

Rational approaches to discover SARS-CoV-2/ACE2 interaction inhibitors: from pharmacophore-based virtual screening to molecular dynamics

Mohsen Yazdani

Laboratory of Bioinformatics and Drug Design, Institute of Biochemistry and Biophysics, University of Tehran, Tehran

Ameneh Jafari

Proteomics Research Center, Shahid Beheshti University of Medical Sciences, Tehran

Soodeh Mahdian (✉ s.mahdian63@gmail.com)

Department of Cellular and Molecular Biology, Faculty of Biological Science, North Tehran Branch, Islamic Azad University, Tehran

Sajjad Gharaghani

Laboratory of Bioinformatics and Drug Design, Institute of Biochemistry and Biophysics, University of Tehran, Tehran

Research Article

Keywords: SARS-CoV-2, drug design, pharmacophore, structure-based modeling, molecular docking, molecular dynamics, protein-protein interaction

Posted Date: April 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-428179/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Rational approaches to discover SARS-CoV-2/ACE2 interaction inhibitors: from pharmacophore-based virtual screening to molecular dynamics

Mohsen Yazdani ^{1†}, Ameneh Jafari ^{2†}, Soodeh Mahdian^{3**}, Sajjad Gharaghani^{1*}

¹ Laboratory of Bioinformatics and Drug Design, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

² Proteomics Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

³ Department of Cellular and Molecular Biology, Faculty of Biological Science, North Tehran Branch, Islamic Azad University, Tehran, Iran

***Corresponding author:** Sajjad Gharaghani

E-mail: s.gharaghani@ut.ac.ir

tel: +98 913 222 9491

****Co-corresponding author:** Soodeh Mahdian

E-mail: s.mahdian63@gmail.com

Abstract

The lack of effective treatment remains the biggest bottleneck in combating the current novel coronavirus (COVID-19) pandemic. Drug discovery is a complex and time-consuming process, while various in silico methods dramatically reduce the time and cost of this process. The infection begins with the binding of the receptor-binding domain (RBD) of Spike (S) glycoprotein from SARS-CoV-2 to the host angiotensin-converting enzyme (ACE2) receptor. We, therefore, present computational screening methods aimed to identify inhibitors potentially able to target RBD. For this purpose, pharmacophore mapping, molecular docking, and molecular dynamics (MD) simulations were performed to screen potential candidates against ACE2/SARS-CoV-2. In particular, pharmacophore-based virtual screening used over a hundred million compounds to construct protein-protein interactions (PPIs) inhibitor library. In conclusion, PubChem-84280085 compound is suggested as a potential PPIs inhibitor for preventing SARS-CoV-2 RBD/ACE2 interactions.

Keywords: SARS-CoV-2, drug design; pharmacophore, structure-based modeling, molecular docking, molecular dynamics, protein-protein interaction

Introduction

The world is currently facing the challenge of the COVID-19 outbreak associated with a novel strain of coronavirus. In the last two decades, two outbreaks have also been caused by coronaviruses, severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS), occurring in 2002 and 2012, respectively [1]. The recent outbreak, severe acute respiratory syndrome (SARS-CoV-2), was first identified in China and spread so rapidly around the world that the World Health Organization (WHO) declared COVID-19 as a pandemic on March 11, 2020 [2]. The prevalence of COVID-19 and the mortality rate from respiratory disease and multiple organ damage have not yet been controlled. To date (December 8, 2020), one year after the start of the pandemic, more than 65.8 million infections and over 1.5 million deaths have been reported worldwide, and the numbers are increasing daily [3]. COVID-19 is currently a global concern for humanity, and it is unclear how long this situation will continue. The pandemic also has tremendous economic damage in most countries around the world [4]. For these reasons, the researchers' efforts are globally trying to discover effective drugs and vaccines to tackle the SARS-CoV-2 infection [5].

The coronavirus genome encodes four structural proteins: spike (S) protein, membrane (M) protein, envelope (E) protein, and nucleocapsid (N) protein [6]. The S protein, located in the outer envelope, promotes the interaction and entry of SARS-CoV-2 into the host cell. The closed state and the open state are two different structures for the S protein [7]. In the closed S state, the three recognition motifs at the interface between protomers are buried. As a result, the open state is necessary for interacting with ACE2 at the host cell surface, facilitating membrane fusion and viral entry [8] (Figure 1).

The spike protein consists of two subunits: S1 contains the receptor-binding domain (RBD), which serves as a receptor-binding region, and S2, which includes the fusion machinery and is

responsible for membrane fusion [9]. The S1 region consists of the N-terminal domain (NTD) and C-terminal domain (CTD) [10]. The first entry step for the SARS-CoV-2 life cycle is binding to the host cellular membrane protein, angiotensin-converting enzyme 2 (ACE2), via CTD [10, 11]. In SARS-CoV-2-CTD, atomic information shows that a series of hydrophilic residues form a strong network of hydrogen bonds and salt bridge interaction substitutions at the binding interface [12]. These protein-protein interactions (PPIs) play a critical role in the virus-receptor interactions [12, 13]. Some PPI inhibitors have entered clinical trials in recent years, and some have been licensed for marketing, showing that PPI inhibitors have a wide perspective. In the anti-COVID-19 drug development context, targeting the PPI interface and blocking coronavirus-host receptor interactions via the discovery of PPI inhibitors could be useful to fight the current coronavirus epidemic. As some specific monoclonal antibodies (e.g., CR3022) have been developed to effectively inhibit the viral invasion by blocking the interactions between RBD domain and ACE2. Since discovering and developing a new drug is a very lengthy and costly process, a combination of computational methods is an excellent replacement to identify potential drug candidates from the large compound libraries [14]. Such in silico approaches are more rational, cost-effective, and time-saving [15]. In this regard, many attempts have recently been made to design new drugs against SARS-CoV-2 infection using computational methods [9, 16, 17]. These approaches including pharmacophore-based virtual screening, molecular docking, and molecular dynamics (MD) simulation. Pharmacophore-based virtual screening over the chemical libraries is an exciting technique that has been successfully applied for supporting in-silico hit discovery, hit-to-lead expansion, and lead optimization [18]. A pharmacophore leads to the specific molecular identification of ligands by biological macromolecules with specific selectivity [19].

In the present study, we designed computational screening methods to identify potential compounds as inhibitors of SARS-CoV-2 and ACE2 in three main stages. Firstly, we applied a rational approach to mapping the pharmacophoric features of the hot spot regions in the PPI interface between SARS-CoV-2 and ACE2. Also, pharmacophore-based virtual screening was performed on over a hundred million compounds to construct the PPI inhibitor library against ACE2/SARS-CoV-2 RBD. Secondly, the library was subjected to molecular docking to select hit compounds as potential inhibitors. Finally, the selected hit compounds were screened by molecular dynamics simulations and binding free energy calculation to confirm the docked protein-ligand complex's stability and affinity.

Method:

Structural analysis of Spike/ACE2 interaction

To investigate the interface between the SARS-CoV-2 RBD and ACE2, the complex crystal structure of ACE2/SARS-CoV-2 RBD (PDB id: 6M0J) was taken from the Protein Data Bank. Moreover, BIOVIA Discovery Studio 2020 software was employed to analyze intramolecular interactions, such as hydrogen bonding, ionic interactions, and hydrophobic contacts of PPI interactions [20].

Pharmacophore query construction

The pharmacophore query for the ACE2-SARS-CoV-2 RBD complex interfaces was built using the PocketQuery feature generation PPI target-based capability. PocketQuery is an online platform that can use machine learning techniques to identify hotspot amino acids and drug binding sites [21]. PocketQuery first predicts and rates the interface residues in terms of druggability for a query

PPI complex. A machine learning technique, the support vector machine (SVM), predicts the druggability of a residue, the change of solvent accessible surface (SASA) area upon complexation (Δ SASA), percentage of the total possible SASA(Δ SASA%), estimation of the change of free energy of an alanine mutation (Rosetta Energy ($\Delta\Delta$ G)), the change of free energy of a residue upon complexation (FastContact Energy (Δ G)), a sequence conservation score, and an evolutionary rate of the residue [22]. The following search filters were applied at this point to find druggable amino acid clusters on the spike: average Δ SASA > 44.6233 average Δ SASA% > 39.6, Cluster Score > 0.55, and Cluster Residue = 3. The clusters were selected based on the highest Cluster Score and the maximum binding site coverage.

Pharmacophore-based virtual screening

Selected clusters were used as a pharmacophore model to conduct a 3D pharmacophore search by the pharmit server [23], which offers both pharmacophore and molecular shape search methods and a rating of results by energy minimization. Since the Pharmit server presented many molecules in the outcome, the resulting hits were refined by applying various filters, such as total hits, hits per conformer, molecular weights, number of rotatable bonds, log P, polar surface area (PSA), number of aromatic rings and RMSD.

Docking-based virtual screening

Hits obtained from the pharmacophore search collected and duplicate compounds were removed by the Babylon tools. The 3D structure of the SARS-CoV-2 spike was selected as the target protein and acquired from Protein Data Bank (PDB Code 6M0J). Ligands and protein were prepared using AutoDock Tools and docked with AutoDock Vina software [24]. Docking was done using a 50 Å°

× 800 Å × 50 Å binding site grid box centered on the position of Gly493, with the exhaustiveness parameter of 32.

Molecular dynamics simulation

MD simulations were done via GROMACS 5.1.4 package [25] using gromos 54a7 force fields and SP3 water model [26]. It was used the ATB server to prepare the coordinates and topology of ligands [27]. The system was neutralized with the appropriate amounts of chloride and sodium ions. Periodic Boundary Condition (PBC) was applied along every simulation box axis in each simulation system [28]. The LINCS algorithms constrained all covalent bonds. A short-range electrostatic interaction in addition to a 1.2 nm distance cutoff for the van der Waals interaction [29] and the Particle Mesh Ewald (PME) algorithm calculated the long-range electrostatic interaction. The steepest descent algorithm fulfilled the energy minimization of all systems, then the NVT ensemble for 500 ps equilibrated all the systems. Afterward, the NPT ensemble progressively directed the equilibration of each system and the Nose-Hoover algorithm temperature [30] preserved and the temperature at 310 K. During the NPT equilibration, the Parrinello-Rahman barostat [31] maintained the pressures at 1 bar. MD simulation was completed for the complexes in 100 ns.

Binding free-energy analyses

Upon performing MD simulations, Gromacs utilities examined and evaluated the results of every trajectory. The nonpolar and polar interactions between the RBD and ligands can be explained via binding free-energy calculation. Through exercising the MM-PBSA method, the binding free-energy was calculated with the `g_mmpbsa` tool [32].

