Computer-Aided Drug Screening Based on the Binding Site Selectivity of ACE2: Machine Learning, Docking, and Molecular Dynamics Simulations

Xinhao Che,a Qilei Liu,a Fang Yu,a Lei Zhanga,\*

aFrontier Science Center for Smart Materials Oriented Chemical Engineering, Institute of Chemical Process Systems Engineering, Dalian University of Technology, Dalian 116024, China.

\*Corresponding Author’s E-mail: keleiz@dlut.edu.cn

Abstract

Since the outbreak of COVID-19, much scientific effort has been made to discover small molecule drugs targeting various stages of the infection of SARS-CoV-2. As the host-cell receptor of SARS-CoV-2, ACE2 is also an important regulatory factor in the human renin-angiotensin system. However, the selectivity of compounds for the two functional sites of ACE2 are not considered in the virtual screening process targeting ACE2. In this work, a virtual screening framework based on the binding site selectivity is developed. The framework integrates two machine learning models, molecular docking, and molecular dynamics simulation methods, which can be used to screen for candidate inhibitors with better pharmaceutical properties and binding site selectivity, so as to reduce potential drug side effects of in humans. Five compounds with better pharmaceutical properties and selectivity than the reported inhibitors are finally selected for experimental assays in future according to the screening results.

**Keywords**: Virtual screening, Binding site selectivity, Machine learning model, Molecular docking, Molecular dynamics simulation

* 1. Introduction

The outbreak of Corona Virus Disease in late 2019 (COVID-2019) (Zhou et al., 2020) has had an unprecedented impact on human society, especially on human health and economic development. Much scientific effort has been made to target various stages of the infection process of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Lu et al., 2020). In this process, the host-cell receptor, angiotensin converting enzyme 2 (ACE2), is of great significance and potential for the reason that it is the “gateway” for SARS (Kuba et al., 2010) and SARS-CoV-2 (Benton et al., 2020) to infect human cells. In March 2020, Yan et al. (2020) determined the full-length structure of the human ACE2 receptor for the first time, which makes it possible to discover small molecule drug targeting ACE2 based on the structure.

Different from vaccines or large molecule drugs, small molecule drugs usually have long development periods, high costs and low success rates (Bhutani et al., 2021). Thus, how to accelerate the development process of lead drugs is a pressing issue at present. Compared to traditional experiment-based exploration of active compounds, computer-aided drug design (CADD) methods, such as virtual screening, can quickly identify a group of promising compounds for focused experiment validation at the early stage of an outbreak of disease. For example, Terali et al. (2020) screened a clinically approved drug library to find drug candidates targeting the catalytic site of ACE2 for stabilizing the closed conformation of ACE2, thereby shifting relative positions of critical exterior residues in ACE2 recognized by SARS-CoV-2. However, inhibitors for ACE2 obtained in this way may cause a potential side effect risk because the catalytic function of ACE2 is essential for the human cardiovascular system. Targeting the binding interface of ACE2 and the receptor binding domain (RBD) to block the binding of SARS-CoV-2 can be a better strategy (Razizadeh et al., 2021). It’s necessary to consider the selectivity of inhibitors to different binding sites on/in the target protein during the virtual screening process, especially for multifunctional targets such as ACE2.

In order to minimize the negative influence of candidate compounds in the early stage of drug discovery, a virtual screening framework based on the binding site selectivity is proposed in this paper, which can be used for small molecule drug discovery targeting those proteins with multiple functional sites.

* 1. Materials and methods

The virtual screening framework proposed in this paper is shown in Figure 1. Details of the steps in the framework are discussed in the following sections.



Figure 1. The virtual screening framework

* + 1. Preparation of drug-like compounds and identification of the binding site

The DrugBank database (v5.1.8, released 2021.01.03) containing 9,137 3D structures of small molecule compounds is used for virtual screening in this work. The compounds that do not meet Lipinski’s Rules are filtered out using RDKit tool (https://www.rdkit.org). The complex of ACE2 and the RBD of the spike protein of SARS-CoV-2 (PDB ID: 6M0J) is selected as the target for virtual screening. The binding site at ACE2-RBD binding interface is determined by Razizadeh et al. (2021). The conserved catalytic site of ACE2 is determined by Towler et al. (2004).

