


 Cite this: *RSC Adv.*, 2022, 12, 7318

# Blocking key mutated hotspot residues in the RBD of the omicron variant (B.1.1.529) with medicinal compounds to disrupt the RBD-hACE2 complex using molecular screening and simulation approaches†

 Abbas Khan,<sup>a</sup> AsfandYar Waheed Randhawa,<sup>b</sup> Ali Raza Balouch,<sup>b</sup> Naila Mukhtar,<sup>c</sup> Abrar Mohammad Sayaf,<sup>d</sup> Muhammad Suleman,<sup>e</sup> Taimoor Khan,<sup>a</sup> Shahid Ali,<sup>e</sup> Syed Shujait Ali,<sup>e</sup> Yanjing Wang,<sup>f</sup> Anwar Mohammad<sup>\*g</sup> and Dong-Qing Wei<sup>†\*ahi</sup>

A new variant of SARS-CoV-2 known as the omicron variant (B.1.1.529) reported in South Africa with 30 mutations in the whole spike protein, among which 15 mutations are in the receptor-binding domain, is continuously spreading exponentially around the world. The omicron variant is reported to be highly contagious with antibody-escaping activity. The emergence of antibody-escaping variants is alarming, and thus the quick discovery of small molecule inhibitors is needed. Hence, the current study uses computational drug screening and molecular dynamics simulation approaches (replicated) to identify novel drugs that can inhibit the binding of the receptor-binding domain (RBD) with hACE2. Screening of the North African, East African and North-East African medicinal compound databases by employing a multi-step screening approach revealed four compounds, namely (–)-pipoxide (C1), 2-(*p*-hydroxybenzyl) benzofuran-6-ol (C2), 1-(4-hydroxy-3-methoxyphenyl)-2-[(*E*)-3-hydroxy-1-propenyl]-2-methoxyphenoxy)-1,3-propanediol (C3), and Rhein (C4), with excellent anti-viral properties against the RBD of the omicron variant. Investigation of the dynamics demonstrates stable behavior, good residue flexibility profiles, and structural compactness. Validation of the top hits using computational bioactivity analysis, binding free energy calculations and dissociation constant ( $K_D$ ) analysis also indicated the anti-viral properties of these compounds. In conclusion, this study will help in the design and discovery of novel drug therapeutics, which may be used against the emerging omicron variant of SARS-CoV-2.

 Received 14th January 2022  
 Accepted 14th February 2022

DOI: 10.1039/d2ra00277a

[rsc.li/rsc-advances](http://rsc.li/rsc-advances)

## Introduction

The pandemic instigated by a member of the coronavirus family, known as SARS-CoV-2, in late 2019 continues to spread chaos around the world. Since the emergence of the first case of COVID-19, many variants of the virus have been reported, increasing the risk of reinfection and immunity evasion.<sup>1</sup> To date, many variants classified as either variants of concern

(VOCs) or variants of interest (VOIs) have been reported to differentially challenge public health. Among the reported variants, the alpha, beta, gamma, delta, and omicron variants have been classified as VOCs.<sup>2</sup> The alpha variant, also known as B.1.1.7, reported in almost 120 countries and harboring 69–70del, N501Y and P681H mutations in the spike proteins, was reported to increase transmissibility by 40–80%.<sup>3</sup> Later on, a new version of B.1.1.7 was reported, known as B.1.1.7+ E484K,

<sup>a</sup>Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China. E-mail: dqwei@sjtu.edu.cn

<sup>b</sup>Central Park Medical College, Lahore, Pakistan

<sup>c</sup>Department of Botany, University of Okara, Punjab, Pakistan

<sup>d</sup>School of Chemical Sciences, Universiti Sains Malaysia, Pulau Pinang, 11800, Malaysia

<sup>e</sup>Centre for Biotechnology and Microbiology, University of Swat, Kanju, Khyber Pakhtunkhwa, Pakistan

<sup>f</sup>Engineering Research Center of Cell and Therapeutics Antibody, School of Pharmacy, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China

<sup>g</sup>Department of Biochemistry and Molecular Biology, Dasman Diabetes Institute, Dasman, Kuwait

<sup>h</sup>State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint Laboratory of International Laboratory of Metabolic and Developmental Sciences, Ministry of Education and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, P.R. China

<sup>i</sup>Peng Cheng Laboratory, Vanke Cloud City Phase 1 Building 8, Xili Street, Nanshan District, Shenzhen, Guangdong, 518055, P.R. China

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d2ra00277a



with an extra mutation in the receptor-binding domain (RBD) of the spike protein along with N501Y. This variant was reported only in 39 confirmed cases and then disappeared.<sup>4</sup> The beta variant (B.1.351), reported in South Africa with mutations K417N, E484K and N501Y, was confirmed to increase the transmissibility as well as cause reduction in T cells and other immune responses triggered against COVID-19 infection. In January 2021, a new variant, P.1, was reported in Brazil harbouring K417T, E484K and N501Y mutations in the RBD of the spike protein. This variant increased the transmission by 38%, mortality by 50% and the immune response against reinfection was also reduced.<sup>5</sup> Abbas *et al.* used structural modeling and protein coupling approaches to discover the mechanism of higher infectivity caused by these variants of SARS-CoV-2.<sup>6–8</sup> They reported that the binding affinity for the host receptor hACE2 increased while the electrostatic interactions were considered the prime factors for enhanced binding. A more devastating variant of SARS-CoV-2 was reported in India in Oct 2020 and officially named B.1.617.2. This variant harbours L452R, T478K and P681R mutations in the RBD with an 87% increase in transmissibility, 85% increment in hospitalization and a mortality increase of 137%. Later on, a new version of this variant known as “delta plus” having an extra mutation, K417N, was reported in India and UK.<sup>9</sup> Likewise, a new variant known as Mu ( $\mu$ ) or B.1.621 was reported to contain mutations in the spike protein among which some are shared with other VOC. The novel mutations in this variant include R346K, Y144T, Y145S, and insertion at 146N position.<sup>10</sup> A novel VOI termed as C.37 or  $\lambda$  variant exhibit L452Q and F490S mutations in RBD, which consequently reduces the antibody mediated neutralization.<sup>11</sup> On the other hand, the B.1.617.1 variant reported having L452R mutation in the RBD is linked with altered antibody neutralization by disrupting the respective conformational epitopes. Moreover, the VOI Iota ( $\iota$ ) (lineage B.1.526) also contains mutation E484K showing resistance to therapeutic monoclonal antibodies and is less susceptible to neutralization.<sup>7</sup>

