinicians provide personalized prescription decisions.

## ADMET

We listed ADMET prediction applications by using AdmetSAR, SwissADME, Lazar, or Toxtree in **Table 11**. For example, Petrescu et al. (2019) used the AdmetSAR computational program to study the cytotoxicity of 15 phenolic compounds. The results showed that these compounds were much less toxic to aquatic life than synthetic pesticides. Roman et al. (2018) used SwissADME, FAFDrugs4, and admetSAR to predict the ADMET profiles and pharmacokinetics of 31 anabolic and androgen steroids in humans. The results revealed that the investigated steroids showed high gastrointestinal absorption and good oral bioavailability, which may be useful in the inhibition of human cytochromes associated with the metabolism of xenobiotics. In addition, the side effects of the studied steroids in humans were also predicted. Silva et al. (2019) predicted the theoretical toxicity of fluconazole (FNZ) by using Lazar to study the toxicity profile of FNZ toward human peripheral blood mononuclear

cells (PBMCs) cultured *in vitro*. The results showed that FNZ had potential mutagenic, tumorigenic, stimulating, and carcinogenic effects (Silva et al., 2019). Zhuang J. et al. (2018) used Toxtree (version 2.6.13) to evaluate the toxicity of extractables from multilayer coextrusion bags, and their prediction results revealed the types of extractables as well as the bioaccumulation factor and mutagenicity.

## DISCUSSION

### Deficiencies in Current *in silico* Methods

Each *in silico* method has its own characteristics and application scope. Hence, we need to select the most proper method for more accurate prediction. However, some methods have obvious deficiencies that may affect the prediction results. For example, molecular modeling plays a major role only in predicting metabolism and can assess only the possible interactions between compounds and metabolic enzymes; it cannot explicitly evaluate the ADMET risks of candidate compounds. The scoring function in molecular docking also affects the accuracy of ADMET prediction (de Graaf et al., 2006; Zhou et al., 2006). For instance, Kemp et al. (2004) and de Graaf et al. (2006) used different scoring functions to evaluate the binding affinities between cytochrome P450 and its inhibitors or substrates, respectively. Kemp et al. (2004) docked 33 compounds to P450, and the results revealed a correlation coefficient of  $R^2 = 0.61$  between the docking scores and active compounds. The docking scores were only able to identify several compounds as CYP2D6 inhibitors. Although de Graaf et al. (2006) integrated six scoring functions to identify the substrates of P450, the highest predicted accuracy (GOLD-Chemscore) identified 60% known substrates in the

**TABLE 10 |** Applications of PBPK in predicting ADMET properties from 2016 to 2018.

Year	Tool	Compound /Preparations	Properties	References
2016	GastroPlus	Alectinib	Oral bioavailability, Aqueous solubility, $C_{max}$ , AUC	Parrott et al., 2016
2016	GastroPlus	Bisoprolol, Nifedipine, Cimetidine, Furosemide	Oral absorption	Hansmann et al., 2016
2016	GastroPlus	Carvedilol loaded nanocapsules (CLN)	AUC, $C_{max}$	George et al., 2016
2016	GastroPlus	Ketoconazole, Erythromycin	AUC, $C_{max}$	Boetsch et al., 2016
2016	GastroPlus	Levofloxacin	logP, Plasma protein binding, AUC, $C_{max}$	Zhu et al., 2016
2016	GastroPlus	Met XR 1000 mg tablets	Dose, logD, AUC, $C_{max}$	Chen W. et al., 2016
2016	PKSim	Azathioprine	Hepatotoxicity	Thiel et al., 2016
2016	PKSim	Morphine and Furosemide	Vss value	Schlender et al., 2016
2016	WinNonlin	6-chloro-9-nitro-5-oxo-5H-benzo-(a)-phenoxazine (CNOB)	AUC, $V_d$ , CL, half-life	Wang J. H. et al., 2016
2016	WinNonlin	Arbekacin	$V_d$	Hagihara et al., 2016
2016	WinNonlin	Busulfan (BU)	AUC, apparent clearance	de Castro et al., 2016
2016	WinNonlin	Cetuximab (CTX), Capecitabine (CCB)	$C_{max}$ , AUC, $V_d$	Rachar et al., 2016
2016	WinNonlin	Glibenclamide	CL, $V_d$ , half-life	Rambiritch et al., 2016
2016	WinNonlin	Imatinib mesylate	Bioavailability, $T_{max}$ , $C_{max}$	Arora et al., 2016
2016	WinNonlin	Midazolam	$C_{max}$ , $V_d$	Vuu et al., 2016
2016	WinNonlin	Midazolam, Irinotecan	Clearance rate	Lee et al., 2016
2016	WinNonlin	Perfluorooctanoic acid (PFOA), Perfluorooctanesulfonic acid (PFOS), Perfluorohexane sulfonic acid (PFHxS)	AUC, renal clearance, $C_{max}$ , $V_d$	Kim et al., 2016
2016	WinNonlin	Propofol	Plasma concentrations, $V_d$	Chen J. Y. et al., 2016
2016	WinNonlin	Sodium succinate, Polysorbate, Arginine, Phosphate-buffered saline (PBS)	CL, Half-life, $C_{max}$	Gupta et al., 2016
2016	WinNonlin	Tacrine hydrochloride	Skin penetration	Patel et al., 2016
2016	WinNonlin	Tilmicosin	AUC, $C_{max}$	Zhang et al., 2016
2016	WinNonlin	Treosulfan (Treo)	$C_{max}$ , AUC	Romanski et al., 2016
2017	WinNonlin	Tacrolimus	Plasma clearance	David-Neto et al., 2017
2017	GastroPlus	Basimisanil	logD, Solubility	Yang et al., 2017
2017	GastroPlus	Buagafuran	Plasma Protein Binding, logP	Yang et al., 2017
2017	GastroPlus	Compound A (CPD A)	Bioavailability, $C_{max}$ , AUC	Stillhart et al., 2017
2017	GastroPlus	Mangiferin	Aqueous solubility, logD, logP, Permeability	Khurana et al., 2017
2017	PKSim	Cefazolin, Cefuroxime, Cefradine	CL	Dallmann et al., 2017
2017	PKSim	Endogenous IgG	CL, $V_d$	Niederalt et al., 2017
2017	PKSim	Fentanyl, Alfentanil, Thiopental, Omadacycline, Amiodarone, Propylthiouracil	Plasma concentration	Pilari et al., 2017
2017	PKSim	Vorinostat	Dose	Moj et al., 2017
2017	WinNonlin	Acetylkritasamycin	$C_{max}$ , $T_{max}$ , AUC	Nan et al., 2017
2017	WinNonlin	Benznidazole	$C_{max}$ , AUC	Molina et al., 2017
2017	WinNonlin	Ceftiofur	CL, $V_d$	Wang J. et al., 2017
2017	WinNonlin	Cloxacillin	Dose	Burmanczuk et al., 2017
2017	WinNonlin	Danofloxacin	AUC	Zhang N. et al., 2017
2017	WinNonlin	Diaveridine	Oral bioavailability, $C_{max}$ , AUC	Li Y. F. et al., 2017
2017	WinNonlin	Enrofloxacin	Plasma concentration	Shan et al., 2017
2017	WinNonlin	Iohexol	CL	Zhang C. et al., 2017
2017	WinNonlin	Lurasidone	$C_{max}$ , AUC	Hu et al., 2017

(Continued)

TABLE 10 | Continued

Year	Tool	Compound /Preparations	Properties	References
2017	WinNonlin	Meropenem	Plasma concentration	Kong et al., 2017
2017	WinNonlin	Metolazone	AUC, $C_{max}$ , $T_{max}$	Li X. et al., 2017
2017	WinNonlin	Psilocybin	Nephrotoxicity	Brown et al., 2017
2017	WinNonlin	Pyrazinamide	$C_{max}$	Voelkner et al., 2017
2017	WinNonlin	Sarafloxacin	Dose	Yu et al., 2017
2017	WinNonlin	Sildenafil	CL, $T_{max}$ , $V_d$	Olguin et al., 2017
2017	WinNonlin	Tenofovir	CL, Plasma concentration	Du X. et al., 2017
2017	WinNonlin	Tilmicosin	AUC, $T_{max}$ , half-life	Zhang L. et al., 2017
2017	WinNonlin	Treosulfan	Liver, Brain toxicity	Romanski et al., 2017
2017	WinNonlin	Tulathromycin	Dose	Zhou et al., 2017
2018	GastroPlus	Cefadroxil	Permeability, logP, Aqueous solubility, Distribution volume	Hu and Smith, 2018
2018	GastroPlus	Compound-A	Bioavailability, logP, Permeability, Aqueous solubility	Kou et al., 2018
2018	GastroPlus	Dasatinib	AUC, $C_{max}$ , $T_{max}$	Vaidhyanathan et al., 2018
2018	GastroPlus	DPP-4 inhibitors	Bioavailability	Daga et al., 2018
2018	GastroPlus	Lanabecestat (AZD3293)	Bioavailability	Ye et al., 2018
2018	GastroPlus	Tramadol	Renal clearance	T'Jollyn et al., 2018
2018	PKSim	Carvedilol	Plasma concentration	Ibarra et al., 2018
2018	PKSim	Escitalopram	Plasma exposure	Delaney et al., 2018
2018	PKSim	Indomethacin, Felodipine	Plasma exposure	Keemink et al., 2018
2018	PKSim	Pregabalin	Plasma concentrations	Idkaidek et al., 2018b
2018	WinNonlin	Cefquinome	$C_{max}$ , AUC	Shan et al., 2018
2018	WinNonlin	Dexmedetomidine	CL, $V_d$	Song et al., 2018
2018	WinNonlin	Gentamicin	AUC	Gestrich et al., 2018
2018	WinNonlin	Gliclazide	$V_d$ , CL	Shaik et al., 2018
2018	WinNonlin	Letrozole	Dose	Arora et al., 2018
2018	WinNonlin	Moxidectin	CL, $T_{max}$ , $C_{max}$	Xiao et al., 2018
2018	WinNonlin	Penicillin G	$V_d$ , CL, half-life	Padari et al., 2018
2018	WinNonlin	Seroquel XR, Quesero XR	AUC, $C_{max}$ , CL	Huang et al., 2018
2018	WinNonlin	Sitagliptin	$C_{max}$ , $T_{max}$ , AUC, $T_{1/2}$	Sangle et al., 2018
2018	WinNonlin	Vitacoxib	Plasma concentrations, $C_{max}$ , AUC	Wang et al., 2018a
2018	WinNonlin	Amoxicillin/clavulanic acid tablets	$C_{max}$ , $T_{max}$ , AUC, $T_{1/2}$	De Velde et al., 2018
2018	WinNonlin	Phenylbutyric acid (PBA), phenylacetic acid (PAA), and phenylacetylglutamine (PAGN), UPAGN	Plasma concentrations, AUC, $T_{max}$	Berry et al., 2018
2018	WinNonlin	Siponimod	$T_{1/2}$	Jin et al., 2018
2018	WinNonlin	Lacosamide	$T_{1/2}$ , $V_d$ , $C_{max}$ , CL, AUC	Franquiz et al., 2018
2018	WinNonlin	Tildipirosin (TD)	$C_{max}$ , AUC, $T_{max}$	Wang et al., 2018b
2018	WinNonlin	Streptomycin	$T_{1/2}$ , AUC, $T_{max}$ , $C_{max}$	Chen and Gao, 2018
2018	WinNonlin	Pantoprazole	$T_{1/2}$ , AUC, $V_d/F$ , CL/F	Shakhnovich et al., 2018
2018	WinNonlin	FVIII concentrates	$T_{1/2}$	Cheng X. et al., 2018
2018	WinNonlin	Adherence to tenofovir disoproxil fumarate/emtricitabine (TDF/FTC)	$C_{max}$ , AUC	Ibrahim et al., 2018
2018	WinNonlin	Valsartan, Hydrochlorothiazide (HCT)	$C_{max}$ , $T_{max}$ , AUC, $T_{1/2}$ , $K_{el}$	Idkaidek et al., 2018a
2018	WinNonlin	Individual total serum cortisol, unbound serum cortisol and salivary cortisone	$C_{max}$ , AUC, $T_{max}$ , Bioavailability	Johnson et al., 2018

LogP, octanol-water partition coefficient; LogD, octanol-water distribution coefficient; AUC, area under the curve;  $C_{max}$ , peak concentration;  $V_{ss}$ , steady-state;  $V_d$ , volume of distribution;  $T_{1/2}$ , half-life; logP, octanol-water partition coefficient; logD, octanol-water distribution coefficient; CL, clearance; CL/F, body clearance corrected for bioavailability;  $V_d/F$ , volume of distribution corrected for bioavailability.  $K_{el}$ , elimination rate constant; F, bioavailability;  $T_{max}$ , time to maximum concentration.

**TABLE 11 |** Applications of predicting ADMET properties using software from 2016 to 2018.

Year	Tool	Compound/Preparations	Properties	References
2016	admetSAR	GSK-3 targeting ligands	Mutagenicity, BBB, HIA, Caco-2, MDCK, PPB, AMES test, carcinogenicity, rat acute toxicity, P-gp substrate/inhibitor probability	Nisha et al., 2016
2016	Lazar	A dataset of air toxins (332 chemicals), A subset of the gold carcinogenic potency database (480 chemicals)	Carcinogenicity, mutagenicity	Pradeep et al., 2016
2016	Lazar	The Schiff bases of Benzothiazol-2-ylamine, Thiazolo [5, 4-b] pyridin-2-ylamine	Max, daily dose, acute toxicity, LC50	Shukla et al., 2016
2016	ToxTree	Cinnamaldehyde, Eugenol	Biodegradability, genotoxicity, carcinogenicity, bioaccumulation, developmental toxicity, mutagenicity, LD50, LC50	Absalan et al., 2016
2016	ToxTree	PC-replacement products—the 48 substances chosen based on the publications of (Simoneau et al., 2012) and (Onghena et al., 2014, 2015)	Genotoxic carcinogenicity	Mertens et al., 2016
2016	ToxTree	Vaccine constituents	Carcinogenicity, mutagenicity, genotoxicity, LD50	White et al., 2016
2017	admetSAR	3-bromopyruvate, Dibromopyruvate (DBPA), Propionic acid (PA)	BBB, Human intestinal absorption, Caco-2 permeability, P-gp substrate, AMES toxicity, Acute oral toxicity, Acute toxicity, CYP450 substrate and inhibitor, hERG	Yadav et al., 2017
2017	admetSAR	Histone deacetylase (HDACs) inhibitors	LogS, Caco-2 permeability	Uba and Yelekci, 2017
2017	admetSAR	Vernonia anthelmintica (L.)	Bioavailability, HIA, Caco-2, metabolism CYP	Wang J. Y. et al., 2017
2017	Lazar	2-amino-6-methylpyridine, 6-heptenoic acid, 2-methylphenol	Carcinogenicity, mutagenicity	Frenzel et al., 2017
2017	Lazar, ToxTree	8 volatile organic compounds (VOC)	Carcinogenicity, mutagenicity	Guerra et al., 2017
2017	SwissADME	3,7-dimethyl-2,6-octadienal, 2-pentene-2-methyl	Physicochemical properties, lipophilicity, hydrophilicity	Simhadri Vsdna et al., 2017
2017	SwissADME	338 different chemical pesticides	Lipophilicity, TPSA, molar refractivity, BBB permeant, GI absorption	Chedik et al., 2017
2017	SwissADME	Ginger	GI absorption, BBB, skin permeability, P-gp substrate	Sanni and Fatoki, 2017
2017	SwissADME	Polyphenols	Lipophilicity, water solubility	Yugandhar et al., 2017
2017	SwissADME	Tributyltin (IV) complex carboxylic acid derivative	GI absorption, BBB, LogP, water solubility, GI, Caco-2 Cells, Ames Test	Waseem et al., 2017
2017	SwissADME	Xeronine	Lipophilicity, GI absorption, solubility, bioavailability	Sanni et al., 2017
2017	ToxTree	400 compounds	skin/eye irritation, corrosion	Verheyen et al., 2017
2017	ToxTree	80 commercially available chemicals (38 liquids and 42 solids)	Eye irritation, corrosion	Geerts et al., 2017
2018	Lazar	(-)-Asimilobine, Aloin, Anorectine, Chrysothron, Coptisine, Elymoclavine, Thaliminine	Genotoxicity, carcinogenicity, mutagenicity	Gluck et al., 2018
2018	Lazar	Newly proposed heterocyclic derivatives	Carcinogenicity, mutagenicity	Azad et al., 2018
2018	Lazar	The synthesized 1,3,5-trisubstituted-2-pyrazoline derivatives (5a-5t)	Maximum recommended daily dose, reproductive toxicity, carcinogenicity, Mutagenicity, Acute toxicity, LC50	Tripathi et al., 2018
2018	SwissADME	21 Organosilicone compounds	TPSA, logP, GI absorption, BBB	Shaaban et al., 2018
2018	SwissADME	Mycotoxins (DON,3-AcDON,15-AcDON)	HIA, BBB penetration, Mutagenicity, Carcinogenicity, Acute toxicity	Taroncher et al., 2018
2018	SwissADME	NAZ2329	Lipophilicity, Water Solubility, GI absorption, BBB permeant	Agoni et al., 2018
2018	SwissADME, admetSAR, ToxTree	31 anabolic hormones and androgen hormones	Gastrointestinal absorption, Blood brain barrier, P-gp substrate, Skin permeability, Carcinogenicity, hERG, Ames toxicity	Roman et al., 2018
2018	ToxTree	48 selected sensitizing and non-sensitizing AS	Skin sensitization	Braeuning et al., 2018

(Continued)



TABLE 11 | Continued

Year	Tool	Compound/Preparations	Properties	References
2018	ToxTree	Bis [2,4-bis(2-methyl-2-propanyl) phenyl] hydrogen phosphate	LD50, Bioaccumulation factor, Mutagenicity	Zhuang J. et al., 2018
2018	ToxTree	Printed paper and board FCM substances	Ames/bacterial mutagenicity	Van Bossuyt et al., 2018
2018	SwissADME, ToxTree	Dominant phytochemicals from <i>Rheum palmatum</i> , <i>Rubus coreanus</i> and <i>Sanguisorba officinalis</i>	logP, logD, TPSA, log S, GI absorption, CYP450 isoforms inhibitor probability, Genotoxicity	Nosrati et al., 2018
2018	SwissADME	8 mPGES-1 binders	PAINS	Lauro et al., 2018
2018	SwissADME	A new series of synthesized quinazoline derivatives	Molecular weight, logP, HBA, HBD, TPSA, Lipinski's RO5, Leadlikeness	Nasab et al., 2018
2018	SwissADME	107 Compounds containing biaryl scaffold	Molecular weight, logP, HBA, HBD, Lipinski's RO5, TPSA, BBB, GI absorption	Khalid et al., 2018
2018	SwissADME	Genistein, Daidzein and Glycitein	Lipinski's RO5, TPSA, Num. rotatable bonds	Shaji, 2018
2018	SwissADME	Four series of diphyllin-related compounds	LogP, PAINS	Lindstrom et al., 2018

BBB, blood brain barrier; P-gp, P glycoprotein; HIA, human intestinal absorption; MDCK, madin Darby canine kidney cell line; PPB, plasma protein binding; GI absorption, gastrointestinal absorption; TPSA, topological polar surface area; LogP, octanol-water partition coefficient; LogD, octanol-water distribution coefficient; logS, aqueous solubility; F, Bioavailability; hERG, human ether-a-go-go related gene block; PAINS, pan-assay interference structures; LC50, median lethal concentration; LD50, median lethal dose; RO5, rule of five.

top 5% results during the virtual screening. Only high-affinity CYP2D6 ligands could be predicted. Therefore, docking methods with scoring functions are mostly applied for coarse screening of a series of compounds.