The total amount of binding free-energy (ΔG) is realized by adding up the nonpolar interaction free energy ($\Delta G_{\text{nonpolar}}$) and the polar interaction free energy (ΔG_{polar}) that can be explained as follows:

$$\Delta G_{\text{total}} = \Delta G_{\text{nonpolar}} (\Delta G_{\text{nonpolar solvation energy}} + \Delta G_{\text{vdW van der Waals energy}}) \\ + \Delta G_{\text{polar}} (\Delta G_{\text{polar solvation energy}} + \Delta G_{\text{electrostatic energy}})$$

Results

Analysis of the interface between the SARS-CoV-2 RBD and ACE2

As illustrated in Figure 2, the core structure of SARS-CoV-2 RBD consists of a twisted five-stranded antiparallel β -sheets ($\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$, and $\beta 7$) with short α helices connected by flexible loops. In addition, between the $\beta 4$ and $\beta 7$ strands in the core, there is an extended insertion containing short β -sheets ($\beta 5$ and $\beta 6$), α -helices ($\alpha 6$ and $\alpha 7$), and loops (A-B-C) (Figure 2a). This groove-shaped extended insertion is the receptor-binding motif (RBM) that slides and interacts with the claw-like structure of the ACE2 receptor protease domain (PD). ACE2 binds to RBD through N-terminal $\alpha 1$ and $\alpha 2$ antiparallel helices as well as the $\beta 3$ – $\beta 4$ loop (Figure 2b). In the contact area, a total of 29 residues, including 16 residues of RBM in contact with 13 residues of ACE2, are involved in their interactions. This contact surface can be divided into three regions **(Error! Reference source not found.)**:

1-At the C-terminus of $\alpha 1$ helix (Asp38, Tyr41, and Gln42) and $\beta 3$ – $\beta 4$ loop (Gly354 and Lys353) from the AEC2 form a network of H-bonds with loop A (Gly446, Tyr449,) loop C (Gly496, Gln498, Thr500, Asn501, Gly502,) and short $\alpha 7$ helix (Tyr505) of the RBD.

2-Middle of $\alpha 1$ helix ACE2 (Asp30, Lys31, His34, and Glu35) which form H- bonds with $\beta 4$ strands (Tyr453, Lue455, Phe456), and $\beta 5$ strands (Gln493) from the RBD, the salt bridge between

Asp30 of ACE2 and Lys417 of RBD and π - π interaction between His34 of ACE2 and Tyr453 of RBD.

3- At the N-terminus of α 1 helix and C-terminus α 2 helix (Gln24, Thr27, Met82) of ACE2 form H-bonds with loop B (Ala475, Phe486, Asn487, and Tyr489) from the RBD, and Tyr83 of ACE2 form π - π interaction with Phe486 of RBD

Pharmacophore query construction

The pharmacophore model reconstruction results on the SARS-CoV-2/ACE2 complex were visually evaluated, and six clusters with three amino acid members were selected based on the score, the maximum coverage of the complex contact surface, and the presence of hotspot amino acids (Table 1). Cluster 1 (Asp30, Lys31, And His34) with a score of 0.863 covers the only region (2) of the complex. Clusters 2 (Thr27; Lys31; His34), cluster 3 (Thr27; Asp30; His34), and cluster 5 (Gln24; Thr27; Asp30) with scores of 0.821, 0.804, and 0.780, respectively, cover regions (2) and (3). Also, regions (1) and (2) were covered by clusters 4 (Lys31; His34; Asp38) and 6 (His34; Asp38; Lys353) with a score of 0.78 and 0.58. Details of these clusters, which include amino acids, type of pharmacophore feature, position, and radius of the feature sphere, are shown in Supplementary Table S1 and Figure S1.

Pharmacophore-based virtual screening

Clusters obtained from PocketQuery were used as a query for pharmacophore-based virtual screening by Pharmit. Virtual screening for each cluster was performed on six public databases: PubChem [33], ChEMBL [34], Zinc [35], Mcule [36], ChemSpace [34], and MolPort [37]. During the virtual search, the Lipinski Rule of five was applied. After the pharmacophore search's minimization stage, hits were selected based on the RMSD less than three and the binding energy

less than -6 KJ. The number of obtained hits for all clusters was 358 compounds. After removing duplicate compounds, 350 unique hits were identified. The number of obtained hits for each cluster is shown in Supplementary Table S2.

Docking-based virtual screening

The best hits obtained by pharmacophore-based screening were further screened using molecular docking. Molecular docking is a well-known method to predict the binding-conformation of ligands to the appropriate target binding site [38]. The docking processes were carried out using autodock vina software to dock all 350 compounds. After molecular docking, the top-ranked 50 molecules were selected for manual observation according to docking scores and binding mods. Following manual selection, the ten molecules with a high docking score, low RMSD, and suitable binding mode were selected as potential active molecules for molecular dynamics (Supplementary Figure S2). Docking results revealed that the top 10 compounds interacted with key residues of the contact surface of RBD (RBM) and exhibited a high affinity between -7 and -7.9 kcal/mol. Details of the top 10 compounds, including Docking scores, minimized scores, minimized RMSD, and interactions, are presented in Table 2.

RMSD and RMSF

To recognize the complexes' stability and flexibility, RMSD was employed. The RMSD of complexes is shown in Figure 3. Six complexes including PubChem-84280085, PubChem-6694569, PubChem-136654447, PubChem-117741619, PubChem-46282789, and ChEMBL2172478 had lower RMSD values than the other complexes.

Four complexes, including PubChem-136107209, PubChem-8948287, ZINC000014748612,

IBYKSDUFFDTFQG-UHFFFAOYSA-N showed high RMSD and instability (Figure 3a). The RMSD plot illustrates that the average RMSD of PubChem-89448287 increased from 0.15 to 0.5 nm around 50 ns and reaches equilibrium around 50 ns to 100 ns (Figure 3b). The RMSD values of the PubChem-136107209 complex had oscillation in the beginning, and after 20 ns, it stabilized at 0.3 nm and then increased to 0.5 nm at around 75 ns to 100 ns. The average RMSD of ZINC000014748612 increased from 0.13 to 0.65 nm at around 70 ns and then oscillated between 0.5 and 0.65 nm at around 70 ns to 100. The average RMSD of bound MIBYKSDUFFDTFQG-UHFFFAOYSA-N increased from 0.19 to 0.64 nm at around 70 ns, then oscillated between 0.46 and 0.65 nm at around 70 ns to 100 ns. The RMSF shows how the binding of ligands impacts residues of the protein. Residues having low RMSF values are more stable due to the limited movements during simulations [39]. High fluctuation can be noticed at positions, including short β 5 and β 6 strands, α 4 and α 5 helices, and loops that have been spotted in Figure 4. The study of the loop C – β 5 (488-505 residues) of the RBD, which contains most of the contacting residues of SARS-CoV-2 with ACE2, is important in terms of flexibility and stability. Low RMSF values were observed for PubChem-84280085, PubChem-136107209, and PubChem-46282789 complexes, while PubChem-136654447 complex was showed a high RMSF value in this area. Compared to the other complexes, this is a positive indication that the PubChem-84280085, PubChem-136107209, and PubChem-46282789 complexes have more residual and structural stability. Considerably, higher RMSF values (0.3-1 nm) were also observed in loop B position between residue 473 and 493 in all complexes as an exposed loop domain tends to be highly flexible, leading to structural instability during MD simulation [40]. This high flexibility seems to make it difficult for binding inhibitors as well as antibodies in this region.

The number of hydrogen bonds and the free binding energy

The number of hydrogen bonds and the free binding energy between the protein and the ligands play a key role in stabilizing the protein/ligand complexes [41]. The average number of H-bonds between RBD and the ligands at 310 K has been calculated that shown in Figure 5. The compounds, including PubChem-89448287, ZINC000014748612, PubChem-136107209, and PubChem-84280085, formed the most significant number of hydrogen bonds with an average of 2.5, 2, 1.7, and 1.6, respectively

MM-PBSA Binding free energy calculation

Calculation of ΔG for the polar & nonpolar interactions between RBD and ligands is shown in Table 3. Among ligands, PubChem-84280085, PubChem-89448287, PubChem-136654447, and PubChem-117741619 had the most favorable binding energy with -141.49 kJ/mol, -130.566 kJ/mol, -91.363 kJ/mol, and -82.094 kJ/mol ΔG binding energy, respectively. According to our calculation, van der Waals interactions played a critical role during the binding process as the ΔE_{vdW} of compounds was higher than other interactions. The hydrophobic interactions were also important for the binding affinity. Meanwhile, ΔE_{ele} made a significant contribution to binding free energy. In contrast, polar solvation energy was found to be unfavorable.

Discussion

In the process of finding PPI inhibitors, combining pharmacophore-based virtual screening with docking is helpful to understand how these ligands bind to the PPI interface and to predict the affinity using a scoring function [42]. In addition, molecular dynamics and MM-PBSA can be used to optimize the structures of the final complexes from docking, calculate more detailed interaction energies, and provide information about the ligand-binding mechanism and binding free energy calculations [43, 44]. In order to discover inhibitors of the spike-ACE2 interaction interface, we

established the workflow combining pharmacophore virtual screening, docking, molecular dynamics, and mmpbs free energy calculation techniques. Firstly, hot spot regions in the PPI interface between SARS-CoV-2 and ACE2 were investigated. Our analysis detected 29 amino acids that were involved in interactions between ACE2 and the RBD that were divided into three parts (Figure 2). Secondly, Pocket Query was employed to map the pharmacophore features onto the PPI interface's amino acids. Therefore, six pharmacophore models were obtained based on the score, maximum coverage of the complex contact surface, and the presence of hotspot amino acids (Table 1). Following this, for each model, the 3D pharmacophore search was performed using the Pharmit server on 6 large databases over a hundred million compounds. Subsequently, 356 hits were obtained from Pharmit server by applying various filtering, and ranking criteria, including drug-like properties, Lipinski's "rule of five", the energy minimization, and RMSD described in detail in the methods section. After that, the hits that match a well-defined pharmacophore were analyzed through molecular docking by autodock vina. The top ten molecules with high docking scores, low RMSD, and suitable binding mode were selected as potential active compounds for molecular dynamics analysis and MMPBSA Free energy calculation (Table 3). Consequently, four compounds, including PubChem-84280085, PubChem-89448287, PubChem-136654447, and PubChem-117741619 with high total binding free energy, chose to analyze the binding mode of interactions (Figure 6).