The latest variant, designated as the omicron variant and officially named as B.1.1.529, reported in November 2021 in South Africa, harbours 30 mutations in the spike protein (A67V,  $\Delta$ 69–70, T95I, G142D,  $\Delta$ 143–145,  $\Delta$ 211, L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F), among which, 15 mutations lie in the RBD.<sup>12</sup> Among the others, such as nsp3 (K38R, V1069I,  $\Delta$ 1265, L1266I, A1892T), nsp4 (T492I), nsp5 (P132H), nsp6 ( $\Delta$ 105–107, A189V), nsp12 (P323L), while 142V was reported in NSP14. Three sub-lineages of this variant such as BA.1/B.1.1.529.1, BA.2/B.1.1.529.2 and BA.3/B.1.1.529.3 are circulating. WHO (World Health Organization) reported this variant as the most serious threat to public health that evades the immune response.<sup>13</sup> The variant poses a significant health threat to the world and the therapeutic efficacy of already developed vaccine therapeutics against it remains elusive.<sup>14,15</sup> Further investigations are required to depict the molecular mechanism of pathogenicity related to the omicron variant. Moreover, improved strategies

are needed to design novel therapeutics against the newly emerged SARS-CoV-2 variants.

To deal with this worrisome pandemic situation, further research is needed for the rapid development of safe and effective therapeutics. In this regard, the spike protein is deemed as the most effective druggable target for the development of drugs for COVID-19 treatment.<sup>16,17</sup> The increasing prevalence of mutations in the RBD of SARS-CoV-2, which is linked with an enhanced viral affinity for host receptors and pathogenicity, is crucial to be considered as a therapeutic target. Thus, herein, we employed computational molecular screening against the RBD of the omicron (B.1.1.529) variant and evaluated the binding affinity of drug-like molecules against it. The current study uses *in silico* approaches such as molecular docking, simulation and free energy calculations for the identification of phytomedicines from North African, East African and North-East African medicinal compounds databases against the SARS-CoV-2. The results demonstrate critical information about the anti-viral potency of the screened drug-like molecules against the mutations in the RBD domain of the omicron variant. The study will help in the design and discovery of novel drug therapeutics, which may be deployed against the emerging omicron variant of SARS-CoV-2.

## Materials and methods

### Structures retrieval and modelling

The wild-type structure of the RBD was obtained from RCSB using 6M0J PDB ID.<sup>18</sup> The sequence of RBD was retrieved from UniProt, which ranges from 313–526.<sup>19</sup> Fifteen mutations reported in the RBD of omicron variants were manually integrated into the primary sequence and subjected to molecular modelling. A PSI-BLAST was carried out, which also results in 6M0J as the closest template, which shares a higher identity with the omicron RBD. A Modeller program was used for homology modeling of the RBD of the omicron variant of SARS-CoV-2.<sup>20,21</sup> The modelled structure was minimized, relaxed, prepared and cleaned for virtual screening.

### Virtual drugs screening and re-scoring of the top hits

North African, East African and North-East African were downloaded from the African Natural Products Databases (ANPDB) website (<http://african-compounds.org/anpdb/>) in a 3D-SDF format which were prepared for screening.<sup>22</sup> These databases are a collection of natural products from South African natural compounds, which have diverse medicinal features. Prior to the computational screening of these databases, the FAF-Drugs4 webserver was used to obtain only for non-toxic, drug-like molecules obeying Lipinski's rule of five.<sup>23</sup> The filtered databases were then screened against the interface site of the RBD. All the drugs were converted to .pdbqt format prior to screening. For virtual drug screening EasyDock Vina 2.0, a GUI interface for virtual screening of the three databases was used. EasyDock Vina uses the AUTODOCK4 algorithm to screen and order the best drug-like molecules. In the initial stage, 16 exhaustiveness was used for quick screening. The top-scoring compounds were



then selected again and screened using 64 exhaustiveness. The purpose of the second screening was to remove false-positive results and re-evaluate the best compounds. The top 10% from each database were then selected for induced-fit docking (IFD) using AutoDockFR, which enables receptor flexibility and also supports covalent docking.<sup>24</sup> This approach also uses AutoDock4 and Vina modes for docking but is very fast and accurate. The final best hits were then subjected to bioactivity evaluation, dissociation constant and molecular dynamics simulation analysis.

### Dynamic evaluation of the final drugs-RBD complexes

The obtained top natural compounds in complex with the omicron RBD were tested for dynamic feature evaluation using AMBER20 package solvated with OPC water model. For this purpose, the FF19SB force field and (GAFF) "General Amber Force Field" were used to parameterize both the protein and small drug molecules, respectively.<sup>25,26</sup> Prior to force field parametrization, the drug topologies were generated with an antechamber module. To neutralize the effect of any charge Na<sup>+</sup> and Cl<sup>-</sup> ions were inoculated followed by energy minimization in two steps (algorithms: steepest descent and conjugate gradient) was achieved. Next, the heating and equilibration steps were completed. Finally, for each complex 200 ns (two replicas) of the production run was completed. To treat the long-range electrostatic interactions with a 10.0 Å cutoff distance, the particle mesh Ewald algorithm was used. However, the covalent bonds, if any, were treated with the SHAKE algorithm. Finally, the CPPTRAJ package was used to analyze the trajectories while the PMEMD.cuda was used for running the simulations.<sup>27</sup>

### Post-simulation validation of the top hits

**The binding free energy calculations.** The evaluation of the strength of small molecules binding through binding free energy (BFE) using MM/GBSA is the most widely employed approach used in different research investigations. Keeping in view the importance of this approach in re-ranking the binding conformations, we also used the MMPBSA.py script to compute the binding free energy of the RBD-ligand complexes by considering 2500 snapshots. For this purpose, the following equation was used to estimate the BFE:

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - [\Delta G_{\text{receptor}} + \Delta G_{\text{ligand}}]$$

The  $\Delta G_{\text{bind}}$  represent the total binding energy, while  $\Delta G_{\text{receptor}}$ ,  $\Delta G_{\text{ligand}}$ , and  $\Delta G_{\text{complex}}$  represent the binding energy of protein, drug, and complex, respectively. The following equation was used to estimate the individual binding energies like bonded ( $G_{\text{bond}}$ ), electrostatic ( $G_{\text{ele}}$ ), polar ( $G_{\text{pol}}$ ) and non-polar ( $G_{\text{npol}}$ ), which contribute to the total binding free energy.

$$G = G_{\text{bond}} + G_{\text{electrostatic}} + G_{\text{van der Waal}} + G_{\text{polar}} + G_{\text{non-polar}}$$

### Prediction of bioactivity and determination of $K_D$ (dissociation constant)

To estimate the IC<sub>50</sub> value as a bioactivity predictor each shortlisted compound was subjected to a cheminformatics tool, Molinspiration (<https://www.molinspiration.com/cgi-bin/properties>). Hundreds of studies have leveraged molinspiration to forecast bioactivity values.<sup>28–32</sup> In addition, the PRODIGY (PROtein binDIng enerGY) (<https://wenmr.science.uu.nl/prodigy/>) webserver was utilized to calculate the dissociation constant values for various biological complexes in order to offer convincing information of the strength of binding as  $K_D$ .<sup>33,34</sup>