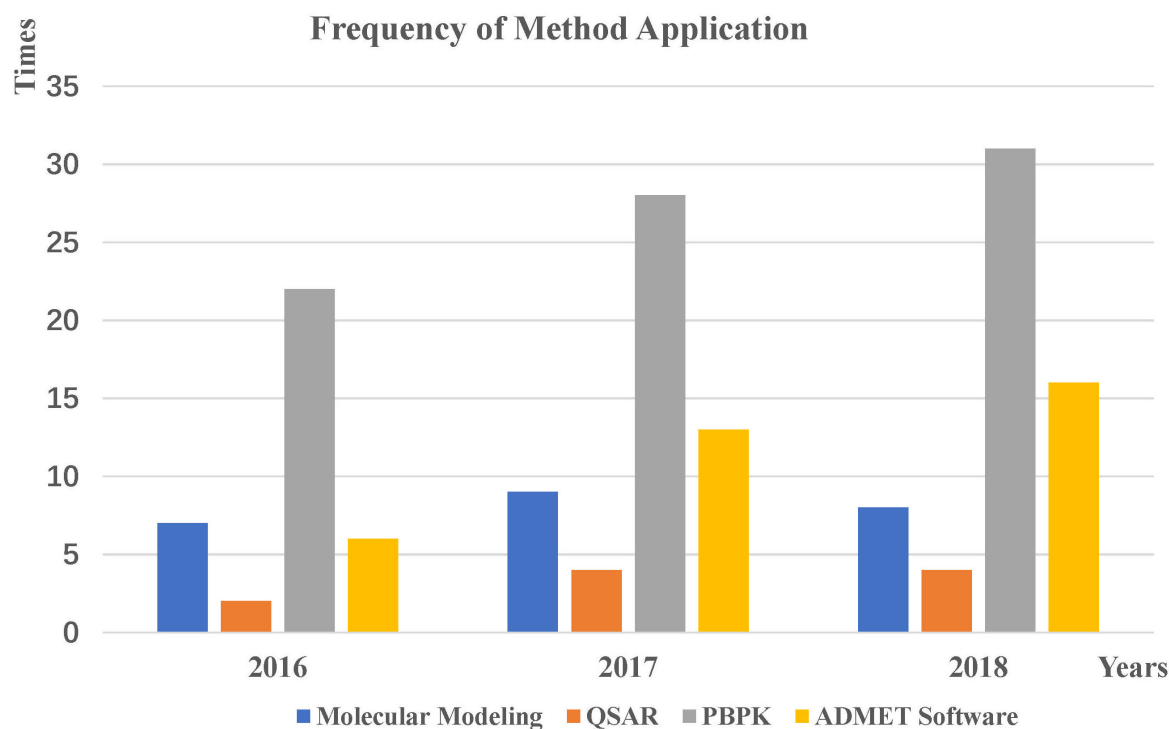
Compared with molecular modeling, data modeling can predict more properties, but its prediction accuracy depends on the quality and quantity of data. The QSAR method, the main strategy in data modeling, has limited value without an estimated model applicability domain for predicting biological or physicochemical properties (Sushko et al., 2010). The predictive ability of models will be limited if the predicted chemical is outside the chemical space where the models were developed (Sheridan et al., 2004). Furthermore, the descriptors used in the model construction for structural transformation are too simplistic and inadequate to predict the behavior of a drug in a whole organism. Therefore, there is a need to develop molecular descriptors containing more information. Many developed QSAR models have been validated only by internal validation without sufficient external validation, which is considered a necessary factor to build a reliable QSAR model (Roy et al., 2012). QSAR prediction is based on the principle that similar molecules have similar properties (Patterson et al., 1996), but in some special cases, such as CYP metabolism, similar molecules may have different activities, which are known as activity cliffs (Guha, 2011). Therefore, the ADMET properties of the compounds in the human body are not independent but are also affected by other factors. The PBPK method can predict multiple properties, but only provides common information about the biological behavior of organs or tissues. It is also limited by the mathematical form of the PBPK equations, which ignores the structural and physical properties of drug compounds (Huynh et al., 2009). Moreover, a large amount of experimental data is required when constructing models. Due to the lack of proper and easily accessible databases related to physiological properties, the data used to build models can be obtained only

from the literature (Rowland et al., 2004). However, the obtained data are relatively limited, which reduces the predictability of the models. Finally, although some PBPK software programs have been developed, most of them are commercial, and users must participate in the training of software companies to make them more useful (Lave et al., 2007).

The existing ADMET software can perform faster and more convenient predictions of multiple properties to obtain more comprehensive prediction results. However, we can see in practical applications that the software is applicable only for qualitative analysis of compounds and cannot accurately predict the quantitative values of some properties. Moreover, issues are observed in the data quality and quantity of these software. While more experimental data are needed to further optimize the software, integrating unconfirmed data into the software to predict new compounds will decrease the prediction accuracy.

## Comparison of the Applications of Three Methods in Predicting ADMET Properties

To compare the application trends of each method more intuitively, we counted the number of applications of these three methods from 2016 to 2018 (Figure 6). Applications using PBPK modeling software have exhibited an upward trend in recent years, which means that this approach will be the mainstream pharmacokinetic evaluation in the future. We believe that more user-friendly software will be developed, making the prediction process more convenient for users. Applications using ADMET software for forecasting are also on the rise. The main reason for these changes may be the demand for multi-property prediction and the phenomenon of drug recall, leading to the hope of predicting more pharmacokinetic properties in advance to reduce drug development costs. This increase also confirms that researchers have gradually integrated software predictions into the early stages of drug discovery to improve the success rate of drug development through complementarity



**FIGURE 6 |** Application trends for each method from 2016 to 2018.

and collaboration. In addition, we found that the QSAR method is rarely used alone, potentially due to the limitations of the QSAR method prediction. Therefore, we can increase or integrate the data volume of some specific QSAR models to generate a more comprehensive prediction model. Researchers can also combine different modeling methods and then perform relevant predictions. If a combined model can predict the properties well, then it can be used as a consensus approach to improve ADMET prediction accuracy.

## Previous Review and Prospective Studies of ADMET and PBPK Simulations

To date, there have been many reviews on ADMET and PBPK, which provide summaries from different perspectives. We found 11 articles that are closely related to this review after careful retrieval of the published literature. We have classified these articles into three categories based on the perspective of the descriptions: (1) directly related to our review, (2) machine learning (ML) methods, and (3) advances in *in silico* ADMET modeling.

Five reviews are directly related to this article. Yamashita and Hashida (2004) reviewed the application of structure-based methods and QSAR methods in predicting ADMET in detail. Cheng et al. (2013) introduced the recent progress and current challenges of QSAR in predicting ADMET and then discussed several new promising research directions that could be employed for systemic *in silico* ADMET prediction. Alqahtani

(2017) reviewed the *in silico* models for predicting the ADMET properties of compounds and provided a comprehensive overview of the latest modeling methods and algorithms, as well as the application prospects of PBPK in predicting pharmacokinetics. Wang et al. (2015) briefly introduced the development of *in silico* models for ADMET prediction. They also focused on the modeling approaches, related applications, and potential advantages or disadvantages of these models used in drug discovery. Wishart (2007) highlighted ADMET property prediction, as well as ADME-related databases and software, and briefly introduced the application of PBPK and related software in ADMET.

Three out of 11 reviews are about machine learning methods for predicting ADMET. Ferreira and Andricopulo (2019) provided a detailed description of current machine learning approaches to ADMET modeling, focusing on key advances from 2017 to 2018. Tao et al. (2015) reviewed the progress of machine learning methods in ADMET prediction and discussed the performance, applications, and challenges in developing machine learning methods. Maltarollo et al. (2015) described the applications of ML methods in ADMET prediction.

The remaining three reviews addressed the progress of ADMET modeling. Specifically, Lin et al. (2003) introduced the desirable properties of new chemical entities (NCEs) from an ADMET perspective and discussed basic concepts, important tools, reagents, and experimental approaches used by researchers in predicting human pharmacokinetics. van de Waterbeemd and Gifford (2003) summarized the endpoints of pharmacokinetics,

metabolism, and toxicity. Wang and Hou (2009) introduced the properties of ADMET, discussed the latest corresponding *in silico* models, and provided a brief summary of some software and databases. These articles introduced ADMET, PBPK, and related research progress from different perspectives.

## CONCLUSION

In this review, we provide a detailed and comprehensive introduction to currently used approaches or tools in predicting ADMET properties, including the basic principles, classification, and applications. In addition, we collect related applications from published articles over the past 3 years and analyze the trends in these applications. The purpose of this review is to help readers quickly understand these approaches and the characteristics of the related tools (databases and software). It may also provide readers with a better understanding of how existing tools can be applied to pharmacokinetic predictions. We are convinced that more accurate predictions due to users' familiarity with existing online services will increase the importance of *in silico* ADMET prediction in pharmacokinetics. In the future, we expect not only a reduced failure rate in drug development and drug recalls but

also a faster timeline from R&D to market, as well as decreased costs during the late stage of development.

## AUTHOR CONTRIBUTIONS

ZH, FW, and YZ contributed to the design and conception of the study. LL, YZ, FW, XS, GC, XW, XL, and MT performed information retrieval and analysis. FW, YZ, LL, and XS wrote the manuscript. YZ, FW, LL, XS, GC, XW, XL, and MT created the tables and figures. ZH and FW guided the manuscript writing and revised the manuscript. ZH provided financial support. All authors contributed to manuscript revision and have read and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (31770774), the Provincial Major Project of Basic or Applied Research in Natural Science, the Guangdong Provincial Education Department (2016KZDXM038), and the Higher Education Reform Project of Guangdong Province (2019268).

## REFERENCES

- Abdelmonsef, A. H., Dulapalli, R., Dasari, T., Padmarao, L. S., Mukkera, T., and Vuruputuri, U. (2016). Identification of novel antagonists for rab38 protein by homology modeling and virtual screening. *Comb. Chem. High Throughput Screen.* 19, 875–892. doi: 10.2174/1386207319666161026153237
- Abid, S. M. A., Younus, H. A., Al-Rashida, M., Arshad, Z., Maryum, T., Gilani, M. A., et al. (2017). Sulfonyl hydrazones derived from 3-formylchromone as non-selective inhibitors of MAO-A and MAO-B: synthesis, molecular modelling and *in-silico* ADME evaluation. *Bioorg. Chem.* 75, 291–302. doi: 10.1016/j.bioorg.2017.10.001
- Absalan, A., Mesbah-Namin, S. A., Tiraihi, T., and Taheri, T. (2016). The effects of cinnamaldehyde and eugenol on human adipose-derived mesenchymal stem cells viability, growth and differentiation: a cheminformatics and *in vitro* study. *Avicenna J. Phytomed.* 6, 643–657.
- Agoni, C., Ramharack, P., and Soliman, M. E. S. (2018). Allosteric inhibition induces an open WPD-loop: a new avenue towards glioblastoma therapy. *RSC Adv.* 8, 40187–40197. doi: 10.1039/C8RA08427K
- Ajay Kumar, T. V., Athavan, A. A. S., Loganathan, C., Saravanan, K., Kabilan, S., and Parthasarathy, V. (2018). Design, 3D QSAR modeling and docking of TGF-beta type I inhibitors to target cancer. *Comput. Biol. Chem.* 76, 232–244. doi: 10.1016/j.compbiolchem.2018.07.011
- Alarn, S., and Khan, F. (2019). 3D-QSAR, docking, ADME/Tox studies on flavone analogs reveal anticancer activity through Tankyrase inhibition. *Sci. Rep.* 9:5414. doi: 10.1038/s41598-019-41984-7
- Alqahtani, S. (2017). *In silico* ADME-Tox modeling: progress and prospects. *Expert Opin. Drug Metab. Toxicol.* 13, 1147–1158. doi: 10.1080/17425255.2017.1389897
- Andrade, C. H., Silva, D. C., and Braga, R. C. (2014). *In silico* prediction of drug metabolism by P450. *Curr. Drug Metab.* 15, 514–525. doi: 10.2174/1389200215666140908102530
- Arora, P., Adams, C. H., Gudelsky, G., DasGupta, B., and Desai, P. B. (2018). Plasma and brain pharmacokinetics of letrozole and drug interaction studies with temozolomide in NOD-scid gamma mice and sprague dawley rats. *Cancer Chemother. Pharmacol.* 83, 81–89. doi: 10.1007/s00280-018-3705-6
- Arora, R., Sharma, M., Monif, T., and Iyer, S. (2016). A multi-centric bioequivalence trial in Ph+ chronic myeloid leukemia patients to assess bioequivalence and safety evaluation of generic imatinib mesylate 400 mg tablets. *Cancer Res. Treat.* 48, 1120–1129. doi: 10.4143/crt.2015.436
- Aswathy, L., Jisha, R. S., Masand, V. H., Gajbhiye, J. M., and Shibi, I. G. (2018). Design of novel amyloid beta aggregation inhibitors using QSAR, pharmacophore modeling, molecular docking and ADME prediction. *In Silico Pharmacol.* 6:12. doi: 10.1007/s40203-018-0049-1
- Azad, I., Nasibullah, M., Khan, T., Hassan, F., and Akhter, Y. (2018). Exploring the novel heterocyclic derivatives as lead molecules for design and development of potent anticancer agents. *J. Mol. Graph. Model.* 81, 211–228. doi: 10.1016/j.jmgm.2018.02.013
- Basheer, L., Schultz, K., Guttman, Y., and Kerem, Z. (2017). *In silico* and *in vitro* inhibition of cytochrome P450 3A by synthetic stilbenoids. *Food Chem.* 237, 895–903. doi: 10.1016/j.foodchem.2017.06.040
- Basheer, L., Schultz, K., and Kerem, Z. (2016). Inhibition of cytochrome P450 3A by acetoxyated analogues of resveratrol in *in vitro* and *in silico* models. *Sci. Rep.* 6:31557. doi: 10.1038/srep31557
- Belekar, V., Lingineni, K., and Garg, P. (2015). Classification of breast cancer resistant protein (BCRP) inhibitors and non-inhibitors using machine learning approaches. *Comb. Chem. High Throughput Screen.* 18, 476–485. doi: 10.2174/1386207318666150525094503
- Berry, S. A., Vockley, J., Vinks, A. A., Dong, M., Diaz, G. A., McCandless, S. E., et al. (2018). Pharmacokinetics of glycerol phenylbutyrate in pediatric patients 2 months to 2 years of age with urea cycle disorders. *Mol. Genet. Metab.* 125, 251–257. doi: 10.1016/j.ymgme.2018.09.001
- Bhatia, S., Schultz, T., Roberts, D., Shen, J., Kromidas, L., and Marie Api, A. (2015). Comparison of cramer classification between toxtree, the OECD QSAR toolbox and expert judgment. *Regul. Toxicol. Pharmacol.* 71, 52–62. doi: 10.1016/j.yrtph.2014.11.005
- Bhatarai, B., Wilson, D. M., Parks, A. K., Carney, E. W., and Spencer, P. J. (2016). Evaluation of TOPKAT, toxtree, and derek nexus *in silico* models for ocular irritation and development of a knowledge-based framework to improve the prediction of severe irritation. *Chem. Res. Toxicol.* 29, 810–822. doi: 10.1021/acs.chemrestox.5b00531
- Boetsch, C., Parrott, N., Fowler, S., Poirier, A., Hainzl, D., Banken, L., et al. (2016). Effects of cytochrome P450 3A4 inhibitors-ketoconazole and erythromycin-on bitopertin pharmacokinetics and comparison with physiologically based modelling predictions. *Clin. Pharmacokinet.* 55, 237–247. doi: 10.1007/s40262-015-0312-0

- Boland, J. W., Johnson, M., Ferreira, D., and Berry, D. J. (2018). *In silico* (computed) modelling of doses and dosing regimens associated with morphine levels above international legal driving limits. *Palliat. Med.* 32, 1222–1232. doi: 10.1177/0269216318773956
- Bouzon, F., Ball, K., Perdaens, N., and Walther, B. (2012). Physiologically based pharmacokinetic (PBPK) modelling tools: how to fit with our needs? *Biopharm. Drug Dispos.* 33, 55–71. doi: 10.1002/bdd.1767
- Bouzon, F., and Walther, B. (2008). Pharmacokinetic predictions in children by using the physiologically based pharmacokinetic modelling. *Fundam. Clin. Pharmacol.* 22, 579–587. doi: 10.1111/j.1472-8206.2008.00648.x
- Bowen, J. P., and Guener, O. F. (2013). A perspective on quantum mechanics calculations in ADMET predictions. *Curr. Top. Med. Chem.* 13, 1257–1272. doi: 10.2174/15680266113139990032
- Braeuning, C., Braeuning, A., Mielke, H., Holzwarth, A., and Peiser, M. (2018). Evaluation and improvement of QSAR predictions of skin sensitization for pesticides. *SAR QSAR Environ. Res.* 29, 823–846. doi: 10.1080/1062936X.2018.1518261
- Brown, R. T., Nicholas, C. R., Cozzi, N. V., Gassman, M. C., Cooper, K. M., Muller, D., et al. (2017). Pharmacokinetics of escalating doses of oral psilocybin in healthy adults. *Clin. Pharmacokinet.* 56, 1543–1554. doi: 10.1007/s40262-017-0540-6
- Bujak, R., Struck-Lewicka, W., Kaliszan, M., Kaliszan, R., and Markuszewski, M. J. (2015). Blood-brain barrier permeability mechanisms in view of quantitative structure-activity relationships (QSAR). *J. Pharm. Biomed. Anal.* 108, 29–37. doi: 10.1016/j.jpba.2015.01.046
- Burmanczuk, A., Grabowski, T., Osypuk, M., Polska, B., and Kowalski, C. (2017). Determination of cloxacillin residues in dairy cows after intramammary administration. *J. Vet. Pharmacol. Ther.* 40, 552–560. doi: 10.1111/jvp.12391
- Byun, J. H., Han, D. G., Cho, H. J., Yoon, I. S., and Jung, I. H. (2020). Recent advances in physiologically based pharmacokinetic and pharmacodynamic models for anticancer nanomedicines. *Arch. Pharm. Res.* 43, 80–99. doi: 10.1007/s12272-020-01209-2
- Caldwell, G. W., Yan, Z., Tang, W., Dasgupta, M., and Hasting, B. (2009). ADME optimization and toxicity assessment in early- and late-phase drug discovery. *Curr. Top. Med. Chem.* 9, 965–980. doi: 10.2174/156802609789630929
- Cao, D., Wang, J., Zhou, R., Li, Y., Yu, H., and Hou, T. (2012). ADMET evaluation in drug discovery. 11. Pharmacokinetics knowledge base (PKKB): a comprehensive database of pharmacokinetic and toxic properties for drugs. *J. Chem. Inf. Model.* 52, 1132–1137. doi: 10.1021/ci300112j
- Chedik, L., Mias-Lucquin, D., Bruyere, A., and Fardel, O. (2017). *In silico* prediction for intestinal absorption and brain penetration of chemical pesticides in humans. *Int. J. Environ. Res. Public Health* 14:708. doi: 10.3390/ijerph14070708
- Chen, J. Y., Yi, M., Yao, S. L., and Zhang, X. P. (2016). Propofol target-controlled infusion modeling in rabbits: pharmacokinetic and pharmacodynamic analysis. *J. Huazhong Univ. Sci. Technol. Med. Sci.* 36, 428–433. doi: 10.1007/s11596-016-1604-9
- Chen, L., and Gao, Z. (2018). Continuous hypoxia reduces the concentration of streptomycin in the blood. *BMC Infect. Dis.* 18:120. doi: 10.1186/s12879-018-3027-7
- Chen, W., Desai, D., Good, D., Crison, J., Timmins, P., Paruchuri, S., et al. (2016). Mathematical model-based accelerated development of extended-release metformin hydrochloride tablet formulation. *AAPS PharmSciTech.* 17, 1007–1013. doi: 10.1208/s12249-015-0423-9
- Chen, Y., Liu, Z. L., Fu, T. M., Li, W., Xu, X. L., and Sun, H. P. (2015). Discovery of new acetylcholinesterase inhibitors with small core structures through shape-based virtual screening. *Bioorg. Med. Chem. Lett.* 25, 3442–3446. doi: 10.1016/j.bmcl.2015.07.026
- Cheng, F., Li, W., Liu, G., and Tang, Y. (2013). *In silico* ADMET prediction: recent advances, current challenges and future trends. *Curr. Top. Med. Chem.* 13, 1273–1289. doi: 10.2174/15680266113139990033
- Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., et al. (2012). admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. *J. Chem. Inf. Model.* 52, 3099–3105. doi: 10.1021/ci300367a
- Cheng, P., Li, J., Wang, J., Zhang, X., and Zhai, H. (2018). Investigations of FAK inhibitors: a combination of 3D-QSAR, docking, and molecular dynamics simulations studies. *J. Biomol. Struct. Dyn.* 36, 1529–1549. doi: 10.1080/07391102.2017.1329095
- Cheng, T., Li, Q., Zhou, Z., Wang, Y., and Bryant, S. H. (2012). Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J.* 14, 133–141. doi: 10.1208/s12248-012-9322-0
- Cheng, X., Li, P., Chen, Z., Zhang, N., Zhen, Y., Zhao, L., et al. (2018). Break-through bleeding in relation to pharmacokinetics of factor VIII in paediatric patients with severe haemophilia A. *Haemophilia* 24, 120–125. doi: 10.1111/hae.13373
- Chow, E. C. Y., Talattof, A., Tsakalozou, E., Fan, J., Zhao, L., and Zhang, X. (2016). Using physiologically based pharmacokinetic (PBPK) modeling to evaluate the impact of pharmaceutical excipients on oral drug absorption: sensitivity analyses. *AAPS J.* 18, 1500–1511. doi: 10.1208/s12248-016-9964-4
- Cook, D., Brown, D., Alexander, R., March, R., Morgan, P., Satterthwaite, G., et al. (2014). Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* 13, 419–431. doi: 10.1038/nrd4309
- Cui, Q., and Karplus, M. (2003). Catalysis and specificity in enzymes: a study of triosephosphate isomerase and comparison with methyl glyoxal synthase. *Adv. Protein Chem.* 66, 315–372. doi: 10.1016/S0065-3233(03)66008-0
- Cumming, J. G., Davis, A. M., Muresan, S., Haeblerlein, M., and Chen, H. (2013). Chemical predictive modelling to improve compound quality. *Nat. Rev. Drug Discov.* 12, 948–962. doi: 10.1038/nrd4128
- Daga, P. R., Bolger, M. B., Haworth, I. S., Clark, R. D., and Martin, E. J. (2018). Physiologically based pharmacokinetic modeling in lead optimization. 1. Evaluation and adaptation of gastroplus to predict bioavailability of medchem series. *Mol. Pharm.* 15, 821–830. doi: 10.1021/acs.molpharmaceut.7b00972
- 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
- Dallmann, A., Ince, I., Solodenko, J., Meyer, M., Willmann, S., Eissing, T., et al. (2017). Physiologically based pharmacokinetic modeling of renally cleared drugs in pregnant women. *Clin. Pharmacokinet.* 56, 1525–1541. doi: 10.1007/s40262-017-0538-0
- Danishuddin, and Khan, A. U., (2016). Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discov. Today* 21, 1291–1302. doi: 10.1016/j.drudis.2016.06.013
- Das, S., Laskar, M. A., Sarker, S. D., Choudhury, M. D., Choudhury, P. R., Mitra, A., et al. (2017). Prediction of anti-alzheimer's activity of flavonoids targeting acetylcholinesterase *in silico*. *Phytochem. Anal.* 28, 324–331. doi: 10.1002/pca.2679
- David-Neto, E., Romano, P., Kamada Triboni, A. H., Ramos, F., Agena, F., Almeida Rezende Ebner, P., et al. (2017). Longitudinal pharmacokinetics of tacrolimus in elderly compared with younger recipients in the first 6 months after renal transplantation. *Transplantation* 101, 1365–1372. doi: 10.1097/TP.0000000000001369
- Davis, A. P., Grondin, C. J., Johnson, R. J., Siciak, D., McMorran, R., Wiegiers, J., et al. (2019). The comparative toxicogenomics database: update 2019. *Nucleic Acids Res.* 47, D948–D954. doi: 10.1093/nar/gky868
- De Buck, S., and Mackie, C. E. (2007). Physiologically based approaches towards the prediction of pharmacokinetics: *in vitro-in vivo* extrapolation. *Expert Opin. Drug Metab. Toxicol.* 3, 865–878. doi: 10.1517/17425255.3.6.865
- de Castro, F. A., Simoes, B. P., Godoy, A. L., Bertagnoli Trigo, F. M., Coelho, E. B., and Lanchote, V. L. (2016). Use of an oral busulfan test dose in patients undergoing hematopoietic stem cell transplantation treated with or without fludarabine. *J. Clin. Pharmacol.* 56, 1555–1562. doi: 10.1002/jcph.758
- de Graaf, C., Oostenbrink, C., Keizers, P. H., van der Wijst, T., Jongejan, A., and Vermeulen, N. P. (2006). Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* 49, 2417–2430. doi: 10.1021/jm0508538
- de Groot, M. J. (2006). Designing better drugs: predicting cytochrome P450 metabolism. *Drug Discov. Today* 11, 601–606. doi: 10.1016/j.drudis.2006.05.001
- de Groot, M. J., Kirton, S. B., and Sutcliffe, M. J. (2004). *In silico* methods for predicting ligand binding determinants of cytochromes P450. *Curr. Top. Med. Chem.* 4, 1803–1824. doi: 10.2174/1568026043387061
- De Velde, F., De Winter, B. C. M., Koch, B. C. P., Van Gelder, T., Mouton, J. W., and COMBACTE-NET consortium (2018). Highly variable absorption of clavulanic acid during the day: a population pharmacokinetic analysis. *J. Antimicrob. Chemother.* 73, 469–476. doi: 10.1093/jac/dkx376