In the context of binding mode analysis, PubChem-84280085 interacts with the residues in the (L1, L4, α 4, β 5, β 4) RBM through hydrogen bonding with GLN493, carbon-hydrogen bond with ASN501, π -Donor hydrogen bond with GLY496, π -cation with ARG403, π -sulfur with TYR495, π -alkyl hydrophobic with PRO507, π - π stake with TYR449 and TYR 505, while forming multiple Van der Waals interactions with TYR453, ASN493, SER494, PHE497, GLN498, THR500,

GLY502, and GLN506. Compound PubChem-89448287 formed H-bond interactions with LUE492 and VAL445, besides other interactions such as π -Donor hydrogen bond with GLY447, Unfavorable acceptor-acceptor with SER494, π -donor hydrogen bond with GLY447 and Van der Waals interactions with GLY446, ASN448, LEU452, PHE490, GLN493, GLY496, TYR 505. Compound PubChem-136654447 formed H-bond interactions with GLU406 and GLN409, carbon-hydrogen bond with LUE492, π -Donor hydrogen bond with TYR473, π -Alkyl with ILU418, LUE455, ALA475, TYR489 and GLN493, π - π -T shaped with PHE 456 and Van der Waals interactions with ARG403, ASP405, LYS 417, TYR453, PHE490, and TYR495. Compound PubChem-117741619 formed H-bond interactions with TYR453 and GLN493 and carbon-hydrogen bond with ASN448 and PHE497, π - π -T shaped with TYR449 and TYR505, π -sulfur with ARG403 and TYR453, and Van der Waals interactions with SER494, TYR495, GLY496, and GLN498. Considering obtained results, compounds PubChem-84280085, PubChem-89448287, and PubChem-117741619 cover regions one and two of spike-ACE2 interaction surface, while compound PubChem-136654447 bind to region two and three (Figure 2c). Also, the high number of van der Waals bonds of these four compounds with spike confirms the result of mmpbs calculation that the ΔE_{vdW} plays an essential role in the $\Delta G_{binding}$. Besides, ΔE_{ele} made a significant contribution to the binding free energy for compounds PubChem-84280085 and PubChem-117741619 due to their strong H-bonds (Table 3). Likewise, the RMSD plot of PubChem-84280085, PubChem-136654447, and PubChem-117741619 compounds shows RMSD value lower than 4 Å led to complex stability during MD simulation (

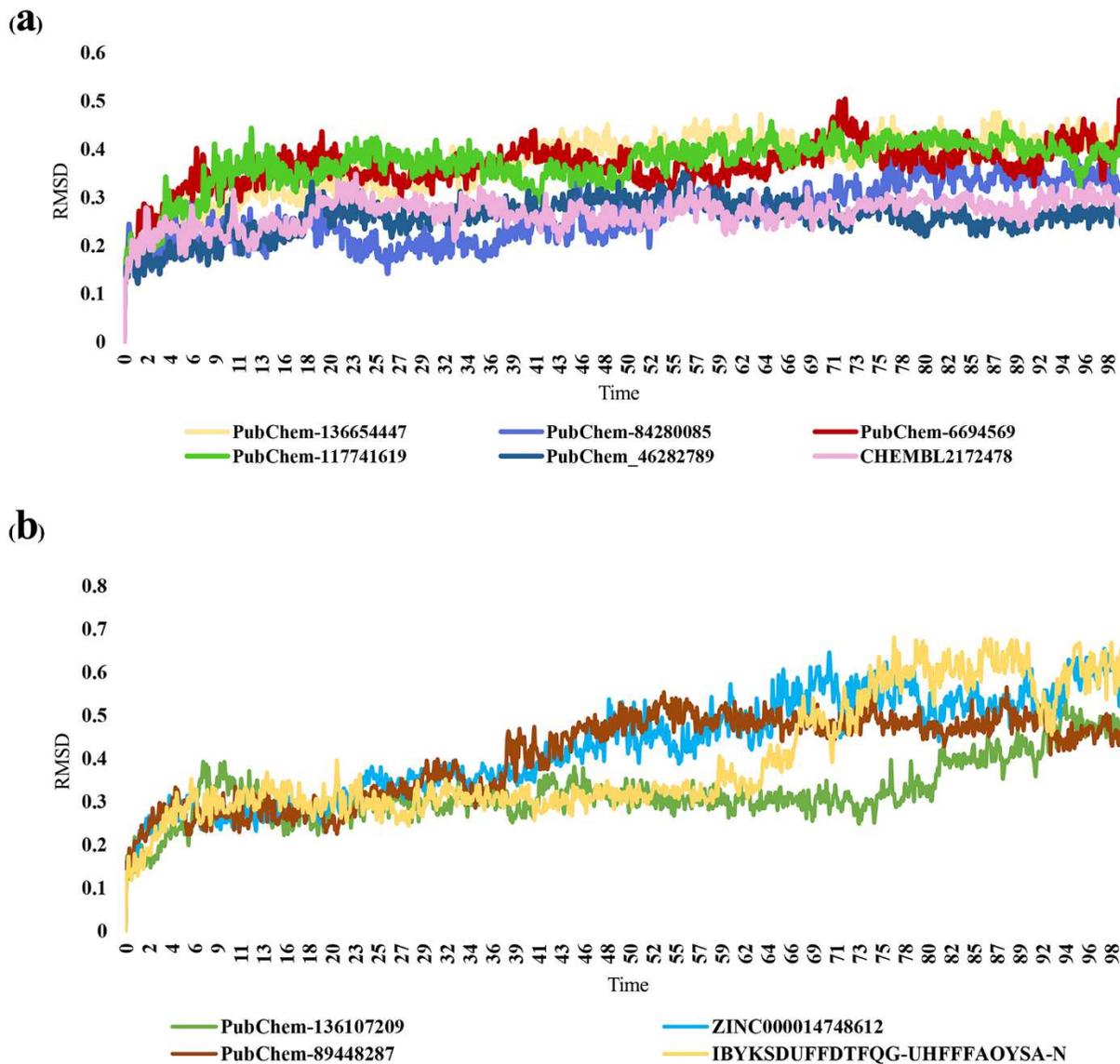


Figure 3). The RMSF result reveals that the complex of PubChem-84280085 is more stable than other complexes (Figure 4).

Conclusion

In the current study, pharmacophore virtual screening, docking, molecular dynamics and MM-PBSA approaches were utilized to discover inhibitors of the SARS-CoV-2 spike-ACE2 interaction. Taking together, high binding energy, a large number of bonds with hot spot amino

acids, maximum coverage of hot spot area, as well as the structural stability of the complex during MD simulation, PubChem-84280085 compound is proposed as a potential PPI inhibitor for preventing interactions between SARS-CoV-2 RBD and ACE2. However, further investigations are needed to confirm these findings.

Ethical Declaration/Conflict of Interest

This work does not require any ethical statement, and the authors declare no competing interests.

Contributions

M.Y and A.J contributed to the conception and design of the study. M.Y and S.M performed bioinformatics analyses; M.Y and A.J Collected and/or assembled data and wrote the manuscript. S.Gh supervised the study; All authors read and approved the manuscript.

References:

1. Lee, P.-I. and P.-R. Hsueh, *Emerging threats from zoonotic coronaviruses-from SARS and MERS to 2019-nCoV*. Journal of microbiology, immunology, and infection, 2020.
2. Jafari, A., et al., *Cancer Care Management During the COVID-19 Pandemic*. Risk Management and Healthcare Policy, 2020. **13**: p. 1711.
3. <https://www.who.int/publications/m/item/weekly-epidemiological-update-8-december-2020>. Accessed 12 December.
4. Chojnacka, K., et al., *Phytochemicals containing biologically active polyphenols as an effective agent against Covid-19-inducing coronavirus*. Journal of Functional Foods, 2020: p. 104146.
5. Molavi, Z., et al., *Identification of FDA Approved Drugs Against SARS-CoV-2 RNA Dependent RNA Polymerase (RdRp) and 3-chymotrypsin-like Protease (3CLpro), Drug Repurposing Approach*. Biomedicine & Pharmacotherapy, 2021: p. 111544.
6. Schoeman, D. and B.C. Fielding, *Coronavirus envelope protein: current knowledge*. Virology journal, 2019. **16**(1): p. 1-22.
7. Li, F., et al., *Structure of SARS coronavirus spike receptor-binding domain complexed with receptor*. Science, 2005. **309**(5742): p. 1864-1868.
8. Walls, A.C., et al., *Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein*. Cell, 2020. **181**(2): p. 281-292. e6.
9. Arun, K., et al., *Drug repurposing against SARS-CoV-2 using E-pharmacophore based virtual screening, molecular docking and molecular dynamics with main protease as the target*. Journal of Biomolecular Structure and Dynamics, 2020: p. 1-12.
10. Huang, Y., et al., *Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19*. Acta Pharmacologica Sinica, 2020. **41**(9): p. 1141-1149.

11. Jyotisha, S. Singh, and I.A. Qureshi, *Multi-epitope vaccine against SARS-CoV-2 applying immunoinformatics and molecular dynamics simulation approaches*. Journal of Biomolecular Structure and Dynamics, 2020: p. 1-17.
12. Mittal, A., et al., *COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2*. PLoS pathogens, 2020. **16**(8): p. e1008762.
13. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor*. Nature, 2020. **581**(7807): p. 215-220.
14. Keretsu, S., S.P. Bhujbal, and S.J. Cho, *Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation*. Scientific reports, 2020. **10**(1): p. 1-14.
15. Vora, J., et al., *Pharmacophore modeling, molecular docking and molecular dynamics simulation for screening and identifying anti-dengue phytochemicals*. Journal of Biomolecular Structure and Dynamics, 2019.
16. Law, W.Y., et al., *Pharmacophore modelling of vanillin derivatives, favipiravir, chloroquine, hydroxychloroquine, monolaurin and tetrodotoxin as M Pro inhibitors of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)*. BMC research notes, 2020. **13**(1): p. 1-8.
17. Gulotta, M.R., et al., *Targeting SARS-CoV-2 RBD interface: a supervised computational data-driven approach to identify potential modulators*. ChemMedChem, 2020. **15**(20): p. 1921-1931.
18. Daoud, S., S.J. Alabed, and L.A. Dahabiyeh, *Identification of potential COVID-19 main protease inhibitors using structure-based pharmacophore approach, molecular docking and repurposing studies*. Acta Pharmaceutica, 2021. **71**(2): p. 163-174.
19. Mishra, V. and C. Pathak, *Structural insights into pharmacophore-assisted in silico identification of protein-protein interaction inhibitors for inhibition of human toll-like receptor 4-myeloid differentiation factor-2 (hTLR4-MD-2) complex*. Journal of Biomolecular Structure and Dynamics, 2019. **37**(8): p. 1968-1991.
20. Dassault Systèmes BIOVIA. Discovery Studio Visualizer, v.S.D., CA: Dassault Systèmes; 2016.
21. Koes, D.R. and C.J. Camacho, *PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure*. Nucleic acids research, 2012. **40**(W1): p. W387-W392.
22. Shin, W.-H., C.W. Christoffer, and D. Kihara, *In silico structure-based approaches to discover protein-protein interaction-targeting drugs*. Methods, 2017. **131**: p. 22-32.
23. Sunseri, J. and D.R. Koes, *Pharmit: interactive exploration of chemical space*. Nucleic acids research, 2016. **44**(W1): p. W442-W448.
24. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading*. Journal of computational chemistry, 2010. **31**(2): p. 455-461.
25. Berendsen, H.J., D. van der Spoel, and R. van Drunen, *GROMACS: a message-passing parallel molecular dynamics implementation*. Computer physics communications, 1995. **91**(1-3): p. 43-56.
26. Schmid, N., et al., *Definition and testing of the GROMOS force-field versions 54A7 and 54B7*. European biophysics journal, 2011. **40**(7): p. 843-856.
27. Malde, A.K., et al., *An automated force field topology builder (ATB) and repository: version 1.0*. Journal of chemical theory and computation, 2011. **7**(12): p. 4026-4037.
28. Hess, B., et al., *GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation*. Journal of chemical theory and computation, 2008. **4**(3): p. 435-447.
29. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems*. The Journal of chemical physics, 1993. **98**(12): p. 10089-10092.
30. Evans, D.J. and B.L. Holian, *The nose-hoover thermostat*. The Journal of chemical physics, 1985. **83**(8): p. 4069-4074.
31. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*. J. Appl. Phys., 2005. **71**82(1981).