## Results and discussion

The emergence of the omicron variant has created alarming situations around the world and worsened the pandemic situation further. It has been reported that this variant is more transmissible and increases the chances of re-infection. Despite all these, the evasion of almost all the neutralization antibodies and reduction in the efficacy of the different vaccines is a greater threat.<sup>35</sup> Thus, further insightful studies are required to overcome the problem of vaccine therapeutic evasion. Thus, small molecules development is of great interest for effective treatment. Hence in this study, we used computational drugs screening and simulation approaches to identify potential drugs for omicron (B.1.1.529) RBD. Screening of multiple databases, *i.e.* North African, East African and North-East African databases shortlisted potential anti-viral drugs. The overall workflow of the work is given in Fig. 1.

The spike protein is an essential determinant of the COVID-19 infection initiation and entry into the host cell. It is a multi-domain protein with the most important RBD, which comes in direct contact with the host receptor and fuses with the cellular membrane for entry. It is deemed an important drug and vaccine target for the neutralization of the virus. The structure of the spike protein and its domain organization is shown in Fig. 2A. The structure of RBD of the omicron variant was modelled and the active site residues were specified for docking. The mutations in the RBD of the omicron variant are shown in Fig. 2B. The superimposed structure of the wild type and omicron RBD revealed 0.841 Å of RMSD difference, which demonstrates that the new variant acquired conformational variations through mutations fitness (Fig. 2C). Prior to the screening of the databases, the interface residues shown in Fig. 2D were considered to generate the receptor grid (6.04 × -70.97 × 24.028).

Prior to database screening, filtering of drugs-like molecules using Lipinski's rule of five was performed. Among the total 1871 compounds in the East African database, only 1131 compounds were selected to follow the R5. In the North African database, a total of 4924 compounds were screened for R5 rules and only 357 compounds were excluded. In the northeast African database among 6512, only 5422 compounds were accepted to obey Lipinski's rule of five. Thus, we screened a total of 11 102 compounds by using a multi-step screening approach.



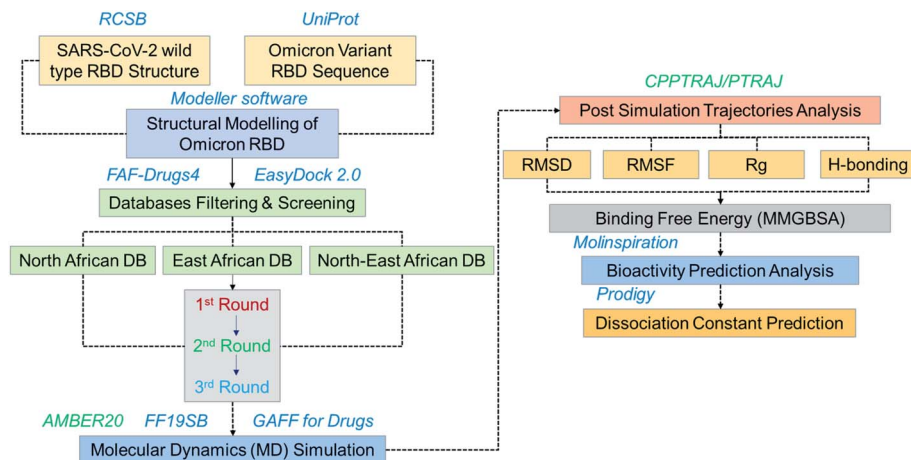


Fig. 1 Overall workflow of this work. The figure explains the structural modeling, screening, simulation, post-simulation and validation steps.

The first step of virtual screening results in docking scores ranging from  $-4.5$  to  $-2.5$  kcal mol $^{-1}$ . Among these, 2664 compounds reported a docking score less than  $-3.0$  while 8438 reported more than  $-3.0$  kcal mol $^{-1}$ . Among 2664, only the top 20% of compounds (266) were selected for the third round of screening. An induced-fit docking approach was then employed to screen the top 418 compounds, which resulted in docking scores ranging from  $-9.63$  to  $-6.71$  kcal mol $^{-1}$ . Among these, only the top ten were selected for the docking using Auto Dock Vina. Among the 42 compounds (10%), only seven compounds were reported to have a docking score higher than  $-5.0$  kcal mol $^{-1}$  and good interaction profiles. Among the seven compounds (–)-pipoxide (C1), 2-(*p*-hydroxybenzyl) benzofuran-6-ol (C2), 1-(4-hydroxy-3-methoxyphenyl)-2-{4-[(*E*)-3-hydroxy-1-propenyl]-2-methoxyphenoxy}-1,3-propanediol (C3), and Rhein

(C4) were selected for further analysis. The top four compounds were selected for detailed investigation and are given in Table 1.

Detailed interaction analysis of the top four compounds revealed information about hydrogen bonding, pie-pie interactions, salt-bridges and pie-stacking. In the case of C1, a total of five hydrogen bonds with Tyr453, Arg493, Ser494 and Ser496 were established. Among the others, Tyr501 and His505 also formed pie-pie interactions. It can be seen that this compound blocked the key mutated residues such as Arg493, Ser494, Ser496, Tyr501 and His505, which consequently halted the binding of RBD to the hACE2. The docking score for (–)-pipoxide was reported to be  $-5.78$  kcal mol $^{-1}$ . The 3D interaction pattern of C1 is given in Fig. 3A while the 2D interactions are shown in Fig. 3B. C2 on the other hand with the docking score  $-5.12$  kcal mol $^{-1}$  established three hydrogen bonds, including