* + 1. Pre-screening by the deep learning model

A deep learning model based on the binding site level is developed to predict which compounds have the binding potential to the specific binding site of a target protein, to quickly and effectively pre-screen a large number of compounds. The model construction steps are shown in Figure 2. By calculating the matrix descriptors of the compounds (active compounds and decoys) from DUD-E database and their binding sites as input to a 2D-convolutional neural network (CNN), the deep learning model is trained and the virtual screening performance of the model is evaluated (given in section 3.1). Then, the drug-like compounds obtained in section 2.1 and the binding site at ACE2-RBD binding interface are used as input to the model. The compounds that are more likely to bind to ACE2-RBD binding interface according to the model prediction results are selected for subsequent molecular docking.



Figure 2. The construction steps of the deep learning model

* + 1. Screening by molecular docking

Molecular docking is performed using Autodock Vina 1.2.3 (Eberhardt et al., 2021). Two different scoring functions, Vina (Eberhardt et al., 2021) and Vinardo (Quiroga and Villarreal, 2016), are selected to cross-validate the pre-screening results. The compounds with binding potential predicted by the developed CNN model are docked to ACE2-RBD binding interface. Then, the top compounds ranked by docking scores, which represent the strength of the binding affinity between the compound and the target protein, are used for binding site selectivity screening.

* + 1. Screening by binding site selectivity

Besides the binding tendency of the compounds evaluated by the binding potential prediction model developed in section 2.2 and two affinity scores in section 2.3, an artificial neural network (ANN) model for binding site prediction (Che et al., 2022) is also used for binding tendency analysis. The ANN model can predict the possibility of a compound in its true binding site, which provides an additional complement to the reliability of the above binding potential prediction model and affinity scores. The screening steps are shown in Figure 3. By comparing and analyzing the above four binding metrics, the compounds with higher selectivity to ACE2-RBD binding interface are obtained.



Figure 3. Screening by binding site selectivity

* + 1. Verification of bind by MD simulations

The complexes consist of the docking conformation of each potential compound screened in section 2.4 and ACE2-RBD binding interface are used for MD simulations. MD simulations are performed using *Desmond* module in the Schrödinger software package. The possible inhibitory mechanism of the compounds as candidate inhibitors and binding process are analyzed using *Simulation interactions diagram* module. In addition, to quantitatively evaluate the dynamic binding strength of compounds to a target protein, binding free energy is calculated using the MM-GBSA method.

* + 1. Evaluation of ADMET properties of candidate inhibitors

In order to further evaluate the pharmaceutical properties, the properties of absorption, distribution, metabolism, excretion, and toxicity (ADMET) of the 5 candidate inhibitors selected in section 2.5 are evaluated using the PharmaMind platform of Infinite Intelligence Pharma (http://www.iipharma.com.cn).

* 1. Results and discussion
		1. Performance of the developed CNN model

The virtual screening performance of the developed CNN model is given in Table 1, which shows the excellent early enrichment capability of the model.

Table 1. The virtual screening performance of the developed CNN model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Adjusted logAUC** | **ROC enrichment** **(RE)** | **Early hit rate** **(Hit)** | **Boltzmann-enhanced** **discrimination of** **ROC****(BEDROC)** | **Enrichment factor** **(EF)** |
| Adjusted logAUC0.5% | Adjusted logAUC1% | Adjusted logAUC2% | RE0.5% | RE1% | RE2% | Hit2% | BEDROC80.5 | EF2% |
| 0.62 | 0.64 | 0.65 | 89.2 | 55.6 | 34.5 | 52.8% | 0.673 | 27.0 |

* + 1. Virtual screening

6,876 drug-like compounds are obtained by filtering the DrugBank database with Lipinski’s Rules. Then, 1,735 compounds with high binding potential according to the prediction results of the CNN model are docked to ACE2-RBD binding interface. 128 compounds ranked in the top 300 by the CNN model, at the same time, with a Vina score below -6.4 kcal/mol and a Vinardo score below -4.8 kcal/mol are used for binding site selectivity screening. Next, their binding tendency to ACE2-RBD binding interface and the catalytic site of ACE2 is evaluated. 11 compounds with better binding site selectivity for ACE2-RBD binding interface are finally selected. Detailed selectivity metrics are given in Table 2. The binding site selectivity of two reported active compounds, Nilotinib and SSAA09E2 (Razizadeh et al., 2021), are also evaluated for comparison.