32. Kumari, R., et al., *g_mmpbsa* – A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of chemical information and modeling*, 2014. **54**(7): p. 1951-1962.
33. Kim, S., et al., *PubChem in 2021: new data content and improved web interfaces*. *Nucleic Acids Research*, 2021. **49**(D1): p. D1388-D1395.
34. Gaulton, A., et al., *The ChEMBL database in 2017*. *Nucleic acids research*, 2017. **45**(D1): p. D945-D954.
35. Irwin, J.J. and B.K. Shoichet, *ZINC* – a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 2005. **45**(1): p. 177-182.
36. Kiss, R., M. Sandor, and F.A. Szalai, *http://McuLe.com: a public web service for drug discovery*. *Journal of cheminformatics*, 2012. **4**(1): p. 1-1.
37. Bajorath, J., *Integration of virtual and high-throughput screening*. *Nature Reviews Drug Discovery*, 2002. **1**(11): p. 882-894.
38. Meng, X.-Y., et al., *Molecular docking: a powerful approach for structure-based drug discovery*. *Current computer-aided drug design*, 2011. **7**(2): p. 146-157.
39. Kumar, C.V., et al., *Computational analysis reveals the association of threonine 118 methionine mutation in PMP22 resulting in CMT-1A*. *Advances in bioinformatics*, 2014. **2014**.
40. Shesham, R.D., L.J. Bartolotti, and Y. Li, *Molecular dynamics simulation studies on Ca²⁺-induced conformational changes of annexin I*. *Protein Engineering, Design & Selection*, 2008. **21**(2): p. 115-120.
41. Gohlke, H. and G. Klebe, *Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors*. *Angewandte Chemie International Edition*, 2002. **41**(15): p. 2644-2676.
42. Wu, K.-J., et al., *Mimicking strategy for protein–protein interaction inhibitor discovery by virtual screening*. *Molecules*, 2019. **24**(24): p. 4428.
43. Wang, E., et al., *End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design*. *Chemical reviews*, 2019. **119**(16): p. 9478-9508.
44. Poli, G., et al., *Application of MM-PBSA methods in virtual screening*. *Molecules*, 2020. **25**(8): p. 1971.

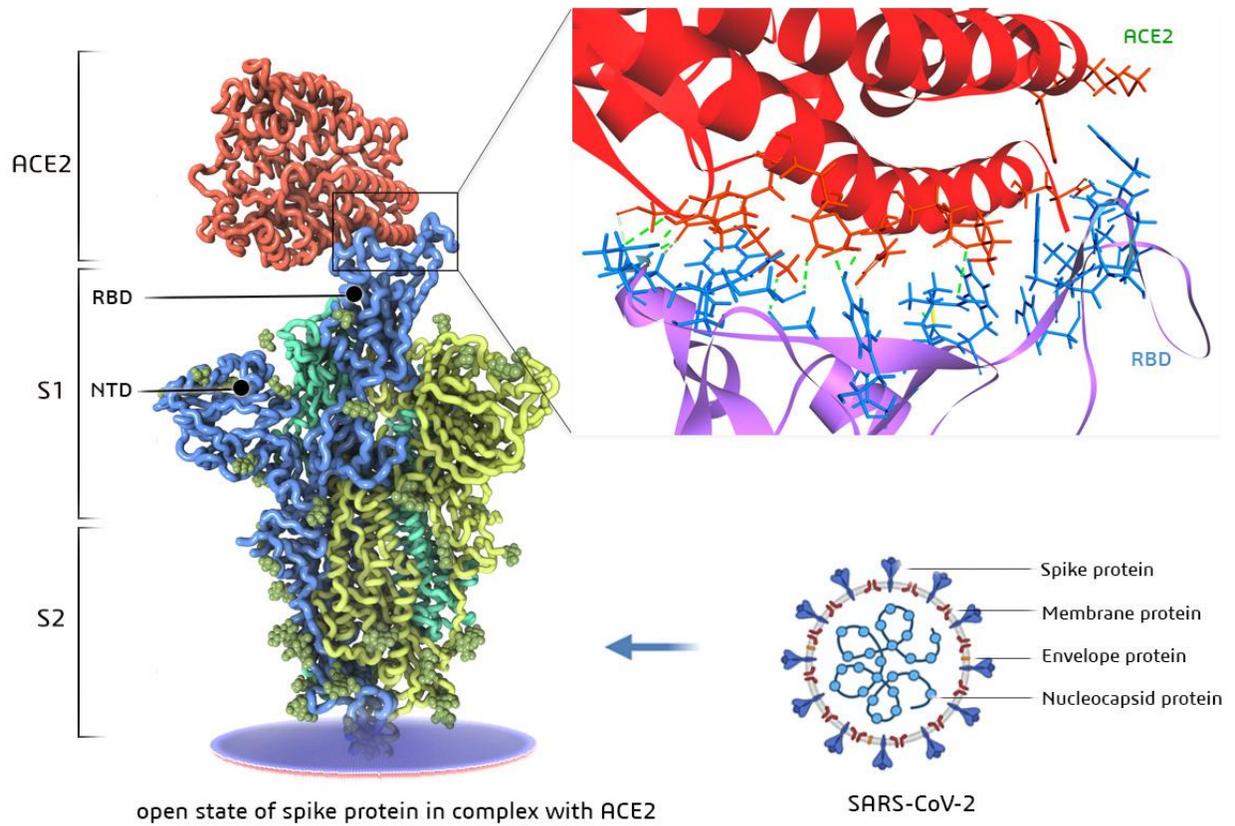


Figure 1

Figure 1. Schematic of SARS-CoV-2 structure and its major proteins: nucleocapsid (N) protein, membrane protein (M), spike protein (S), and envelope protein (E). The ribbon views represent the open state of spike protein in complex with ACE2 (illustrated by Protein Imager). Spike protein is composed of S1 subunit (which includes the N-terminal domain (NTD) and the receptor-binding domain (RBD)) and S2 subunits. With a magnification of the RBD-ACE2 binding interface. ACE2 is colored red, SARS-CoV-2 RBD core is colored blue.

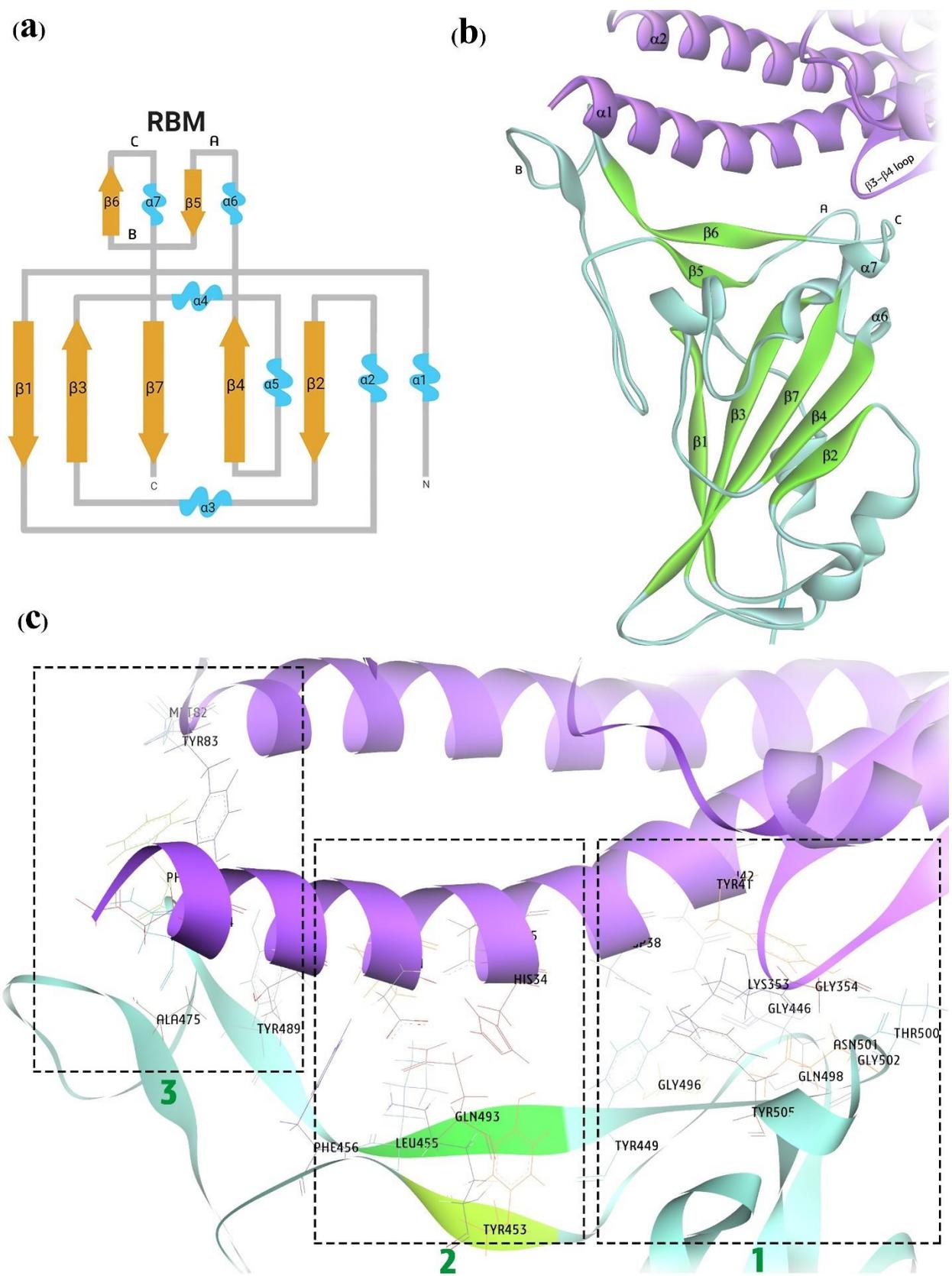


Figure 2

Figure 2.