- Dearden, J. C. (2007). *In silico* prediction of ADMET properties: how far have we come? *Expert Opin. Drug Metab. Toxicol.* 3, 635–639. doi: 10.1517/17425255.3.5.635
- Delaney, S. R., Malik, P. R. V., Stefan, C., Edgington, A. N., Colantonio, D. A., and Ito, S. (2018). Predicting escitalopram exposure to breastfeeding infants: integrating analytical and *in silico* techniques. *Clin. Pharmacokinet.* 57, 1603–1611. doi: 10.1007/s40262-018-0657-2
- Dickins, M., and van de Waterbeemd, H. (2004). Simulation models for drug disposition and drug interactions. *Drug Discov. Today BIOSILICO* 2, 38–45. doi: 10.1016/S1741-8364(04)02388-1
- Dong, J., Wang, N. N., Yao, Z. J., Zhang, L., Cheng, Y., Ouyang, D., et al. (2018). ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J. Cheminform.* 10:29. doi: 10.1186/s13321-018-0283-x
- Doss, C. G. P., Chakraborty, C., Narayan, V., and Kumar, D. T. (2014). “Computational approaches and resources in single amino acid substitutions analysis toward clinical research,” in *Advances in Protein Chemistry and Structural Biology*, Vol. 94, ed R. Donev (San Diego, CA: Elsevier Academic Press), 365–423. doi: 10.1016/B978-0-12-800168-4.00010-X
- Du, H., Li, J., Cai, Y., Zhang, H., Liu, G., Tang, Y., et al. (2017). Computational investigation of ligand binding to the peripheral site in CYP3A4: conformational dynamics and inhibitor discovery. *J. Chem. Inf. Model.* 57, 616–626. doi: 10.1021/acs.jcim.7b00012
- Du, X., Kou, H., Fu, Q., Li, Y., Zhu, Z., and Li, T. (2017). Steady-state pharmacokinetics of tenofovir disoproxil fumarate in human immunodeficiency virus-infected Chinese patients. *Expert Rev. Clin. Pharmacol.* 10, 783–788. doi: 10.1080/17512433.2017.1321480
- Dutour, R., Cortes-Benitez, F., Roy, J., and Poirier, D. (2017). Structure-based design and synthesis of new estrane-pyridine derivatives as cytochrome P450 (CYP) 1B1 inhibitors. *ACS Med. Chem. Lett.* 8, 1159–1164. doi: 10.1021/acsmedchemlett.7b00265
- Edgington, A. N., Theil, F. P., Schmitt, W., and Willmann, S. (2008). Whole body physiologically-based pharmacokinetic models: their use in clinical drug development. *Expert Opin. Drug Metab. Toxicol.* 4, 1143–1152. doi: 10.1517/17425255.4.9.1143
- El-Zahabi, H. S. A., Khalifa, M. M. A., Gado, Y. M. H., Farrag, A. M., Elaasser, M. M., Safwat, N. A., et al. (2019). New thiobarbituric acid scaffold-based small molecules: synthesis, cytotoxicity, 2D-QSAR, pharmacophore modelling and *in-silico* ADME screening. *Eur. J. Pharm. Sci.* 130, 124–136. doi: 10.1016/j.ejps.2019.01.023
- Espie, P., Tytgat, D., Sargentini-Maier, M. L., Poggesi, I., and Watelet, J. B. (2009). Physiologically based pharmacokinetics (PBPK). *Drug Metab. Rev.* 41, 391–407. doi: 10.1080/10837450902891360
- Evangelista, W., Weir, R. L., Ellingson, S. R., Harris, J. B., Kapoor, K., Smith, J. C., et al. (2016). Ensemble-based docking: from hit discovery to metabolism and toxicity predictions. *Bioorg. Med. Chem.* 24, 4928–4935. doi: 10.1016/j.bmc.2016.07.064
- Fan, J., and de Lannoy, I. A. (2014). Pharmacokinetics. *Biochem. Pharmacol.* 87, 93–120. doi: 10.1016/j.bcp.2013.09.007
- Ferreira, L. L. G., and Andricopulo, A. D. (2019). ADMET modeling approaches in drug discovery. *Drug Discov. Today* 24, 1157–1165. doi: 10.1016/j.drudis.2019.03.015
- Franquiz, M. J., Kalaria, S. N., Armahizer, M. J., Gopalakrishnan, M., McCarthy, P. J., and Badjatia, N. (2018). Lacosamide pharmacokinetics in a critically ill patient receiving continuous venovenous hemofiltration. *Pharmacotherapy* 38, E17–E21. doi: 10.1002/phar.2063
- Fravef, F., and Benfenati, E. (2008). A combination of 3D-QSAR, docking, local-binding energy (LBE) and GRID study of the species differences in the carcinogenicity of benzene derivatives chemicals. *J. Mol. Graph. Model.* 27, 147–160. doi: 10.1016/j.jmgm.2008.04.004
- Frenzel, F., Buhrke, T., Wenzel, I., Andrack, J., Hielscher, J., and Lampen, A. (2017). Use of *in silico* models for prioritization of heat-induced food contaminants in mutagenicity and carcinogenicity testing. *Arch. Toxicol.* 91, 3157–3174. doi: 10.1007/s00204-016-1924-3
- Fu, L., Liu, L., Yang, Z. J., Li, P., Ding, J. J., Yun, Y. H., et al. (2020). Systematic modeling of logD7.4 based on ensemble machine learning, group contribution, and matched molecular pair analysis. *J. Chem. Inf. Model.* 60, 63–76. doi: 10.1021/acs.jcim.9b00718
- Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., et al. (2017). The ChEMBL database in 2017. *Nucleic Acids Res.* 45, D945–D954. doi: 10.1093/nar/gkw1074
- Geerts, L., Adriaens, E., Alepee, N., Guest, R., Willoughby, J. A. Sr., Kandarova, H., et al. (2017). CON4EI: evaluation of QSAR models for hazard identification and labelling of eye irritating chemicals. *Toxicol. In Vitro* 49, 90–98. doi: 10.1016/j.tiv.2017.09.004
- George, J. K., Singh, S. K., and Verma, P. R. (2016). *In vivo in silico* pharmacokinetic simulation studies of carvedilol-loaded nanocapsules using GastroPlus. *Ther. Deliv.* 7, 305–318. doi: 10.4155/tde-2015-0004
- Gestrich, A., Bedenice, D., Ceresia, M., and Zaghloul, I. (2018). Pharmacokinetics of intravenous gentamicin in healthy young-adult compared to aged alpacas. *J. Vet. Pharmacol. Ther.* 41, 581–587. doi: 10.1111/jvp.12506
- Gluck, J., Buhrke, T., Frenzel, F., Braeuning, A., and Lampen, A. (2018). *In silico* genotoxicity and carcinogenicity prediction for food-relevant secondary plant metabolites. *Food Chem. Toxicol.* 116, 298–306. doi: 10.1016/j.fct.2018.04.024
- Gong, E. C., Chea, S., Balupuri, A., Kang, N. S., Chin, Y. W., and Choi, Y. H. (2018). Enzyme kinetics and molecular docking studies on cytochrome 2B6, 2C19, 2E1, and 3A4 activities by saquinone. *Molecules* 23:555. doi: 10.3390/molecules23030555
- Guerra, L. R., de Souza, A. M. T., Cortes, J. A., Lione, V. O. F., Castro, H. C., and Alves, G. G. (2017). Assessment of predictivity of volatile organic compounds carcinogenicity and mutagenicity by freeware *in silico* models. *Regul. Toxicol. Pharmacol.* 91, 1–8. doi: 10.1016/j.yrtph.2017.09.030
- Guha, R. (2011). The ups and downs of structure-activity landscapes. *Methods Mol. Biol.* 672, 101–117. doi: 10.1007/978-1-60761-839-3\_3
- Gupta, P., Kamath, A. V., Park, S., Chiu, H., Lutman, J., Maia, M., et al. (2016). Preclinical pharmacokinetics of MHAA4549A, a human monoclonal antibody to influenza A virus, and the prediction of its efficacious clinical dose for the treatment of patients hospitalized with influenza A. *MAbs* 8, 991–997. doi: 10.1080/19420862.2016.1167294
- Hagihara, M., Kato, H., Hamada, Y., Hirai, J., Sakanashi, D., Suematsu, H., et al. (2016). Population pharmacokinetics of arbekacin in different infectious disease settings and evaluation of dosing regimens. *J. Infect. Chemother.* 22, 436–443. doi: 10.1016/j.jiac.2016.03.008
- Han, Y., Zhang, J. P., Hu, C. Q., Zhang, X., Ma, B. F., and Zhang, P. P. (2019). *In silico* ADME and toxicity prediction of ceftazidime and its impurities. *Front. Pharmacol.* 10:434. doi: 10.3389/fphar.2019.00434
- Hansch, C. (1981). The physicochemical approach to drug design and discovery (QSAR). *Drug Dev. Res.* 1, 267–309. doi: 10.1002/ddr.430010403
- Hansmann, S., Darwich, A., Margolske, A., Aarons, L., and Dressman, J. (2016). Forecasting oral absorption across biopharmaceutics classification system classes with physiologically based pharmacokinetic models. *J. Pharm. Pharmacol.* 68, 1501–1515. doi: 10.1111/jphp.12618
- Haque, A., Leong, K. H., Lo, Y. L., Awang, K., and Nagoor, N. H. (2017). *In vitro* inhibitory mechanisms and molecular docking of 1'-S-1'-acetoxychavicol acetate on human cytochrome P450 enzymes. *Phytomedicine* 31, 1–9. doi: 10.1016/j.phymed.2017.05.002
- Hardy, B., Douglas, N., Helma, C., Rautenberg, M., Jeliakova, N., Jeliakov, V., et al. (2010). Collaborative development of predictive toxicology applications. *J. Cheminform.* 2:7. doi: 10.1186/1758-2946-2-7
- Helma, C. (2006). Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and salmonella mutagenicity. *Mol. Divers.* 10, 147–158. doi: 10.1007/s11030-005-9001-5
- Hu, C., Wang, Y., Song, R., Yu, C., Luo, X., and Jia, J. (2017). Single- and multiple-dose pharmacokinetics, safety and tolerability of lurasidone in healthy Chinese subjects. *Clin. Drug Investig.* 37, 861–871. doi: 10.1007/s40261-017-0546-8
- Hu, Y., and Smith, D. E. (2018). *In silico* prediction of the absorption and disposition of cefadroxil in humans using an intestinal permeability method scaled from humanized PepT1 mice. *Drug Metab. Dispos.* 47, 173–183. doi: 10.1124/dmd.118.084236
- Huang, J., and Fan, X. (2011). Why QSAR fails: an empirical evaluation using conventional computational approach. *Mol. Pharm.* 8, 600–608. doi: 10.1021/mp100423u
- Huang, X., Zhang, S., Ma, Y., Yang, H., He, C., Tian, R., et al. (2018). Bioequivalence of two quetiapine extended release tablets in Chinese healthy volunteers under fasting and fed conditions and effects of food on pharmacokinetic profiles. *Drug Des. Devel. Ther.* 13, 255–264. doi: 10.2147/DDDT.S182965

- Huynh, L., Masereeuw, R., Friedberg, T., Ingelman-Sundberg, M., and Manivet, P. (2009). *In silico* platform for xenobiotics ADME-T pharmacological properties modeling and prediction. Part I: beyond the reduction of animal model use. *Drug Discov. Today* 14, 401–405. doi: 10.1016/j.drudis.2009.01.009
- Ibarra, M., Valiente, C., Sopena, P., Schiavo, A., Lorier, M., Vazquez, M., et al. (2018). Integration of *in vitro* biorelevant dissolution and *in silico* PBPK model of carvedilol to predict bioequivalence of oral drug products. *Eur. J. Pharm. Sci.* 118, 176–182. doi: 10.1016/j.ejps.2018.03.032
- Ibrahim, M. E., Brooks, K. M., Castillo-Mancilla, J. R., McHugh, C., Morrow, M., Brothers, J., et al. (2018). Short communication: bioequivalence of tenofovir and emtricitabine after coencapsulation with the proteus ingestible sensor. *AIDS Res. Hum. Retroviruses* 34, 835–837. doi: 10.1089/aid.2018.0081
- Idkaidek, N., Agha, H., and Arafat, T. (2018a). Saliva versus plasma bioequivalence of valsartan/hydrochlorothiazide in humans: validation of classes II and IV drugs of the salivary excretion classification system. *Drug Res.* 68, 54–59. doi: 10.1055/s-0043-117775
- Idkaidek, N., Hamadi, S., El-Assi, M., Al-Shalalfeh, A., and Al-Ghawazi, A. (2018b). Saliva versus plasma therapeutic drug monitoring of pregabalin in jordanian patients. *Drug Res.* 68, 596–600. doi: 10.1055/a-0600-2113
- Issa, N. T., Wathieu, H., Ojo, A., Byers, S. W., and Dakshanamurthy, S. (2017). Drug metabolism in preclinical drug development: a survey of the discovery process, toxicology, and computational tools. *Curr. Drug Metab.* 18, 556–565. doi: 10.2174/1389200218666170316093301
- Jamei, M., Marciniak, S., Feng, K., Barnett, A., Tucker, G., and Rostami-Hodjegan, A. (2009). The simcyp population-based ADME simulator. *Expert Opin. Drug Metab. Toxicol.* 5, 211–223. doi: 10.1517/17425250802691074
- Jiang, D., Lei, T., Wang, Z., Shen, C., Cao, D., and Hou, T. (2020). ADMET evaluation in drug discovery. 20. Prediction of breast cancer resistance protein inhibition through machine learning. *J. Cheminform.* 12:16. doi: 10.1186/s13321-020-00421-y
- Jin, Y., Borell, H., Gardin, A., Ufer, M., Huth, F., and Camenisch, G. (2018). *In vitro* studies and *in silico* predictions of fluconazole and CYP2C9 genetic polymorphism impact on siponimod metabolism and pharmacokinetics. *Eur. J. Clin. Pharmacol.* 74, 455–464. doi: 10.1007/s00228-017-2404-2
- Johnson, T. N., Whitaker, M. J., Keevil, B., and Ross, R. J. (2018). Bioavailability of oral hydrocortisone corrected for binding proteins and measured by LC-MS/MS using serum cortisol and salivary cortisone. *J. Bioequivalence Bioavailab.* 10, 001–003. doi: 10.4172/jbb.1000365
- Jones, H., Chen, Y., Gibson, C., Heimbach, T., Parrott, N., Peters, S., et al. (2015). Physiologically based pharmacokinetic modeling in drug discovery and development: a pharmaceutical industry perspective. *Clin. Pharmacol. Ther.* 97, 247–262. doi: 10.1002/cpt.37
- Kar, S., and Leszczynski, J. (2017). Recent advances of computational modeling for predicting drug metabolism: a perspective. *Curr. Drug Metab.* 18, 1106–1122. doi: 10.2174/1389200218666170607102104
- Kato, H., Yamaotsu, N., Iwazaki, N., Okamura, S., Kume, T., and Hirono, S. (2017). Precise prediction of activators for the human constitutive androstane receptor using structure-based three-dimensional quantitative structure-activity relationship methods. *Drug Metab. Pharmacokinet.* 32, 179–188. doi: 10.1016/j.dmpk.2017.02.001
- Kavitha, R., Karunakaran, S., Chandrasekhar, S. S., Lee, K. W., and Meganathan, C. (2015). Pharmacophore modeling, virtual screening, molecular docking studies and density functional theory approaches to identify novel ketohexokinase (KHK) inhibitors. *Biosystems* 138, 39–52. doi: 10.1016/j.biosystems.2015.10.005
- Kazmi, S. R., Jun, R., Yu, M.-S., Jung, C., and Na, D. (2019). *In silico* approaches and tools for the prediction of drug metabolism and fate: a review. *Comput. Biol. Med.* 106, 54–64. doi: 10.1016/j.combiomed.2019.01.008
- Keemink, J., Sjogren, E., Holm, R., and Bergstrom, C. A. S. (2018). Does the intake of ethanol affect oral absorption of poorly soluble drugs? *J. Pharm. Sci.* 108, 1765–1771. doi: 10.1016/j.xphs.2018.12.006
- Kemp, C. A., Flanagan, J. U., van Eldik, A. J., Marechal, J. D., Wolf, C. R., Roberts, G. C., et al. (2004). Validation of model of cytochrome P450 2D6: an *in silico* tool for predicting metabolism and inhibition. *J. Med. Chem.* 47, 5340–5346. doi: 10.1021/jm049934e
- Khalid, S., Zahid, M. A., Ali, H., Kim, Y. S., and Khan, S. (2018). Biaryl scaffold-focused virtual screening for anti-aggregatory and neuroprotective effects in alzheimer's disease. *BMC Neurosci.* 19:74. doi: 10.1186/s12868-018-0472-6
- Khan, M. S., Majid, A. M., Iqbal, M. A., Majid, A. S., Al-Mansoub, M., and Haque, R. S. (2016). Designing the angiogenic inhibitor for brain tumor via disruption of VEGF and IL17A expression. *Eur. J. Pharm. Sci.* 93, 304–318. doi: 10.1016/j.ejps.2016.08.032
- Khan, M. T. (2010). Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. *Curr. Drug Metab.* 11, 285–295. doi: 10.2174/138920010791514306
- Khurana, R. K., Kaur, R., Kaur, M., Kaur, R., Kaur, J., Kaur, H., et al. (2017). Exploring and validating physicochemical properties of mangiferin through GastroPlus(R) software. *Future Sci. OA* 3:FSO167. doi: 10.4155/foa-2016-0055
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47, D1102–D1109. doi: 10.1093/nar/gky1033
- Kim, S. J., Heo, S. H., Lee, D. S., Hwang, I. G., Lee, Y. B., and Cho, H. Y. (2016). Gender differences in pharmacokinetics and tissue distribution of 3 perfluoroalkyl and polyfluoroalkyl substances in rats. *Food Chem. Toxicol.* 97, 243–255. doi: 10.1016/j.fct.2016.09.017
- Kirchmair, J., Williamson, M. J., Tyzack, J. D., Tan, L., Bond, P. J., Bender, A., et al. (2012). Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. *J. Chem. Inf. Model.* 52, 617–648. doi: 10.1021/ci200542m
- Kola, I., and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–715. doi: 10.1038/nrd1470
- Kollar, J., and Frece, V. (2018). Diarylcyclopropane hydroxamic acid inhibitors of histone deacetylase 4 designed by combinatorial approach and QM/MM calculations. *J. Mol. Graph Model.* 85, 97–110. doi: 10.1016/j.jmgm.2018.08.008
- Kong, L., Tang, Y., Zhang, X., Lu, G., Yu, M., Shi, Q., et al. (2017). Pharmacokinetic/pharmacodynamic analysis of meropenem for the treatment of nosocomial pneumonia in intracerebral hemorrhage patients by monte carlo simulation. *Ann. Pharmacother.* 51, 970–975. doi: 10.1177/1060028017719715
- Kou, D., Zhang, C., Yiu, H., Ng, T., Lubach, J. W., Janson, M., et al. (2018). *In vitro*, *in silico*, and *in vivo* assessments of intestinal precipitation and its impact on bioavailability of a BCS class 2 basic compound. *Mol. Pharm.* 15, 1607–1617. doi: 10.1021/acs.molpharmaceut.7b01143
- Kumar, R., Garg, P., and Bharatam, P. V. (2015). Shape-based virtual screening, docking, and molecular dynamics simulations to identify Mtb-ASADH inhibitors. *J. Biomol. Struct. Dyn.* 33, 1082–1093. doi: 10.1080/07391102.2014.929535
- Lapins, M., Arvidsson, S., Lampa, S., Berg, A., Schaal, W., Alvarsson, J., et al. (2018). A confidence predictor for logD using conformal regression and a support-vector machine. *J. Cheminform.* 10:17. doi: 10.1186/s13321-018-0271-1
- Lauro, G., Cantone, V., Potenza, M., Fischer, K., Koeberle, A., Werz, O., et al. (2018). Discovery of 3-hydroxy-3-pyrrolin-2-one-based mPGES-1 inhibitors using a multi-step virtual screening protocol. *Medchemcomm* 9, 2028–2036. doi: 10.1039/C8MD00497H
- Lave, T., Parrott, N., Grimm, H. P., Fleury, A., and Reddy, M. (2007). Challenges and opportunities with modelling and simulation in drug discovery and drug development. *Xenobiotica* 37, 1295–1310. doi: 10.1080/00498250701534885
- 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, D1091–D1097. doi: 10.1093/nar/gkt1068
- Lea, I. A., Gong, H., Paleja, A., Rashid, A., and Fostel, J. (2017). CEBS: a comprehensive annotated database of toxicological data. *Nucleic Acids Res.* 45, D964–D971. doi: 10.1093/nar/gkw1077
- Lee, L. S., Seng, K. Y., Wang, L. Z., Yong, W. P., Hee, K. H., Soh, T. I., et al. (2016). Phenotyping of UGT1A1 activity using raltegravir predicts pharmacokinetics and toxicity of irinotecan in FOLFIRI. *PLoS ONE* 11:e0147681. doi: 10.1371/journal.pone.0147681
- Li, D., Wang, Y., and Han, K. (2012). Recent density functional theory model calculations of drug metabolism by cytochrome P450. *Coord Chem. Rev.* 256, 1137–1150. doi: 10.1016/j.ccr.2012.01.016
- Li, J., Du, H., Wu, Z., Su, H., Liu, G., Tang, Y., et al. (2016). Interactions of omeprazole-based analogues with cytochrome P450 2C19: a computational study. *Mol. Biosyst.* 12, 1913–1921. doi: 10.1039/C6MB00139D