Table 2. Detailed selectivity metrics of the 11 compounds from screening and two reported active compounds

|  |  |
| --- | --- |
| **DrugBank****ID** | **Prediction values of the two deep learning models and the docking scores of Vina & Vinardo** |
| **For the catalytic site of ACE2** | **For the ACE2-RBD binding interface** |
| CNN | ANN | Vina | Vinardo | CNN | ANN | Vina | Vinardo |
| **DB06837** | 0.002 | 0.086 | -7.5 | -5.674 | 0.959 | 0.660 | -8.5 | -5.725 |
| **DB08029** | 0.017 | 0.404 | -6.6 | -5.322 | 0.949 | 0.662 | -7.1 | -5.559 |
| **DB08409** | 0.128 | 0.103 | -5.8 | -4.817 | 0.946 | 0.575 | -6.8 | -5.254 |
| **DB04371** | 0.031 | 0.091 | -6.6 | -5.261 | 0.924 | 0.584 | -7.3 | -5.394 |
| **DB07579** | 0.152 | 0.112 | -6.4 | -4.544 | 0.923 | 0.543 | -8.2 | -5.492 |
| **DB08394** | 0.384 | 0.060 | -5.8 | -4.587 | 0.910 | 0.550 | -7.2 | -5.826 |
| **DB12574** | 0.030 | 0.131 | -7.2 | -5.745 | 0.907 | 0.528 | -7.9 | -5.961 |
| **DB08302** | 0.317 | 0.231 | -6.4 | -5.362 | 0.903 | 0.526 | -7.9 | -6.392 |
| **DB03313** | 0.019 | 0.015 | -6.6 | -5.045 | 0.891 | 0.720 | -7.3 | -5.547 |
| **DB01139** | 0.054 | 0.144 | -6.5 | -4.486 | 0.884 | 0.678 | -7.2 | -5.142 |
| **DB08397** | 0.050 | 0.072 | -6.3 | -5.167 | 0.873 | 0.528 | -7.4 | -5.586 |
| **Nilotinib** | 0.008 | 0.068 | -8.8 | -6.862 | 0.000 | 0.698 | -9.2 | -5.868 |
| **SSAA09E2** | 0.056 | 0.739 | -6.8 | -4.836 | 0.391 | 0.193 | -7.3 | -4.907 |

* + 1. Verification of bind and final inhibitor selection

The possible inhibitory mechanism and binding process of the above 11 compounds are analyzed according to the MD simulation results. Four main aspects are as follows: (1) protein-compound RMSD analysis. RMSD analysis not only indicates if a simulation has equilibrated, but also gives insights into conformation changes of the compound and its target protein throughout the simulation. (2) protein RMSF analysis. RMSF characterizes the flexibility of different amino acid residues in a protein, which is used to compare the conformational differences of the protein before and after the binding of compounds. (3) protein-compound interaction analysis. By calculating the non-bonding interactions of binding between compounds and the protein, possible inhibitory mechanism of the compounds as candidate inhibitors is analyzed. (4) protein-compound binding free energy (ΔG). ΔG is used to quantitatively evaluate the dynamic binding strength between different compounds and the target protein. Besides the above verification of bind, the binding tendency of compounds to different binding sites are also analyzed according to the visualization of MD simulation trajectories. Finally, the five most promising candidate inhibitors are selected and their ADMET properties are given in Table 3.