a. Schematic of SARS-CoV-2 RBD structure. Different structures are shown by different colors. The core structure of SARS-CoV-2 RBD consists of a twisted five-stranded antiparallel β -sheets ($\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$, and $\beta 7$) with short α helices connected by flexible loops. receptor-binding motif (RBM), an extended insertion containing short β -sheets ($\beta 5$ and $\beta 6$), α -helices ($\alpha 6$ and $\alpha 7$), and loops (A-B-C), located between the $\beta 4$ and $\beta 7$ strands in the core.

b. Overall structure of the SARS-CoV-2 RBD (green and gray) bound to ACE2(purple). The ACE2 binds to RBD through N-terminal $\alpha 1$ and $\alpha 2$ antiparallel helices as well as the $\beta 3$ - $\beta 4$ loop.

c. Ribbon representation of the RBD-ACE2 binding interface. ACE2 is shown in blue. This contact surface can be divided into three regions: Region one includes residues in C-terminus of $\alpha 1$ and $\beta 3$ - $\beta 4$ loop from the AEC2 and loop A, loop C, and short $\alpha 7$ residues of the RBD. Region two: includes residues in the middle $\alpha 1$ helix from AEC2 and $\beta 4$ strands and $\beta 5$ strands residues from the RBD. Region three: includes residues in N-terminus of $\alpha 1$ and C-terminus $\alpha 2$ helix of ACE2 and loop B residues from the RBD.

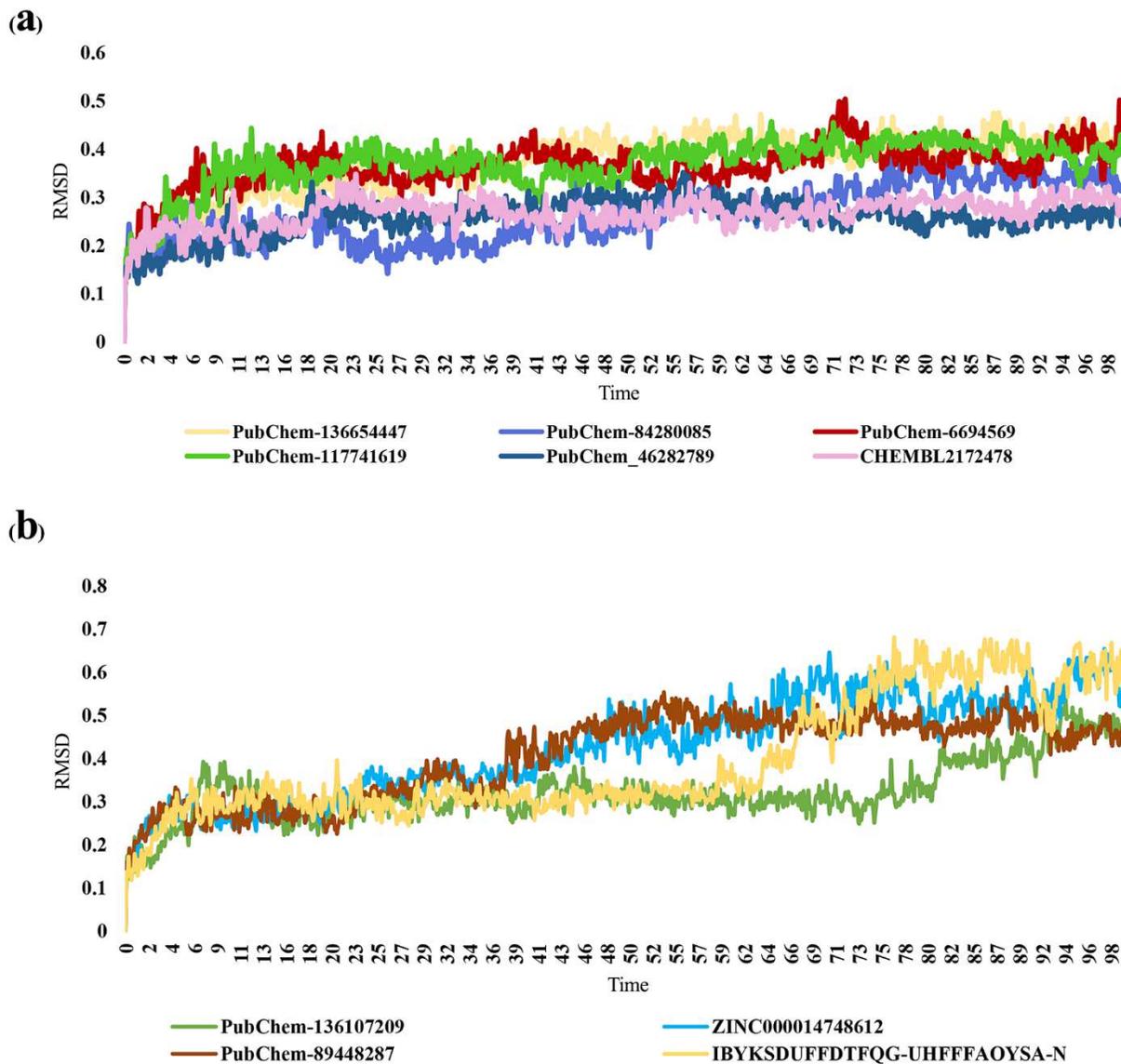


Figure 3

Figure 3. The root mean square deviation (RMSD) of SARS-CoV-2 RBD in complex with ten compounds during 100 ns molecular dynamics simulation. **a.** complexes with RMSD > 0.4 nm. **b.** complexes with RMSD < 0.4 nm.

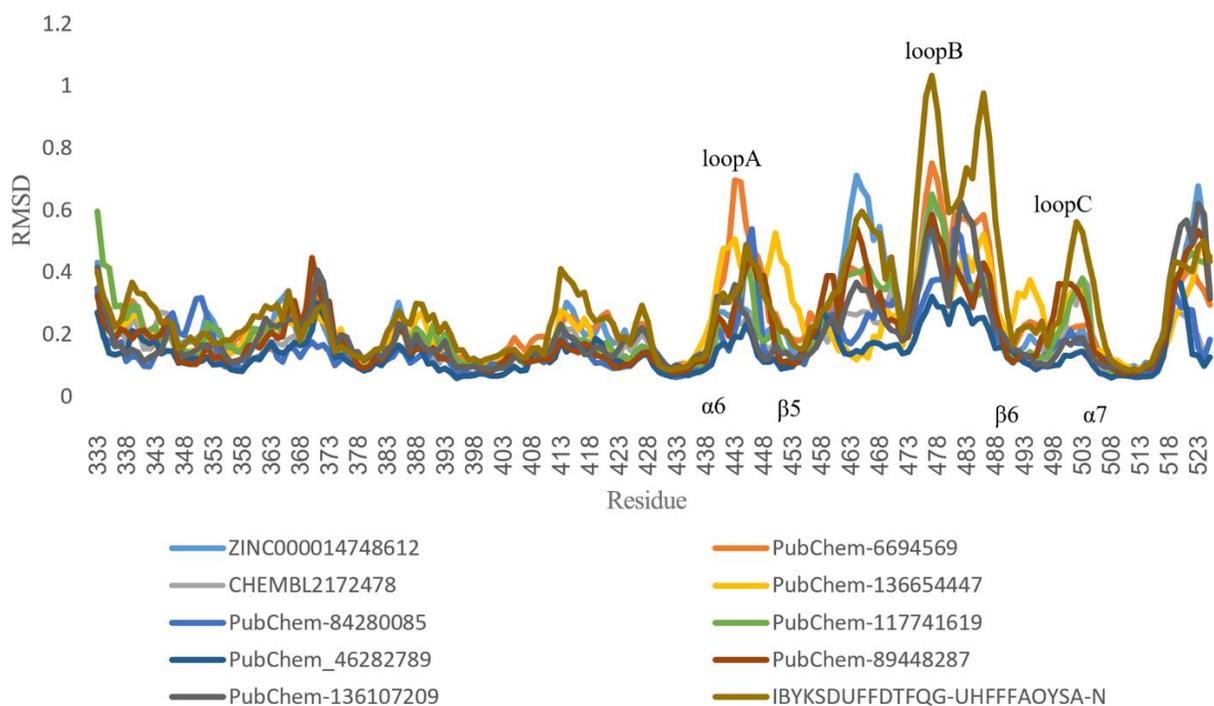


Figure 4

Figure 4. The root mean square fluctuation (RMSF) values of SARS-CoV-2 RBD in complex with ten compounds during 100 ns plotted against residue numbers. Residues with values above 0.5 nm are labeled.

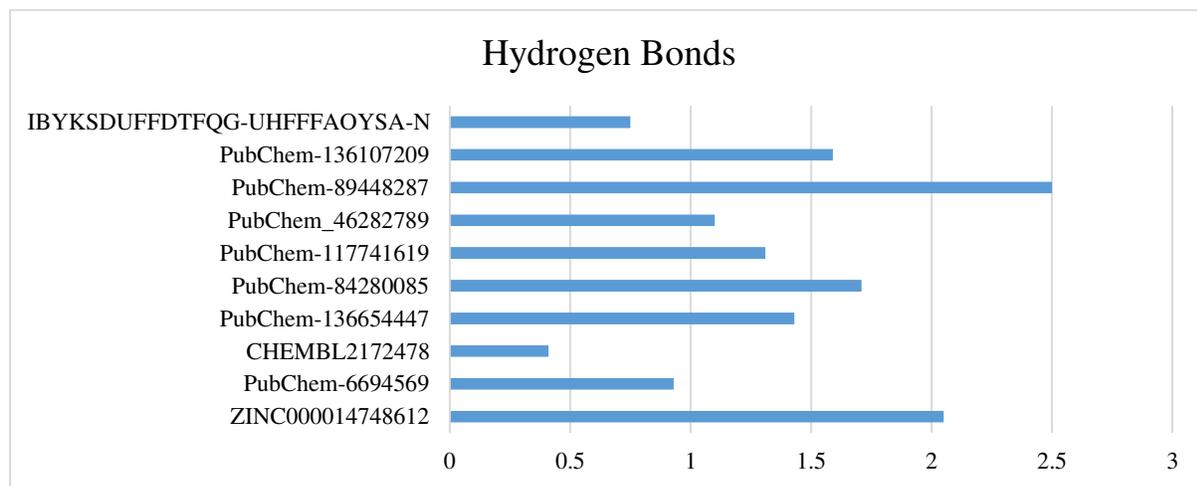


Figure 5

Figure 5. The average number of hydrogen bonds between top ten compounds and RBD during 100 ns of MD simulations.