- Li, M., Zou, P., Tyner, K., and Lee, S. (2017). Physiologically based pharmacokinetic (PBPK) modeling of pharmaceutical nanoparticles. *AAPS J.* 19, 26–42. doi: 10.1208/s12248-016-0010-3
- Li, X., Wang, R., Liu, Y., Liu, Y., Zheng, H., Feng, Y., et al. (2017). Pharmacokinetic study of single- and multiple-dosing with metolazone tablets in healthy Chinese population. *BMC Pharmacol. Toxicol.* 18:73. doi: 10.1186/s40360-017-0178-x
- Li, Y. F., Guo, H. Y., Yang, F., Zhou, L. G., Huang, X. H., Ding, H. Z., et al. (2017). Comparative pharmacokinetics of diaveridine in pigs and chickens following single intravenous and oral administration. *J. Vet. Pharmacol. Ther.* 40, 500–504. doi: 10.1111/jvp.12384
- Lin, J., Sahakian, D. C., de Moraes, S. M., Xu, J. J., Polzer, R. J., and Winter, S. M. (2003). The role of absorption, distribution, metabolism, excretion and toxicity in drug discovery. *Curr. Top. Med. Chem.* 3, 1125–1154. doi: 10.2174/1568026033452096
- Lindstrom, A., Anantpadma, M., Baker, L., Raghavendra, N. M., Davey, R., and Davisson, V. J. (2018). Phenotypic prioritization of diphyllin derivatives that block filoviral cell entry by vacuolar (H<sup>+</sup>)-ATPase inhibition. *Chemmedchem* 13, 2664–2676. doi: 10.1002/cmdc.201800587
- Luo, J., Lai, T., Guo, T., Chen, F., Zhang, L., Ding, W., et al. (2018). Synthesis and acaricidal activities of scopoletin phenolic ether derivatives: QSAR, molecular docking study and *in Silico* ADME predictions. *Molecules* 23:995. doi: 10.3390/molecules23050995
- Ma, Y., Wei, H. Y., Zhang, Y. Z., Jin, W. Y., Li, H. L., Zhou, H., et al. (2017). Synthesis, bioactivity, 3D-QSAR studies of novel dibenzofuran derivatives as PTP-MEG2 inhibitors. *Oncotarget* 8, 38466–38481. doi: 10.18632/oncotarget.16595
- Macalino, S. J., Gosu, V., Hong, S., and Choi, S. (2015). Role of computer-aided drug design in modern drug discovery. *Arch. Pharm. Res.* 38, 1686–1701. doi: 10.1007/s12272-015-0640-5
- Malik, R., Mehta, P., Srivastava, S., Choudhary, B. S., and Sharma, M. (2016). Pharmacophore modeling, 3D-QSAR, and *in silico* ADME prediction of N-pyridyl and pyrimidine benzamides as potent antiepileptic agents. *J. Recept. Signal. Transduct. Res.* 37, 259–266. doi: 10.1080/10799893.2016.1217883
- Maltarollo, V. G., Gertrudes, J. C., Oliveira, P. R., and Honório, K. M. (2015). Applying machine learning techniques for ADME-Tox prediction: a review. *Expert Opin. Drug Metab. Toxicol.* 11, 259–271. doi: 10.1517/17425255.2015.980814
- Maunz, A., Gutlein, M., Rautenberg, M., Vorgrimmler, D., Gebele, D., and Helma, C. (2013). lazar: a modular predictive toxicology framework. *Front. Pharmacol.* 4:38. doi: 10.3389/fphar.2013.00038
- Merlot, C. (2010). Computational toxicology—a tool for early safety evaluation. *Drug Discov. Today* 15, 16–22. doi: 10.1016/j.drudis.2009.09.010
- Mertens, B., Simon, C., Van Bossuyt, M., Onghena, M., Vandermarken, T., Van Langenhove, K., et al. (2016). Investigation of the genotoxicity of substances migrating from polycarbonate replacement baby bottles to identify chemicals of high concern. *Food Chem. Toxicol.* 89, 126–137. doi: 10.1016/j.fct.2016.01.009
- Michielan, L., and Moro, S. (2010). Pharmaceutical perspectives of nonlinear QSAR strategies. *J. Chem. Inf. Model.* 50, 961–978. doi: 10.1021/ci100072z
- Modi, S. (2003). Computational approaches to the understanding of ADMET properties and problems. *Drug Discov. Today* 8, 621–623. doi: 10.1016/S1359-6446(03)02739-9
- Mohan, A., Kirubakaran, R., Parray, J. A., Sivakumar, R., Muruges, E., and Govarthanan, M. (2020). Ligand-based pharmacophore filtering, atom based 3D-QSAR, virtual screening and ADME studies for the discovery of potential ck2 inhibitors. *J. Mol. Struct.* 1205:127670. doi: 10.1016/j.molstruc.2019.127670
- Moj, D., Britz, H., Burhenne, J., Stewart, C. F., Egerer, G., Haefeli, W. E., et al. (2017). A physiologically based pharmacokinetic and pharmacodynamic (PBPK/PD) model of the histone deacetylase (HDAC) inhibitor vorinostat for pediatric and adult patients and its application for dose specification. *Cancer Chemother. Pharmacol.* 80, 1013–1026. doi: 10.1007/s00280-017-3447-x
- Molina, I., Salvador, F., Sanchez-Montalva, A., Artaza, M. A., Moreno, R., Perin, L., et al. (2017). Pharmacokinetics of benznidazole in healthy volunteers and implications in future clinical trials. *Antimicrob. Agents Chemother.* 61, e01912–16. doi: 10.1128/AAC.01912-16
- Mondal, S., Mandal, S. M., Mondal, T. K., and Sinha, C. (2017). Spectroscopic characterization, antimicrobial activity, DFT computation and docking studies of sulfonamide Schiff bases. *J. Mol. Struct.* 1127, 557–567. doi: 10.1016/j.molstruc.2016.08.011
- Moroy, G., Martiny, V. Y., Vayer, P., Villoutreix, B. O., and Miteva, M. A. (2012). Toward *in silico* structure-based ADMET prediction in drug discovery. *Drug Discov. Today* 17, 44–55. doi: 10.1016/j.drudis.2011.10.023
- Muster, W., Breidenbach, A., Fischer, H., Kirchner, S., Müller, L., and Pähler, A. (2008). Computational toxicology in drug development. *Drug Discov. Today* 13, 303–310. doi: 10.1016/j.drudis.2007.12.007
- Nan, J., Hao, H., Xie, S., Pan, Y., Xi, C., Mao, F., et al. (2017). Pharmacokinetic and pharmacodynamic integration and modeling of acetylkisamycin in swine for clostridium perfringens. *J. Vet. Pharmacol. Ther.* 40, 641–655. doi: 10.1111/jvp.12404
- Nandekar, P. P., Khomane, K., Chaudhary, V., Rathod, V. P., Borkar, R. M., Bhandi, M. M., et al. (2016). Identification of leads for antiproliferative activity on MDA-MB-435 human breast cancer cells through pharmacophore and CYP1A1-mediated metabolism. *Eur. J. Med. Chem.* 115, 82–93. doi: 10.1016/j.ejmech.2016.02.061
- Nasab, R. R., Mansourian, M., Hassanzadeh, F., and Shahlaei, M. (2018). Exploring the interaction between epidermal growth factor receptor tyrosine kinase and some of the synthesized inhibitors using combination of *in-silico* and *in-vitro* cytotoxicity methods. *Res. Pharm. Sci.* 13, 509–522. doi: 10.4103/1735-5362.245963
- Nestorov, I. (2003). Whole body pharmacokinetic models. *Clin. Pharmacokinet.* 42, 883–908. doi: 10.2165/00003088-200342100-00002
- Niederalt, C., Kuepfer, L., Solodenko, J., Eissing, T., Siegmund, H. U., Block, M., et al. (2017). A generic whole body physiologically based pharmacokinetic model for therapeutic proteins in PK-sim. *J. Pharmacokinet. Pharmacodyn.* 45, 235–257. doi: 10.1007/s10928-017-9559-4
- Nisha, C. M., Kumar, A., Vimal, A., Bai, B. M., Pal, D., and Kumar, A. (2016). Docking and ADMET prediction of few GSK-3 inhibitors divulges 6-bromoindirubin-3-oxime as a potential inhibitor. *J. Mol. Graph. Model.* 65, 100–107. doi: 10.1016/j.jmgm.2016.03.001
- Niu, L., Ding, L., Lu, C., Zuo, F., Yao, K., Xu, S., et al. (2016). Flavokawain A inhibits Cytochrome P450 in *in vitro* metabolic and inhibitory investigations. *J. Ethnopharmacol.* 191, 350–359. doi: 10.1016/j.jep.2016.06.039
- Nosrati, M., Shakeran, Z., and Shakeran, Z. (2018). Frangulosid as a novel hepatitis B virus DNA polymerase inhibitor: a virtual screening study. *In Silico Pharmacol.* 6:10. doi: 10.1007/s40203-018-0047-3
- Olguin, H. J., Martinez, H. O., Perez, C. F., Mendiola, B. R., Espinosa, L. R., Pacheco, J. L. C., et al. (2017). Pharmacokinetics of sildenafil in children with pulmonary arterial hypertension. *World J. Pediatr.* 13, 588–592. doi: 10.1007/s12519-017-0043-4
- Onghena, M., Van Hoeck, E., Van Loco, J., Ibanez, M., Cherta, L., Portoles, T., et al. (2015). Identification of substances migrating from plastic baby bottles using a combination of low-resolution and high-resolution mass spectrometric analysers coupled to gas and liquid chromatography. *J. Mass Spectrom.* 50, 1234–1244. doi: 10.1002/jms.3644
- Onghena, M., van Hoeck, E., Vervliet, P., Scippo, M. L., Simon, C., van Loco, J., et al. (2014). Development and application of a non-targeted extraction method for the analysis of migrating compounds from plastic baby bottles by GC-MS. *Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess.* 31, 2090–2102. doi: 10.1080/19440049.2014.979372
- Overington, J. (2009). ChEMBL: an interview with John Overington, team leader, chemogenomics at the European bioinformatics institute outstation of the European molecular biology laboratory (EMBL-EBI). interview by Wendy A. Warr. *J. Comput. Aided Mol. Des.* 23, 195–198. doi: 10.1007/s10822-009-9260-9
- Padari, H., Metsvaht, T., Germovsek, E., Barker, C. I., Kipper, K., Herodes, K., et al. (2018). Pharmacokinetics of penicillin G in preterm and term neonates. *Antimicrob. Agents Chemother.* 62, 1995–2000. doi: 10.1128/AAC.02238-17
- Parrott, N. J., Yu, L. J., Takano, R., Nakamura, M., and Morcos, P. N. (2016). Physiologically based absorption modeling to explore the impact of food and gastric pH changes on the pharmacokinetics of alectinib. *AAPS J.* 18, 1464–1474. doi: 10.1208/s12248-016-9957-3
- Patel, C. N., Kumar, S. P., Rawal, R. M., Patel, D. P., Gonzalez, F. J., and Pandya, H. A. (2020). A multiparametric organ toxicity predictor for drug discovery. *Toxicol. Mech. Methods* 30, 159–166. doi: 10.1080/15376516.2019.1681044
- Patel, D. B., Darji, D. G., Patel, K. R., Rajani, D. P., Rajani, S. D., and Patel, H. D. (2020). Synthesis of novel quinoline-thiosemicarbazide hybrids and evaluation of their biological activities, molecular docking, molecular



- dynamics, pharmacophore model studies, and ADME-tox properties. *J. Heterocyclic Chem.* 57, 1183–1200. doi: 10.1002/jhet.3855
- Patel, N., Jain, S., and Lin, S. (2016). Transdermal iontophoretic delivery of tacrine hydrochloride: correlation between *in vitro* permeation and *in vivo* performance in rats. *Int. J. Pharm.* 513, 393–403. doi: 10.1016/j.ijpharm.2016.09.038
- Pathak, S. M., Schaefer, K. J., Jamei, M., and Turner, D. B. (2019). Biopharmaceutic IVIVE-mechanistic modeling of single- and two-phase *in vitro* experiments to obtain drug-specific parameters for incorporation into PBPK models. *J. Pharm. Sci.* 108, 1604–1618. doi: 10.1016/j.xphs.2018.11.034
- Patlewicz, G., Jeliakova, N., Safford, R. J., Worth, A. P., and Aleksiev, B. (2008). An evaluation of the implementation of the cramer classification scheme in the toxtree software. *SAR QSAR Environ. Res.* 19, 495–524. doi: 10.1080/10629360802083871
- Patterson, D. E., Cramer, R. D., Ferguson, A. M., Clark, R. D., and Weinberger, L. E. (1996). Neighborhood behavior: a useful concept for validation of “molecular diversity” descriptors. *J. Med. Chem.* 39, 3049–3059. doi: 10.1021/jm960290n
- Perdaems, N., Blasco, H., Vinson, C., Chenel, M., Whalley, S., Cazade, F., et al. (2010). Predictions of metabolic drug-drug interactions using physiologically based modelling: two cytochrome P450 3A4 substrates coadministered with ketoconazole or verapamil. *Clin. Pharmacokinet.* 49, 239–258. doi: 10.2165/11318130-000000000-00000
- Perez-Nueno, V. I., and Ritchie, D. W. (2011). Using consensus-shape clustering to identify promiscuous ligands and protein targets and to choose the right query for shape-based virtual screening. *J. Chem. Inf. Model.* 51, 1233–1248. doi: 10.1021/ci100492r
- Petrescu, A. M., Paunescu, V., and Ilia, G. (2019). The antiviral activity and cytotoxicity of 15 natural phenolic compounds with previously demonstrated antifungal activity. *J. Environ. Sci. Health B* 54, 498–504. doi: 10.1080/03601234.2019.1574176
- Pilari, S., Gaub, T., Block, M., and Gortlitz, L. (2017). Development of physiologically based organ models to evaluate the pharmacokinetics of drugs in the testes and the thyroid gland. *CPT Pharmacometr. Syst. Pharmacol.* 6, 532–542. doi: 10.1002/psp4.12205
- Pires, D. E., Blundell, T. L., and Ascher, D. B. (2015). pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* 58, 4066–4072. doi: 10.1021/acs.jmedchem.5b00104
- Poirier, D., Roy, J., Cortes-Benitez, F., and Dutour, R. (2016). Targeting cytochrome P450 (CYP) 1B1 with steroid derivatives. *Bioorg. Med. Chem. Lett.* 26, 5272–5276. doi: 10.1016/j.bmcl.2016.09.046
- Polepally, A. R., Brundage, R. C., Rummel, R. P., Leppik, I. E., Pennell, P. B., White, J. R., et al. (2018). Lamotrigine pharmacokinetics following oral and stable-labeled intravenous administration in young and elderly adult epilepsy patients: effect of age. *Epilepsia* 59, 1718–1726. doi: 10.1111/epi.14519
- Prabhu, S. V., and Singh, S. K. (2019). Identification of potential dual negative allosteric modulators of group I mGluR family: a shape based screening, ADME prediction, induced fit docking and molecular dynamics approach against neurodegenerative diseases. *Curr. Top. Med. Chem.* 19, 2687–2707. doi: 10.2174/1568026619666191105112800
- Pradeep, P., Povinelli, R. J., White, S., and Merrill, S. J. (2016). An ensemble model of QSAR tools for regulatory risk assessment. *J. Cheminform* 8:48. doi: 10.1186/s13321-016-0164-0
- Prokop, J. W., Lazar, J., Crapitto, G., Smith, D. C., Worthey, E. A., and Jacob, H. J. (2017). Molecular modeling in the age of clinical genomics, the enterprise of the next generation. *J. Mol. Model.* 23:75. doi: 10.1007/s00894-017-3258-3
- Putta, S., and Beroza, P. (2007). Shapes of things: computer modeling of molecular shape in drug discovery. *Curr. Top. Med. Chem.* 7, 1514–1524. doi: 10.2174/156802607782194770
- Rachar, V., Czejka, M., Kitzmueller, M., Buchner, P., Lichtneckert, M., Greil, R., et al. (2016). Assessment of pharmacokinetic interaction between capecitabine and cetuximab in metastatic colorectal cancer patients. *Anticancer Res.* 36, 4715–4723. doi: 10.21873/anticancer.11026
- Ramatenki, V., Dumpati, R., Vadija, R., Vellanki, S., Potlapally, S. R., Rondla, R., et al. (2016). Targeting the ubiquitin-conjugating enzyme E2D4 for cancer drug discovery—a structure-based approach. *J. Chem. Biol.* 10, 51–67. doi: 10.1007/s12154-016-0164-6
- Ramatenki, V., Dumpati, R., Vadija, R., Vellanki, S., Potlapally, S. R., Rondla, R., et al. (2017). Identification of new lead molecules against UBE2NL enzyme for cancer therapy. *Appl. Biochem. Biotechnol.* 182, 1497–1517. doi: 10.1007/s12010-017-2414-7
- Rambirith, V., Naidoo, P., Maharaj, B., and Pillai, G. (2016). Population pharmacokinetic modeling of glibenclamide in poorly controlled South African type 2 diabetic subjects. *Clin. Pharmacol.* 8, 83–92. doi: 10.2147/CPAA.S102676
- Rawat, R., and Verma, S. M. (2020). High-throughput virtual screening approach involving pharmacophore mapping, ADME filtering, molecular docking and MM-GBSA to identify new dual target inhibitors of PfDHODH and PfCytbc1 complex to combat drug resistant malaria. *J. Biomol. Struct. Dyn.* doi: 10.1080/07391102.2020.1784288. [Epub ahead of print].
- Reddy, K. K., Singh, S. K., Tripathi, S. K., Selvaraj, C., and Suryanarayanan, V. (2013). Shape and pharmacophore-based virtual screening to identify potential cytochrome P450 sterol 14 alpha-demethylase inhibitors. *J. Recept. Signal. Transd.* 33, 234–243. doi: 10.3109/10799893.2013.789912
- Richard, A. M., and Williams, C. R. (2002). Distributed structure-searchable toxicity (DSSTox) public database network: a proposal. *Mutat. Res.* 499, 27–52. doi: 10.1016/S0027-5107(01)00289-5
- Roman, M., Roman, D. L., Ostafe, V., Ciorsac, A., and Isvoran, A. (2018). Computational assessment of pharmacokinetics and biological effects of some anabolic and androgen steroids. *Pharm. Res.* 35:41. doi: 10.1007/s11095-018-2353-1
- Romanski, M., Kasprzyk, A., Karbownik, A., and Glowka, F. K. (2016). Ocular disposition of treosulfan and its active epoxy-transformers following intravenous administration in rabbits. *Drug Metab. Pharmacokinet.* 31, 356–362. doi: 10.1016/j.dmpk.2016.07.001
- Romanski, M., Kasprzyk, A., Walczak, M., Ziolkowska, A., and Glowka, F. (2017). Disposition of treosulfan and its active monooxide in a bone marrow, liver, lungs, brain, and muscle: studies in a rat model with clinical relevance. *Eur. J. Pharm. Sci.* 109, 616–623. doi: 10.1016/j.ejps.2017.09.011
- Rostami-Hodjegan, A. (2012). Physiologically based pharmacokinetics joined with *in vitro-in vivo* extrapolation of ADME: a marriage under the arch of systems pharmacology. *Clin. Pharmacol. Ther.* 92, 50–61. doi: 10.1038/clpt.2012.65
- Rowland, M., Balant, L., and Peck, C. (2004). Physiologically based pharmacokinetics in drug development and regulatory science: a workshop report (Georgetown University, Washington, DC, May 29–30, 2002). *AAPS PharmSci.* 6:E6. doi: 10.1208/ps060106
- Rowland, M., Peck, C., and Tucker, G. (2011). Physiologically-based pharmacokinetics in drug development and regulatory science. *Annu. Rev. Pharmacol. Toxicol.* 51, 45–73. doi: 10.1146/annurev-pharmtox-010510-100540
- Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., and Kabir, H. (2012). Comparative studies on some metrics for external validation of QSPR models. *J. Chem. Inf. Model.* 52, 396–408. doi: 10.1021/ci200520g
- Saba, N., and Seal, A. (2018). Identification of a less toxic vinca alkaloid derivative for use as a chemotherapeutic agent, based on *in silico* structural insights and metabolic interactions with CYP<sub>3A4</sub> and CYP<sub>3A5</sub>. *J. Mol. Model.* 24:82. doi: 10.1007/s00894-018-3611-1
- Sangle, G. V., Patil, M., Deshmukh, N. J., Shengule, S. A., Kamble, S., Vuppulavanchu, K. K., et al. (2018). Evaluation of pharmacokinetic and pharmacodynamic parameters following single dose of sitagliptin in healthy Indian males. *Eur. J. Clin. Pharmacol.* 74, 561–569. doi: 10.1007/s00228-018-2433-5
- Sanni, D. M., and Fatoki, T. H. (2017). Computational evaluation of pharmacokinetics and potential protein targets of ginger (*Zingiber officinale*). *J. Microbiol. Biotechnol. Res.* 7:14. doi: 10.24896/jmbr.2017713
- Sanni, D. M., Fatoki, T. H., Kolawole, A. O., and Akinmoladun, A. C. (2017). Xerone structure and function: computational comparative mastery of its mystery. *In Silico Pharmacol.* 5:8. doi: 10.1007/s40203-017-0028-y
- Santes-Palacios, R., Romo-Mancillas, A., Camacho-Carranza, R., and Espinosa-Aguirre, J. J. (2016). Inhibition of human and rat CYP1A1 enzyme by grapefruit juice compounds. *Toxicol. Lett.* 258, 267–275. doi: 10.1016/j.toxlet.2016.07.023
- Sasahara, K., Mashima, A., Yoshida, T., and Chuman, H. (2015). Molecular dynamics and density functional studies on the metabolic selectivity of antipsychotic thioridazine by cytochrome P450 2D6: connection with crystallographic and metabolic results. *Bioorg. Med. Chem.* 23, 5459–5465. doi: 10.1016/j.bmc.2015.07.043