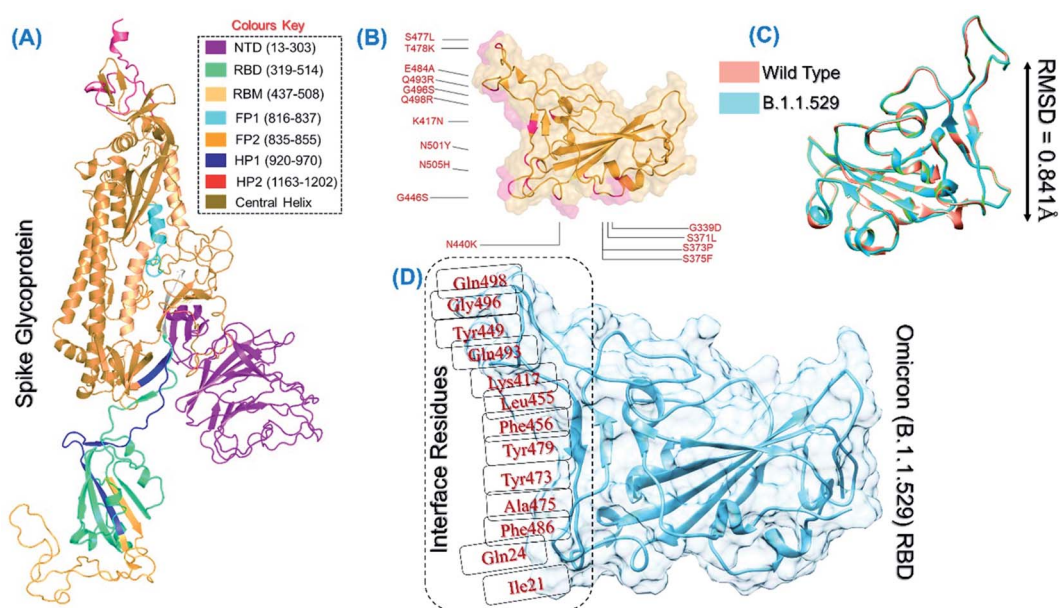
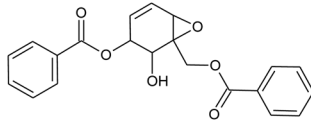
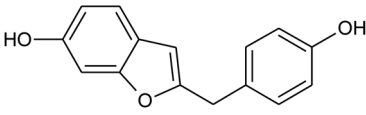
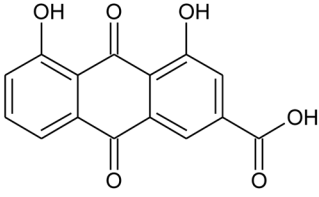
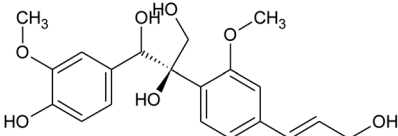


Fig. 2 (A) Structural organization of the whole spike protein. (B) RBD domain of the B.1.1.529 variant with the reported mutations. (C) Superimposed wild type and B.1.1.529-RBD structures. (D) Interface residues on RBD used for grid generation and screening.



Table 1 List of the top four hits with their scientific names, 2D structures and docking scores

S. no.	Compound name	2D structure	Docking score
1	(-)-Pipoxide		-5.794
2	2-( <i>p</i> -Hydroxybenzyl)benzofuran-6-ol		-5.124
3	Rhein		-5.127
4	1-(4-Hydroxy-3-methoxyphenyl)-2-{4-[( <i>E</i> )-3-hydroxy-1-propenyl]-2-methoxyphenoxy}-1,3-propanediol		-7.601

Glu406, Ser496 and His505. Tyr453 and His505 also reported two pi-pi stacking contact with the ligand. Moreover, the three pi-cation interactions are formed by the only residue Arg493, which is mutated in the omicron variant. Herein, the same residues are targeted, which are novel mutations in the RBD of the omicron variant. The 3D interaction pattern of C2 is given in Fig. 4A while the 2D interactions are shown in Fig. 4B.

Next, the binding pattern of C3 demonstrated the strongest binding of the other three compounds. The docking scores of  $-7.601 \text{ kcal mol}^{-1}$  for five hydrogen binds and one pi-pi interaction were reported. The residues Glu406, Tyr453, Ser494 and Ser496 established hydrogen bonds while the only pi-pi interaction was established with the C3. The interaction pattern of C3 is similar to those of C1, C2 and C4, which also block these

residues to disrupt the binding between RBD and hACE2. The 3D interaction pattern of C3 is given in Fig. 5A while the 2D interactions are shown in Fig. 5B.

The molecular interaction analysis of C4 reported only four hydrogen bonds with Tyr453, Ser494 and Ser496. The docking score for C4 was reported to be  $-5.127 \text{ kcal mol}^{-1}$ . The interaction pattern of C4 is similar to those of C1 and C2, which also block these residues to disrupt the binding between RBD and hACE2. The 3D interaction pattern of C4 is given in Fig. 6A while the 2D interactions are shown in Fig. 6B.

### Dynamic and binding stability analysis

Investigation of dynamic stability using simulation trajectories is important to reveal the binding of the small molecular ligand

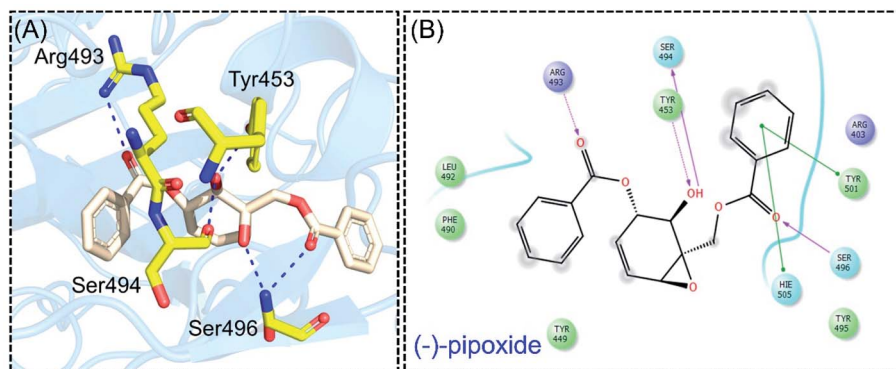


Fig. 3 Interaction pattern of (-)-pipoxide with the RBD of the omicron variant. (A) 3D interaction pattern of (-)-pipoxide; (B) the 2D interaction pattern generated with Maestro free Academic version 2018-1 (for visualization only).



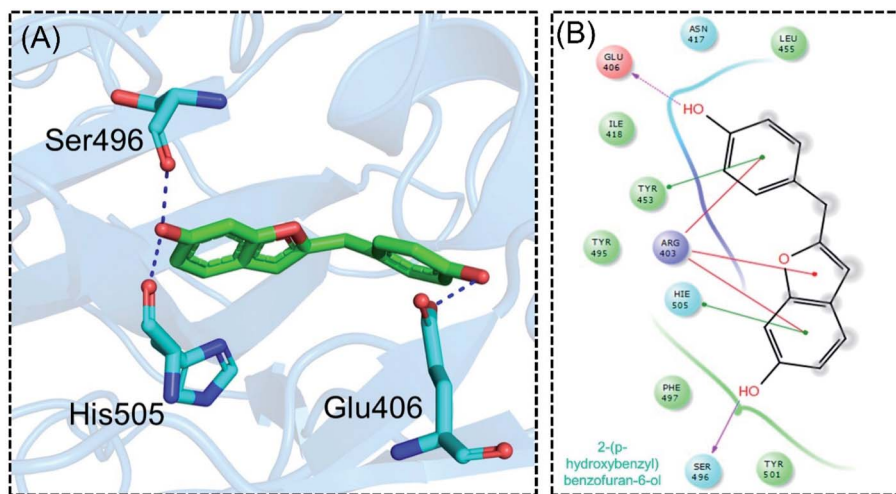


Fig. 4 Interaction pattern of 2-(*p*-hydroxybenzyl) benzofuran-6-ol with the RBD of the omicron variant. (A) 3D interaction pattern of 2-(*p*-hydroxybenzyl) benzofuran-6-ol; (B) the 2D interaction pattern generated with Maestro free Academic version 2018-1 (for visualization only).