Table 3. ADMET properties of the 5 candidate inhibitors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Compound drugBank** **ID** | **Adsorption** | **Distribution** | **Metabolism** | **Excretion** | **Toxicity** |
| Human oral Bioavailability(F20) |  Caco2 cellpermeability | Human intestinal absorption | Plasma protein binding | Volume of distribution at steady state | CYP Substrate/Inhibitor | Clearance in Hepatocyte | Half-life | LD50 |
| probability | Log cm/s | probability | % | L/kg |  | uL/(min·106 cells) | h | Mg/kg |
| DB06837 | 0.80 | -6.10 | 0.80 | 76.65 | 1.14 | CYP2C19 inhibitionCYP3A4inhibitionCYP3A4substrate | 29.28 | 7.95 | 902.94 |
| DB08029 | 0.88 | -5.26 | 0.90 | 71.65 | 4.07 | - | 37.34 | 14.26 | 2519.97 |
| DB07579 | 0.53 | -5.73 | 0.47 | 63.40 | 5.83 | - | 30.26 | 17.89 | 1708.90 |
| DB01139 | 0.74 | -5.84 | 0.49 | 66.03 | 0.59 | CYP3A4inhibition | 19.55 | 1.33 | 16118.83 |
| DB08397 | 0.84 | -5.12 | 0.83 | 91.75 | 1.69 | - | 22.29 | 15.53 | 468.84 |

* 1. Conclusions

In this paper, a virtual screening framework based on the binding site selectivity is developed to discover small molecule inhibitors targeting ACE2-RBD binding interface. The binding potential, binding affinity, and binding tendency of the candidate inhibitors are evaluated. Five compounds with better pharmaceutical properties and selectivity than the reported inhibitors are finally selected for experimental assays in future according to the screening results.

Acknowledgements

This work was supported by the NSFC [22278053, 22078041]. The authors are grateful for the University of Minnesota for their generous assistance of the Schrödinger software package used for molecular dynamics simulation in this paper.

References

D.J. Benton, A.G. Wrobel, P. Xu, C. Roustan, S.R. Martin, P.B. Rosenthal, J.J. Skehel, S.J. Gamblin, 2020. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. Nature 588, 327-330.

P. Bhutani, G. Joshi, N. Raja, N. Bachhav, P.K. Rajanna, H. Bhutani, A.T. Paul, R. Kumar, 2021. US FDA Approved Drugs from 2015-June 2020: A Perspective. Journal of medicinal chemistry 64, 2339-2381.

X.H. Che, S.Y. Chai, Z.Z. Zhang, L. Zhang, 2022. Prediction of ligand binding sites using improved blind docking method with a Machine Learning-Based scoring function. Chemical Engineering Science 261, 10.

J. Eberhardt, D. Santos-Martins, A.F. Tillack, S. Forli, 2021. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model 61, 3891-3898.

K. Kuba, Y. Imai, T. Ohto-Nakanishi, J.M. Penninger, 2010. Trilogy of ACE2: a peptidase in the renin-angiotensin system, a SARS receptor, and a partner for amino acid transporters. Pharmacol Ther 128, 119-128.

R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W.J. Liu, D. Wang, W. Xu, E.C. Holmes, G.F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet 395, 565-574.

R. Quiroga, M.A. Villarreal, 2016. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. PLoS One 11, e0155183.

M. Razizadeh, M. Nikfar, Y. Liu, 2021. Small molecule therapeutics to destabilize the ACE2-RBD complex: A molecular dynamics study. Biophys J 120, 2793-2804.

K. Terali, B. Baddal, H.O. Gulcan, 2020. Prioritizing potential ACE2 inhibitors in the COVID-19 pandemic: Insights from a molecular mechanics-assisted structure-based virtual screening experiment. Journal of Molecular Graphics & Modelling 100, 10.

P. Towler, B. Staker, S.G. Prasad, S. Menon, J. Tang, T. Parsons, D. Ryan, M. Fisher, D. Williams, N.A. Dales, M.A. Patane, M.W. Pantoliano, 2004. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. J Biol Chem 279, 17996-18007.

R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 367, 1444-1448.

Q.A. Zhou, J. Kato-Weinstein, Y. Li, Y. Deng, R. Granet, L. Garner, C. Liu, D. Polshakov, C. Gessner, S. Watkins, 2020. Potential Therapeutic Agents and Associated Bioassay Data for COVID-19 and Related Human Coronavirus Infections. ACS Pharmacol Transl Sci 3, 813-834.