- Schlender, J. F., Meyer, M., Thelen, K., Krauss, M., Willmann, S., Eissing, T., et al. (2016). Development of a whole-body physiologically based pharmacokinetic approach to assess the pharmacokinetics of drugs in elderly individuals. *Clin. Pharmacokinet.* 55, 1573–1589. doi: 10.1007/s40262-016-0422-3
- Schmidt, U., Struck, S., Gruening, B., Hossbach, J., Jaeger, I. S., Parol, R., et al. (2009). SuperToxic: a comprehensive database of toxic compounds. *Nucleic Acids Res.* 37, D295–D299. doi: 10.1093/nar/gkn850
- Schmitt, W., and Willmann, S. (2004). Physiology-based pharmacokinetic modeling: ready to be used. *Drug Discov. Today Technol.* 1, 449–456. doi: 10.1016/j.ddtec.2004.09.006
- Schrödinger Suite (2012). Schrödinger, LLC, New York, NY.
- Schyman, P., Liu, R., Desai, V., and Wallqvist, A. (2017). vNN web server for ADMET predictions. *Front. Pharmacol.* 8:889. doi: 10.3389/fphar.2017.00889
- Selick, H. E., Beresford, A. P., and Tarbit, M. H. (2002). The emerging importance of predictive ADME simulation in drug discovery. *Drug Discov. Today* 7, 109–116. doi: 10.1016/S1359-6446(01)02100-6
- Shaaban, S., Vervandier-Fasseur, D., Andreoletti, P., Zarrouk, A., Richard, P., Negm, A., et al. (2018). Cytoprotective and antioxidant properties of organic selenides for the myelin-forming cells, oligodendrocytes. *Bioorg. Chem.* 80, 43–56. doi: 10.1016/j.bioorg.2018.05.019
- Shaik, M., Shaik, S., and Kilari, E. K. (2018). Population pharmacokinetics of glimepiride in normal and diabetic rabbits. *Biopharm. Drug Dispos.* 39, 265–274. doi: 10.1002/bdd.2132
- Shaji, D. (2018). Molecular docking studies of human MCT8 protein with soy isoflavones in allan-herndon-dudley syndrome (AHDS). *J. Pharm. Anal.* 8, 318–323. doi: 10.1016/j.jppha.2018.07.001
- Shakhnovich, V., Smith, P. B., Gupta, J. T., James, L. P., Collier, D. N., Wu, H., et al. (2018). Obese children require lower doses of pantoprazole than nonobese peers to achieve equal systemic drug exposures. *J. Pediatr.* 193, 102–108.e101. doi: 10.1016/j.jpeds.2017.10.011
- Shan, Q., Fan, J., Wang, J., Zhu, X., Yin, Y., and Zheng, G. (2017). Pharmacokinetics of enrofloxacin after oral, intramuscular and bath administration in crucian carp (*Carassius auratus gibelio*). *J. Vet. Pharmacol. Ther.* 41, 159–162. doi: 10.1111/jvp.12428
- Shan, Q., Wang, J., Yang, F., Ma, L., Yin, Y., Liu, S., et al. (2018). Pharmacokinetics of cefquinome in crucian carp (*Carassius auratus gibelio*) after oral, intramuscular, intraperitoneal, and bath administration. *J. Vet. Pharmacol. Ther.* 41, 734–738. doi: 10.1111/jvp.12510
- Shang, J., Sun, H., Liu, H., Chen, F., Tian, S., Pan, P., et al. (2017). Comparative analyses of structural features and scaffold diversity for purchasable compound libraries. *J. Cheminform.* 9:25. doi: 10.1186/s13321-017-0212-4
- Sheridan, R. P., Feuston, B. P., Maiorov, V. N., and Kearsley, S. K. (2004). Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. *J. Chem. Inf. Comput. Sci.* 44, 1912–1928. doi: 10.1021/ci049782w
- Shukla, R., Singh, A. P., Sonar, P. K., Mishra, M., and Saraf, S. K. (2016). Schiff bases of benzothiazol-2-ylamine and thiazolo[5,4-b]pyridin-2-ylamine as anticonvulsants: synthesis, characterization and toxicity profiling. *Cent. Nerv. Syst. Agents Med. Chem.* 16, 240–248. doi: 10.2174/1871524916666160428110728
- Silva, G., Zuravski, L., Duarte, M., Machado, M., and Oliveira, L. (2019). Fluconazole induces genotoxicity in cultured human peripheral blood mononuclear cells via immunomodulation of TNF- $\alpha$ , IL-6, and IL-10: new challenges for safe therapeutic regimens. *Immunopharmacol. Immunotoxicol.* 41:123–129. doi: 10.1080/08923973.2019.1566357
- Silva-Junior, E. F., Aquino, T. M., and Araujo-Junior, J. X. (2017). Quantum mechanical (QM) calculations applied to ADMET drug prediction: a review. *Curr. Drug Metab.* 18, 511–526. doi: 10.2174/1389200218666170316094514
- Simhadri Vsdna, N., Muniappan, M., Kannan, I., and Viswanathan, S. (2017). Phytochemical analysis and docking study of compounds present in a polyherbal preparation used in the treatment of dermatophytosis. *Curr. Med. Mycol.* 3, 6–14. doi: 10.29252/cmm.3.4.6
- Simoneau, C., Van den Eede, L., and Valzacchi, S. (2012). Identification and quantification of the migration of chemicals from plastic baby bottles used as substitutes for polycarbonate. *Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess.* 29, 469–480. doi: 10.1080/19440049.2011.644588
- Song, J. C., Gao, H., Qiu, H. B., Chen, Q. B., Cai, M. H., Zhang, M. Z., et al. (2018). The pharmacokinetics of dexmedetomidine in patients with obstructive jaundice: a clinical trial. *PLoS ONE* 13:e0207427. doi: 10.1371/journal.pone.0207427
- Song, L., Yu, Z., Xu, Y., Li, X., Liu, X., Liu, D., et al. (2020). Preliminary physiologically based pharmacokinetic modeling of renally cleared drugs in chinese pregnant women. *Biopharm. Drug Dispos.* 41:248–267. doi: 10.1002/bdd.2243
- Stillhart, C., Parrott, N. J., Lindenberg, M., Chalus, P., Bentley, D., and Szepes, A. (2017). Characterising drug release from immediate-release formulations of a poorly soluble compound, bismisanil, through absorption modelling and dissolution testing. *AAPS J.* 19, 827–836. doi: 10.1208/s12248-017-0060-1
- Sun, H., and Scott, D. O. (2010). Structure-based drug metabolism predictions for drug design. *Chem. Biol. Drug Des.* 75, 3–17. doi: 10.1111/j.1747-0285.2009.00899.x
- Sushko, I., Novotarskyi, S., Korner, R., Pandey, A. K., Cherkasov, A., Li, J., et al. (2010). Applicability domains for classification problems: benchmarking of distance to models for ames mutagenicity set. *J. Chem. Inf. Model.* 50, 2094–2111. doi: 10.1021/ci100253r
- Tabeshpour, J., Sahebkar, A., Zirak, M. R., Zeinali, M., Hashemzaei, M., Rakhshani, S., et al. (2018). Computer-aided drug design and drug pharmacokinetic prediction: a mini-review. *Curr. Pharm. Des.* 24, 3014–3019. doi: 10.2174/1381612824666180903123423
- Takac, M. J.-M., Casimiro Magina, J. D., and Takac, T. (2019). Evaluation of phenylethylamine type entactogens and their metabolites relevant to ecotoxicology - a QSAR study. *Acta Pharm.* 69, 563–584. doi: 10.2478/acph-2019-0038
- Tan, B. H., Ahemad, N., Pan, Y., Palanisamy, U. D., Othman, I., Yip, B. C., et al. (2018). Cytochrome P450 2C9-natural antiarthritic interactions: evaluation of inhibition magnitude and prediction from *in vitro* data. *Biopharm. Drug Dispos.* 39, 205–217. doi: 10.1002/bdd.2127
- Tang, Y., Zhu, W., Chen, K., and Jiang, H. (2006). New technologies in computer-aided drug design: toward target identification and new chemical entity discovery. *Drug Discov. Today Technol.* 3, 307–313. doi: 10.1016/j.ddtec.2006.09.004
- Tao, L., Zhang, P., Qin, C., Chen, S. Y., Zhang, C., Chen, Z., et al. (2015). Recent progresses in the exploration of machine learning methods as *in-silico* ADME prediction tools. *Adv Drug Deliv. Rev.* 86, 83–100. doi: 10.1016/j.addr.2015.03.014
- Taroncher, M., Tolosa, J., Prosperini, A., and Ruiz, M. J. (2018). *In silico* and *in vitro* prediction of the toxicological effects of individual and combined mycotoxins. *Food Chem. Toxicol.* 122, 194–202. doi: 10.1016/j.fct.2018.09.055
- Taxak, N., Prasad, K. C., and Bharatam, P. V. (2013). Mechanistic insights into the bioactivation of phenacetin to reactive metabolites: a DFT study. *Comput. Theor. Chem.* 1007, 48–56. doi: 10.1016/j.comptc.2012.11.018
- Teorell, T. (1935). Studies on the “Diffusion Effect” upon ionic distribution. Some theoretical considerations. *Proc. Natl. Acad. Sci. U.S.A.* 21, 152–161. doi: 10.1073/pnas.21.3.152
- Thiel, C., Cordes, H., Conde, I., Castell, J. V., Blank, L. M., and Kuepfer, L. (2016). Model-based contextualization of *in vitro* toxicity data quantitatively predicts *in vivo* drug response in patients. *Arch. Toxicol.* 91, 865–883. doi: 10.1007/s00204-016-1723-x
- Tian, S., Li, Y., Wang, J., Zhang, J., and Hou, T. (2011). ADME evaluation in drug discovery. 9. Prediction of oral bioavailability in humans based on molecular properties and structural fingerprints. *Mol. Pharm.* 8, 841–851. doi: 10.1021/mp100444g
- T’Jollyn, H., Vermeulen, A., and Van Bocxlaer, J. (2018). PBPK and its virtual populations: the impact of physiology on pediatric pharmacokinetic predictions of tramadol. *AAPS J.* 21:8. doi: 10.1208/s12248-018-0277-7
- Tripathi, A. C., Upadhyay, S., Paliwal, S., and Saraf, S. K. (2018). N1-benzenesulfonyl-2-pyrazoline hybrids in neurological disorders: syntheses, biological screening and computational studies. *EXCLI J.* 17, 126–148. doi: 10.17179/excli2017-871
- Trivedi, A., Dave, S., Luhana, K., and Trivedi, S. (2020). Design and prediction of absorption, distribution, metabolism, excretion and toxicity properties of drugs for H1N1 Flu (Swine Flu) using *in-silico* approaches. *Biosci. Biotech. Res. Commun.* 13, 326–334. doi: 10.21786/bbrc/13.1/52
- Trivedi, A., Dave, S. M., and Patel, J. S. (2019). Design and prediction of absorption, distribution, metabolism, excretion and toxicity properties of drugs for dengue

- fever using *in-silico* approaches. *Biosci. Biotech. Res. Commun.* 12, 1201–1209. doi: 10.21786/bbrc/12.4/45
- Uba, A. I., and Yeleki, K. (2017). Identification of potential isoform-selective histone deacetylase inhibitors for cancer therapy: a combined approach of structure-based virtual screening, ADMET prediction and molecular dynamics simulation assay. *J. Biomol. Struct. Dyn.* 36, 3231–3245. doi: 10.1080/07391102.2017.1384402
- Vaidhyanathan, S., Wang, X., Crison, J., Varia, S., Gao, J. Z. H., Saxena, A., et al. (2018). Bioequivalence comparison of pediatric dasatinib formulations and elucidation of absorption mechanisms through integrated PBPK modeling. *J. Pharm. Sci.* 108, 741–749. doi: 10.1016/j.xphs.2018.11.005
- Van Bossuyt, M., Van Hoeck, E., Raitano, G., Vanhaecke, T., Benfenati, E., Mertens, B., et al. (2018). Performance of *in silico* models for mutagenicity prediction of food contact materials. *Toxicol. Sci.* 163, 632–638. doi: 10.1093/toxsci/kfy057
- van de Waterbeemd, H., and Gifford, E. (2003). ADMET *in silico* modelling: towards prediction paradise? *Nat. Rev. Drug Discov.* 2, 192–204. doi: 10.1038/nrd1032
- Vaz, R. J., Li, Y., Munson, M., Elliot, M., and Thuraiatnam, S. (2018). Amelioration of mechanism-based inactivation of CYP3A4 by a H-PGDS inhibitor. *Bioorg. Med. Chem. Lett.* 28, 3046–3049. doi: 10.1016/j.bmcl.2018.07.049
- Verheyen, G. R., Braeken, E., Van Deun, K., and Van Miert, S. (2017). Evaluation of existing (Q)SAR models for skin and eye irritation and corrosion to use for REACH registration. *Toxicol. Lett.* 265, 47–52. doi: 10.1016/j.toxlet.2016.11.007
- Verma, S. K., and Thareja, S. (2017). Structure based comprehensive modelling, spatial fingerprints mapping and ADME screening of curcumin analogues as novel ALR2 inhibitors. *PLoS ONE* 12:e0175318. doi: 10.1371/journal.pone.0175318
- Voelckner, N. M. F., Voelckner, A., Costa, J., Sy, S. K. B., Hermes, J., Weitzel, J., et al. (2017). Dermal pharmacokinetics of pyrazinamide determined by microdialysis sampling in rats. *Int. J. Antimicrob. Agents* 51, 190–196. doi: 10.1016/j.ijantimicag.2017.10.001
- Vuu, I., Coles, L. D., Maglalang, P., Leppik, I. E., Worrell, G., Crepeau, D., et al. (2016). Intravenous topiramate: pharmacokinetics in dogs with naturally occurring epilepsy. *Front. Vet. Sci.* 3:107. doi: 10.3389/fvets.2016.00107
- Wang, J., and Hou, T. (2009). Recent advances on *in silico* ADME modeling. *Annu. Rep. Comput. Chem.* 5, 101–127. doi: 10.1016/S1574-1400(09)00505-2
- Wang, J., Peng, H., Kong, J., Zhao, T., Zhang, S., and Cao, X. (2017). Pharmacokinetic profile of ceftiofur hydrochloride injection in lactating holstein dairy cows. *J. Vet. Pharmacol. Ther.* 41, 301–306. doi: 10.1111/jvp.12469
- Wang, J., Xue, J., Kong, J., Li, J., Zhang, S., and Cao, X. (2018a). Evaluation of pharmacokinetic properties of vitacoxib in fasted and fed horses. *J. Vet. Pharmacol. Ther.* 41, 843–847. doi: 10.1111/jvp.12703
- Wang, J., Zhao, T., Sun, X., Liu, Y., Zhu, J., Zhang, S., et al. (2018b). Pharmacokinetics of tildipirosin in beagle dogs. *J. Vet. Pharmacol. Ther.* 41, E49–E52. doi: 10.1111/jvp.12453
- Wang, J. H., Endsley, A. N., Green, C. E., and Martin, A. C. (2016). Utilizing native fluorescence imaging, modeling and simulation to examine pharmacokinetics and therapeutic regimen of a novel anticancer prodrug. *BMC Cancer* 16:524. doi: 10.1186/s12885-016-2508-6
- Wang, J. Y., Chen, H., Wang, Y. Y., Wang, X. Q., Chen, H. Y., Zhang, M., et al. (2017). Network pharmacological mechanisms of *Vernonia anthelmintica* (L.) in the treatment of vitiligo: isorhamnetin induction of melanogenesis via up-regulation of melanin-biosynthetic genes. *BMC Syst. Biol.* 11:103. doi: 10.1186/s12918-017-0486-1
- Wang, L., Hai, Y., An, L., Chen, J., Liang, R., and He, X. (2017). Rapid screening the potential mechanism-based inhibitors of CYP3A4 from Tripterygium wilfordii based on computer approaches combined with *in vitro* bioassay. *Bioorg. Med. Chem.* 25, 2689–2700. doi: 10.1016/j.bmc.2017.03.037
- Wang, L., Yue, H., Huang, N., Xue, G., Liu, W., and Xin, H. (2018). Human cytochrome P450 enzyme inhibition profile of three flavonoids isolated from *Psoralea corylifolia*: *in silico* predictions and experimental validation. *N. J. Chem.* 42, 10922–10934. doi: 10.1039/C7NJ00884H
- Wang, S., Sun, H., Liu, H., Li, D., Li, Y., and Hou, T. (2016). ADMET evaluation in drug discovery. 16. predicting hERG blockers by combining multiple pharmacophores and machine learning approaches. *Mol. Pharm.* 13, 2855–66. doi: 10.1021/acs.molpharmaceut.6b00471
- Wang, Y., Lin, W., Wu, N., He, X., Wang, J., Feng, Z., et al. (2018). An insight into paracetamol and its metabolites using molecular docking and molecular dynamics simulation. *J. Mol. Model.* 24:243. doi: 10.1007/s00894-018-3790-9
- Wang, Y., Xing, J., Xu, Y., Zhou, N., Peng, J., Xiong, Z., et al. (2015). *In silico* ADME/T modelling for rational drug design. *Q. Rev. Biophys.* 48, 488–515. doi: 10.1017/S0033583515000190
- Waseem, D., Butt, A. F., Haq, I. U., Bhatti, M. H., and Khan, G. M. (2017). *Daru* 25:8. doi: 10.1186/s40199-017-0174-0
- Watanabe, Y., Fukuyoshi, S., Kato, K., Hiratsuka, M., Yamaotsu, N., Hirono, S., et al. (2017). Investigation of substrate recognition for cytochrome P450 1A2 mediated by water molecules using docking and molecular dynamics simulations. *J. Mol. Graph. Model.* 74, 326–336. doi: 10.1016/j.jmgm.2017.04.006
- Watford, S., Ly, P., Wignall, J., Shin, R., Martin, M. T., and Friedman, K. P. (2019). ToxRefDB version 2.0: improved utility for predictive and retrospective toxicology analyses. *Reprod. Toxicol.* 89, 145–158. doi: 10.1016/j.reprotox.2019.07.012
- White, J., Wrzesinski, C., Green, M., Johnson, G. T., McCluskey, J. D., Abrits, A., et al. (2016). A novel method for deriving thresholds of toxicological concern for vaccine constituents. *Toxicol. Mech. Methods* 26, 270–275. doi: 10.3109/15376516.2016.1170250
- Williams, A. J., Grulke, C. M., Edwards, J., McEachran, A. D., Mansouri, K., Baker, N. C., et al. (2017). The comptox chemistry dashboard: a community data resource for environmental chemistry. *J. Cheminform.* 9:61. doi: 10.1186/s13321-017-0247-6
- Wishart, D. S. (2007). Improving early drug discovery through ADME modelling: an overview. *Drugs R D* 8, 349–362. doi: 10.2165/00126839-200708060-00003
- 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
- Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., et al. (2008). DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 36, D901–D906. doi: 10.1093/nar/gkm958
- Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* 34, D668–D672. doi: 10.1093/nar/gkj067
- Wu, Z., Lei, T., Shen, C., Wang, Z., Cao, D., and Hou, T. (2019). ADMET evaluation in drug discovery. 19. reliable prediction of human cytochrome P450 inhibition using artificial intelligence approaches. *J. Chem. Inf. Model.* 59, 4587–4601. doi: 10.1021/acs.jcim.9b00801
- Xiao, H., Peng, H., Zhao, T., Kong, J., Xue, J., Wang, J., et al. (2018). The pharmacokinetics of moxidectin following intravenous and topical administration to swine. *J. Vet. Pharmacol. Ther.* 42, 111–115. doi: 10.1111/jvp.12693
- Xu, Y., Pei, J., and Lai, L. (2017). Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction. *J. Chem. Inf. Model.* 57, 2672–2685. doi: 10.1021/acs.jcim.7b00244
- Yadav, S., Pandey, S. K., Singh, V. K., Goel, Y., Kumar, A., and Singh, S. M. (2017). Molecular docking studies of 3-bromopyruvate and its derivatives to metabolic regulatory enzymes: implication in designing of novel anticancer therapeutic strategies. *PLoS ONE* 12:e0176403. doi: 10.1371/journal.pone.0176403
- Yamashita, F., and Hashida, M. (2004). *In silico* approaches for predicting ADME properties of drugs. *Drug Metab. Pharmacokinet.* 19, 327–338. doi: 10.2133/dmpk.19.327
- Yang, F., Wang, B., Liu, Z., Xia, X., Wang, W., Yin, D., et al. (2017). Prediction of a therapeutic dose for buagafuran, a potent anxiolytic agent by physiologically based pharmacokinetic/pharmacodynamic modeling starting from pharmacokinetics in rats and human. *Front. Pharmacol.* 8:683. doi: 10.3389/fphar.2017.00683
- Yang, H., Lou, C., Sun, L., Li, J., Cai, Y., Wang, Z., et al. (2018a). admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics* 35, 1067–1069. doi: 10.1093/bioinformatics/bty707
- Yang, H., Sun, L., Li, W., Liu, G., and Tang, Y. (2018b). *In silico* prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Front. Chem.* 6:30. doi: 10.3389/fchem.2018.00129
- Yang, H., Sun, L., Wang, Z., Li, W., Liu, G., and Tang, Y. (2018c). ADMETopt: a web server for ADMET optimization in drug design via scaffold hopping. *J. Chem. Inf. Model.* 58, 2051–2056. doi: 10.1021/acs.jcim.8b00532