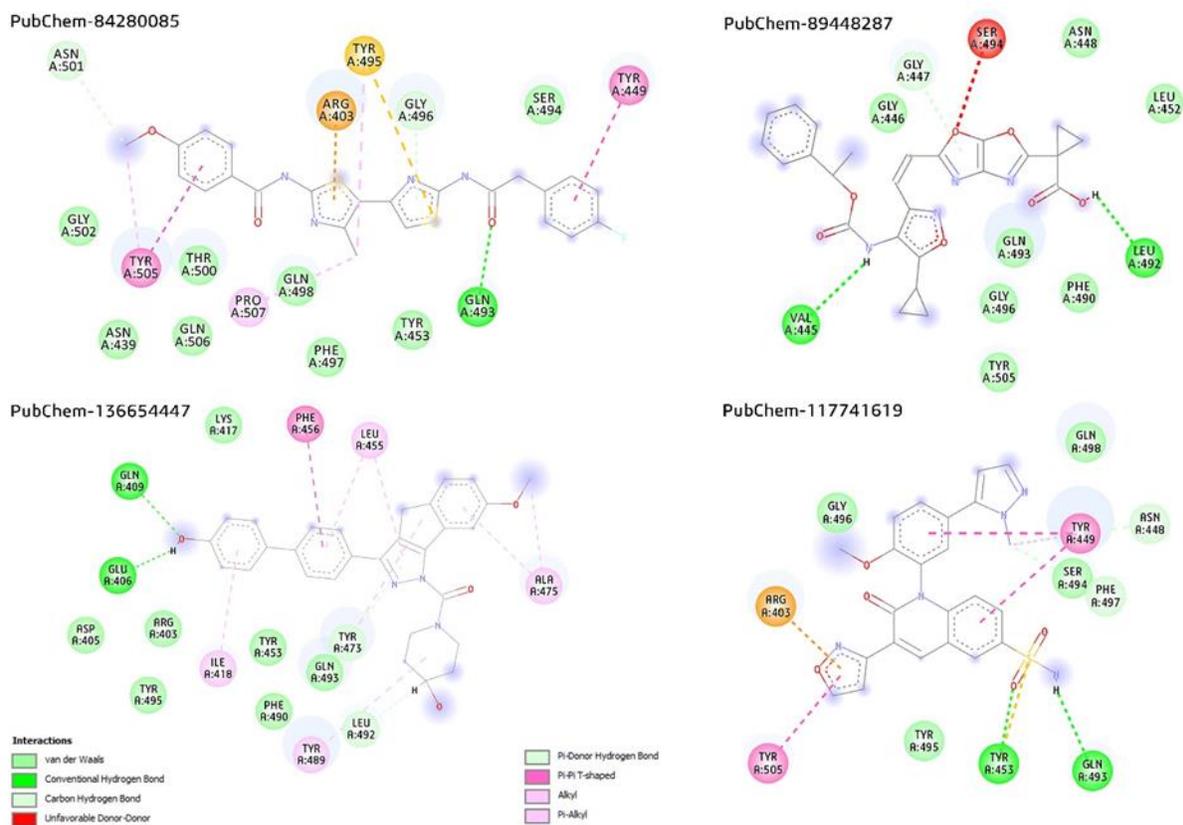


Figure 6

Figure 6. 2D interaction diagrams of four candidates. The snapshot structures were taken at time 100 ns of four ligand-protein complexes from MD simulation. Ligands showed strong interactions like conventional hydrogen and carbon-hydrogen bonds, as well as weak interactions like van der Waals, alkyl, and pi-alkyl bonds. Dashed lines represent the different interactions and their color the interaction type. Colored circles represent residues with their three-letter code, chain identifier, and residue number.

Table 1. Summary of 6 clusters obtained from PocketQuery. The terms consist of residue type and number, FastContact Energy (ΔG_{FC}), Rosetta Energy ($\Delta\Delta G_R$), Absolute $\Delta SASA$, Relative $\Delta SASA\%$, Conservation (Con), Evolutionary Rate (Evol Rate), Score.

Cluster	Residues	ΔG_{FC}	$\Delta\Delta G_R$	$\Delta SASA$	$\Delta SASA\%$	Con	Evol Rate	Score
1	Asp30; Lys31; His34	-4.33	-0.246	69.13	48.56	0.527	3.81	0.863
2	Thr27; Lys31; His34	-1.76	-0.162	78.95	58.33	0.564	3.09	0.821
3	Thr27; Asp30; His34	-2.71	0.261	59.2	50.83	0.544	3.33	0.804
4	Lys31; His34; Asp38	-2.57	-0.180	70.82	50.2	0.592	3.27	0.781
5	Gln24; Thr27; Asp30	-3.45	0.058	55.1	48.66	0.544	3.34	0.780
6	His34; Asp38; Lys353	-0.99	0.354	60.97	44.16	0.678	2.41	0.586

Table 1

Table 2. Molecular Docking Results and Interaction Analysis of top ten ligands with RBD.

Ligand	Score	MinimizeRMSD	Interactions
PubChem-46282789	-7.93	2.58	ARG403, TYR449, TYR453, GLN493, SER494, TYR495, GLY496, PHE 497, ASN501, TYR505
PubChem-136107209	-7.82	1.96	TYR449, LEU452, LEU492, SER494, THR500, GLY502, TYR505
CHEMBL2172478	-7.77	2.25	ARG403, TYR453, TYR449, GLY493, SER494, TYR495, PHE497, TYR505
IBYKSDUFFDTFQG-UHFFFAOYSA-N	-7.75	2.31	ARG403, TYR449, TYR453, GLN493, TYR495, GLY496, ASN501, TYR505
PubChem-6694569	-7.67	2.03	GLU406, TYR453, TYR495, GLY496, ASN501, TYR505
PubChem-89448287	-7.59	1.98	ARG403, GLN493, SER494, TYR495, GLY496, TYR505
PubChem-84280085	-7.46	1.32	TYR449, TYR453, SER494, TYR495, PHE497, GLN498, TYR505
ZINC000014748612	-7.43	1.17	TYR453, TYR449, SER 494
PubChem-117741619	-7.248	2.65	LEU455, GLN493, PHE497, GLY502, TYR505
PubChem-136654447	-7.05	2.25	GLU406, LUE455, PHE456, ALA475, TYR489

Table 2

Table 3. Ranked calculated binding free energies using MM/PBSA for the ten protein-ligand systems (the values presented are in kcal/mol). The terms consist of van der Waals(ΔE_{vdW}), electrostatic(ΔE_{ele}), polar solvation(ΔG_{PB}), solvent-accessible surface area (ΔG_{SA}), and binding energies($\Delta G_{binding}$).

Ligand	ΔE_{vdW}	ΔE_{ele}	ΔG_{PB}	ΔG_{SA}	$\Delta G_{binding}$
PubChem-84280085	-117.126	-74.728	62.319	-11.954	-141.49
PubChem-89448287	-139.108	-28.04	51.479	-14.898	-130.566
PubChem-136654447	-100.711	-21.209	41.855	-11.298	-91.363
PubChem-117741619	-76.717	-39.858	42.664	-8.184	-82.094
PubChem-6694569	-72.884	-22.713	34.642	-7.695	-68.651
IBYKSDUFFDTFQG- UHFFFAOYSA-N	-64.137	-25.054	35.853	-6.833	-60.171
ZINC000014748612	-47.483	-17.2	20.117	-5.203	-49.769
CHEMBL2172478	-58.879	-11.992	38.146	-6.09	-38.816
PubChem-46282789	-52.525	-7.178	40.416	-5.777	-25.065
PubChem-136107209	-17.351	-5.233	5.847	-1.718	-18.455