in the binding cavity. It is essential to demonstrate the dynamic stability of the interacting molecules to determine the binding strength. Hence, here, we also calculated the dynamic stability of the top complexes as root mean square deviation (RMSD). The stability of each complex was calculated as a function of time using the simulation trajectory. As given in Fig. 7A, the C1-RBD demonstrated stable dynamics and no significant deviation was observed. The complex reached stability at 1.68 Å and equilibrium at 3 ns. The RMSD of this complex increased a little between 22–30 ns and then again decreased back with a uniform pattern until 80 ns. A smaller increase in the RMSD was experienced between 81–88 ns and then again the RMSD decreased. During the last 100 ns (101–200 ns) the RMSD decreased gradually and was stabilized in the long-run simulation. No significant convergence was observed during the last 100 ns. Similarly in replica 2 of the C1 complex, a more similar trend in the RMSD graph with few deviations until 80 ns was observed. However, the RMSD then decreased gradually and

followed a similar pattern as that of replica 1. The complex demonstrated a stable binding during the simulation and the average RMSD for this complex replica 1 was reported to be 1.63 Å while for replica 2 the average RMSD was calculated to be 1.65 Å. The RMSD graph of C1 is given in Fig. 7A. On the other hand, the RMSD graph of C2 demonstrated a more stable behavior comparatively but overall the complex remained stable. The structure gained stability at 1.2 Å and then a uniform pattern of RMSD was followed. The complex reported minor deviations such as between 61–72 ns and 91–95 ns but no major convergence was observed. The average RMSD for C2 was reported to be 1.22 Å for replica 1, while for replica 2, a similar average RMSD was reported. The RMSD graph of C2 is given in Fig. 7B. Furthermore, the RMSD of C3 also reported a little unstable dynamic behavior and reported only minor acceptable deviations. In both replicas, 1 and 2, the complex initially demonstrated a stable uniform pattern until 100 ns, however, at 101–130 ns both the replicas reported an increase in the RMSD and

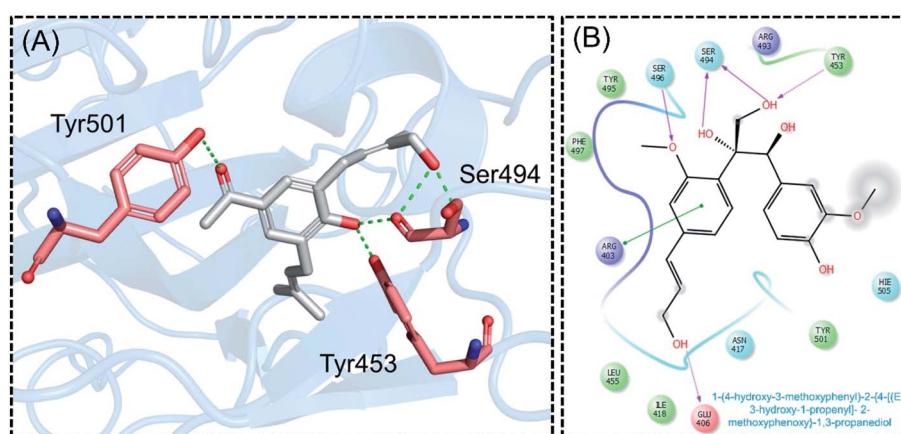


Fig. 5 Interaction pattern of C3 with the RBD of the omicron variant. (A) 3D interaction pattern of C3; (B) the 2D interaction pattern generated with Maestro free Academic version 2018-1 (for visualization only).



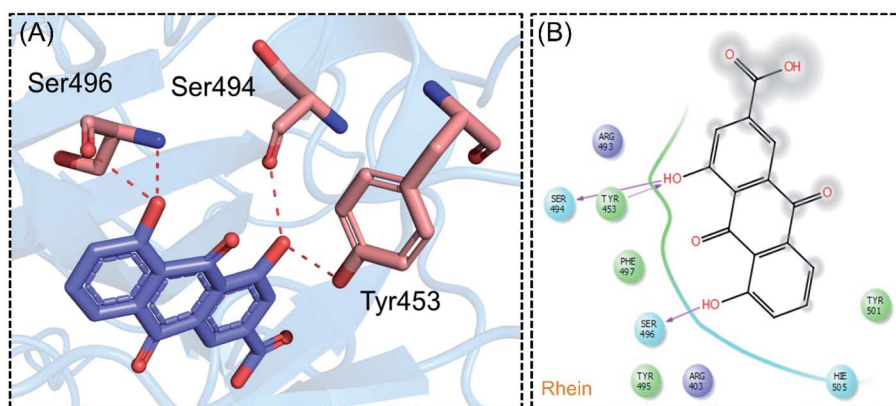


Fig. 6 Interaction pattern of Rhein with the RBD of the omicron variant. (A) 3D interaction pattern of Rhein; (B) the 2D interaction pattern generated with Maestro free Academic version 2018-1 (for visualization only).

minor structural perturbation was reported. Afterwards, both the replicas gained stability again and no significant deviation was observed particularly in replica 1. Though replica 2 demonstrated a little deviation in the later part of simulation overall the average RMSD remained lower than replica 1. The complex reached stability at 1.9 Å and the pattern continues until 100 ns of simulation. The average RMSD for C3 was reported to be 1.96 Å for replica 1, while for replica 2, it was reported to be 1.95 Å. The RMSD graph of C3 is given in Fig. 7C. Finally, the RMSD of C4 was also demonstrated to see the binding strength of the Rhein compound. As shown in Fig. 7D, the complex remained stable, however, an increment in the RMSD value over the simulation time was observed. The structure did not report any significant perturbation except a decrease–increase pattern between 40–100 ns. The average RMSD for C4 was

reported to be 2.84 Å for replica 1 while for replica 2 the average RMSD was 2.90 Å. The RMSD graph of C4 is given in Fig. 7D. Overall the simulation of these complexes reported stronger binding and consequently potential robust inhibitory properties against the RBD of omicron and other variants. These results explain the strong inhibitory features, of C1 and C2 particularly, owned by the shortlisted drugs.