- Yang, Z. Y., Yang, Z. J., Dong, J., Wang, L. L., Zhang, L. X., Ding, J. J., et al. (2019). Structural analysis and identification of colloidal aggregators in drug discovery. *J. Chem. Inf. Model.* 59, 3714–3726. doi: 10.1021/acs.jcim.9b00541
- Ye, N., Monk, S. A., Daga, P., Bender, D. M., Rosen, L. B., Mullen, J., et al. (2018). Clinical bioavailability of the novel BACE1 inhibitor lanabecestat (AZD3293): assessment of tablet formulations versus an oral solution and the impact of gastric pH on pharmacokinetics. *Clin. Pharmacol. Drug Dev.* 7, 233–243. doi: 10.1002/cpdd.422
- Yongye, A. B., and Medina-Franco, J. L. (2013). Systematic characterization of structure-activity relationships and ADME-T compliance: a case study. *Drug Discov. Today* 18, 732–739. doi: 10.1016/j.drudis.2013.04.002
- Yu, H., and Adedoyin, A. (2003). ADME-Tox in drug discovery: integration of experimental and computational technologies. *Drug Discov. Today* 8, 852–861. doi: 10.1016/S1359-6446(03)02828-9
- Yu, Y., Zhou, Y. F., Sun, J., Shi, W., Liao, X. P., and Liu, Y. H. (2017). Pharmacokinetic and pharmacodynamic modeling of sarafloxacin against avian pathogenic *Escherichia coli* in muscovy ducks. *BMC Vet. Res.* 13:47. doi: 10.1186/s12917-017-0964-0
- Yugandhar, P., Kumar, K. K., Neeraja, P., and Savithramma, N. (2017). Isolation, characterization and *in silico* docking studies of synergistic estrogen receptor a anticancer polyphenols from *Syzygium alternifolium* (Wt.) Walp. *J. Interact. Ethnopharmacol.* 6, 296–310. doi: 10.5455/jice.20170709031835
- Zhang, C., Ding, S., Fang, Y., Zhang, L., Hu, W., Lu, J., et al. (2017). Iohexol clearance for determination of glomerular filtration rate in cynomolgus monkeys (*Macaca fascicularis*). *J Am Assoc Lab Anim Sci.* 56, 330–333.
- Zhang, L., Zhang, H., Ai, H. X., Hu, H., Li, S. M., Zhao, J., et al. (2018). Applications of machine learning methods in drug toxicity prediction. *Curr. Top. Med. Chem.* 18, 987–997. doi: 10.2174/1568026618666180727152557
- Zhang, L., Zhao, L., Liu, Y., Liu, J., and Li, X. (2017). Pharmacokinetics of tilmicosin in healthy pigs and in pigs experimentally infected with *Haemophilus parasuis*. *J. Vet. Sci.* 18, 431–437. doi: 10.4142/jvs.2017.18.4.431
- Zhang, N., Wu, Y., Huang, Z., Yao, L., Zhang, L., Cai, Q., et al. (2017). The PK-PD relationship and resistance development of danofloxacin against *Mycoplasma gallisepticum* in an *in vivo* infection model. *Front. Microbiol.* 8:926. doi: 10.3389/fmicb.2017.00926
- Zhang, P., Hao, H., Li, J., Ahmad, I., Cheng, G., Chen, D., et al. (2016). The epidemiologic and pharmacodynamic cutoff values of tilmicosin against *Haemophilus parasuis*. *Front. Microbiol.* 7:385. doi: 10.3389/fmicb.2016.00385
- Zhao, P., Zhang, L., Grillo, J. A., Liu, Q., Bullock, J. M., Moon, Y. J., et al. (2011). Applications of physiologically based pharmacokinetic (PBPK) modeling and simulation during regulatory review. *Clin. Pharmacol. Ther.* 89, 259–267. doi: 10.1038/clpt.2010.298
- Zhou, D., Afzelius, L., Grimm, S. W., Andersson, T. B., Zauhar, R. J., and Zamora, I. (2006). Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions. *Drug Metab. Dispos.* 34, 976–983. doi: 10.1124/dmd.105.008631
- Zhou, Q., Zhang, G., Wang, Q., Liu, W., Huang, Y., Yu, P., et al. (2017). Pharmacokinetic/pharmacodynamic modeling of tulathromycin against *Pasteurella multocida* in a porcine tissue cage model. *Front. Pharmacol.* 8:392. doi: 10.3389/fphar.2017.00392
- Zhu, J., Wang, J., Yu, H., Li, Y., and Hou, T. (2011). Recent developments of *in silico* predictions of oral bioavailability. *Comb. Chem. High Throughput Screen.* 14, 362–374. doi: 10.2174/138620711795508368
- Zhu, L., Zhang, Y., Yang, J., Wang, Y., Zhang, J., Zhao, Y., et al. (2016). Prediction of the pharmacokinetics and tissue distribution of levofloxacin in humans based on an extrapolated PBPK model. *Eur. J. Drug Metab. Pharmacokinet.* 41, 395–402. doi: 10.1007/s13318-015-0271-8
- Zhuang, J., Xing, X., Wang, D., Du, Z., Wang, J., Dong, Y., et al. (2018). Toxicity assessment of the extractables from multi-layer coextrusion polyethylene bags exposed to pH=5 solution containing 4% benzyl alcohol and 0.1M sodium acetate. *Regul. Toxicol. Pharmacol.* 94, 47–56. doi: 10.1016/j.yrtph.2018.01.015
- Zhuang, S., Zhang, L., Zhan, T., Lu, L., Zhao, L., Wang, H., et al. (2018). Binding specificity determines the cytochrome P450 3A4 mediated enantioselective metabolism of metconazole. *J. Phys. Chem. B* 122, 1176–1184. doi: 10.1021/acs.jpcc.7b11170
- Zhuang, X., and Lu, C. (2016). PBPK modeling and simulation in drug research and development. *Acta Pharm. Sin. B* 6, 430–440. doi: 10.1016/j.apsb.2016.04.004

**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.

Copyright © 2020 Wu, Zhou, Li, Shen, Chen, Wang, Liang, Tan and Huang. 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.



# Virtual Screening and Bioactivity Evaluation of Novel Androgen Receptor Antagonists From Anti-PCa Traditional Chinese Medicine Prescriptions

Wenya Han, Yuqi Shi, Jie Su, Zhennan Zhao, Xin Wang, Jiazhong Li\* and Huanxiang Liu

School of Pharmacy, Lanzhou University, Lanzhou, China

## OPEN ACCESS

### Edited by:

Zunnan Huang,  
Guangdong Medical University, China

### Reviewed by:

Tingting Feng,  
University of Pittsburgh, United States  
Abdullahi Ibrahim Uba,  
Beijing Computational Science  
Research Center, China

Ruo-Xu Gu,  
Max Planck Institute for Biophysical  
Chemistry, Germany

### \*Correspondence:

Jiazhong Li  
lijiazhong@lzu.edu.cn

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 13 July 2020

**Accepted:** 12 August 2020

**Published:** 17 September 2020

### Citation:

Han W, Shi Y, Su J, Zhao Z, Wang X,  
Li J and Liu H (2020) Virtual Screening  
and Bioactivity Evaluation of Novel  
Androgen Receptor Antagonists From  
Anti-PCa Traditional Chinese Medicine  
Prescriptions. *Front. Chem.* 8:582861.  
doi: 10.3389/fchem.2020.582861

Prostate cancer (PCa), a type of malignancy that arises in the prostate gland, is the most commonly diagnosed neoplasm and the second leading cause of cancer-related deaths in men. Acquisition of resistance to conventional therapy is a major problem for PCa patient treatment. Androgen receptor (AR) signaling pathway is necessary in the pathogenesis of prostate cancer, and there is a heightened interest in finding novel AR antagonists that target AR and its regulatory pathways. In our search for novel androgen receptor antagonists, we focus on the Traditional Chinese Medicine (TCM), which has been used for thousands of years to prove effective in the treatment of cancer. In this study, we collected 653 traditional Chinese medicine prescriptions that have certain therapeutic effect to prostate cancer, including the prescriptions and even the folk prescriptions. After summarizing the frequency of herbs and gathering the natural products contained in these prescriptions, we built a natural products database to do computer-aided virtual screening and drug-like evaluation to find potential AR antagonists. Totally 25 compounds were submitted to experimental biological activity tests. Through the MTT cell proliferation experiment, 5 chemicals were found to inhibit the proliferation of LNCaP cells in a concentration-dependent manner. Especially, MoL\_11 was found to have good antagonistic activity and significantly inhibit fluorescence enzyme activity by the AR reporter gene experiment. Finally, the molecular dynamics simulation method was used to study the interaction between the most active compound MoL\_11 and the wild-type and F876L mutant androgen receptor (WT/F876L AR), and it was found that F876L AR could not cause resistance to MoL\_11.

**Keywords:** prostate cancer, androgen receptor, traditional Chinese medicine, virtual screening, activity evaluation

## INTRODUCTION

Prostate cancer (PCa), a common non-cutaneous cancer, accounts for almost one-fifth of new cancer diagnoses, and is the second leading cause of cancer-related deaths in men in the United States (Siegel et al., 2018). According to Cancer statistics, there will be an estimated 191,930 new cases, 33,330 deaths in 2020 (Siegel et al., 2020). 5% of global deaths of prostate cancer come from China (Ferlay et al., 2015). Aging and westernization of lifestyles have increased the incidence



of prostate cancer (Møller et al., 2015). The people in the developed areas and the elderly people have a higher incidence, and the survival rate of prostate cancer is not optimistic, especially in rural areas.

Androgen receptor (AR) is a ligand-dependent trans-regulatory protein that belongs to the steroid receptor in the nuclear receptor superfamily (Gao et al., 2005). Androgen receptor can be activated through the binding of endogenous androgens, such as testosterone and its metabolite dihydrotestosterone (DHT), to regulate the expression of a series of key downstream genes, causing uncontrolled proliferation of prostate cells and triggering prostate cancer (Tan et al., 2015). So AR is a highly explored target in the development and progression of PCa, regarded as the crucial therapeutic target for PCa (Brinkmann et al., 1999; Augello et al., 2014; Attard et al., 2016). AR antagonists have been commonly used to treat PCa for decades, such as flutamide, bicalutamide, enzalutamide, darolutamide, and apalutamide. These drugs have excellent effects initially, but long-term use of the drugs has resulted in the emergence of point mutations in the ligand binding domain (LBD) of AR, which eventually cause the AR to be resistant to antiandrogens, such as point mutations of T877A, W741L, and F876L etc. (Taplin et al., 1999; Hara et al., 2003; Tran et al., 2009).

Natural extracts have great potential in traditional medicines for the treatment of diseases and are also an essential resource for new drug discovery with their novel, rich and diverse chemical structures. Traditional Chinese Medicine (TCM) has been extensively used as an alternate treatment for prostate cancer. Many herbal extracts have been shown to inhibit the development of prostate cancer. Generally, TCM achieves therapeutic effects by targeting multiple physiological pathways, including AR, exerting pharmacological activities by numerous active agents and components. The complex active substances in

herbs are still elusive and not accurately detected solely by routine methods (Li et al., 2015; Zhang et al., 2015).

In recent years, virtual screening (VS) has widely used in lead compound identification, which exhibits undefeatable advantage and has become very popular in drug discovery and design based on its time, cost, resources, and labor saving (Hou and Xu, 2004; Kitchen et al., 2004; Xu et al., 2014; Huang and Wong, 2016).

In this study, we aim to screen novel AR antagonists in Chinese medicine prescriptions using virtual screening and bioassays. At the beginning, the traditional Chinese medicine prescriptions for prostate cancer were collected, and the herbs contained in them were summarized. Then, the ingredients of each herb were summarized, and the compound structure database was established. Then we conduct virtual screening protocol (Figure 1) to select potential compounds by using CDOCKER method and drug-likeness filters. The obtained 25 potential ingredients were further analyzed for bioassays, including MTT assay, AR reporter gene assay and immunofluorescence assay. The molecular simulation method was used to study the interaction mechanism between the most promising ingredient and WT 1T65/F876L AR, and the effect of F876L mutation on its AR antagonistic effect was studied.

## MATERIALS AND METHODS

### Computer-Aided Virtual Screening

#### Collection of Prescriptions for PCa

The Chinese medicine prescriptions to treat prostate cancer mainly come from the books “Anti-Tumor Chinese Medicine Highlights,” “Anti-Cancer Chinese Herbal Medicine,” “Anti-Cancer Chinese Herbal Medicine and Prescription,” “Materia Medica New,” “Materia Medica,” and other books for systematic

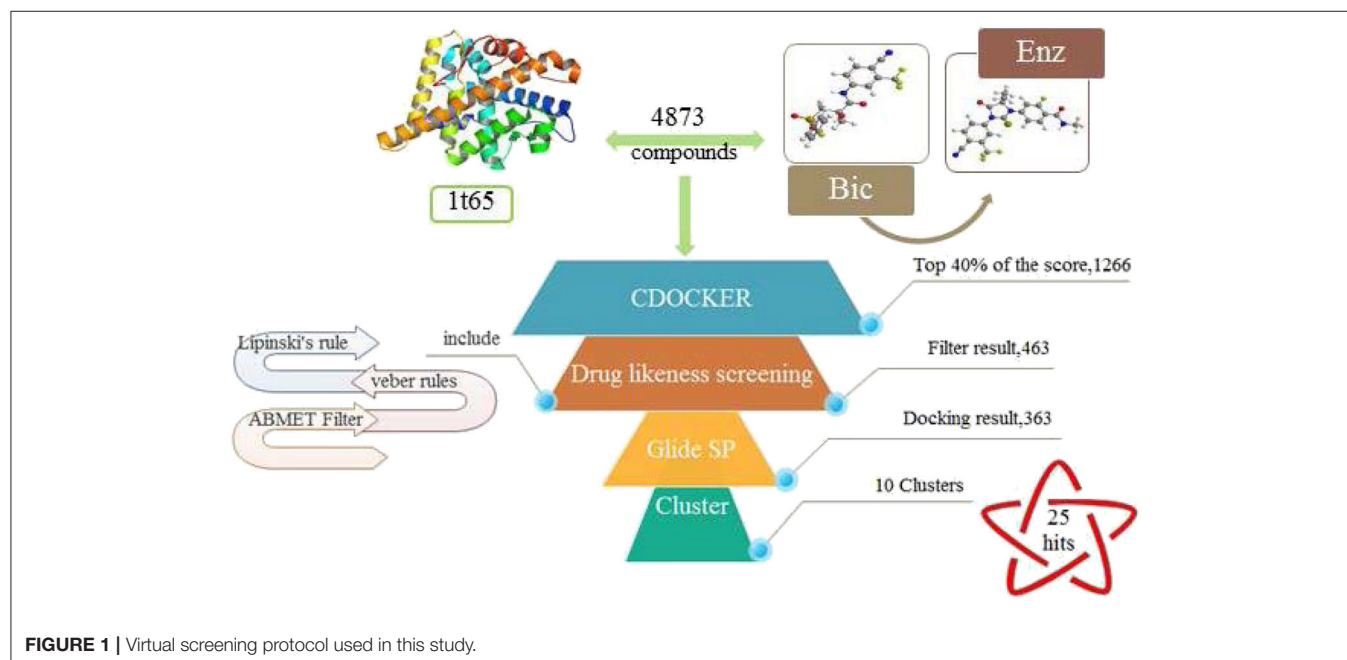


FIGURE 1 | Virtual screening protocol used in this study.

summing up prescriptions. As a result, we collected a total of 653 prescriptions.

### Chemical Database Construction

We counted the herbs contained in the prescriptions collected above, and reserved the herbs with high frequency ( $\geq 10$ ) for the succeeding research. Among different prescriptions, the herbs with high frequency should play important roles in anti-prostate cancer, including anti-AR effects. Then the existing chemical ingredients of each Chinese medicine were inquired by using Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP) (<https://tcmspwp.com/tcmsp.php>).

### Rough Docking Screening

The X-ray crystal structure of androgen receptor in complex with DHT (PDB entry: 1T65) was downloaded from RCSB Protein Data Bank (Estébanez-Perpi et al., 2005). The protein was prepared by using Prepare Protein tools in Accelrys Discovery Studio 2.5 software (Accelrys, 2010) and the active binding site was identified the same as endogenous androgen DHT present in the protein. Subsequently all these compounds were prepared using Prepare Ligands tools. CDOCKER is a grid-based molecular semi-flexible docking method that uses a CHARMM force field-based molecular dynamics (MD) search algorithm to dock ligands to receptor binding site (Sobhy et al., 2019). The torsions for ligands and receptors were set to be rotatable and rigid, respectively, Fast filtering was implemented using the CDOCKER protocol, the binding site radius was set to 12, the Top hits was 1, the pose was 1 and the other parameters were set as default, 3,173 small molecules docked in were obtained. The top 40% of the docked compound based on the final docking scores were retained, a total of 1,266 compounds were used for the following study.

### Drug Likeness Screening

Initially, compounds were filtered through Lipinski's rule of five, which means that when there are more than five hydrogen-bond donors, ten hydrogen-bond acceptors, molecular weight (MWT)  $> 500$  and calculated Log P (CLogP)  $> 5$ , compounds are more likely to have poor absorption or permeability (Lipinski et al., 2001). Therefore, these compounds were excluded. Subsequently, the remaining compounds were further screened by the Verb rules, which need to meet two criteria of (1) 10 or fewer rotatable bonds and (2) polar surface area equal to or  $< 140 \text{ \AA}^2$  (Veber et al., 2002) to guarantee the good oral bioavailability. Ultimately, ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties were used to select molecules precisely (Lagorce et al., 2008; Tian et al., 2015).

### High Precision Screening

To further screen compounds and get more reliable docking results, Glide XP (extra Precision) docking was performed to execute the additional molecular docking. The structures of these compounds obtained above were given to the corresponding OPLS\_2005 force field, and then were converted to their most probable charged states at  $\text{pH } 7.0 \pm 2$  and different tautomer were generated using the LigPrep module in Schrödinger

(Schrödinger, LLC, New York, NY). The compounds with docking scores higher than enzalutamide were chosen for further analysis.

### Novelty Analysis

The aim of this study is to obtain potential AR antagonists with rich diversity, so the hierarchical clustering method is used by the canvas module. The tanimoto coefficient based on binary fingerprints is used as a molecular similarity measure. The clustering similarity threshold was set to 0.98. Then we consider the following factors to choose suitable compounds, receptor-ligand binding mode, ligand steric hindrance, molecular skeleton, hydrocarbon group, N-heterocyclic compound, amino and amide, halogen, carboxyl, ester groups and so on.

### MD Simulations and MM/PBSA Calculations

The wild-type androgen receptor (WT AR) 2AXA and mutant receptor F876L AR were docked with the small molecules Enz and the most promising compound, respectively, by the CDOCKER module in Discovery Studio 2.5 software. We used Pymol (<http://www.pymol.org>) to mutate position 876 from phenylalanine to leucine based on the WT AR to obtain mutant F876L AR. In the docking parameters, the radius of the active binding site is set to 12, and the docking pose is set to 10. Finally, the best conformation with the most optimal docking score is selected for molecular dynamics (MD) studies.

The Gaussian09 software is used to calculate the atomic charge for small molecules, and the acpype script is used to convert it into a small molecule force field file that can be recognized by the software (Sousa da Silva and Vranken, 2012), giving the protein an Amber99SB force field (Nerenberg et al., 2011). Then the coordinate files of the small molecule and protein are combined, and the topology file of the complex system is constructed. The entire system is immersed in the cube box of the spc water model (Alsharif et al., 2014). The protein is at least 1.0 nm away from the box.  $\text{Na}^+$  and  $\text{Cl}^-$  ions are added to neutralize the system charge and make the whole system neutral.

The entire molecular dynamics simulation process is completed in Gromacs 5.1.4 software (Abraham et al., 2015). After the energy of the system is minimized to ensure the stability of the protein structure, the position of the heavy atoms is restricted, the molecules around the protein are balanced, and the entire system is gradually heated from 0 to 310 K. Finally 100 ns molecular dynamics simulation analysis was conducted under 1 atm pressure and 310 K. The time step is set to 2 fs, and the track coordinates are reserved every 5,000 steps.