Table 3

Figures

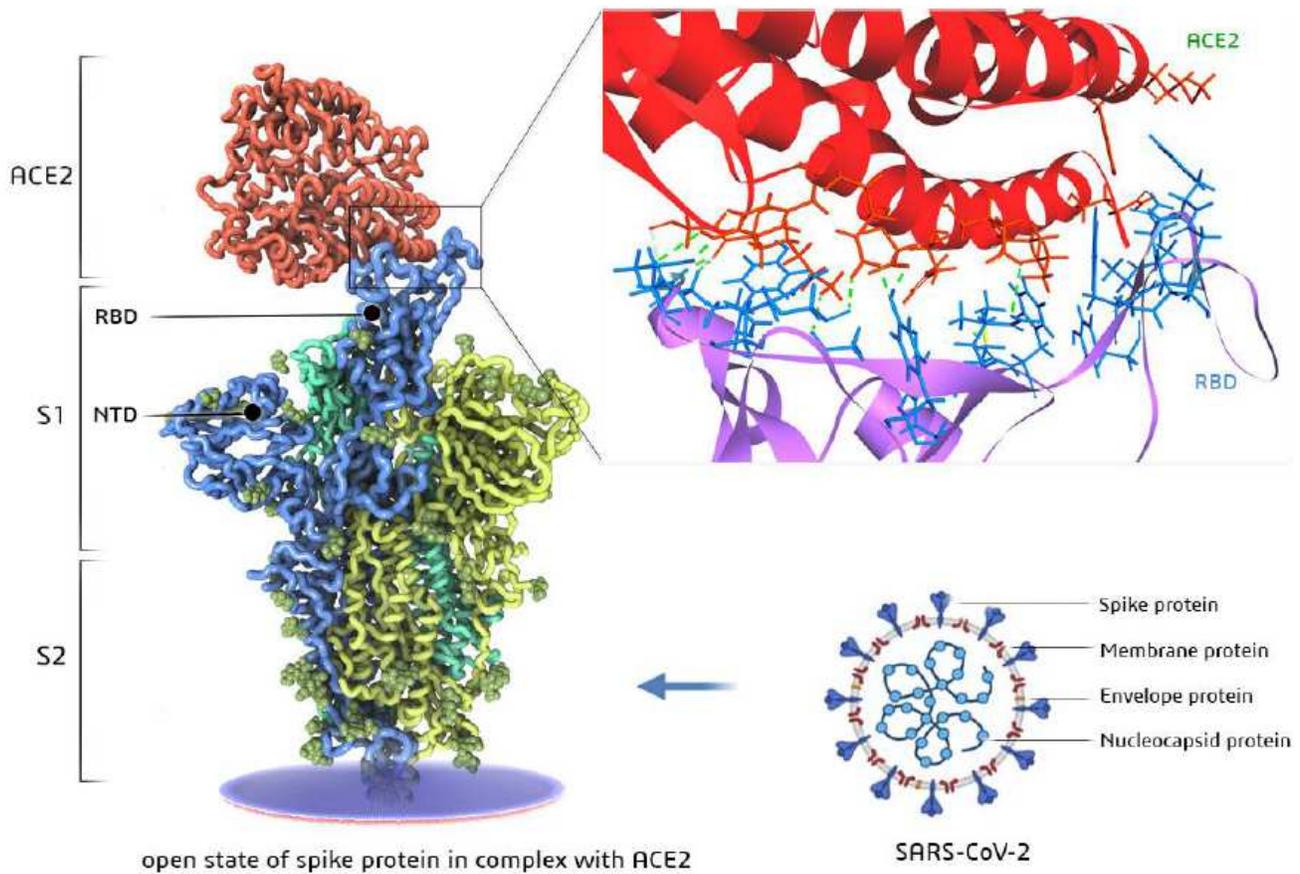


Figure 1

Schematic of SARS-CoV-2 structure and its major proteins: nucleocapsid (N) protein, membrane protein (M), spike protein (S), and envelope protein (E). The ribbon views represent the open state of spike protein in complex with ACE2 (illustrated by Protein Imager). Spike protein is composed of S1 subunit (which includes the N-terminal domain (NTD) and the receptor-binding domain (RBD)) and S2 subunits. With a magnification of the RBD-ACE2 binding interface. ACE2 is colored red, SARS-CoV-2 RBD core is colored blue.

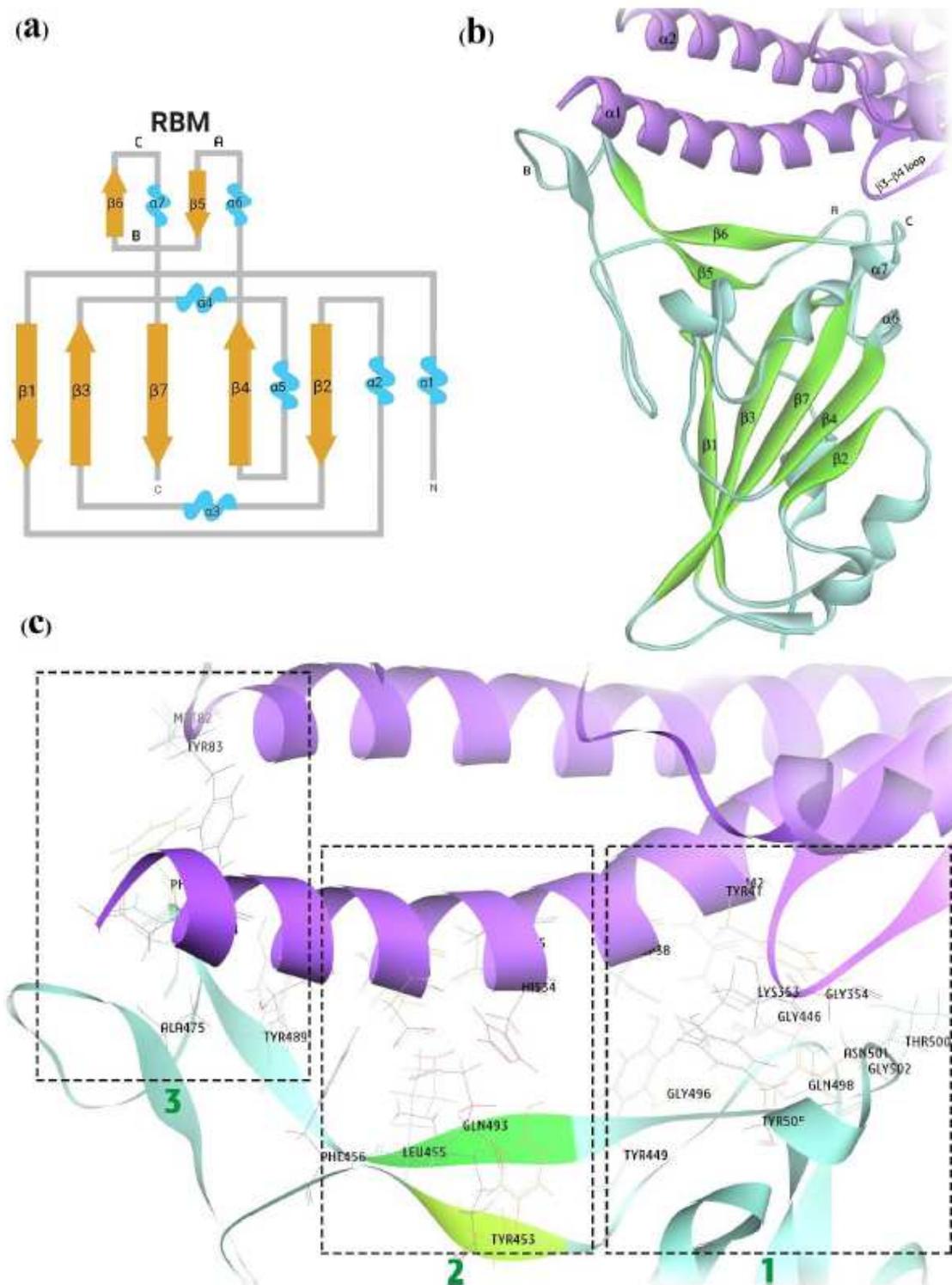


Figure 2

a. Schematic of SARS-CoV-2 RBD structure. Different structures are shown by different colors. The core structure of SARS-CoV-2 RBD consists of a twisted five-stranded antiparallel β -sheets ($\beta 1, \beta 2, \beta 3, \beta 4,$ and $\beta 7$) with short α helices connected by flexible loops. receptor-binding motif (RBM), an extended insertion containing short β -sheets ($\beta 5$ and $\beta 6$), α -helices ($\alpha 6$ and $\alpha 7$), and loops (A-B-C), located between the $\beta 4$ and $\beta 7$ strands in the core. b. Overall structure of the SARS-CoV-2 RBD (green and gray) bound to

ACE2(purple). The ACE2 binds to RBD through N-terminal $\alpha 1$ and $\alpha 2$ antiparallel helices as well as the $\beta 3$ – $\beta 4$ loop. c. Ribbon representation of the RBD-ACE2 binding interface. ACE2 is shown in blue. This contact surface can be divided into three regions: Region one includes residues in C-terminus of $\alpha 1$ and $\beta 3$ – $\beta 4$ loop from the AEC2 and loop A, loop C, and short $\alpha 7$ residues of the RBD. Region two: includes residues in the middle $\alpha 1$ helix from AEC2 and $\beta 4$ strands and $\beta 5$ strands residues from the RBD. Region three: includes residues in N-terminus of $\alpha 1$ and C-terminus $\alpha 2$ helix of ACE2 and loop B residues from the RBD.

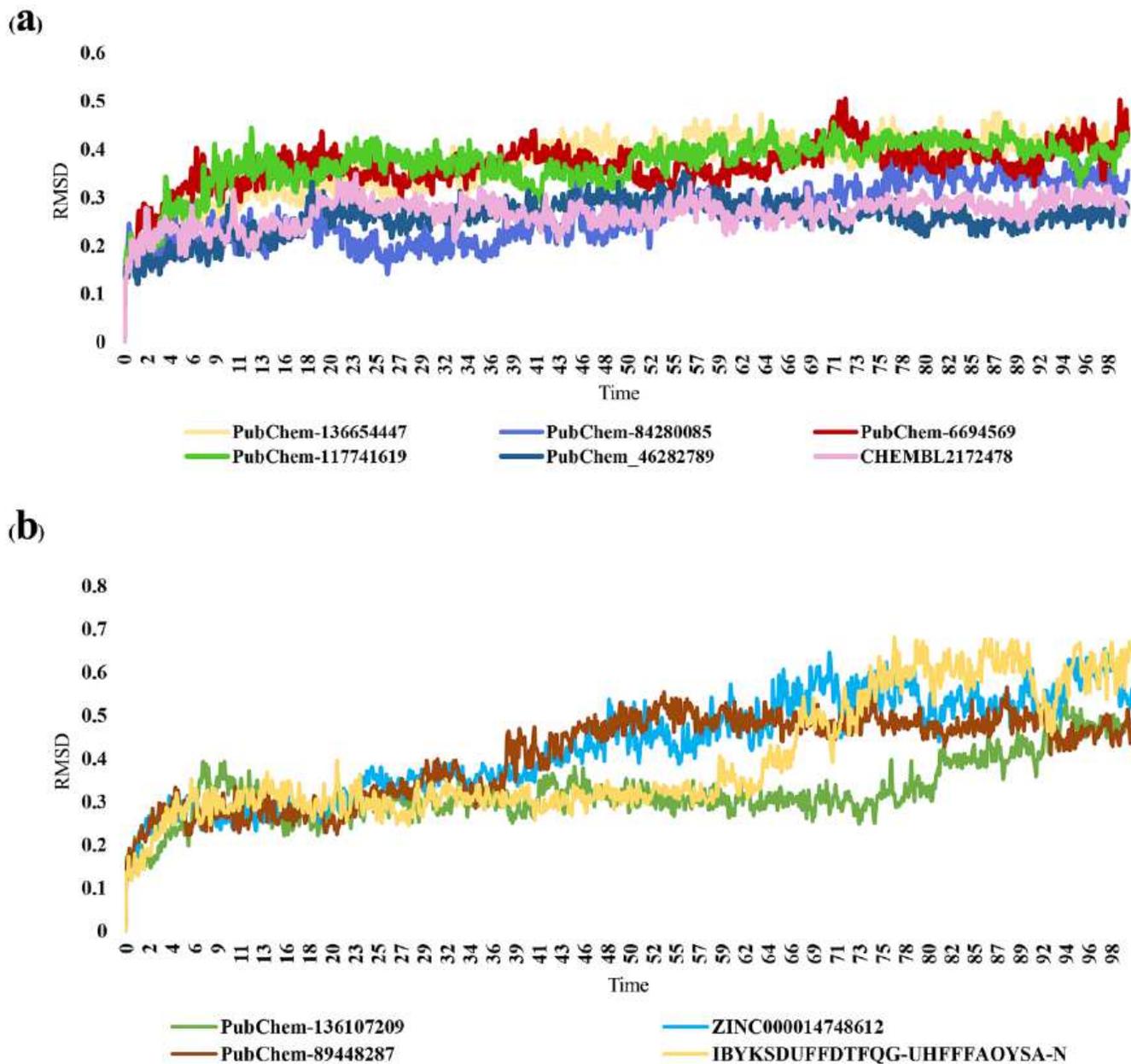


Figure 3

The root mean square deviation (RMSD) of SARS-CoV-2 RBD in complex with ten compounds during 100 ns molecular dynamics simulation. a. complexes with RMSD > 0.4 nm. b. complexes with RMSD < 0.4 nm.

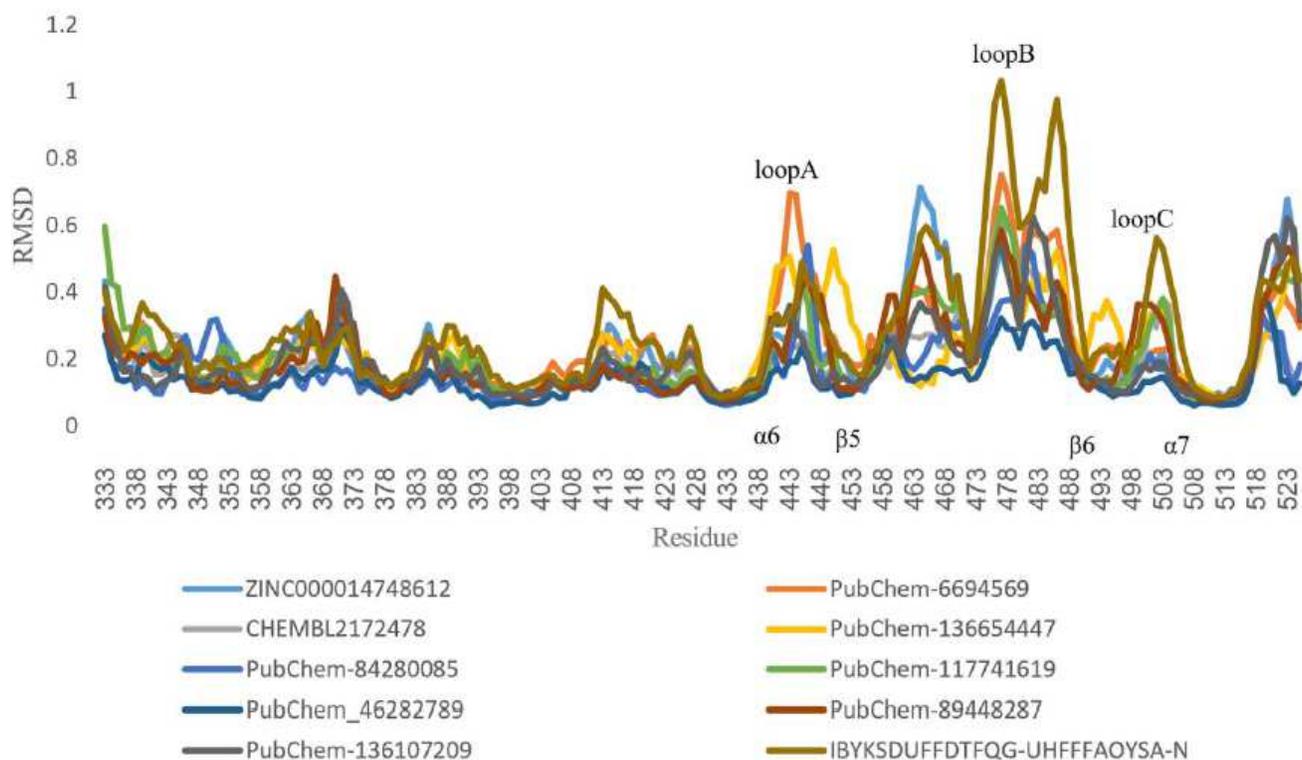


Figure 4

The root mean square fluctuation (RMSF) values of SARS-CoV-2 RBD in complex with ten compounds during 100 ns plotted against residue numbers. Residues with values above 0.5 nm are labeled.

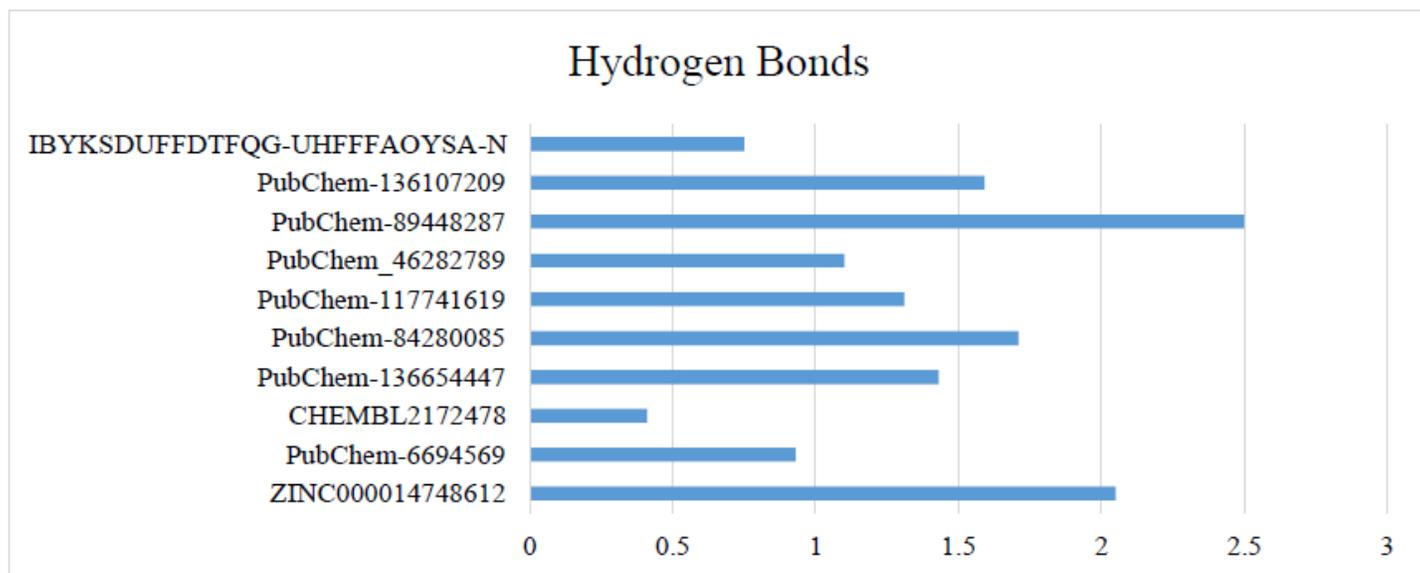


Figure 5

The average number of hydrogen bonds between top ten compounds and RBD during 100 ns of MD simulations.

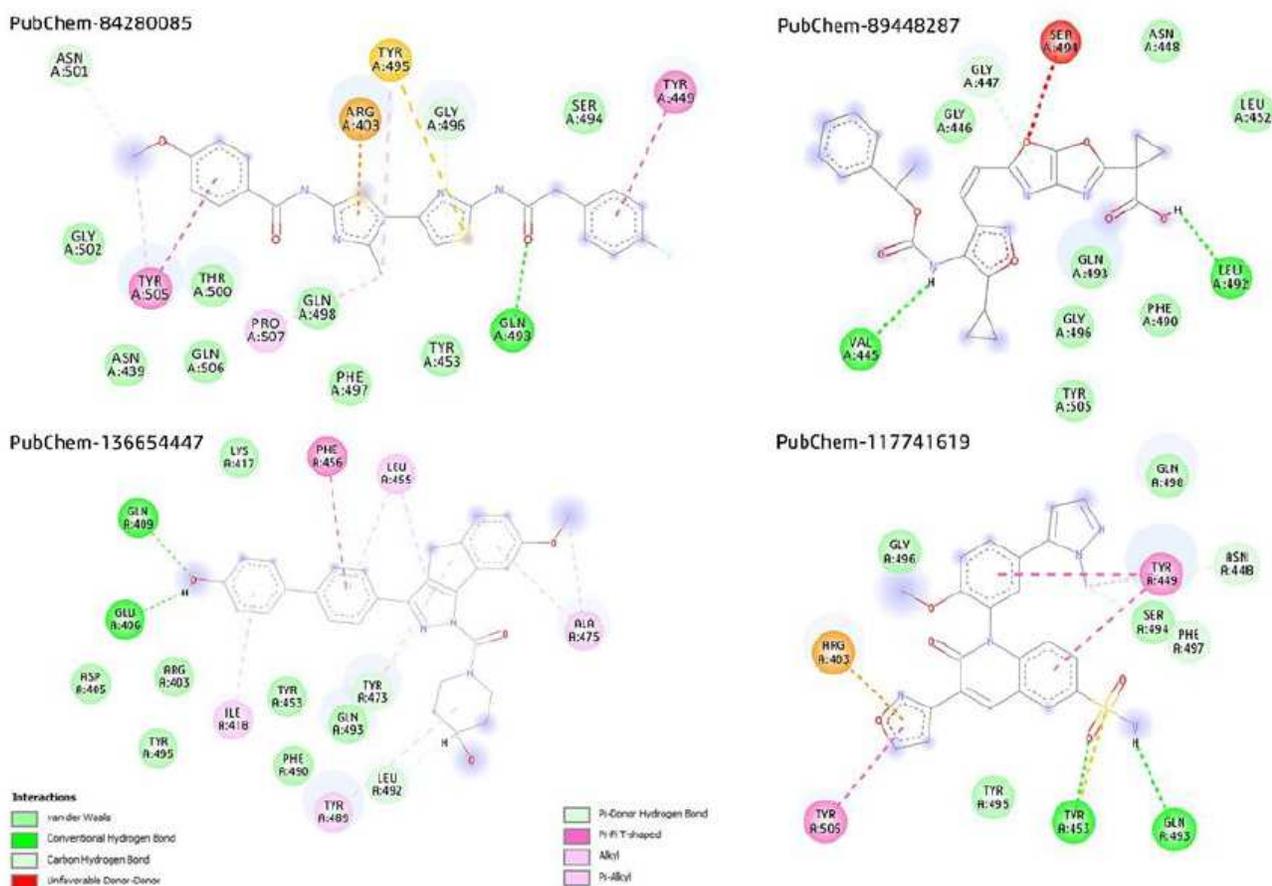


Figure 6

2D interaction diagrams of four candidates. The snapshot structures were taken at time 100 ns of four ligand-protein complexes from MD simulation. Ligands showed strong interactions like conventional hydrogen and carbon-hydrogen bonds, as well as weak interactions like van der Waals, alkyl, and pi-alkyl bonds. Dashed lines represent the different interactions and their color the interaction type. Colored circles represent residues with their three-letter code, chain identifier, and residue number.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementary.pdf](#)