ORIGINAL RESEARCH

Hybrid drug-screening strategy identifies potential SARS-CoV-2 cell-entry inhibitors targeting human transmembrane serine protease

Yufei Feng[1] · Xiaoning Cheng[2] · Shuilong Wu[2] · Konda Mani Saravanan[3] · Wenxin Liu[2]

Abstract

The spread of coronavirus infectious disease (COVID-19) is associated with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has risked public health more than any other infectious disease. Researchers around the globe use multiple approaches to identify an effective approved drug (drug repurposing) that treats viral infections. Most of the drug repurposing approaches target spike protein or main protease. Here we use transmembrane serine protease 2 (TMPRSS2) as a target that can prevent the virus entry into the cell by interacting with the surface receptors. By hypothesizing that the TMPRSS2 binders may help prevent the virus entry into the cell, we performed a systematic drug screening over the current approved drug database. Furthermore, we screened the Enamine REAL fragments dataset against the TMPRSS2 and presented nine potential drug-like compounds that give us clues about which kinds of groups the pocket prefers to bind, aiding future structure-based drug design for COVID-19. Also, we employ molecular dynamics simulations, binding free energy calculations, and well-tempered metadynamics to validate the obtained candidate drug and fragment list. Our results suggested three potential FDA-approved drugs against human TMPRSS2 as a target. These findings may pave the way for more drugs to be exposed to TMPRSS2, and testing the efficacy of these drugs with biochemical experiments will help improve COVID-19 treatment.

Keywords SARS-CoV-2 · Serine protease · Transmembrane proteins · Molecular docking and dynamics · Drug repurposing

Introduction

In December 2019, a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with a high mortality rate occurred in Hubei Province, People's Republic of China [1–3]. The World Health Organization (WHO) announced SARS-CoV-2 as a global pandemic on March 11, 2020, because of the high frequency of contamination and exponential spread of infections across six continents and over a hundred countries [4]. According to WHO, the disease has caused ~6.17 million deaths, ~495 million confirmed cases, and ~211 countries as of April 7, 2022 [5]. Comparatively, with two other acute human coronaviruses, namely Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV), SARS-CoV-2 shows a high transmission level and is very contagious [6–8]. Despite the severe lack of SARS-CoV-2 specific medicines, many promising therapeutic targets are being investigated by researchers around the globe, and many carefully planned clinical experiments are being performed in a systematic way [9].

During this pandemic, researchers urge to identify preclinical or clinical drugs that target three essential proteins: main protease, spike protein, and RNA-dependent RNA polymerase (RdRp), respectively [10]. The spike proteins in the course of entry, main protease in proteolytic activation, and RdRp in transcription are three distinct targets that play essential roles in the SARS-CoV-2 replication [11]. Our previous work used a hybrid drug screening strategy

✉ Wenxin Liu
  liuwenxin2009@163.com

1 Life Science and Technology School, Lingnan Normal University, Zhanjiang 524048, Guangdong Province, China

2 Central People's Hospital of Zhanjiang, Zhanjiang 524045, Guangdong Province, China

3 Department of Biotechnology, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu 600073, India

targeting RdRp and found that pralatrexate and azithromycin efficiently prevent SARS-CoV-2 replication in vitro [12]. Simultaneously, many studies in the literature revealed that SARS-CoV-2 recognizes angiotensin-converting enzyme 2 (ACE2) as the cell entrance receptor, interacting in tandem with the TMPRSS2, a transmembrane serine protease 2 encoding gene on the membrane cell surface [13–15]. A 492-amino acid protein that anchors to the plasma membrane is encoded by the TMPRSS2 gene [16], and unfortunately, the protein's crystal structure is still unresolved. Recent research revealed that invading the host cell requires spike protein priming, which is feasible due to the production of TMPRSS2 by the host cell [13]. Matsuyama et al. show that SARS-CoV-2 infection is enhanced by TMPRSS2 [17]. This invasion into the cell could be stopped by spike protein neutralizing antibodies and TMPRSS2 inhibitors like camostat mesylate [13]. Given the importance of TMPRSS2 as a potential target, we believe that serine protease inhibitors can aid in the development of novel strategies to prevent SARS-CoV-2 viral entry and pathogenesis [18–20].

New advancements in computational drug repurposing are enabled by the increased FDA (Food and Drug Administration)-authorized small molecule drugs in public repositories [21]. An example is the ZINC database which contains a vast number of different molecules, as well as links to merchants where such molecules can be purchased and obtained physically [22]. The other is the popular drug bank, a database of more than 10,000 drugs [22]. Databases of approved drugs are the most crucial aspect of drug discovery since they constitute the foundation for drug repurposing approaches and are reviewed in the literature [23]. Drug repurposing strategies have become a popular way to test for prospective authorized drugs that may also have effectiveness for other indications [21, 24, 25]. In this method, approved drugs for sets of diseases are considered safe for a human prescription. The remaining task is to establish their efficiency toward the disease under study [26, 27]. In addition, Enamine REAL fragments database (https://enamine.net) contains 38.2 million different collections of compounds that reflect the REAL drug-like space (substances that meet the "rule of 5" and Veber criteria: MW500, SlogP5, HBA10, HBD5, rotatable bonds10, and TPSA140) and are free of PAINS and hazardous compounds.

The hidden therapeutic potential of existing drugs is identified by using a variety of methodologies in drug repurposing methods [28]. Molecular modeling and data mining methodologies have recently been presented as effective drug repurposing strategies and are discussed in our published reports [12, 29–32]. We recently presented a hybrid drug virtual screening technique based on deep machine learning, molecular docking, and molecular dynamics simulation for discovering prospective RdRp targeting therapeutic candidates from 1906 market-available pharmaceuticals [12, 33–36]. This computational study analyzes potential approved FDA drugs and others that can block TMPRSS2 priming ACE2 after infecting the host cell. Three computational approaches are combined to enhance the accuracy and efficiency of the prediction [37–39]; (i) virtual screening over the current approved drugs database and fragments database to analyze their prospective for repurposing [40–42], (ii) machine learning [43], and (iii) molecular dynamics and well-tempered metadynamics [44].

Materials and methods

Target information

The target used in the present study is a typical transmembrane serine protease type 2 (TMPRSS2). This is because the SARS-CoV-2 epidemic is dependent on the host cell molecules like ACE2 and TMPRSS2, which can be prevented by a therapeutically validated protease inhibitor [13]. The amino acid sequence of TMPRSS2 was reported in the UniProt database (UniProt accession: O15393) and is made up of 492 amino acid residues. As per UniProt, the target protein comprises a catalytic (1–255) and non-catalytic domain (256–492). Around 35% sequence identity was found between the templates of known structures in Protein Data Bank (PDB) and TMPRSS2, which is evident from similarity searching of UniProt sequence against PDB using BLAST (Basic Local Alignment Search Tool) online program [45]. I-TASSER, iterative threading assembly refinement software, generated the atomic coordinates of the 3D structural model of TMPRSS2 [46]. Plasma kallikrein A was the template with the highest overall sequence identity (PDB: 1ZOM). It uses secondary-structure enhanced profile-profile threading alignment and iterative structure assembly simulations using a threading assembly refinement tool [47]. Although the crystal structure of transmembrane protease serine 2 (PDB Identifier: 7MEQ) is available now, no PDB structure was available when we started this project. Since our predicted structure is highly similar to the experimental ones, we used the model for further research. Moreover, the I-TASSER predicted ligand-binding site and cofactor methods are reliable for our screening procedure. The quality of the model was evaluated by examining the model's stereochemical quality on the Ramachandran map by using ProSA and Rampage [48, 49]. Furthermore, PROCHECK [50], ERPAT [51], and QMEAN [52] tests were done to evaluate the quality of the predicted models. The final optimized structural model is taken into account for further investigation.

Ligand library

The virtual screening library was the TargetMol-Approved Drug Library, consisting of 2356 compounds (https://www.targetmol.com/compound-library/Approved-Drugs-Library). These 2356 compounds present in the TargetMol repository are drugs approved by Food and Drug Administration (FDA), the European Medicine Agency (EMA), or China Food and Drug Administration (CFDA) or included in the US Pharmacopeia (USP) Dictionary, the British Pharmacopoeia (BP), the European Pharmacopoeia (EP), the Japanese Pharmacopoeia (JP), or Chinese Pharmacopoeia (CP) Dictionary. All approved drugs in the database have well-known bioactivities, safety, and bioavailability, and these drugs are structurally diverse, medicinally active, and cell permeable.

DFCNN-based protein-drug binding prediction

Dense fully connected neural network (DFCNN), a deep learning-based technique for predicting protein-drug interactions, was previously reported in our paper and was utilized in this research for primary drug screening [12]. Our recently published articles describe developing the deep learning model [12, 30]. Our DFCNN model takes its training data from the PDBbind database [53], where crystallized high-resolution protein–ligand complexes are positive, and cross-docked complexes are negative. DFCNN uses a concatenated molecular vector of protein pocket and ligand as an input representation. The molecular vector is created by Mol2vec [54], a natural language processing model inspired by the word2vec model. The DFCNN model has advantages over other methods: it does not rely on the docking simulation results (returns no docking pose). It incorporates nonbinding decoys in the training dataset. The docking simulation is exceptionally quick due to its independence, and the addition of nonbinding decoys during training conceives the model robust in real-world circumstances. Since the model does not rely on the protein-drug complex conformation, the model is around 100,000 times faster than Autodock Vina in estimating the protein–ligand binding probability (range 0–1). This model is used to predict protein-drug binding because the approach has been proven to have greater accuracy and efficiency and is well suited for use in an emerging disease outbreak. The DFCNN model's source code can be found at https://github.com/haiping1010/DeepBindRG.

Docking and deep learning-based prediction

For large-scale drug screening, DeepBindBC, an efficient and accurate deep machine learning-based model, is employed. Autodock Vina provided structural information for protein-drug complexes is used as input [55]. DeepBindBC can attain higher accuracy because it combines physical–chemical and spatial parameters between the protein–ligand interfaces. The PDBbind database was used to train the DeepBindBC ResNet model. The protein–ligand interface parameters in DeepBindBC will be transformed into a figure-like representation [53, 56]. Since DeepBindBC uses top docking poses generated by Autodock Vina and DFCNN needs molecular vector information, the two approaches are complementary, and DeepBindBC consumes significantly longer time than DFCNN. The DeepBindBC model can achieve higher accuracy, but it requires Autodock Vina's protein-drug complex structure information as input. We use a DFCNN score, DeepBindBC score, and Autodock Vina score, respectively, to rank the top binding protein–ligand complexes. The source code of the DeepBindBC model is available at https://github.com/haiping1010/DeepBindBC.

Enamine REAL fragments screening

We have screened the Enamine REAL fragments dataset against the target, obtained from an online link (https://enamine.net/compound-collections/real-compounds/real-compound-libraries), a size of 15,635,761 compound fragments. We first do the DFCNN-based screening, select compound fragments with a DFCNN score larger than 0.8, and then carry docking for those selected compounds over the protein target using the procedure described above. After docking, we chose a ligand with a docking score ≤ -7.4 kcal/mol.

MD simulations and well-tempered metadynamics

Force field-based molecular dynamics (MD) simulations conducted additional drug screening for the protein–ligand complexes with the top score. The initial protein-drug complexes were generated using Autodock Vina docking, and the ligand was modified using pymol software [57] to place it in the correct protonation state. The MD simulations were performed using the AMBER-99SB force field in Gromacs [36]. ACPYPE was used to generate the ligand topology and the partial charges of the ligand [58]. First, we made a dodecahedron box and placed the structural complex in the center, then filled with TIP3P water molecules [59]. A minimum distance of 1 nm was fixed between the protein and the box edge. The Gromacs program tool added counter ions to neutralize the total charge. For non-bonded van der Waals interactions, a threshold of 14 Å was used. The LINCS algorithm restricted covalent bonds using hydrogen atoms [60]. We used a 0.001-ns step size for energy minimization, a 100-ps simulation with the isothermal-isovolumetric ensemble (NVT), and a 10-ns simulation with the isothermalisobaric ensemble (NPT) for water equilibrium. A 100-ns NPT production run was completed (step size 2 fs). With

a fixed temperature of 300 K and a pressure of 1 atm, the Parrinello-Rahman barostat and the modified Berendsen thermostat were applied for simulation. Gromacs tools were used to calculate the trajectory's root mean square deviation (RMSD) and hydrogen bond number. Finally, the compounds were used for MD simulation. Further, the obtained candidate fragments were subjected to molecular dynamics simulations. To reduce the simulation time without affecting the accuracy, we consider a domain (the amino acid 250 to 492 part) for MD simulation, which should cover most of the amino acids involved in protein–ligand interaction.

Metadynamics simulations have been employed to compute binding free energy to estimate if a protein–ligand will bind. The number of atom contacts correlates with the number of protein–ligand interface coordination numbers, and a higher number implies that the protein–ligand is in a binding state. The coordination number C is as follows:

$$C = \sum_{i \in A} \sum_{j \in B} S_{ij}$$

and

$$S_{ij} = \frac{1 - \frac{(r_{ij} - d_0)^n}{r_0}}{1 - \frac{(r_{ij} - d_0)^m}{r_0}}$$

During simulation, n was 6, m was 12, d_0 was 0 nm, and r_0 was 0.5 nm. d_0 is a parameter of the switching function. The distance between atoms i and j is given by r_{ij}. The above function can calculate the degrees of contact between two groups of atoms. The community-developed PLUgin for MolEcular Dynamics (PLUMED) was used to execute a 100-ns metadynamics simulation for each protein–ligand system [61]. During the metadynamics simulation, Gaussian values with a height of 0.3 kJ/mol and SIGMA (Gaussian width) of 5 were deposited every 1 ps. The Plumed program was used to construct the free energy landscapes of the metadynamics simulations, which were then visualized using Gnuplot. The detailed procedures for conducting simulations were similar to our recently published work [12].

Computation of binding free energy by MM-PBSA

The molecular mechanics/Poisson-Boltzmann surface area (MMPBSA) technique was used to calculate the free binding energy of TMPRSS2-drug and TMPRSS2-fragment complex structures from molecular dynamics trajectories [62]. For the top three TMPRSS2-drug complexes and nine TMPRSS2-fragment complexes, the MM-PBSA free energy was computed. The MM/PBSA method in the g mmpbsa tool was used to compute the binding free energy using the last 20 ns trajectory (20 frames from each nanosecond) of the 40 ns typical NPT MD simulation [63].

Table 1 List of selected approved drug candidates after screening. The protein–ligand has Autodock Vina score above −7.0 kcal/mol, DFCNN score above 0.9, and DeepBindBC above 0.8

| Approved drug name	Autodock Vina (kcal/mol)	DFCNN	DeepBindBC
ABT-199	−9.9	0.970649	0.994248
Anidulafungin	−8.0	0.990461	0.967462
Carfilzomib	−8.5	0.988915	0.908972
Micafungin_sodium	−7.9	0.982125	0.895186
Vancomycin_hydrochloride	−7.8	0.967362	0.894179

Results

Structural model and drug screening by different methods

The theoretical structural model built by using ITASSER has been used in this study. COFACTOR algorithm within I-TASSER was used to extract the ligand from the template structure (PDB ID: 1ZOM) using structural comparison and protein–protein networks [46, 64]. The model's quality is validated using standard procedures, where around 98% of amino acid residues fall inside the allowed region of the Ramachandran map. The model is also ranked according to its lowest DOPE (Discrete Optimized Protein Energy) score of −42,586.89. The binding pocket of the target is defined by extracting amino acid residues within 1 nm of the ligand. The protein molecule and drugs are prepared in an appropriate format to perform Autodock Vina docking, DFCNN-based screening, and DeepBindBC-based screening. The results suggest that by observing protein–ligand interactions from diverse perspectives, these three techniques (Autodock Vina, DFCNN, and DeepBindBC) complement one another. We eliminate drugs that have poor prediction by any of the three

Table 2 The hits obtained by Enamine_REAL_fragments screening with their docking score and DFCNN score. We select docking Score ≤ −7.4 kcal/mol, DFCNN score > 0.85

Name	Docking score (kcal/mol)	DFCNN score
Z3521035556	−7.4	0.949045
Z2712431854	−7.4	0.931763
Z3521678515	−7.4	0.921708
Z2873112960	−7.6	0.912652
Z2364211566	−8	0.898588
Z3045856817	−7.4	0.873949
Z3510403924	−7.7	0.8516255
Z3047170929	−7.7	0.8513706
Z2482433032	−7.6	0.8513706

Fig. 1 Three-dimensional view of the top three protein-drug complexes resulted from Autodock Vina, DFCNN, and DeepBindBC. The top three drugs such as ABT-199 (**A**), anidulafungin (**B**), and carfilzomib (**C**), respectively, binding to the non-catalytic domain of human TMPRSS2 is presented. The protein is shown as cartoon whereas drug is shown as sticks
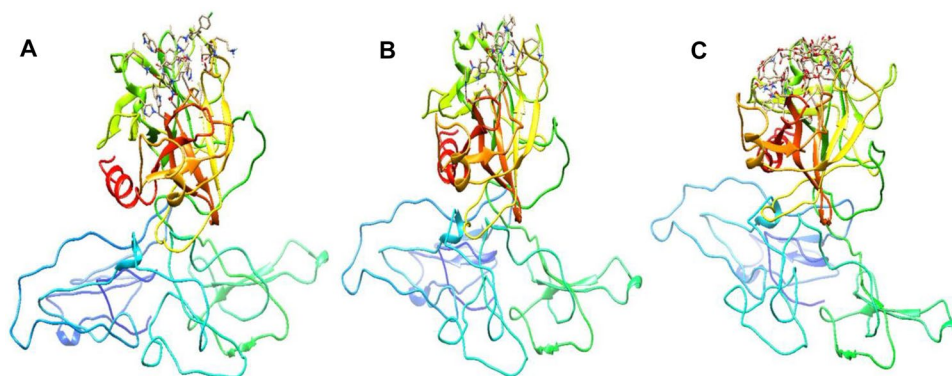
approaches. Table 1 shows the 5 drugs with an Autodock Vina score of more than −7.0 kcal/mol, DFCNN score of more than 0.9, and DeepBindBC score of more than 0.8. We considered the top three protein-drug complexes that have Autodock Vina score greater than or equal to 8 for further studies.

Our drug screening against TMPRSS2 as a target resulted in some top FDA-approved drugs. ABT-199

(Autodock Vina score: 9.9, DFCNN: 0.97, DeepBindBC: 0.99) is a top-in-class small molecular inhibitor approved by FDA that inhibits the B-cell lymphoma-2 (BCL-2) gene and is used to treat chronic lymphoid leukemia [65]. Furthermore, the drug is found to be effective against other leukemia types [66]. Interestingly, we found an antifungal drug "anidulafungin" to treat candida infections in the top
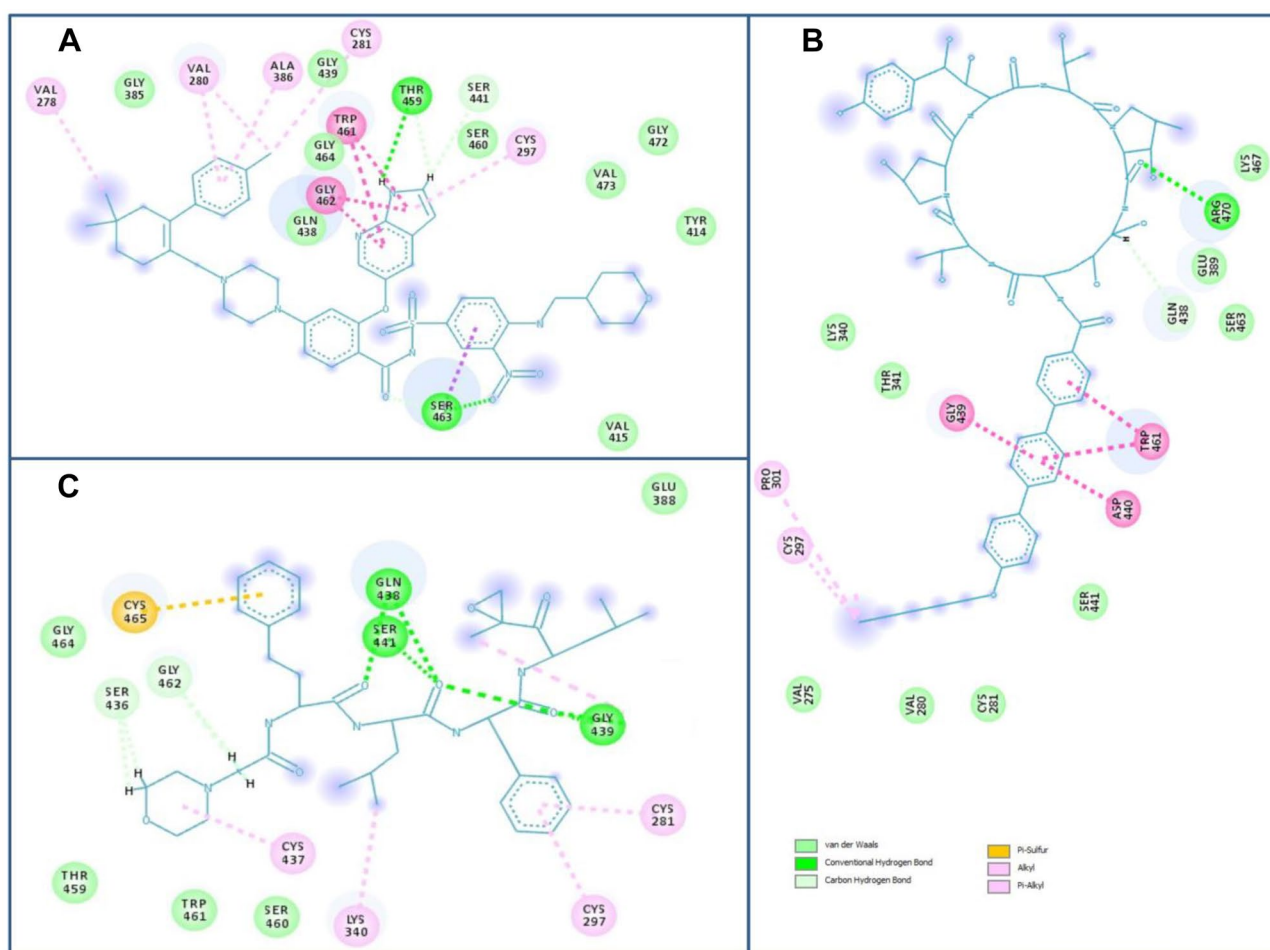

Fig. 2 Two-dimensional protein-drug interaction diagrams for top three complexes (TMPRSS2-ABT-199 (**A**), TMPRSS2-anidulafungin (**B**), and TMPRSS2-carfilzomib (**C**). The key interactions between protein and drug are shown as dotted lines
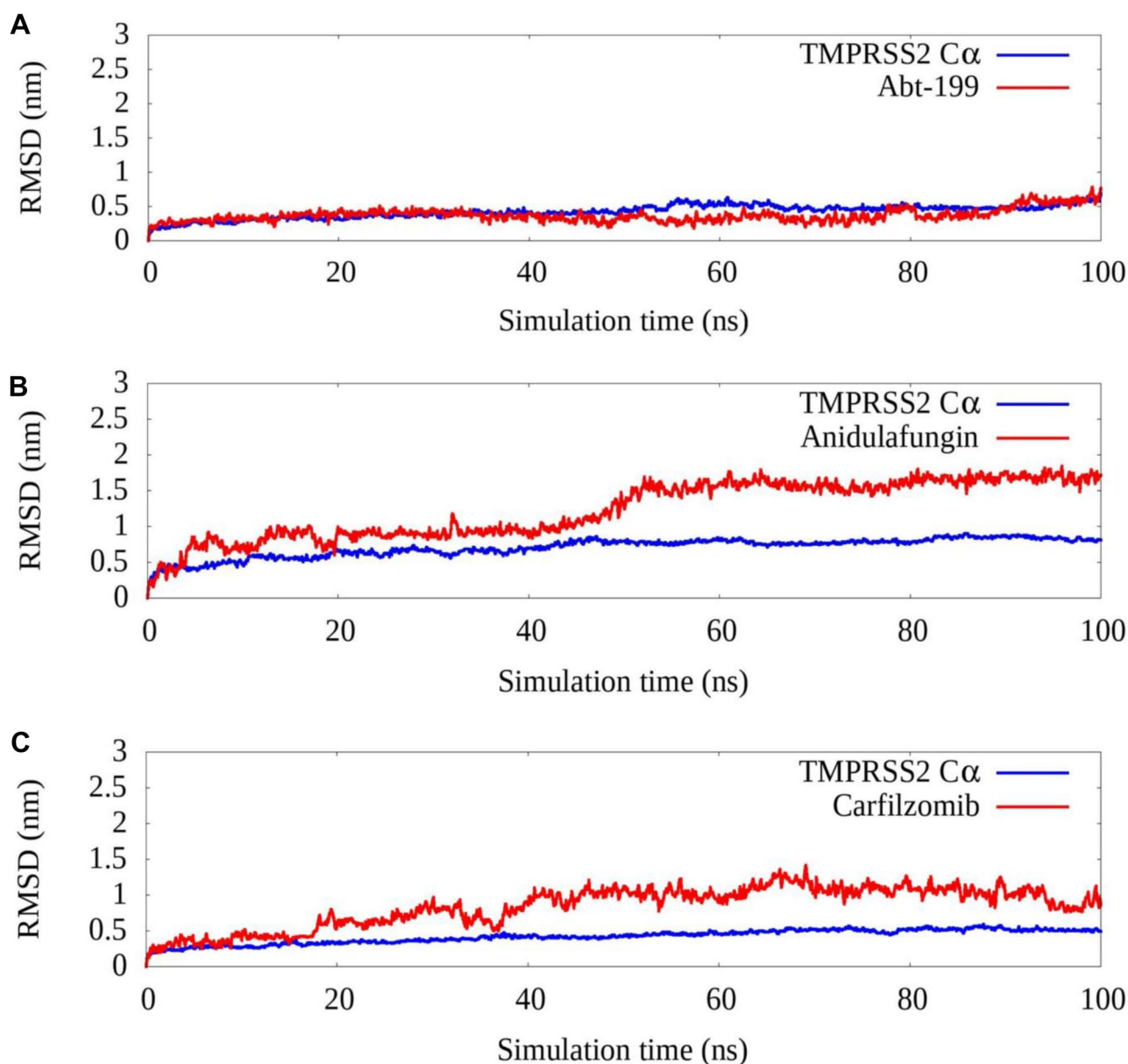
Fig. 3 The root mean square deviation of the top three protein-drug complexes (TMPRSS2-ABT-199 (**A**), TMPRSS2-anidulafungin (**B**), and TMPRSS2-carfilzomib (**C**) for 100-ns atomic molecular dynamics simulations

three of our drug screening procedures. Pfizer's anidulafungin is an antifungal medicine that the FDA permitted on February 21, 2006. The drug carfilzomib is known to treat small lung cancer. Interestingly, it is reported that carfilzomib is a promising proteasome inhibitor for the treatment of relapsed and refractory multiple myeloma [67]. The three-dimensional representation of the ligand binding to the pocket of TMPRSS2 is presented in Fig. 1. The figure shows that the ligand is binding to the target protein's non-catalytic domain (256–492). The Enamine REAL fragments screening results were tabulated and found to have different compounds as top hits by each

method (Table 2 drugs with a DFCNN score greater than 0.9). The top docking poses of the nine fragments binding with TMPRSS2 are presented in Supplementary Fig. 1.

The two-dimensional chemical interactions of protein and ligands are shown in Fig. 2. The drug ABT-199 was found to form more interactions with the amino acid residues of TMPRSS2. The drug carfilzomib forms 15 contacts with amino acid residues of the protein. The cyclic compound with antifungal effect forms 16 contacts with the amino acid of the target protein. The figure shows that van der Waals interactions and conventional hydrogen bonds dominate the protein–ligand complex formation.
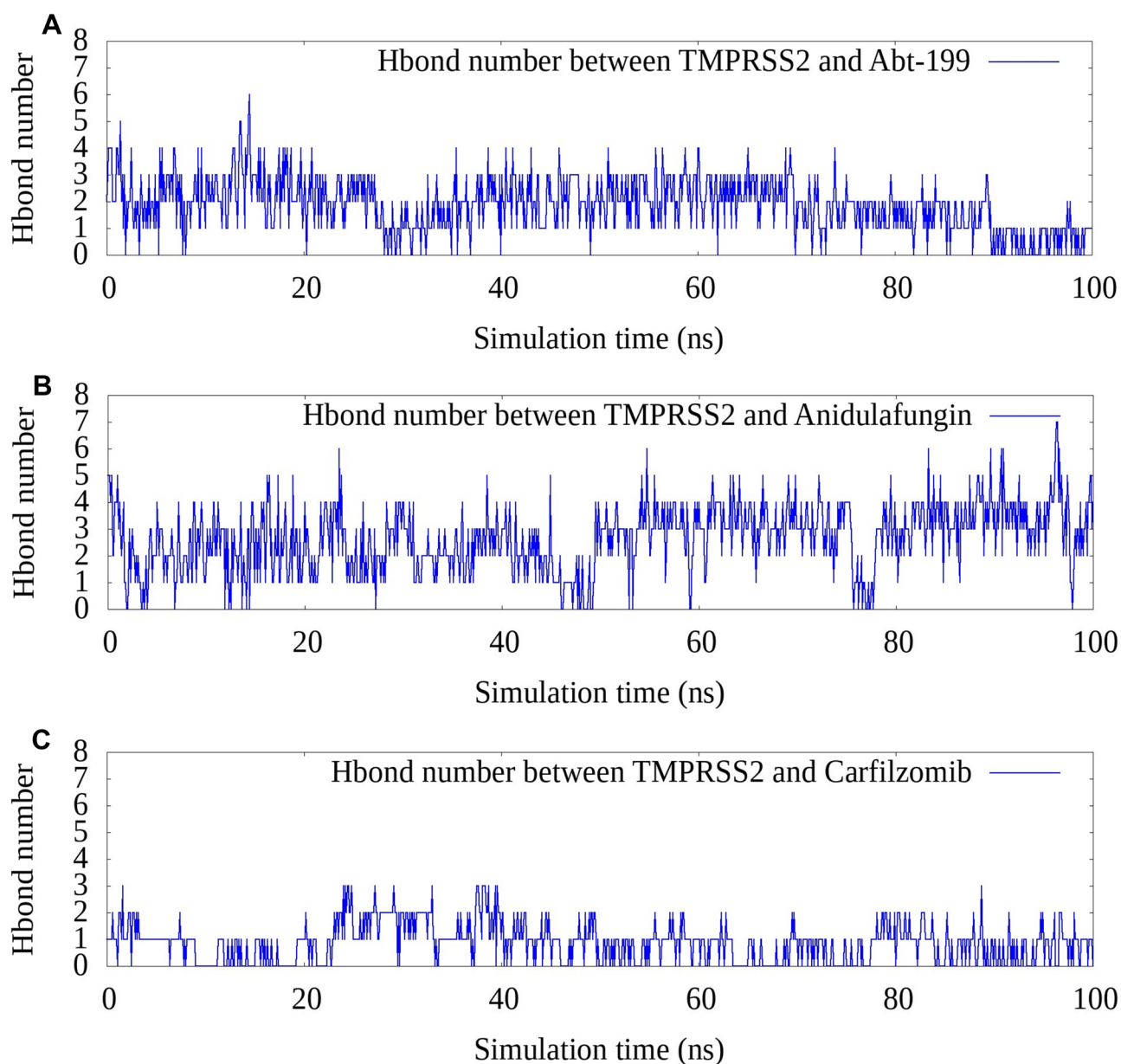
Fig. 4 The number of hydrogen bonds formed between TMPRSS2 and top three drugs (TMPRSS2-ABT-199 (**A**), TMPRSS2-anidulafungin (**B**), and TMPRSS2-carfilzomib (**C**)) during the simulation time

The figure indicates the van der Waals and conventional hydrogen bonds as light and dark green colors. It is observed that the glycine and tryptophan residues at different positions of TMPRSS2 form van der Waals interactions with drugs upon complex formation. Since van der Waals interactions are highly distance dependent and occur when adjacent atoms form close contact, the dominant interactions between protein and drug indicate a strong binding [68]. The ABT-199 and anidulafungin form amide-pi stacked interactions with TMPRSS2, indicating a dark pink color. The amino acid tryptophan at 461st position is the critical residue involved in amide-pi stacked interactions. The cysteine residue (297th and 437th positions) forms alkyl interactions with ABT-199, carfilzomib, and anidulafungin.

Molecular dynamics simulations

We used atomic molecular dynamics simulations on TMPRSS2-drug complexes to further validate the top three selected drug complexes and investigate their interactions and stability. The protein-drug complex stability is validated
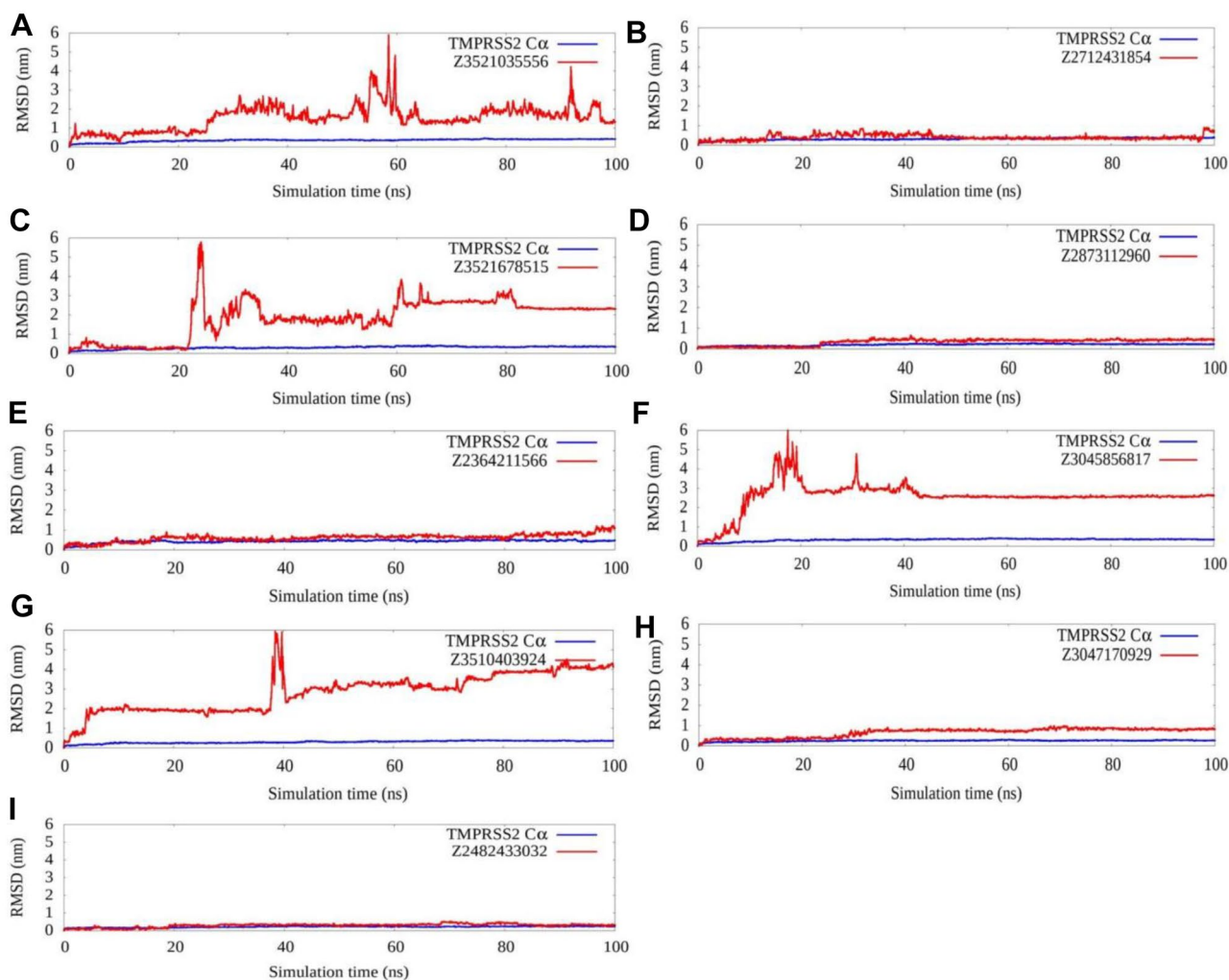
Fig. 5 The root mean square deviation of the top nine protein-fragment complexes for 100-ns molecular dynamics simulations. The panels **A**–**H** represents the RMSD values of protein with nine fragments presented in Table 2

by computing the root mean square deviation (RMSD) over the simulation time. We performed 100-ns atomic MD simulations with a fixed temperature of 300 K and a pressure of 1 atm. The drugs such as ABT-199 and carfilzomib are found to be stable as observed by the minimum RMSD fluctuations (less than 1 nm) of ligand binding to the protein, presented in Fig. 3. The snapshot of protein-drug binding poses after molecular dynamics simulations is presented in Supplementary Fig. 2. The drug anidulafungin was relatively stable over most simulation time but had more fluctuation after 80 ns time (greater than 1.5 nm). Overall, the drug ABT-199 was stable over most of the simulation time with minimum RMSD values during the 100-ns simulation time. The larger values of RMSD indicate the obtaining of reliable binding conformation during the simulation. If the fluctuation in RMSD values is low, it still shows stable binding.

Hydrogen bonding, hydrophobic, and electrostatic interactions are typically used to stabilize protein–ligand interactions. The hydrogen bond numbers were calculated using Gromacs tools by inputting trajectory files. The distance (≥ 3.5 Å) and the hydrogen donor-hydrogen acceptor angle (30°) were used to define hydrogen bond formation. Figure 4 depicts the formation of hydrogen bonds between protein and drug over a 100-ns simulation time. Among the three top TMPRSS2-drug complexes, ABT-199 and anidulafungin form more hydrogen bond numbers (> 5) with the protein than others. Further, carfilzomib forms fewer hydrogen bonds to stabilize the complex (average of > 3). As a result, RMSD estimates and hydrogen bond numbers between protein and drug suggest ABT-199 and anidulafungin are forming stabilizing interactions than the other complex. The top nine Enamine REAL fragments' RMSD and hydrogen bond numbers are shown in Figs. 5 and 6, respectively. The potential real drug-like compounds (Z2712431854, Z2873112960, Z2364211566, Z3047170929, and Z2482433032) are found to be relatively stable (RMSF
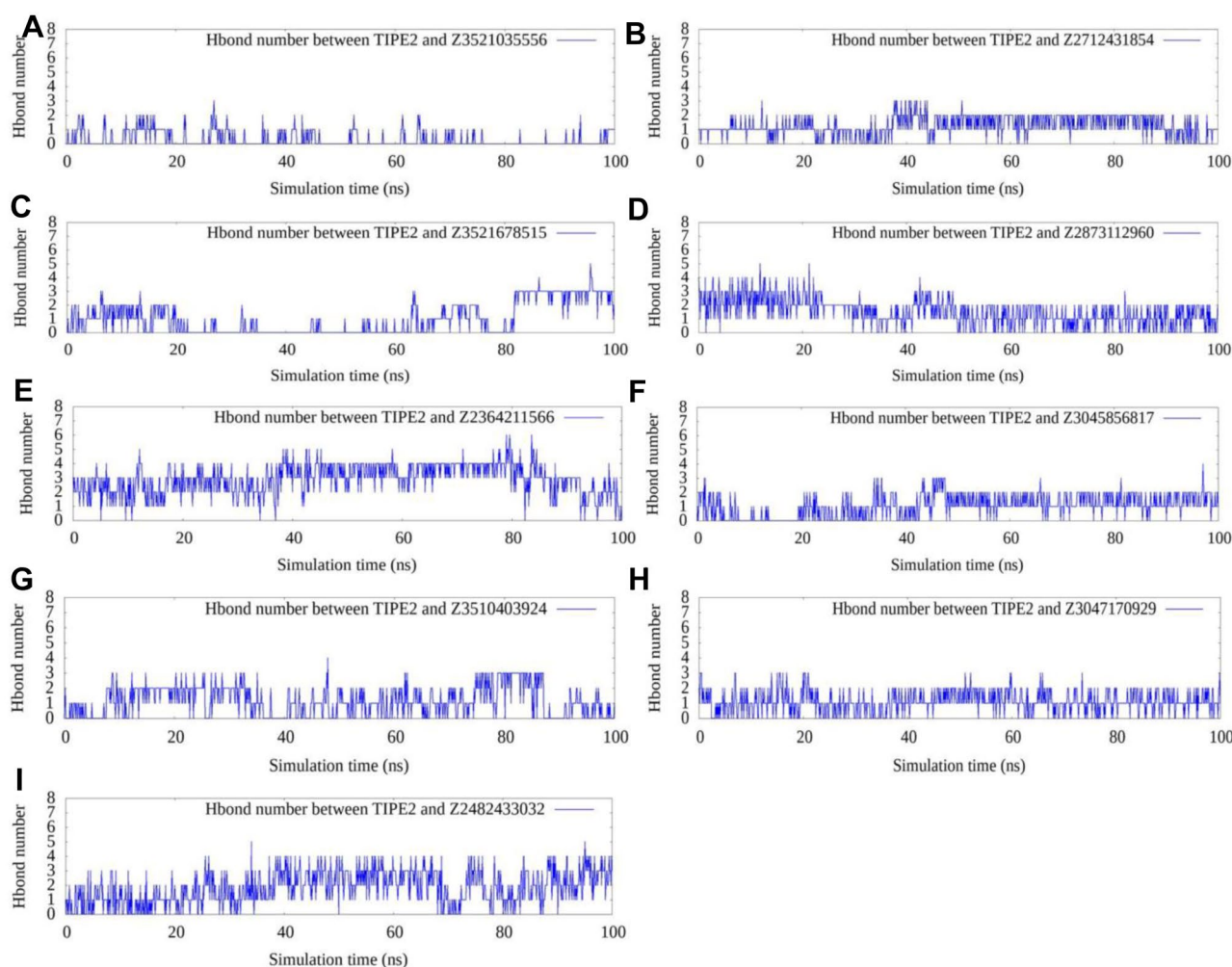
Fig. 6 The number of hydrogen bonds formed between TMPRSS2 and top nine fragments during the simulation time. The panel **A**–**H** denotes the number of hydrogen bonds fromed between protein and nine fragments presented in Table 2

less than 1 nm) by observing RMSF values obtained during MD simulations. All other compounds have high RMSF values, as evident from Fig. 5A, C, F, and G. The hydrogen bond numbers between target and compounds support the above finding presented in Fig. 6. After molecular dynamics simulations, the snapshot of TMPRSS2-fragment's binding poses is presented in Supplementary Fig. 3. However, the Enamine REAL fragments may not be very strong binders, which is understandable because smaller compounds usually have weak binding. But the preferred fragments can give us a clue about which kinds of groups the pocket prefers to bind, hence helping future structure-based drug design. Since Enamine REAL fragments are relatively small, it is possible to modify or combine different fragments to achieve stronger binding.

Free energy landscape of complexes

The most efficient technique to support virtual screening results is to apply metadynamics approaches to compute the binding free energy landscape of the protein and drug complex [44]. The molecular dynamics trajectories obtained from molecular dynamics simulations were used to compute the binding free energies. In the metadynamics approach, the number of atoms at the interface of the protein-drug complex was used as a collective variable (CV). The number of atom contacts correlates with the number of protein-drug interface coordination numbers, and a higher number implies that the protein–ligand interface is in a binding state. Figure 7 shows the binding free energy vs. coordination number (CV: collective variable) of the top three TMPRSS2-drug complexes from metadynamics simulations. Suppose the lowest energy basin falls close to the coordination number of 0. In that case, it indicates the compounds are unbinding, while if the free energy basin falls at a relatively large coordination number value, it prefers to bind. Figure 7 shows that the lowest energy basins for all three complexes (ABT-199, anidulafungin, and carfilzomib) are at a relatively large
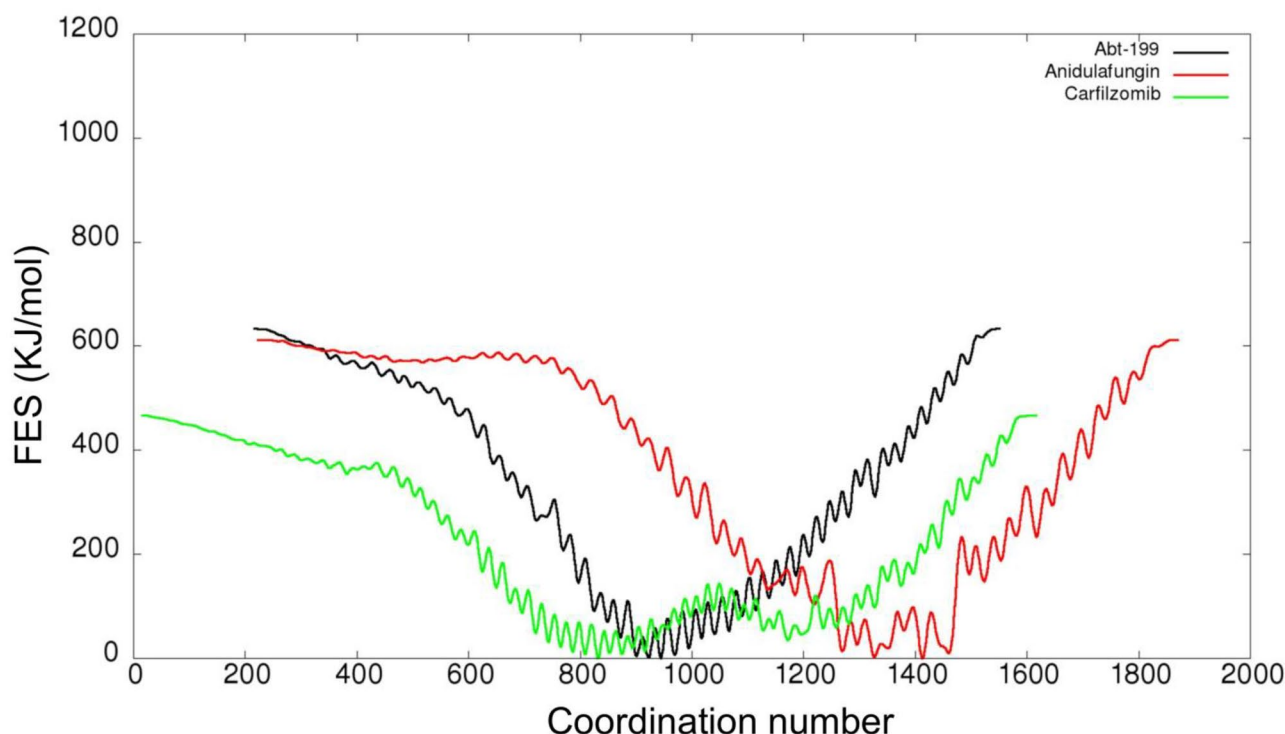
Fig. 7 Free energy landscape of top three TMPRSS2-drug complexes (TMPRSS2-ABT-199, TMPRSS2-anidulafungin, and TMPRSS2-carfilzomib), computed using well-tempered metadynamics approaches

coordination number value, especially anidulafungin. This indicates the compounds are strongly binding with the protein. The lowest energy poses of protein-drug complexes for ABT-199, anidulafungin, and carfilzomib had more interactions in the interface region, as specified by high coordination numbers.

Binding free energy of protein–ligand complexes

The most often utilized approach to support molecular docking results is computing the binding free energy of protein–ligand complexes using molecular mechanics-Poisson-Boltzmann surface area continuum salvation (MM-PBSA). The MM PBSA method was used to calculate the physical properties of protein and ligand complexes. The MD trajectories were used to predict the binding energies. Table 3 shows that the free binding energy of protein with ABT-199 (−145.096 kJ/mol) and carfilzomib (−115.071 kJ/

mol) is higher than the other protein-anidulafungin complex (−10.899 kJ/mol). Compared to TMPRSS2-drug complexes, the TMPRSS2-Z2482433032 fragment complex has a comparatively low free energy of binding (−104.142 kJ/mol). Tables 3 and 4 also include stabilizing physical free energies such as van der Waals, electrostatic, and solvent accessible surface area (SASA). The van der Waals, electrostatic, and SASA free energies of the TMPRSS2-drug complexes are all higher. According to our findings, TMPRSS2 complexed with drugs has improved projected free energies, as indicated by MMPBSA calculations.

Discussion

The SARS-CoV-2 is enclosed and includes single-stranded RNA beta coronaviruses, making them highly pathogenic. The 16 non-structural proteins (NSP1-16) and four primary

Table 3 The calculated energy between protein and drugs by MMPBSA method

Name	Van der Waals energy (kJ/mol)	Electrostatic energy (kJ/mol)	Polar solvation energy (kJ/mol)	SASA energy (kJ/mol)	Binding energy (kJ/mol)
Abt-199	−259.737+/−17.456	−40.684+/−16.568	183.762+/−18.652	−28.438+/−1.626	−145.096+/−18.484
Anidulafungin	−203.312+/−25.079	−121.213+/−50.707	361.222+/−77.932	−25.798+/−2.430	−10.899+/−38.857
Carfilzomib	−210.235+/−19.260	−58.914+/−13.028	175.178+/−20.469	−21.101+/−1.937	−115.071+/−16.898

Table 4 The calculated energy between protein and fragments by MMPBSA method

Name	Van der Waals energy (kJ/mol)	Electrostatic energy (kJ/mol)	Polar solvation energy (kJ/mol)	SASA energy (kJ/mol)	Binding energy (kJ/mol)
Z2364211566	−132.494+/−17.864	−65.026+/−18.593	138.575+/−24.120	−13.790+/−1.366	−72.734+/−18.206
Z2482433032	−193.428+/−12.811	−37.105+/−12.357	142.855+/−14.637	−16.464+/−0.769	−104.142+/−19.390
Z2712431854	−113.922+/−9.756	−27.201+/−11.703	77.696+/−14.627	−10.933+/−1.044	−74.359+/−10.206
Z2873112960	−149.846+/−13.618	−23.398+/−8.336	103.319+/−12.791	−14.983+/−0.971	−84.907+/−13.977
Z3045856817	−104.026+/−25.503	−33.621+/−14.637	82.724+/−20.257	−11.836+/−2.565	−66.760+/−25.330
Z3047170929	−149.440+/−16.873	−21.139+/−10.159	119.500+/−23.026	−16.884+/−1.499	−67.962+/−14.633
Z3510403924	−77.930+/−22.707	−22.551+/−16.299	66.987+/−33.124	−9.473+/−2.757	−42.967+/−28.213
Z3521035556	−53.292+/−27.061	−9.747+/−11.123	32.447+/−34.716	−6.842+/−3.239	−37.434+/−25.963
Z3521678515	−79.239+/−38.009	−18.880+/−19.611	59.932+/−38.213	−9.594+/−3.624	−47.781+/−26.771

structural proteins (spike, membrane, envelope, and nucleocapsid) are coded by the SARS-CoV-2 genome [69]. The structural proteins are required for SARS-CoV-2 replication, whereas the non-structural proteins are involved in viral assembly and transmission. The SARS-CoV-2 infection progression begins with viral entrance mediated by the spike glycoprotein's interaction with the host ACE2 receptor molecule and breaking of the spike protein by the TMPRSS2 before fusion with the host cell membrane [70, 71]. We targeted the TMPRSS2 protein as a therapeutic target to inhibit viral entrance because of its crucial role in viral pathogenesis. We recently published several papers on recommending possible small molecular chemical compounds against the SARS-CoV-2 main protease and RdRp proteins using our novel pipeline that includes deep learning–based drug screening [12, 72]. The resulting top three drugs and top nine drug-like compounds were used for further analysis. MD simulations examined the top three TMPRSS2-ligand complexes for 100 ns to validate the drug-screening results further. During the simulation, the structure of the docking complexes altered. However, ligands stayed mainly within the TMPRSS2-binding region. Multiple van der Waals and hydrogen bonding interactions help to stabilize the TMPRSS2-ligand complexes. MD simulations revealed that the compounds are relatively stable, based on a careful examination of hydrogen bonding during simulation time. The well-tempered metadynamics and MMPBSA free energies further support the results of drug screening.

Conclusions

Deep learning–based large-scale drug screening and MD simulations suggest that the drugs ABT-199, anidulafungin, and carfilzomib are the most promising candidates for blocking viral entry in cells. In addition, this research provides nine potential drug-like compounds from the database that can facilitate drug discovery. The validation of recommended possible drugs through experimental investigations should be the next stage in moving the research forward.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Wu F, Zhao S, Yu B et al (2020) A new coronavirus associated with human respiratory disease in China. Nature 579:265–269. https://doi.org/10.1038/s41586-020-2008-3
2. Zhou P, Yang X-L, Wang X-G et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273. https://doi.org/10.1038/s41586-020-2012-7

3. Zhao S, Lin Q, Ran J et al (2020) Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. Int J Infect Dis 92:214–217. https://doi.org/10.1016/j.ijid.2020.01.050

4. Cucinotta D, Vanelli M (2020) WHO Declares COVID-19 a pandemic. Acta Biomed 91:157–160. https://doi.org/10.23750/abm.v91i1.9397

5. Manski CF, Molinari F (2021) Estimating the COVID-19 infection rate: anatomy of an inference problem. J Econom 220:181–192. https://doi.org/10.1016/j.jeconom.2020.04.041

6. Fehr AR, Channappanavar R, Perlman S (2017) Middle East respiratory syndrome: emergence of a Pathogenic Human Coronavirus. Annu Rev Med 68:387–399. https://doi.org/10.1146/annurev-med-051215-031152

7. Lau SKP, Woo PCY, Li KSM et al (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci 102:14040–14045. https://doi.org/10.1073/pnas.0506735102

8. de Wit E, van Doremalen N, Falzarano D, Munster VJ (2016) SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 14:523–534. https://doi.org/10.1038/nrmicro.2016.81

9. Song L-G, Xie Q-X, Lao H-L, Lv Z-Y (2021) Human coronaviruses and therapeutic drug discovery. Infect Dis Poverty 10:28. https://doi.org/10.1186/s40249-021-00812-9

10. Kumar D, Chauhan G, Kalra S et al (2020) A perspective on potential target proteins of COVID-19: comparison with SARS-CoV for designing new small molecules. Bioorg Chem 104:104326. https://doi.org/10.1016/j.bioorg.2020.104326

11. Krumm ZA, Lloyd GM, Francis CP et al (2021) Precision therapeutic targets for COVID-19. Virol J 18:66. https://doi.org/10.1186/s12985-021-01526-y

12. Zhang H, Yang Y, Li J et al (2020) A novel virtual screening procedure identifies Pralatrexate as inhibitor of SARS-CoV-2 RdRp and it reduces viral replication in vitro. PLOS Comput Biol 16:e1008489. https://doi.org/10.1371/journal.pcbi.1008489

13. Hoffmann M, Kleine-Weber H, Schroeder S et al (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181:271-280.e8. https://doi.org/10.1016/j.cell.2020.02.052

14. Tai W, He L, Zhang X et al (2020) Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cell Mol Immunol 17:613–620. https://doi.org/10.1038/s41423-020-0400-4

15. Liu C, Zhou Q, Li Y et al (2020) Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. ACS Cent Sci 6:315–331. https://doi.org/10.1021/acscentsci.0c00272

16. Shen LW, Mao HJ, Wu YL et al (2017) TMPRSS2: a potential target for treatment of influenza virus and coronavirus infections. Biochimie 142:1–10. https://doi.org/10.1016/j.biochi.2017.07.016

17. Matsuyama S, Nao N, Shirato K et al (2020) Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proc Natl Acad Sci 117:7001–7003. https://doi.org/10.1073/pnas.2002589117

18. Mollica V, Rizzo A, Massari F (2020) The pivotal role of TMPRSS2 in coronavirus disease 2019 and prostate cancer. Futur Oncol 16:2029–2033. https://doi.org/10.2217/fon-2020-0571

19. Stopsack KH, Mucci LA, Antonarakis ES et al (2020) TMPRSS2 and COVID-19: serendipity or opportunity for intervention? Cancer Discov 10:779–782. https://doi.org/10.1158/2159-8290.CD-20-0451

20. Asselta R, Paraboschi EM, Mantovani A, Duga S (2020) ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. Aging (Albany NY) 12:10087–10098. https://doi.org/10.18632/aging.103415

21. Li J, Zheng S, Chen B et al (2016) A survey of current trends in computational drug repositioning. Brief Bioinform 17:2–12. https://doi.org/10.1093/bib/bbv020

22. Irwin JJ, Sterling T, Mysinger MM et al (2012) ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 52:1757–1768. https://doi.org/10.1021/ci3001277

23. Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, Masoudi-Nejad A (2020) Drug databases and their contributions to drug repurposing. Genomics 112:1087–1095. https://doi.org/10.1016/j.ygeno.2019.06.021

24. Ashburn TT, Thor KB (2004) Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 3:673–683. https://doi.org/10.1038/nrd1468

25. Chopra G, Samudrala R (2016) Exploring polypharmacology in drug discovery and repurposing using the CANDO platform. Curr Pharm Des 22:3109–3123. https://doi.org/10.2174/1381612822666160325121943

26. Hurle MR, Yang L, Xie Q et al (2013) Computational drug repositioning: from data to therapeutics. Clin Pharmacol Ther 93:335–341. https://doi.org/10.1038/clpt.2013.1

27. Moorthy V, Henao Restrepo AM, Preziosi M-P, Swaminathan S (2020) Data sharing for novel coronavirus (COVID-19). Bull World Health Organ 98:150–150. https://doi.org/10.2471/BLT.20.251561

28. Fukuoka Y (2019) Machine learning approach for predicting new uses of existing drugs and evaluation of their reliabilities. In: Vanhaelen Quentin (ed) Computational methods for drug repurposing. Humana Press, New York, pp 269–279

29. Saravanan KM, Zhang H, Hossain MT et al (2021) Deep learning-based drug screening for COVID-19 and case studies. In: Roy K (ed) In Silico Modeling of Drugs Against Coronaviruses. Humana, New York, pp 631–660

30. Zhang H, Saravanan KM, Yang Y et al (2020) Deep learning based drug screening for novel coronavirus 2019-nCov. Interdiscip Sci Comput Life Sci 12:368–376. https://doi.org/10.1007/s12539-020-00376-6

31. Zhang H, Saravanan KM, Lin J et al (2020) DeepBindPoc: a deep learning method to rank ligand binding pockets using molecular vector representation. PeerJ 8:e8864. https://doi.org/10.7717/peerj.8864

32. Saravanan KM, Zhang H, Senthil R et al (2022) Structural basis for the inhibition of SARS-CoV2 main protease by Indian medicinal plant-derived antiviral compounds. J Biomol Struct Dyn 40:1970–1978. https://doi.org/10.1080/07391102.2020.1834457

33. Goodsell DS, Sanner MF, Olson AJ, Forli S (2021) The <scp>AutoDock</scp> suite at 30. Protein Sci 30:31–43. https://doi.org/10.1002/pro.3934

34. Xu X, Chen P, Wang J et al (2020) Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci China Life Sci 63:457–460. https://doi.org/10.1007/s11427-020-1637-5

35. Zhang H, Zhang Q, Ju F et al (2019) Predicting protein inter-residue contacts using composite likelihood maximization and deep learning. BMC Bioinform 20:537. https://doi.org/10.1186/s12859-019-3051-7

36. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447. https://doi.org/10.1021/ct700301q

37. Wang T, Wu M-B, Lin J-P, Yang L-R (2015) Quantitative structure–activity relationship: promising advances in drug discovery platforms. Expert Opin Drug Discov 10:1283–1300. https://doi.org/10.1517/17460441.2015.1083006

38. Kadioglu O, Efferth T (2019) A machine learning-based prediction platform for P-glycoprotein modulators and its validation by molecular docking. Cells 8:1286. https://doi.org/10.3390/cells8101286

39. Yang X, Wang Y, Byrne R et al (2019) Concepts of artificial intelligence for computer-assisted drug discovery. Chem Rev 119:10520–10594

40. Cheng T, Li Q, Zhou Z et al (2012) Structure-based virtual screening for drug discovery: a problem-centric review. AAPS J 14:133–141. https://doi.org/10.1208/s12248-012-9322-0

41. Lavecchia A, Giovanni C (2013) Virtual screening strategies in drug discovery: a critical review. Curr Med Chem 20:2839–2860. https://doi.org/10.2174/09298673113209990001

42. Scior T, Bender A, Tresadern G et al (2012) Recognizing pitfalls in virtual screening: a critical review. J Chem Inf Model 52:867–881. https://doi.org/10.1021/ci200528d

43. Kadioglu O, Saeed M, Greten HJ, Efferth T (2021) Identification of novel compounds against three targets of SARS CoV-2 coronavirus by combined virtual screening and supervised machine learning. Comput Biol Med 133:104359. https://doi.org/10.1016/j.compbiomed.2021.104359

44. Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys Rev Lett 100:020603. https://doi.org/10.1103/PhysRevLett.100.020603

45. Escalante DE, Ferguson DM (2021) Structural modeling and analysis of the SARS-CoV-2 cell entry inhibitor camostat bound to the trypsin-like protease TMPRSS2. Med Chem Res 30:399–409. https://doi.org/10.1007/s00044-021-02708-7

46. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738. https://doi.org/10.1038/nprot.2010.5

47. Wu S, Zhang Y (2007) LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Res 35:3375–3382. https://doi.org/10.1093/nar/gkm251

48. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410. https://doi.org/10.1093/nar/gkm290

49. Lovell SC, Davis IW, Arendall WB et al (2003) Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins Struct Funct Bioinforma 50:437–450. https://doi.org/10.1002/prot.10286

50. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291. https://doi.org/10.1107/S0021889892009944

51. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519. https://doi.org/10.1002/pro.5560020916

52. Benkert P, Tosatto SCE, Schomburg D (2008) QMEAN: a comprehensive scoring function for model quality assessment. Proteins 71:261–277. https://doi.org/10.1002/prot.21715

53. Wang R, Fang X, Lu Y et al (2005) The PDBbind database: methodologies and updates. J Med Chem 48:4111–4119. https://doi.org/10.1021/jm048957q

54. Jaeger S, Fulle S, Turk S (2018) Mol2vec: unsupervised machine learning approach with chemical intuition. J Chem Inf Model 58:27–35. https://doi.org/10.1021/acs.jcim.7b00616

55. Brooks BR III, CLB, A. D. Mackerell J, et al (2009) Autodock Vina. J Comput Chem. https://doi.org/10.1002/jcc

56. Zhang H, Liao L, Saravanan KM et al (2019) DeepBindRG: a deep learning based method for estimating effective protein–ligand affinity. PeerJ 7:e7362. https://doi.org/10.7717/peerj.7362

57. Schrödinger (2018) PyMOL, molecular visualization system. https://pymol.org. Accessed 15 Aug 2021

58. Sousa da Silva AW, Vranken WF (2012) ACPYPE - AnteChamber PYthon Parser interfacE. BMC Res Notes 5:367. https://doi.org/10.1186/1756-0500-5-367

59. Jorgensen WL, Chandrasekhar J, Madura JD et al (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935. https://doi.org/10.1063/1.445869

60. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472

61. Bussi G, Tribello GA (2019) Analyzing and biasing simulations with PLUMED. In: Bonomi M, Camilloni C (eds) Biomolecular Simulations. Humana, New York, pp 529–578

62. Wang C, Greene D, Xiao L et al (2018) Recent developments and applications of the MMPBSA method. Front Mol Biosci 4:87

63. Kumari R, Kumar R, Lynn A (2014) G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 54:1951–1962. https://doi.org/10.1021/ci500020m

64. Roy A, Yang J, Zhang Y (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res 40:W471–W477. https://doi.org/10.1093/nar/gks372

65. Pan R, Hogdal LJ, Benito JM et al (2014) Selective BCL-2 inhibition by ABT-199 causes on-target cell death in acute myeloid leukemia. Cancer Discov 4:362–375. https://doi.org/10.1158/2159-8290.CD-13-0609

66. Bodo J, Zhao X, Smith MR, Hsi ED (2014) Activity of ABT-199 and acquired resistance in follicular lymphoma cells. Blood 124:3635–3635. https://doi.org/10.1182/blood.V124.21.3635.3635

67. Jayaweera SPE, Wanigasinghe Kanakanamge SP, Rajalingam D, Silva GN (2021) Carfilzomib: a Promising proteasome inhibitor for the treatment of relapsed and refractory multiple myeloma. Front Oncol. https://doi.org/10.3389/fonc.2021.740796

68. Pollard TD, Earnshaw WC, Lippincott-Schwartz J, Johnson GT (2017) Chapter 4 - Biophysical Principles. In: Pollard TD (ed) Cell Biology, 3rd edn. Elsevier, Amsterdam, pp 53–62

69. Pal M, Berhanu G, Desalegn C, Kandi V (2020) Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): an update. Cureus. https://doi.org/10.7759/cureus.7423

70. Weissenstein A (2020) TMPRSS2-inhibitors play a role in cell entry mechanism of COVID-19: an insight into camostat and nafamostat. J Regen Biol Med 2:1–3. https://doi.org/10.37191/Mapsci-2582-385X-2(2)-022

71. Yesudhas D, Srivastava A, Gromiha MM (2021) COVID-19 outbreak: history, mechanism, transmission, structural studies and therapeutics. Infection 49:199–213. https://doi.org/10.1007/s15010-020-01516-2

72. Zhang H, Bei Z, Xi W et al (2021) Evaluation of residue-residue contact prediction methods: from retrospective to prospective. PLOS Comput Biol 17:e1009027

Springer