The combined free energy and decomposition free energy are calculated using the MM-PBSA method.

### Experimental Test and Verification

#### Cell Culture and Reagents

LNCaP (Human prostate cancer cell line) was purchased from Hangzhou Qiannuo Biotechnology Co., LTD (Hangzhou, China). Cos-7 (African green monkey kidney cell transformed with SV40) was kindly provided by Stem Cell Bank of the Chinese Academy of Sciences (Shanghai, China). LNCaP cells were propagated in RPMI 1640 medium (Hyclone) supplemented

with 10% FBS and 1% penicillin—streptomycin. Cos-7 cells were grown in DMEM medium (HyClone) supplemented with 10% FBS and 1% penicillin-streptomycin. Cell cultures were maintained in a humidified atmosphere (37°C, 5% CO<sub>2</sub>). pcDNA3.1-AR and pMMTV-Luc were reorganized by Changsha Youbao Biological Co., Ltd. Lipofectamine3000 (Invitrogen) and Dual-Glo Luciferase Assay System (Promega) was purchased. Thiazolyl tetrazolium (MTT) was purchased from Solarbio Science & Technology. Enzalutamide (MDV-3100) was purchased from J&K. All other chemicals were obtained from Wuhan Tianzhi Biological Co., Ltd.

### Cell Proliferation Assay

Trypsinized LNCaP cells were seeded in 96-well plates at a density of 5,000 cells per well and incubated for 24 h. When cells were attached to the bottom, the original medium was discarded, and 150 µL of serially diluted compounds solution was added which were dissolved in DMSO. After 120 h of treatment with compounds, 15 µL MTT (5 mg/ml) was supplemented to each well, and followed by another 4 h of incubation. Then 150 µL of three linked dissolved solution per well was added to dissolve the crystal of formazan for overnight (the solution was prepared from 10% SDS, 5% isobutanol and 0.01 mol/L HCl in water). Afterwards, the absorbance at 570 nm was measured by a microplate reader, and the IC<sub>50</sub> values were calculated.

### AR Reporter Gene Assay

AR-antagonist nature of test compounds was exhibited by luciferase reporter gene assay using Cos-7 cells. The Cos-7 cells were seeded in 24-well plates at a density of 40,000 cells per well and incubated for 24 h, pcDNA3.1-AR, pMMTV-Luc and pRL-SV40 vectors were co-transfected into Cos-7 cells using lipofectamine 3000 according to the manufacturer's instructions. After transfection for 24 h, cells in each well were treated with 10 nm DHT or 10 nm DHT with 10 µm compounds, and placed in the incubator for another 24 h. Then cells were harvested with 150 µL of cell passive lysis buffer, Luciferase activity was determined with luciferase assay systems (Promega) following the manufacturer's protocol to detect the AR mediated transcriptional activity.

### Statistical Analysis

The result of the experiment shows the average value of the results from three independent repeated experiments. Statistical analysis was performed using the GraphPad Prism 5.0 statistical software (GraphPad software, Inc., San Diego, CA, USA). Data are expressed as means ± standard deviation (SD). Statistical analysis was carried out using one-way ANOVA followed by Tukey's multiple comparison, with  $p < 0.05$  considered statistically significant.

## RESULTS AND DISCUSSION

### Statistical Analysis of TCM Prescriptions

After statistical analysis of the TCM prescriptions, we found that 82 herbs have high frequency ( $\geq 10$ ) in the prescriptions and we summarized them in Table 1. The natural products contained

**TABLE 1 |** Chinese medicine with frequency  $\geq 10$  in anti-prostate cancer prescription.

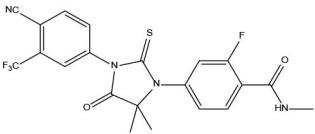
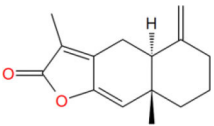
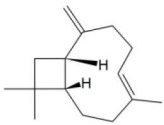
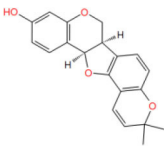
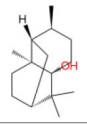
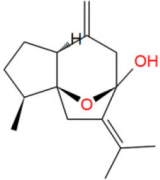
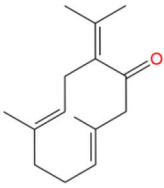
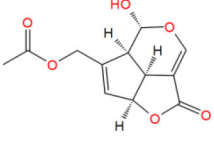
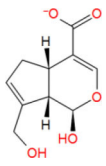
Chinese medicine name	
<i>Poria</i>	<i>Paeonia suffruticosa</i> Andr.
<i>Radix Glycyrrhizae</i>	<i>Malva verticillata</i> L.
<i>Coptis chinensis</i> Franch.	<i>Juncus effuses</i>
<i>Radix Astragali</i>	<i>Chenpi</i>
<i>Angelica sinensis</i>	<i>Eupatorium japonicum</i> Thunb.
<i>Magnolia officinalis</i> Rehd. et Wils	<i>Perotis indica</i> (L.) Kuntze
<i>Alisma orientalis</i> (Sam.) Juzep.	<i>Semen Persicae</i>
<i>Polyporus</i>	<i>Lindera aggregata</i> (Sims) Kosterm.
<i>Atractylodes macrocephala</i> Koidz.	<i>Panax ginseng</i> C. A. Mey.
<i>Cortex Phellodendri Chinensis</i>	<i>Sparganium stolonum</i> Buch. -Ham.
<i>Dioscorea tokoro</i> Makino	<i>Duchesnea indica</i> (Andr.) Focke
<i>Dianthus superbus</i> L.	<i>Gardenia jasminoides</i> Ellis
<i>Solanum nigrum</i> L.	<i>Pinellia ternata</i> (Thunb.) Breit.
<i>Akebia Decne.</i>	<i>Eucommia ulmoides</i>
<i>Dioscoreae Rhizoma</i>	<i>Pyrosia lingua</i> (Thunb.) Farwell
<i>Polygonatum sibiricum</i>	<i>Sargassum</i>
<i>Thallus laminariae</i>	<i>Carthamus tinctorius</i> L.
<i>Forsythiae Fructus</i>	<i>Ligusticum chuanxiong</i> Hort.
<i>Anemarrhena asphodeloides</i> Bge.	<i>Tetrapanax papyriferus</i>
<i>Curcuma zedoaria</i> (Christm.) Rosc	<i>Aconitum carmichaeli</i> Debx
<i>Achyranthes bidentata</i> Blume	<i>Fallopia multiflora</i>
<i>Sophora flavescens</i>	<i>Herba patriniae cum radice</i>
<i>Zingiber officinale</i> Roscoe	<i>Epimedium brevicornu</i> Maxim.
<i>Paeonia lactiflora</i> Pall.	<i>Tulipa edulis</i> (Miq.) Baker
<i>Salvia miltiorrhiza</i> Bunge	<i>Clematis chinensis</i>
<i>Polygonum aviculare</i> L.	<i>Lysimachia christinae</i> Hance
<i>Cinnamomum cassia</i> Presl	<i>Eclipta prostrata</i>
<i>Bupleurum chinense</i>	<i>Verbena officinalis</i> L.
<i>Cornus officinalis</i>	<i>Rhizoma Smilacis Glabrae</i>
<i>Radix Pseudostellariae</i>	<i>Vaccaria segetalis</i> (Neck.) Garcke
<i>Houttuynia cordata</i> Thunb.	<i>Hedyotis diffusa</i>
<i>Schisandrachinensis</i> (Turcz.) Baill	<i>Codonopsis pilosula</i> (Franch.) Nannf.
<i>Cuscuta chinensis</i>	<i>Rheum palmatum</i> L.
<i>Cistanche deserticola</i> Ma	<i>Scutellaria baicalensis</i> Georgi
<i>Morinda officinalis</i> How.	<i>Fructus Ligustri Lucidi</i>
<i>Taxillus sutchuenensis</i> (Lecomte) Danser	<i>Plantago asiatica</i> L.
<i>Scutellaria barbata</i> D. Don	<i>Lycii Fructus</i>
<i>Lobelia chinensis</i> Lour.	<i>Prunella vulgaris</i> L.
<i>Semen coicis</i>	<i>Cynanchum otophyllum</i>
<i>Kadsurae caulis</i>	<i>Lygodium japonicum</i> (Thunb.) Sw.
<i>Psoralea corylifolia</i> Linn.	<i>Rehmannia glutinosa</i> (Gaetn.) Libosch. ex Fisch. et Mey.

in these herbs were searched and collected from TCMSP. As a result, totally 4,873 compounds were obtained to construct a 3D molecular structure database.

### Virtual Screening of AR Antagonists

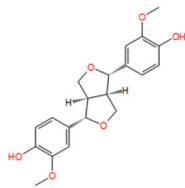
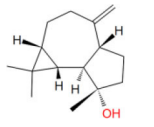
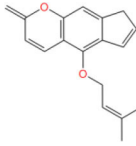
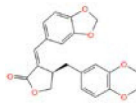
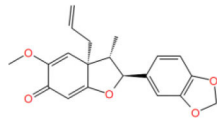
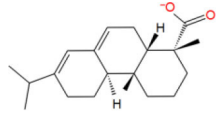
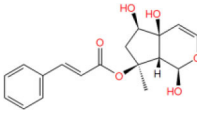
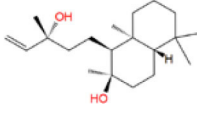
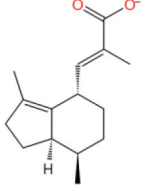
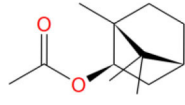
Virtual screening techniques have become a reliable and relatively inexpensive technique for the discovery of lead compounds and

**TABLE 2 |** The obtained potential ingredients by virtual screening.

No.	Name	Structure	cdocker_ interaction_ energy
Enz	Enzalutamide		-23.6075
MOL_1	Atractylenolide III		-32.3774
MOL_2	$\beta$ -Caryophyllene		-30.8992
MOL_3	Phaseolin		-27.8874
MOL_4	Patchouli alcohol		-30.2698
MOL_5	Isocurcumenol		-28.2755
MOL_6	Germacrone		-31.3891
MOL_7	Asperuloside		-34.1738
MOL_8	Geniposidic acid		-27.5956

(Continued)

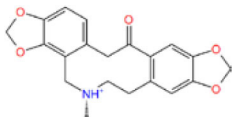
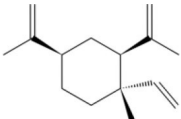
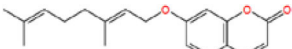
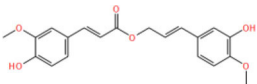
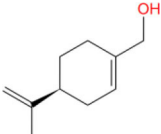
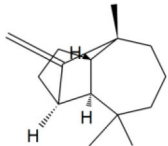
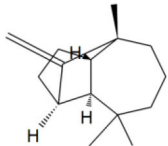
**TABLE 2 |** Continued

No.	Name	Structure	cdocker_ interaction_ energy
MOL_9	(-)-Pinoresinol		-42.336
MOL_10	Spathulenol		-34.1738
MOL_11	Isoimperatorin		-33.9276
MOL_12	Jatrophan		-35.1193
MOL_13	Burchellin		-30.0107
MOL_14	Abietic acid		-32.697
MOL_15	Harpagoside		-39.4463
MOL_16	Sclareol		-35.1172
MOL_17	Valerenic acid		-33.652
MOL_18	Bornyl acetate		-28.9967

(Continued)



TABLE 2 | Continued

No.	Name	Structure	cdocker_ interaction_ energy
MOL_19	Protopine		-38.5926
MOL_20	Beta-Elemene		-30.1592
MOL_21	Auraptene(rg)		-38.5015
MOL_22	Coniferyl ferulate		-10.0322
MOL_23	(-)-Perillyl alcohol		-25.1945
MOL_24	(+)-Longifolene		-26.5226
MOL_25	Alpha-bisabolol		-32.6259

have been widely used in hit identification. Docking power is a critical indicator of docking reliability for reproducing the experimental combination mode of antagonists (Tian et al., 2017). The endogenous ligand DHT was extracted from the protein 1T65 and then re-docked into the corresponding binding pocket using CDOCKER to calculate the RMSD value between the optimal pose and DHT, resulting in 0.3646 Å. below the threshold of 2.0 Å, which indicated that the docking method is reliable, and the docking protocol for virtual screening can be selected. As shown in **Figure 1**, the CDOCKER procedure was used at first to perform molecular docking for a rough screening of compounds. According to the -CDOCKER ENEGRY score, the top 40% chemicals with high docking scores, i.e., 1,266 compounds were retained. Subsequently, these compounds were

filtered according to Lipinski's rule of five, Veber rules and ADMET Filter, and 463 molecules were chosen for subsequent analysis. Finally, the Glide XP protocol with higher docking accuracy was exploited, and 365 satisfactory compounds were harvested. Then using the Hierarchical Clustering of the Canavas module in the Schrödinger software, 10 groups of compounds were gathered. Within these clusters, 25 active ingredients were selected and purchased for subsequent bioassays (as shown in **Table 2**).

## Cell Proliferation Assay

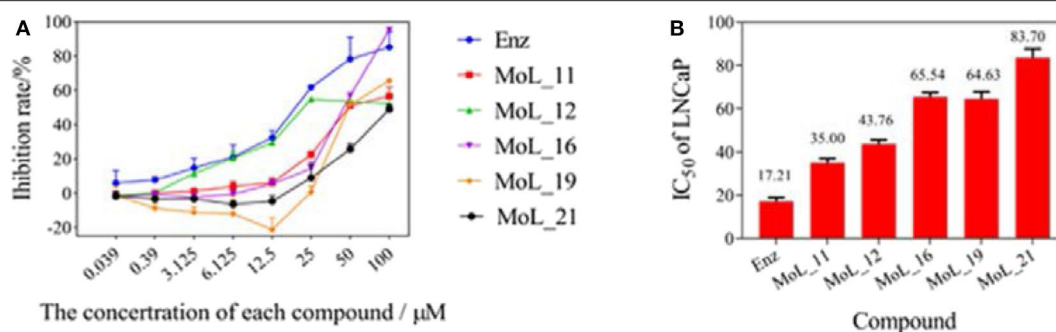
The cytotoxic effect of compounds was first analyzed by MTT assay. Using enzalutamide (Enz) as the positive control, LNCaP cell was treated with different concentrations of enzalutamide and compounds for 120 h.

The MTT results indicated that Enz and the five compounds can inhibit the growth of LNCaP cells in a dose-dependent manner, shown in **Figure 2A**. **Figure 2B** shows that the positive drug enzalutamide has an  $IC_{50}$  value of 17.21  $\mu$ M, while the compound MoL\_11 obtained by virtual screening has an  $IC_{50}$  value of 35.00  $\mu$ M, which is about twice of that for Enz. Among all the tested compounds, the  $IC_{50}$  value for MoL\_11 is the smallest, and the activity is relatively good. MoL\_11 is derived from *Gardenia jasminoides* Ellis, also known as Isoimperatorin. The  $IC_{50}$  values of MoL\_12, MoL\_16, and MoL\_19 increase in sequence. MoL\_12 is suzilactone, which is found in *Cistanche deserticola* Ma and *Panax ginseng* C. A. Mey. MoL\_16 is perillyl alcohol, present in *Salvia miltiorrhiza* Bunge. MoL\_19 is the original opioid, and *Panax ginseng* C. A. Mey., *Cortex Phellodendri Chinensis* all have this ingredient. The  $IC_{50}$  value of MoL\_21 is close to 5 times that of enzalutamide, which is orange peel olein and is present in the *Cortex Phellodendri Chinensis*.

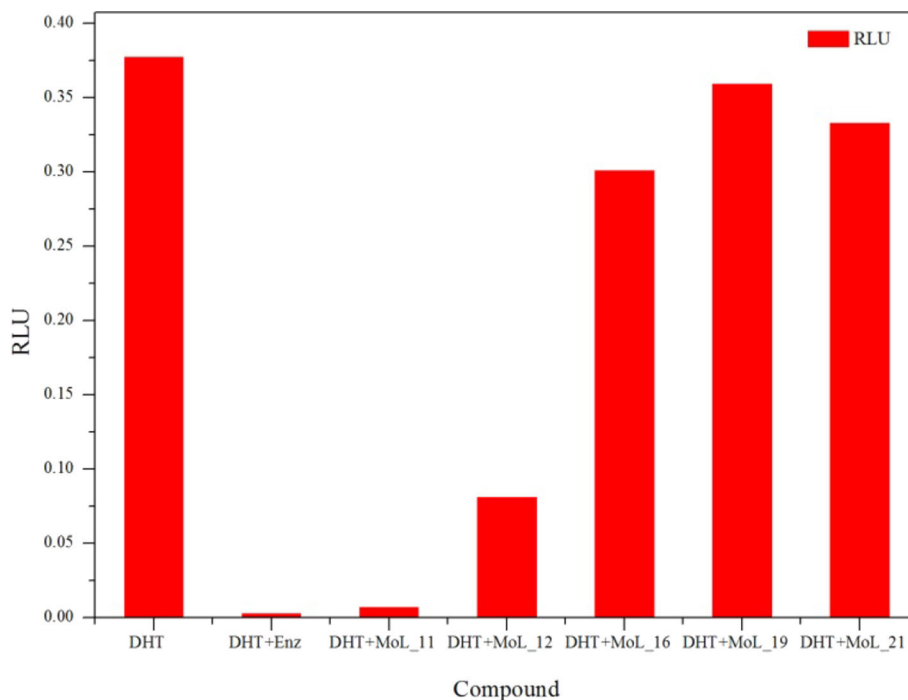
From the cdocker interaction energy in **Table 2**, we can see that among the above 6 compounds with significant  $IC_{50}$  values, the most active enzalutamide has the largest cdocker interaction energy, followed by the second best activity MoL\_11, then MoL\_16 with the third docking value, but its activity value is in the fifth place, At the same time, MoL\_21 with the worst activity value, ranked second to last in cdocker interaction energy. The best cdocker interaction energy of MoL\_10 has no activity value in the MTT experiment among all the compounds in **Table 2**. Therefore, the docking value is not directly proportional to the final activity value, it can only be used as a preliminary judgment whether the compound has an effect on the receptor, but their effects may not be effective in biological experiments.

## AR Antagonistic Effect

The dual luciferase reporter gene can detect the expression and regulation of AR protein. PMMTV-Luc was constructed as a reporter gene plasmid, which contains 5,005 bp base pairs and 216 bp MMTV-LTR promoter sequences. MMTV-LTR is a long terminal repeat sequence of mouse breast tumors. It has four ARE sequences and can be activated as a steroid response element. pcDNA3.1-AR, pMMTV-Luc and pRL-SV40 vectors were co-transfected into Cos-7 cells, and then the dual luciferase reporter gene kit was used to detect whether the compound acted on AR.



**FIGURE 2 |** The graph of LNCaP cell inhibition activities. **(A)** is a graph showing the inhibition rate of the 6 compounds with concentration. **(B)** is the IC<sub>50</sub> value of the 6 compounds for LNCaP.



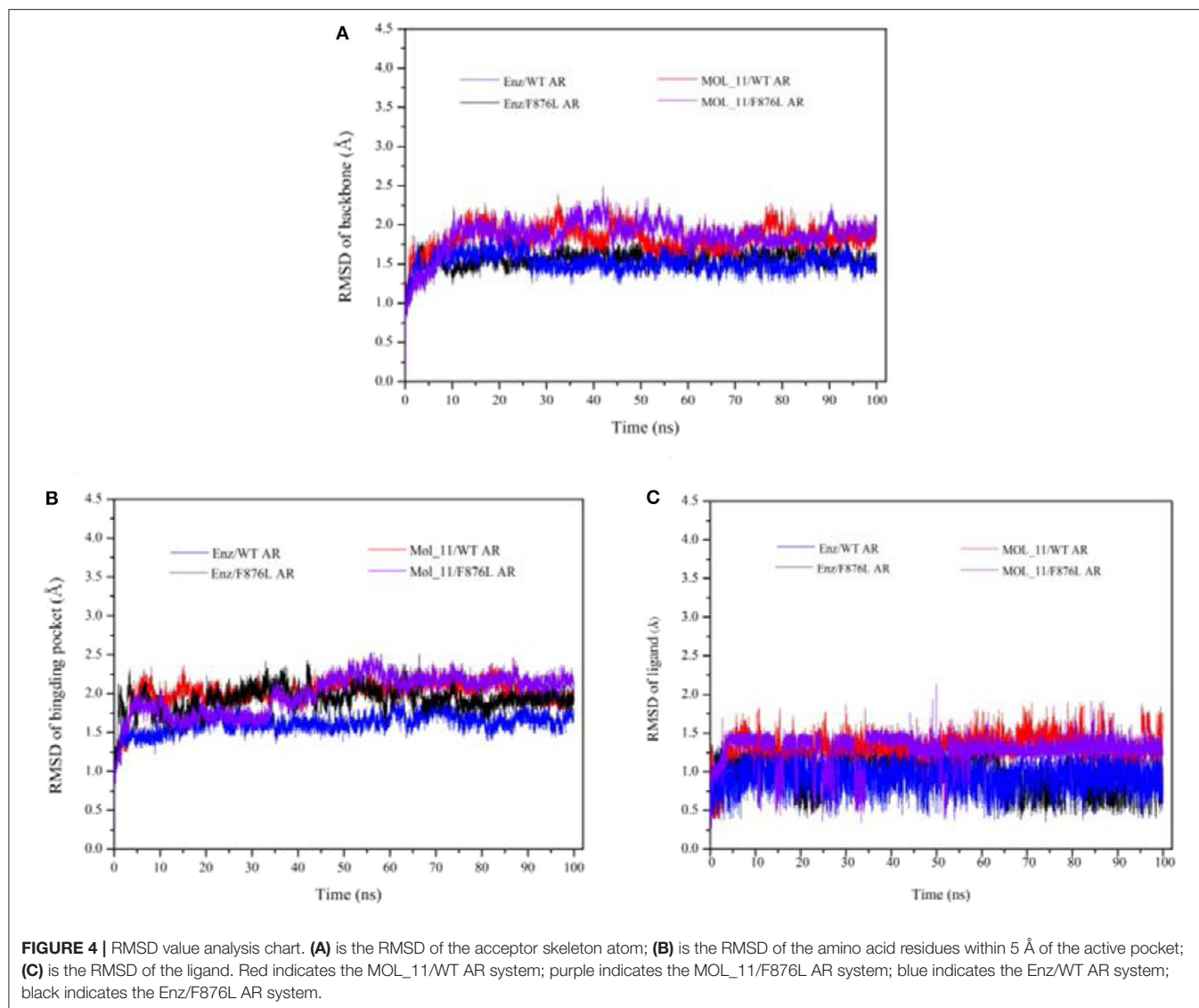
**FIGURE 3 |** Dual luciferase gene report results.