### Structural compactness analysis

Next, we evaluated the structural compactness of each complex in a dynamic environment to reveal the binding and unbinding events that occurred during the simulation. Herein, the structural compactness was computed as the radius of gyration ( $R_g$ ) as a function of time. As given in Fig. 8A, the  $R_g$  of C1 followed a similar pattern as RMSD. The complex initially reported higher  $R_g$  values for a shorter period between 8–16 ns and then the  $R_g$  value decreased back and continues to follow a uniform

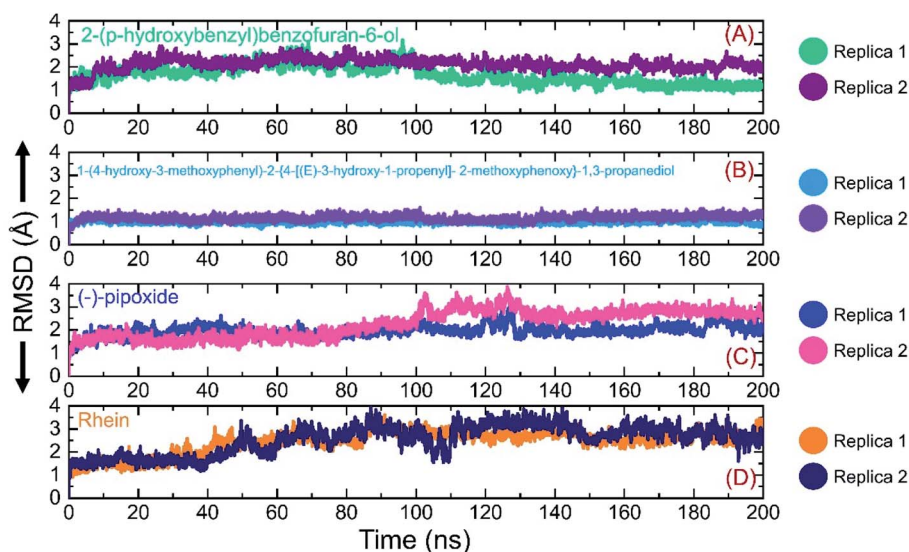


Fig. 7 Binding stability analysis of the top complexes C1–C4. (A) RMSD of C1 during simulation. (B) RMSD of C2 during simulation. (C) RMSD of C3 during simulation. (D) RMSD of C4 during simulation.



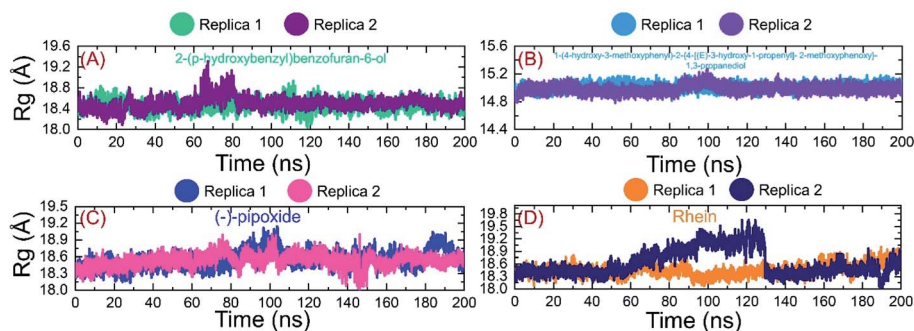


Fig. 8 Structural compactness analysis of the top complexes C1–C4 calculated as  $R_g$ . (A)  $R_g$  of C1 during simulation. (B)  $R_g$  of C2 during simulation. (C)  $R_g$  of C3 during simulation. (D)  $R_g$  of C4 during simulation.

pattern until 100 ns. In the extended simulation until 200 ns a uniform  $R_g$  pattern was observed in replica 1. Replica 2 reported an increment in the  $R_g$  value between 60–80 ns and then decreased back. A uniform pattern was followed until 200 ns with no significant increase or decrease in the  $R_g$  value. The average  $R_g$  value for C1 was calculated to be 18.4 Å for replica 1 while for replica 2 the average  $R_g$  value was 18.0 Å. Likewise, the C2 complex demonstrated a similar  $R_g$  pattern as RMSD, thus determining the tighter binding of C2 during the simulation. The average  $R_g$  values for C2 were reported to be 15.0 and 15.1 Å for replica 1 and replica 2, respectively. On the other hand, C3 demonstrated a little perturbation in the  $R_g$  value, particularly between 1–20 ns and then despite a gradual increment in a uniform pattern was reported. The  $R_g$  for C3 (replica 1 and replica 2) demonstrated a little perturbation between 80–100 ns and then no significant convergence was experienced. The C4 complex also reported a more similar pattern to RMSD. The complex initially reported a flattened graph until 40 ns, then a little decrease and increased at 65 ns. The  $R_g$  then decreased again in the end. The average  $R_g$  for C4 was reported to be 18.42 Å. Conclusively these results report a stable binding of the ligand on the binding surface with minimal unbinding events and favourable inhibition properties. The  $R_g$  graphs of C1–C4 (both replicas) are shown in Fig. 8A–D.

### Residual flexibility analysis

An insight into the residue-level fluctuations of the wild and the variants was further accomplished as such local level flexibility conferred strength to intermolecular binding, negatively impacting molecular recognition, and can potentially influence the overall function of the biological molecule. Herein, we calculated the residual flexibility as root mean square fluctuation (RMSF). Higher and lower RMSF implies flexible and stable regions, respectively (Fig. 9A). All the complexes here demonstrated a similar pattern of residual flexibility. All the complexes except C3 reported a higher RMSF for region 360–375 then the C4 complex only demonstrated a higher RMSF for region 378–385. The region between 472–482 in all the complexes also demonstrated a higher RMSF. Previously the three loops shown in Fig. 9B were reported to confer an important role in the higher affinity and increased transmissibility. Abbas *et al.*

reported that the three loops reported higher flexibility in almost all the variants thus explaining the importance of these loops in the binding.<sup>7,8,36</sup> It can be seen that the residual flexibility index of these loops is decreased and thus reduces the chances of binding with hACE2. This implies that the binding of these ligands produces different internal residual dynamics patterns and causes to stabilize the movement of loops required for direct interaction with hACE2 to minimize the chances of coupling with the receptor. For replica 2 all the complexes demonstrated a more similar pattern of flexibility but a little higher fluctuation in each complex was witnessed. The RMSF graphs of these complexes replica 2 are given in ESI Fig. S1A and B.†

### Hydrogen bonding analysis

Analysis of hydrogen bonds demonstrates the binding strength of the interacting molecules. Hence, to observe how the hydrogen bonding pattern changed over the simulation, we calculated the total number of hydrogen bonds in each trajectory. The calculated hydrogen bonds for each complex as a function of time are given in ESI Fig. S2.† Moreover, we also calculated the hydrogen bonding half-life for each interacting

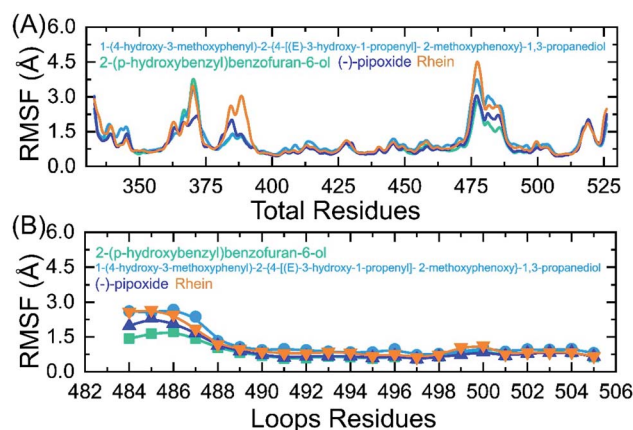


Fig. 9 (A) Residual flexibility analysis of C1, C2, C3 and C4-RBD complexes. (B) Flexibility index of the three binding loops (484–505) required for interaction.





residue in the complex. The interacting residues include Arg403, Glu406, Ser446, Tyr453, Arg493, Ser494, Ser496, Tyr501 and His505. The half-life analysis of the hydrogen bonds revealed that Glu406, Tyr453, Arg493, Ser494 and Ser496 contributed mainly towards the tighter binding while in the case of C1 and C2 the extra strength to the binding is conferred by Tyr501. The hydrogen bonding percentage of each residue in each complex is shown in Fig. 10.

### Binding free energy estimation

The binding strength of small molecules employing the binding free energy method, *i.e.* MM-GBSA is a widely applicable approach to re-demonstrate the docking stability and correct binding. The MM-GBSA approach, mentioned above, is more computationally affordable than the costly alchemical free energy method. It is one of the accurate methods by comparing with the rational scoring functions. Keeping in mind, the applications of this method, we also utilized the same approach to estimate the binding free energy for C1, C2, C3 and C4-RBD complexes. The total binding free energy (TBFE) for the C1-RBD complex was estimated to be  $-60.38$  kcal mol<sup>-1</sup>, for C2-RBD the TBFE was  $-41.44$  kcal mol<sup>-1</sup>, for C3-RBD complex  $-36.22$  kcal mol<sup>-1</sup> while C4-RBD complex the TBFE  $-44.30$  kcal mol<sup>-1</sup>. The vdW values for C1-RBD, C2-RBD, and C3-RBD and C4-RBD complexes were reported to be  $-51.25$ ,  $-36.72$ ,  $-31.85$  and  $-45.77$  kcal mol<sup>-1</sup>, respectively. For these complexes, the electrostatic energy was reported to be  $-28.14$ ,  $-20.51$ ,  $-17.48$  and  $-23.75$  kcal mol<sup>-1</sup>, respectively. Decisively, this demonstrates that these small molecules effectively bind to the interface residues of RBD robustly and could potentially hinder the interaction with the hACE2. The binding free energy results are given in Table 2.

### Bioactivity prediction and dissociation constant ( $K_D$ )

Estimation of IC<sub>50</sub> values for different complexes of druggable proteins is commonly performed using *in silico* bioactivity prediction. Using the molinspiration server, we calculated the bioactivity scores as: C1-RBD (0.39), C2-RBD (0.21), C3-RBD ( $-0.31$ ) and C4-RBD (0.26), respectively, which, demonstrate the stronger bioactivity against the RBD protein. For re-ranking of protein-ligand complexes using  $K_D$  and validation of potential inhibitory properties of these final compounds, the prodigy-

Table 2 MM/GBSA results for all the complexes including vdW, electrostatic and the total binding energy. All the energies are given in kcal mol<sup>-1</sup>

Complexes names	vdW	Electrostatic	SA	GB	Total
C1-RBD	-51.25	-28.14	-16.32	35.33	-60.38
C2-RBD	-36.72	-20.51	-6.55	22.34	-41.44
C3-RBD	-31.85	-17.48	-13.52	26.63	-36.22
C4-RBD	-45.77	-23.75	-11.36	36.58	-44.30

LIG server predicted the  $K_D$  for each complex as C1-RBD ( $-6.73$ ), C2-RBD ( $-524$ ), C3-RBD ( $-4.87$ ) and C4-RBD ( $-6.11$ ), respectively. This demonstrates that these drugs have the ability to suppress the RBD and halt the interaction with hACE2.

## Conclusions

The identification of novel drugs to control the SARS-CoV-2 pandemic is important to overcome the problem of antibodies escaping features of the newly emerging variants. In this regard, computational methods can hasten the identification of such drugs, which could inhibit the variants of SARS-CoV-2. Hence, using virtual screening and simulation approaches potential drugs against the SARS-CoV-2 omicron (B.1.1.529) variant are identified. Simulation and post-simulation validations such as binding free energy calculation, *in silico* bioactivity, and dissociation constant prediction confirmed the potency of these compounds. In conclusion, this study provides a basis for drug designing against the SARS-CoV-2 variants.

## Funding

Dong-Qing Wei is supported by grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Science Foundation of China (Grant No. 32070662, 61832019, 32030063), the Science and Technology Commission of Shanghai Municipality (Grant No.: 19430750600), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2021ZD02).

## Data availability

All the data is available on RCSB, UniProt and any simulation data would be provided on reasonable demand. The accession numbers to access this data are given in the manuscript.

## Conflicts of interest

There are no conflicts of interest.

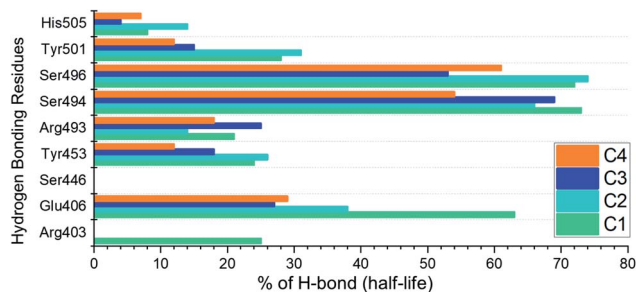


Fig. 10 Hydrogen bonding half-life estimation for the C1–C4 complexes during the 200 ns simulation.



## Acknowledgements

The computations were partially performed at the PengCheng Lab. and the Center for High-Performance Computing, Shanghai Jiao Tong University.

## References

- 1 K. Tao, P. L. Tzou, J. Nouhin, R. K. Gupta, T. de Oliveira, S. L. Kosakovsky Pond, D. Fera and R. W. Shafer, *Nat. Rev. Genet.*, 2021, **22**, 757–773.
- 2 D. Duong, *Can. Med. Assoc. J.*, 2021, E1059–E1060.
- 3 B. Meng, S. A. Kemp, G. Papa, R. Datir, I. A. Ferreira, S. Marelli, W. T. Harvey, S. Lytras, A. Mohamed and G. Gallo, *Cell Rep.*, 2021, **35**, 109292.
- 4 P. Han, C. Su, Y. Zhang, C. Bai, A. Zheng, C. Qiao, Q. Wang, S. Niu, Q. Chen and Y. Zhang, *Nat. Commun.*, 2021, **12**, 1–9.
- 5 A. S. Luring and P. N. Malani, *JAMA*, 2021, **326**, 880.
- 6 A. Khan, J. Gui, W. Ahmad, I. Haq, M. Shahid, A. A. Khan, A. Shah, A. Khan, L. Ali and Z. Anwar, *RSC Adv.*, 2021, **11**, 30132–30147.
- 7 A. Khan, D.-Q. Wei, K. Kousar, J. Abubaker, S. Ahmad, J. Ali, F. Al-Mulla, S. S. Ali, N. Nizam-Uddin and A. M. Sayaf, *ChemBioChem*, 2021, 2641–2649.
- 8 A. Khan, T. Zia, M. Suleman, T. Khan, S. S. Ali, A. A. Abbasi, A. Mohammad and D. Q. Wei, *J. Cell. Physiol.*, 2021, **236**, 7045–7057.
- 9 P. Mlcochova, S. A. Kemp, M. S. Dhar, G. Papa, B. Meng, I. A. Ferreira, R. Datir, D. A. Collier, A. Albecka and S. Singh, *Nature*, 2021, **599**, 114–119.
- 10 S. Messali, A. Bertelli, G. Campisi, A. Zani, M. Ciccozzi, A. Caruso and F. Caccuri, *J. Med. Virol.*, 2021, 6468–6470.
- 11 P. E. Romero, A. Dávila-Barclay, G. Salvatierra, L. González, D. Cuicapuza, L. Solís, P. Marcos-Carbajal, J. Huancachoque, L. Maturrano and P. Tsukayama, *Microbiol. Spectrum*, 2021, **9**, e00789–e00721.
- 12 X. He, W. Hong, X. Pan, G. Lu and X. Wei, *MedComm*, 2021, 838–845.
- 13 S. S. A. Karim and Q. A. Karim, *Lancet*, 2021, **398**, 2126–2128.
- 14 A. Wilhelm, M. Widera, K. Grikscheit, T. Toptan, B. Schenk, C. Pallas, M. Metzler, N. Kohmer, S. Hoehl and F. A. Helfritz, *medRxiv*, 2021.
- 15 A. Roessler, L. Riepler, D. Bante, D. von Laer and J. Kimpel, *medRxiv*, 2021.
- 16 F. Li, *Annu. Rev. Virol.*, 2016, **3**, 237–261.
- 17 F. Li, W. Li, M. Farzan and S. C. Harrison, *Science*, 2005, **309**, 1864–1868.
- 18 P. W. Rose, A. Prlić, A. Altunkaya, C. Bi, A. R. Bradley, C. H. Christie, L. D. Costanzo, J. M. Duarte, S. Dutta and Z. Feng, *Nucleic Acids Res.*, 2016, gkw1000.
- 19 M. Magrane, *Database*, 2011, 2011.
- 20 B. Webb and A. Sali, in *Structural Genomics*, Springer, 2021, pp. 239–255.
- 21 B. Webb and A. Sali, *Curr. Protoc. Bioinf.*, 2016, **54**, 5.6.1–5.6.37.
- 22 F. Ntie-Kang, D. Zofou, S. B. Babiaka, R. Meudom, M. Scharfe, L. L. Lifongo, J. A. Mbah, L. M. a. Mbaze, W. Sippl and S. M. Efange, *PLoS One*, 2013, **8**, e78085.
- 23 D. Lagorce, L. Bouslama, J. Becot, M. A. Miteva and B. O. Villoutreix, *Bioinformatics*, 2017, **33**, 3658–3660.
- 24 P. A. Ravindranath, S. Forli, D. S. Goodsell, A. J. Olson and M. F. Sanner, *PLoS Comput. Biol.*, 2015, **11**, e1004586.
- 25 D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
- 26 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 2001, **222**, U403.
- 27 D. R. Roe and T. E. Cheatham III, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
- 28 A. Khan, S. S. Ali, M. T. Khan, S. Saleem, A. Ali, M. Suleman, Z. Babar, A. Shafiq, M. Khan and D.-Q. Wei, *J. Biomol. Struct. Dyn.*, 2020, 1–12.
- 29 A. Khan, W. Heng, Y. Wang, J. Qiu, X. Wei, S. Peng, S. Saleem, M. Khan, S. S. Ali and D.-Q. Wei, *Phytother. Res.*, 2021, 2841–2845.
- 30 A. Khan, A. C. Kaushik, S. S. Ali, N. Ahmad and D.-Q. Wei, *RSC Adv.*, 2019, **9**, 10326–10339.
- 31 A. Khan, M. Khan, S. Saleem, Z. Babar, A. Ali, A. A. Khan, Z. Sardar, F. Hamayun, S. S. Ali and D.-Q. Wei, *Interdiscip. Sci.: Comput. Life Sci.*, 2020, **12**, 335–348.
- 32 Y. Wang, A. Khan, A. Chandra Kaushik, M. Junaid, X. Zhang and D.-Q. Wei, *J. Biomol. Struct. Dyn.*, 2019, **37**, 4051–4069.
- 33 L. C. Xue, J. P. Rodrigues, P. L. Kastiris, A. M. Bonvin and A. Vangone, *Bioinformatics*, 2016, **32**, 3676–3678.
- 34 A. Vangone, J. Schaarschmidt, P. Koukos, C. Geng, N. Citro, M. E. Trellet, L. C. Xue and A. Bonvin, *Bioinformatics*, 2019, **35**, 1585–1587.
- 35 A. Khan, H. Waris, M. Rafique, M. Suleman, A. Mohammad, S. S. Ali, T. Khan, Y. Waheed, C. Liao and D.-Q. Wei, *Int. J. Biol. Macromol.*, 2022, 438–448.
- 36 A. Khan, J. Gui, W. Ahmad, I. Haq, M. Shahid, A. A. Khan, A. Shah, A. Khan, L. Ali, Z. Anwar, M. Safdar, J. Abubaker, N. N. Uddin, L. Cao, D.-Q. Wei and A. Mohammad, *RSC Adv.*, 2021, **11**, 30132–30147.