Through the MTT cell proliferation experiment, five small molecules with significant inhibitory effects on cells were selected to study AR reporter gene experiments. DHT was used as the model group, and enzalutamide was used as the positive control group. With reference to the model group, the positive control group had a significant inhibition of AR expression, as shown in **Figure 3**. Among all groups, enzalutamide has the strongest antagonistic activity, and the luciferase activity decreased by 93%. The second is MoL\_11, which also has obvious antagonistic activity, with an inhibition rate of 82%, indicating that MoL\_11 exerts antagonistic activity through interaction with AR, rather than other toxic effects. The inhibition rate of MoL\_12 is 20%, and MoL\_16, MoL\_19, and MoL\_21 have almost no antagonistic activity, indicating that they may have an effect on cell proliferation through cytotoxicity or non-AR pathway.

## Interaction Mechanism

From the above experimental results, we know that compound MoL\_11 can be a promising lead compound deserved further study. Then we conducted MD simulations to investigate the interaction of MoL\_11 with AR, taking Enz as a compare. We did four different MD analyses as stated in section 2.1.7, i.e., Enz/WT AR, Enz/F876L AR, MoL\_11/WT AR, and MoL\_11/F876L AR.

After 100 ns molecular dynamics simulation of the systems, the root mean square deviation (RMSD) was used to analyze and evaluate the stability of the entire simulation system. The RMSD values of the protein skeleton structure, amino acid residues within 5 Å of the active pocket and ligand atoms were calculated, as shown in **Figure 4**. From **Figure 4A**, it can be seen that the RMSD of the protein backbone of all systems became stable after 80 ns. The RMSD of the amino acid residues within 5 Å of the



active pocket tends to balance after 60 ns. The RMSD of the ligand atom is also very stable and the fluctuation is small.

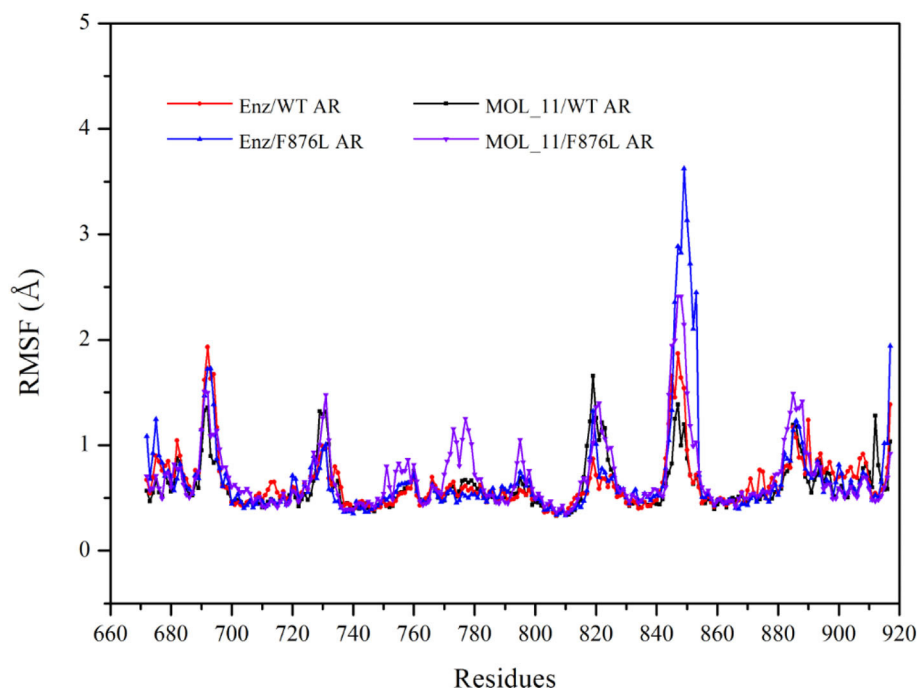
From the RMSF value analysis chart of the last 10 ns MD trajectory in **Figure 5**, it can be seen that RMSF value of the four systems fluctuate similarly. The MOL\_11/F876L AR system fluctuates greatly in Helix-6 (H6, residues 771–778) region that obviously different from other systems. Perhaps this is the reason why MOL\_11 is not resistant to F876L. In the Helix-3 (H3, residues 700–720) and Helix-5 (H5, residues 740–750) regions, the RMSF value fluctuates relatively smoothly, indicating that the amino acids in the active site of the androgen receptor are more stable than the loop region (690–695, 844–854).

Taking the RMSD of the ligand atom as the abscissa and the RMSD of the amino acid residues within 5 Å of the active pocket as the ordinate, the protein conformation was extracted through the lowest energy point to find the dominant conformation.

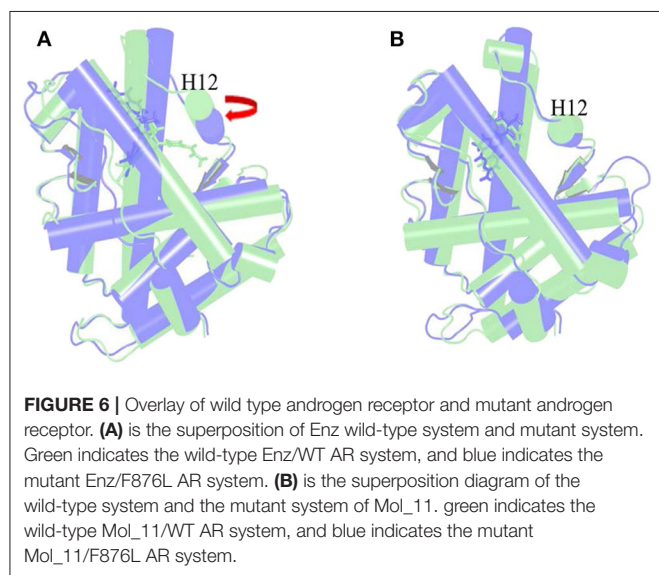
The  $\alpha$  Helix-12 (H12) is a stretch of folding protein (residue 892–908 region) on the ligand binding domain of

androgen receptor, and is a molecular switch for AR activation and inactivation. When AR is combined with androgen, the conformation of the H12 position changes, thereby covering the hormone-binding pocket and forming the activation functional area 2 (AF-2). This surface can recruit co-activators to promote AR transcription. If the small molecule destroys H12 and keeps it away from the binding pocket area, it is difficult to form an activation functional area, and then it will inhibit the transcription of AR and achieve the purpose of treating disease. This process is difficult to find directly in biological experiments, so it is further explained with the help of molecular dynamics simulations. Next, the action mode of the small molecule was observed through the H12 conformational change.

**Figure 6A** is the overlay of Enz wild-type system and mutant system, green indicates the wild-type Enz/WT AR system, and blue indicates the mutant Enz/F876L AR system. We can see that the conformation of H12 has changed significantly. In the Enz/WT AR system, the Enz structure is close to H12, making it



**FIGURE 5 |** RMSF value analysis chart. Red indicates the Enz/WT AR system; blue indicates the Enz/F876L AR system; black indicates MOL\_11/WT AR system; purple indicates the MOL\_11/F876L AR system.



**FIGURE 6 |** Overlay of wild type androgen receptor and mutant androgen receptor. **(A)** is the superposition of Enz wild-type system and mutant system. Green indicates the wild-type Enz/WT AR system, and blue indicates the mutant Enz/F876L AR system. **(B)** is the superposition diagram of the wild-type system and the mutant system of Mol\_11. Green indicates the wild-type Mol\_11/WT AR system, and blue indicates the mutant Mol\_11/F876L AR system.

difficult for H12 to cover the binding site, resulting the disorder of AF2. In the Enz/F876L AR system, Enz is farther away from H12, which causes H12 to cover the binding site, which leads to the formation of AF2 and the drug resistance problem. It also makes Enz change from an antagonist in the original wild-type system to an agonist in the mutant system.

**Figure 6B** is a superposition diagram of the wild-type system and the mutant system of Mol\_11. Green indicates the

**TABLE 3 |** Enz/F876L AR, MOL\_11/WT AR, MOL\_11/F876L AR, Enz/WT AR system combined with free energy analysis (unit: kcal • mol<sup>-1</sup>).

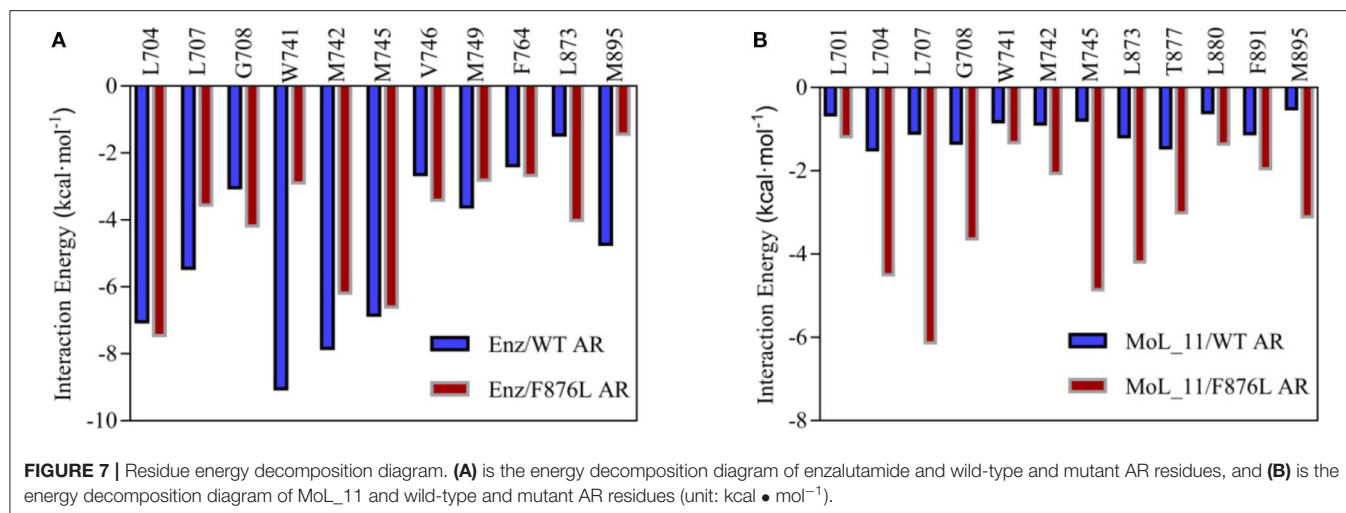
Complex	Contribution				
	$\Delta E_{vdw}$	$\Delta E_{ele}$	$\Delta G_p$	$\Delta G_{np}$	$\Delta G_{bind}$
MOL_11/WT AR	-49.197	0.734	23.850	-4.835	-29.226
MOL_11/F876L AR	-155.862	-9.438	61.042	-15.114	-119.661
Enz/WT AR	-246.292	-39.348	162.254	-22.601	-146.003
Enz/F876L AR	-245.466	-31.071	145.818	-22.619	-153.308

Mol\_11/WT AR system, and blue indicates the Mol\_11/F876L AR system. It can be seen from the figure that the conformation of H12 does not change, and also the conformation of Mol\_11 does not change. So we can deduce that Mol\_11 may not be resistant to the mutant F876L androgen receptor.

The combined free energy of each system is listed in **Table 3**. From the table we know that van der Waals force interaction and non-polar solvation free energy are favorable for the binding of protein and small molecules, and van der Waals force contributes most to its binding. The polar solvation free energy is not conducive to the binding. In each system, when the protein is changed from the wild type to the mutant type, the binding free energy will increase accordingly, indicating that the binding capacity of small molecules and proteins in the mutant system is stronger.

The total binding free energy was decomposed to each amino acid, looking for key amino acid in the interaction between small





molecules and proteins. From **Figure 7A**, it can be found that whether enzalutamide binds to wild-type AR or to mutant AR, the six amino acid residues L704, L707, W741, M742, M745, and M895 contribute greatly to their mutual binding. Compared with WT AR, the energy of M895 residues in F876LAR is significantly reduced (from  $-4.7256$  kcal • mol<sup>-1</sup> to  $-1.4351$  kcal • mol<sup>-1</sup>), and M895 is an amino acid located on H12, which further illustrates that in the F876L AR system, the interaction between enzalutamide and H12 is weak, causing H12 to close and promote transcription, thereby promoting the conversion of enzalutamide from antagonist to agonist.

In **Figure 7B**, when MoL\_11 is combined with F876L AR, the binding energy of all key amino acid residues increases. Among them, the binding energy of M895 amino acid residue located on H12 increases from  $-0.5125$  kcal • mol<sup>-1</sup> to  $-3.1034$  kcal • mol<sup>-1</sup>, which proves that the interaction between MoL\_11 and H12 is enhanced, which makes H12 deviate from the binding pocket, expose the binding site, and destroy the formation of the co-activation site, which is consistent with the analysis of the binding mode diagram of **Figure 5**. Therefore, MoL\_11 may not cause resistance to F876L mutation.

## CONCLUSION

Due to the complex composition of traditional Chinese medicine, the mechanism of action is not clear. In this study, virtual

screening of anti-prostate cancer traditional Chinese medicine ingredients is performed to obtain potential antiandrogens. The biological activity of the selected compounds was verified by biological experiments, and the small molecule MoL\_11 was found to own the best activity. The molecular dynamics simulation method was used to study the mechanism of action between MoL\_11 and wild-type AR/ F876L AR. It was found that MoL\_11 will not cause resistance to F876L theoretically. The newly discovered compound, Isoimperatorin, can be a lead compound used for potential anti-PCa drug discovery.

## DATA AVAILABILITY STATEMENT

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

## AUTHOR CONTRIBUTIONS

WH undertook structure-based virtual screening and *in vitro* evaluation of hit compounds. YS and JS helped to do the *in vitro* evaluation of hit compounds. ZZ carried out molecular-dynamics simulations. XW guided the *in vitro* experiments. JL conceived and coordinated the study. HL helped to discuss the results. All authors analyzed the results and approved the final version of the manuscript.

## REFERENCES

- Abraham, M., Murtola, T., Schulz, R., Páll, Z., Smith, J., Hess, B., et al. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Software* 1-2, 19–25. doi: 10.1016/j.softx.2015.06.001
- Accelrys (2010). *Accelrys Discovery Studio, Version 2.5*. San Diego, CA: Accelrys.
- Alsharif, S. A., Chen, L. Y., Tlahuice-Flores, A., Whetten, R. L., and Yacaman, M. J. (2014). Interaction between functionalized gold nanoparticles in physiological saline. *Phys. Chem. Chem. Phys. Pccp.* 16, 3909–3913. doi: 10.1039/c3cp54503b
- Attard, G., Parker, C., Lees, R. A., Schröder, F., Tomlins, S. A., Tannock, I., et al. (2016). Prostate cancer. *Lancet* 387, 70–82. doi: 10.1016/S0140-6736(14)61947-4
- Augello, M. A., Den, R. B., and Knudsen, K. E. (2014). AR function in promoting metastatic prostate cancer. *Cancer Metast. Rev.* 33, 399–411. doi: 10.1007/s10555-013-9471-3
- Brinkmann, A. O., Blok, L. J., de Ruiter, P. E., Doesburg, P., Steketee, K., Berrevoets, C. A., et al. (1999). Mechanisms of androgen receptor activation and function. *J. Steroid Biochem. Mol. Biol.* 69, 307–313. doi: 10.1016/S0960-0760(99)00049-7
- Estébanez-Perpiñá E, Moore, J. M., Mar, E., Delgado-Rodriguez, E., Nguyen, P., Baxter, J. D., et al. (2005). The molecular mechanisms of coactivator utilization

- in ligand-dependent transactivation by the androgen receptor. *J. Biol. Chem.* 280, 8060–8068. doi: 10.1074/jbc.M407046200
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 136, E359–E386. doi: 10.1002/ijc.29210
- Gao, W., Bohl, C. E., and Dalton, J. T. (2005). Chemistry and structural biology of androgen receptor. *Chem. Rev.* 105, 3352–3370. doi: 10.1021/cr020456u
- Hara, T., Miyazaki, J., Araki, H., Yamaoka, M., Kanzaki, N., Kusaka, M., et al. (2003). Novel mutations of androgen receptor: a possible mechanism of bicalutamide withdrawal syndrome. *Cancer Res.* 63, 149–153. doi: 10.1097/00130404-200301000-00011
- Hou, T., and Xu, X. (2004). Recent development and application of virtual screening in drug discovery: an overview. *Curr. Pharm. Des.* 10, 1011–1033. doi: 10.2174/1381612043452721
- Huang, Z., and Wong, C. F. (2016). Inexpensive method for selecting receptor structures for virtual screening. *J. Chem. Information Model.* 56, 21–34. doi: 10.1021/acs.jcim.5b00299
- Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3, 935–949. doi: 10.1038/nrd1549
- Lagorce, D., Sperandio, O., Galons, H., Miteva, M. A., and Villoutreix, B. O. (2008). FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects. *BMC Bioinform.* 9:396. doi: 10.1186/1471-2105-9-396
- Li, J., Zhao, P., Li, Y., Tian, Y., and Wang, Y. (2015). Systems pharmacology-based dissection of mechanisms of Chinese medicinal formula Bufe Yishen as an effective treatment for chronic obstructive pulmonary disease. *Sci. Rep.* 5:15290. doi: 10.1038/srep15290
- 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, 3–26. doi: 10.1016/S0169-409X(00)00129-0
- Møller, H., Roswall, N., Van Hemelrijck, M., Larsen, S. B., Cuzick, J., Holmberg, L., et al. (2015). Prostate cancer incidence, clinical stage and survival in relation to obesity: a prospective cohort study in Denmark. *Int. J. Cancer* 136, 1940–1947. doi: 10.1002/ijc.29238
- Nerenberg, P. S., So, C., Tripathy, A., and Head-gordon, T. (2011). Evaluation and improvement of the Amber ff99SB Force Field with an Advanced Water Model. *Biophys. J.* 100:311a. doi: 10.1016/j.bpj.2010.12.1900
- Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer statistics, 2018. 68, 7–30. doi: 10.3322/caac.21442
- Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. 70, 7–30. doi: 10.3322/caac.21590
- Sobhy, M. K., Mowafy, S., Lasheen, D. S., Farag, N. A., and Abouzid, K. A. M. (2019). 3D-QSAR pharmacophore modelling, virtual screening and docking studies for lead discovery of a novel scaffold for VEGFR 2 inhibitors: Design, synthesis and biological evaluation. *Bioorgan. Chem.* 89:102988. doi: 10.1016/j.bioorg.2019.102988
- Sousa da Silva, A. W., and Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interface. *BMC Res. Notes* 5:367. doi: 10.1186/1756-0500-5-367
- Tan, M. H., Li, J., Xu, H. E., Melcher, K., and Yong, E. L. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol Sin.* 36, 3–23. doi: 10.1038/aps.2014.18
- Taplin, M. E., Bubley, G. J., Ko, Y. J., Small, E. J., Upton, M., Rajeshkumar, B., et al. (1999). Selection for androgen receptor mutations in prostate cancers treated with androgen antagonist. *Cancer Res.* 59, 2511–2515.
- Tian, S., Wang, J., Li, Y., Li, D., Xu, L., and Hou, T. (2015). The application of *in silico* drug-likeness predictions in pharmaceutical research. *Adv. Drug Deliv. Rev.* 86, 2–10. doi: 10.1016/j.addr.2015.01.009
- Tian, S., Wang, X., Li, L., Zhang, X., Li, Y., Zhu, F., et al. (2017). Discovery of novel and selective adenosine A(2A) receptor antagonists for treating Parkinson's disease through comparative structure-based virtual screening. *J. Chem. Inf. Model.* 57, 1474–1487.
- Tran, C., Ouk, S., Clegg, N. J., Chen, Y., Watson, P. A., Arora, V., et al. (2009). Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* 324, 787–790. doi: 10.1126/science.1168175
- 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, 2615–2623. doi: 10.1021/jm020017n
- Xu, L., Zhang, Y., Zheng, L., Qiao, C., Li, Y., Li, D., et al. (2014). Discovery of novel inhibitors targeting the macrophage migration inhibitory factor via structure-based virtual screening and bioassays. *J. Med. Chem.* 57, 3737–3745. doi: 10.1021/jm401908w
- Zhang, Y., Bai, M., Zhang, B., Liu, C., Guo, Q., Sun, Y., et al. (2015). Uncovering pharmacological mechanisms of Wu-tou decoction acting on rheumatoid arthritis through systems approaches: drug-target prediction, network analysis and experimental validation. *Sci. Rep.* 5:9463. doi: 10.1038/srep09463

**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.

Copyright © 2020 Han, Shi, Su, Zhao, Wang, Li and Liu. 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.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

Visit us: [www.frontiersin.org](http://www.frontiersin.org)

Contact us: [frontiersin.org/about/contact](http://frontiersin.org/about/contact)



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership