

## **P2-P3 $\alpha$ -ketoamide derivative and Pentagastrin: an *in-silico* drug discovery approach**

Ikechukwu Achilonu<sup>1\*</sup>, Emmanuel Amarachi Iwuchukwu<sup>1</sup>, Okechinyere Juliet Achilonu<sup>2</sup>,  
Manuel Antonio Fernandes<sup>3</sup>, and Yasien Sayed<sup>1</sup>

<sup>1</sup>Protein Structure-Function Research Unit, School of Molecular and Cell Biology, Faculty of  
Sciences, University of Witwatersrand, Johannesburg, 2050, South Africa

<sup>2</sup>Division of Biostatistics, School of Public Health, Faculty of Health Sciences, University of  
Witwatersrand, Johannesburg, 2050, South Africa

<sup>3</sup>Molecular Sciences Institute, School of Chemistry, University of Witwatersrand,  
Johannesburg, 2050, South Africa

\*Corresponding author email: [Ikechukwu.Achilonu@wits.ac.za](mailto:Ikechukwu.Achilonu@wits.ac.za)

## Abstract

The SARS-CoV-2 main protease ( $M^{pro}$ ) is an attractive target towards discovery of drugs to treat COVID-19 because of its key role in virus replication. The atomic structure of  $M^{pro}$  in complex with an  $\alpha$ -ketoamide inhibitor (Lig13b) is available (PDB ID:6Y2G). Using 6Y2G and the prior knowledge that protease inhibitors could eradicate COVID-19, we designed a computational study aimed at identifying FDA-approved drugs that could interact with  $M^{pro}$ . We searched the DrugBank and PubChem for analogs and built a virtual library containing ~33000 conformers. Using high-throughput virtual screening and ligand docking, we identified Isavuconazonium, a ketoamide inhibitor ( $\alpha$ -KI) and Pentagastrin as the top three molecules (Lig13b as the benchmark) based on docking energy. The  $\Delta G_{bind}$  of Lig13b, Isavuconazonium,  $\alpha$ -KI, Pentagastrin was -117.68, -191.11, -187.36, -145.76 kJ/mol, respectively. Molecular dynamics simulation revealed that these ligands are stable within the  $M^{pro}$  active site. Binding of these ligands is driven by a variety of non-bonded interaction, including polar bonds, H-bonds, van der Waals and salt bridges. The overall conformational dynamics of the complexed- $M^{pro}$  was slightly altered relative to apo- $M^{pro}$ . This study demonstrates that three distinct classes molecules, Isavuconazonium (triazole),  $\alpha$ -KI (ketoamide) and Pentagastrin (peptide) could serve as potential drugs to treat patients with COVID-19.

## **Keywords**

COVID-19; Antiretroviral; Main protease; Ligand docking; Molecular dynamics simulation, High-throughput virtual screening; FDA; DrugBank; PubChem; Pentagastrin; ketoamide inhibitor; Isavuconazonium

## Introduction

With over 4.53 million infections and 307108 deaths today (15<sup>th</sup> May 2020), the world is witnessing a calamitous viral pandemic caused by a new strain of a coronavirus, scientifically referred to as SARS-Cov2, the causative agent of corona virus disease (COVID-19). Retroviral SARS-CoV-2 shares ~82% genome similarity to the SARS coronavirus<sup>1,2</sup>. The viral infection is believed to have originated in China with the initial epicentre in Wuhan, a city in the Hubei Province of China<sup>3,4</sup>. The actual date of origin of the infection and first patient is still unknown; however, it is believed that this virus may have crossed from wild animals to humans, a possible zoonotic virus that is originally from bats, akin to the African Ebola and the Lassa fever viruses<sup>5</sup>. Over the course of time the epicentre, the epicentre has moved from China to Italy and currently the epicentre in the USA. At the time of writing this article there is no known vaccine or treatment option available. Even though the death rate is lower than the historical coronavirus-associated SARS, the recovery rate seems to be relatively protracted. This has resulted in straining health care systems globally and sub-Saharan Africa may become the next epicentre if the infection is not controlled effectively. In most countries around the world, the only means of control is by instating a nationwide lockdown. Countries, such as South Africa, that took this initiative, however disastrous it was on the economy, have witnessed a slower infection rate, from 45% to less than 4% (between 27<sup>th</sup> March 2020 and 10<sup>th</sup> of April 2020).

As viral disease outbreaks are not often foreseen, COVID-19 is a global emergency and there is a race against time to produce either a vaccine and/or effective drugs to curb the global plague of COVID-19<sup>6</sup>. Several drug treatments have been proposed worldwide. Chloroquine/hydroxychloroquine, an FDA-approved drug that was used to treat malaria has been proposed, in combination with zinc, to be effective in eradicating SARS-CoV-2 in patients<sup>7-12</sup>. This is because chloroquine is an endocytosis blocker and acts as an ionophore that facilitates the entry of  $Zn^{2+}$  into cells such as T-lymphocytes and  $Zn^{2+}$  is known to inhibit coronavirus and arterivirus RNA polymerase<sup>13</sup>. Other antiviral agents including remdesivir, lopinavir, ritonavir and interferon  $\alpha$ , have also been proposed as possible treatment alternatives against COVID-19<sup>14,15</sup>. Currently, several countries have begun chloroquine and antiretroviral treatment trials; however, some of the clinical information emanating from these trials is still anecdotal and cannot be justified as the treatment of choice<sup>16</sup>. As the world continues to grapple with the outbreak of the SARS-Cov-2 virus, the most logical approach to treating this infection is by accelerated rational drug discovery<sup>17</sup> using a combination of computational modelling and empirical studies<sup>18,19</sup>, such as in autoimmune disease drug discovery<sup>20</sup>. However, this is

only possible if there are empirically-determined crystal structures of key druggable targets in the virus. In this instance, Mpro, a viral protease, represents a prime target because it is critical for processing viral polyproteins and viral maturation inside infected host cells<sup>21</sup>. Of concern are unconfirmed reports in Asia of people who have recently recovered from COVID-19 infection and have tested positive for a second time. This may point to the fact that vaccines alone may not be the best strategy for dealing with this pandemic. This, therefore, underscores the need for a rational approach to the discovery of drugs to treat the current scourge of COVID-19.

Proteases are attractive targets in a rational approach to COVID-19 drug discovery. This is because most retroviruses depend on key enzymes, such as proteases, for processing of their polyprotein precursors<sup>22-24</sup>. Zhang et al.<sup>21</sup> recently published a paper on the structure of the SARS-CoV-2 main protease (SARS-Cov-2 Mpro) with accession code 6Y2G and deposited it in the Protein Data Bank in March 2020<sup>21</sup>. A notable feature of this atomic structure is that a derivative of an  $\alpha$ -ketoamide inhibitor is bound to the enzyme and this must, therefore, serve as the seeding point that drives the rational drug design approach towards anti-SARS-CoV-2 drug discovery. This atomic structure, in addition to the other ~112 PDB-deposited structures related to COVID-19, can serve as a template to discover other potential inhibitors using computational/machine learning (artificial intelligence) studies. Currently, several databases are curating several million compounds, some of which are FDA-approved drugs as well as drugs that are at the final stages of clinical trials with published outcomes. Databases such as DrugBank<sup>25,26</sup> and PubChem<sup>27,28</sup> offer structural biologists the opportunity to X-ray millions of compounds that can be validated theoretically as having potential bioactivity towards SARS-CoV2 viral enzymes.

This study typifies the contribution of Computational Biology and Chemistry in the race towards successful rational drug design discovery. Our work showcases the value of biocomputational studies and that it may be used effectively and reliably as a tool to validate the results obtained using other biochemical and biophysical techniques. In light of this, our study validated the study by Zhang et al.<sup>21</sup> and other studies that proposed to re-purposing<sup>29</sup> of currently available drugs such as such as remdesivir, a drug that was initially designed, although ineffective, against the Ebola virus<sup>18,19</sup>. We could have screened the entire DrugBank and other publicly available small molecule databases using high-throughput virtual screening (HTVS), but we chose to begin with two classes of protease inhibitors (i) the ketoamide inhibitors and (ii) known antiretrovirals including remdesivir, saquinavir, atazanavir using

computational modelling approaches. We used HTVS, induced-fit ligand docking and molecular dynamics simulation studies to identify additional classes of plausible FDA-approved drugs as possible drug candidate to treat COVID-19. The conceptual framework of our study is illustrated in Figure 1.

## Results

### *HTVS and ligand docking*

The rationale behind performing molecular docking is to make a systematic prediction of the ideal pose or conformation of a ligand in a protein's binding site, which could be taken further for molecular dynamics simulation studies. In this study, we used HTVS to screen over 33000 (Step 1) conformers of 11 FDA-approved antiretrovirals generated using the Maestro LigPrep algorithm. Specifically, we screened (i) HIV PIs: Atazanavir, Darunavir, Fosamprenavir, Indinavir, Lopinavir, Ritonavir, Saquinavir and Tipranavir; (ii) NNRTIs: Delavirdine and Nevirapine; (iii) A broad spectrum antiviral: Remdesivir. HTVS and Standard Protocol (SP) Glide Docking algorithms implemented in Maestro Schrödinger were used for the library screening. Only eight (Atazanavir, Fosamprenavir, Indinavir, Saquinavir, Tipranavir, Delavirdine, Nevirapine and Remdesivir) out of the 11 classes of Maestro-generated conformers submitted in Step 1 yielded results (See Table S1 and Figure S1). The top-scoring ligands (from Step 1) on the basis of the pose with the lowest Glide Emodel energy values, were submitted for IFD (Step 2) alongside (i) the parent compound (ii) Ligand 13b (benchmark) described by Zhang et al.<sup>21</sup> (iii) Isavuconazonium, (iv) P2-P3  $\alpha$ -ketoamide inhibitor derivative ( $\alpha$ -KI), and (v) Pentagastrin. The properties, docking scores and MM/GBSA  $\Delta G_{\text{bind}}$  of all compounds with Emodel energy less than the benchmark ligand (Lig13b) are shown in Table 1.

We will focus mainly on the four peptidomimetics (indicated in Table 1) that were subsequently submitted for MD simulation studies. Figure S2 illustrates the 2D molecular structure of these four ligands. Overall, all the poses showed that the ligands bind at a single site in the receptor. Besides, the Maestro IFD algorithm produced docking poses comparable (RMSD = 0.75 Å) with the coordinates of the 6Y2G model (Figure S3). Asp 187 and Glu 166 are the two charged amino acid residues within 4 Å of the bound ligand common across all four ligands.

Figure 2 illustrates the interaction between the ligands and the amino acid side chains within 4 Å of the ligand. For the binding of the benchmark ligand 13b (Lig13b) to the M<sup>pro</sup>, the docking simulation showed that the interaction occurred with an average E<sub>model</sub> energy of -486.7 kJ/mol. This interaction was stabilised mainly by H-bonds and van der Waals interactions involving Gly 143, eight polar, seven hydrophobic, and three charged amino acid side chains. There was H-bond contact between Gly 143, Asn 142, Glu 166 and Lig13b. The binding of  $\alpha$ -KI to M<sup>pro</sup> also occurred with E<sub>model</sub> energy of -575.2 kJ/mol. This interaction was stabilised by a  $\pi$ - $\pi^*$  interaction between the active site His 41 and the ligand, and five hydrogen bond interactions with Cys 145, His 163, His 164, Asn 142 and Gly 143 side chains. There were 24 amino acid residues within 4 Å of  $\alpha$ -KI that potentially contribute to the stability of the ligand within the receptor. The activity between M<sup>pro</sup> and Pentagastrin involved 13 polar, 10 hydrophobic and two negatively charged amino acid side chains. Although catalytic Cys 145 and His 41 were within 4 Å of the Pentagastrin, these two amino acids were not directly contributing H-bond contacts with the ligand. The overall E<sub>model</sub> energy and  $\Delta G_{\text{bind}}$  of this interaction were -575.2 kJ/mol, respectively. The interaction between M<sup>pro</sup> and Isavuconazonium was stabilised by four H-bond interactions between the ligand and His 163, Leu 141 and His 41 and one  $\pi$ - $\pi^*$  interaction with catalytic His 41. Eleven polar, nine hydrophobic, three charged amino acids side chains were within 4 Å of the ligand that could potentially form non-bonded interactions with the ligand in real-time. The E<sub>model</sub> binding energy between M<sup>pro</sup> and Isavuconazonium is -507.17 kJ/mol.

#### *Free binding energy and per-residue energy decomposition calculations*

Free energy of interaction between two molecules in a biological system is the crucial measure in the description of the thermodynamics of biological systems, which makes it an important objective of simulation of biomolecular systems. Therefore, it was important that we determined the free energy of binding between the top three ligands and the receptor, in comparison with the benchmark ligand (Lig13b) using the MM/GBSA algorithm. In addition to the free binding energy, it was also important to highlight the contributions each amino acid residue makes by estimating the per-residue energy decomposition of residues that contribute to the binding of the ligands at the active site of M<sup>pro</sup>. Table 2 summarises the energy components of the MM/GBSA calculations. The calculated  $\Delta G_{\text{bind}}$  for the four ligands suggest that the gas phase energy, although counteracted by the solvation free energy, contributes chiefly to the overall  $\Delta G_{\text{bind}}$ . The results show that the binding of the four ligands may occur spontaneously with negative  $\Delta G_{\text{bind}}$ . The association of Lig13b and  $\alpha$ -KI with M<sup>pro</sup> is mainly

driven by van der Waals interactions, whereas the association between Isavuconazonium and M<sup>pro</sup> is largely driven by electrostatic forces. The van der Waals and electrostatic component of Pentagastrin interaction is comparable. Figure S4 shows the per-residue energy decomposition plots of the four ligands. Isavuconazonium appears to have more amino acid residues contributing towards its stability within the active site when compared to the other three ligands.

#### *Molecular dynamics simulation*

Each of the complexes between M<sup>pro</sup> and the four peptidomimetic ligands, as well as the non-liganded (apo-M<sup>pro</sup>) discussed above were subjected to a 50 ns isothermal-isobaric (NPT) ensemble (~ 1 bar and 300 K) MD simulation using the OPLS\_2005 force field and a TIP3P solvent model. The structural fluctuations or stability of each complex was measured by the root-mean-square deviation (RMSD) and the root-mean-square fluctuation (RMSF). Only  $\alpha$ -carbon RMSD and side chain RMSF are reported. The overall quality of the simulation event (Table S2) showed that the overall total energy, potential energy, temperature and volume were stable throughout the 50 ns simulation period (average slope  $\pm$  0.002-unit/ps). All the M<sup>pro</sup> - ligand complexes showed comparable C $\alpha$  RMSD in comparison with the non-liganded system (Figure 3A). Overall, the system reached equilibrium around 10 ns with an average C $\alpha$  RMSD of 1.66 Å ( $\pm$  0.14 Å) after an initial rapid increase from 0 to 2.5 ns (Figure S5). The system converged and remained stable beyond 10 ns. This is further affirmed by the representative 3D structures of the complexes superimposed on the apo-M<sup>pro</sup> (Figure 3B). The RMSD between apo-M<sup>pro</sup> and the M<sup>pro</sup> complexes are between 1.22 and 1.66 Å. This indicates that the binding of the ligand does not severely perturb the global structure of the protease. The RMSF per-residue calculated over the 301 amino acid residues of the M<sup>pro</sup> for the liganded complex in comparison with the non-liganded M<sup>pro</sup> is shown in Figure 4. RMSF evaluates the degree of the displacement of a specific atom, or group of atoms, relative to the crystal structure, which is averaged over the number of atoms. The results show that ligand binding did not induce major side chain perturbations in the receptor, although subtle differences in amino acid side chain RMSF between the liganded and non-liganded system is observable. In order to further examine the side chain RMSF, we plotted the difference in the side chain RMSF between the apo protein and the liganded protein (Figure S6). According to the figure, residues with positive RMSF fluctuated more in comparison with the corresponding residue in the apo-protein (and vice versa). The figure shows that side chains of the M<sup>pro</sup>-Lig13b complex, fluctuated the least in comparison with the apo-protein, while the other three complexes showed almost similar

patterns in the side chain RMSF, relative to the apo-protein. One similarity across the four complexes is that ligand binding induced stability in Gln 273, although this residue is not directly involved in ligand binding.

To further gain insights into the stability of the individual M<sup>Pro</sup>-ligand complexes, we also investigated the conformational dynamics of the ligands within the binding site and their H-bond interactions with the protein and the environment over the 50 ns MD production phase of the simulation. Using RMSD of the ligand with respect to the receptor and an index of ligand conformational stability, our results show that the average RMSD of the ligands is 5.10 Å ( $\pm 1.80$  Å) with respect to the receptor. The plot of the average ligand RMSD (Figure S6) with respect to protein versus time shows that there was an overall increase in fluctuation within the first 20 ns of the simulation (Figure 5) and stable evolution of the ligand within the receptor in the last 30 ns of the simulation. Isavuconazonium showed the least deviation ( $3.63 \pm 0.61$  Å) within M<sup>Pro</sup>, followed by Lig13b, Pentagastrin and  $\alpha$ -KI, with  $4.82 \pm 1.48$  Å,  $5.66 \pm 1.19$  Å and  $6.65 \pm 1.87$  Å, respectively. In addition to the RMSD of the ligand with respect to the protein, other ligand properties that are crucial for interaction with the receptor including intra-H-bond contacts, the radius of gyration, solvent accessible surface area, polar accessible surface area and molecular surface area of the ligands (summarised in Table S3 and shown in Figure S7). Pentagastrin makes the most intra-H-bond contacts ( $\sim 1.4$  contacts) within the 50 ns simulation time, followed by Isavuconazonium, Lig13b and  $\alpha$ -KI (which rarely made any intra-H-bond contact throughout the 50 ns MD production period. Concerning molecular surface area, Pentagastrin showed the largest surface area of approximately ( $637.19 \pm 15.50$  Å<sup>2</sup>), which may arise as a result of high intra-H-bond contacts. Overall, the dynamic polar surface properties of the ligand remained constant throughout the 50 ns MD simulation period. Figure 6 shows the fraction of H-bond contact between the ligand and the side chains. The figure shows that majority of the contacts between the ligands and the active site residues are water bridge, van der Waals and H-bond interactions.

To further affirm and validate our MD simulation in comparison with the IFD, we examined the non-bonded interactions that stabilise the ligand within the M<sup>Pro</sup> active site. By dividing the trajectories into two different conformational clusters based on C $\alpha$  atoms position, the Desmond algorithms reduce such ensembles to a manageable size. Figure 7 is an average of a representative cluster determined using C $\alpha$  RMSD trajectory clustering algorithm implemented in the Desmond MD machine. Our results show (summarised in Table 3) that all four ligands are stabilised by potential non-bonded interactions between the ligand and polar, hydrophobic

and charged amino acid side chains, as well as water and ions within 4 Å of the ligand. Pentagastrin and Lig13b are both stabilised by 10 H-bond contacts with water molecules and amino acid side chains, whereas Isavuconazonium and  $\alpha$ -KI are stabilised by seven H-bond contacts with water molecules and amino acid side chains. In addition to H-bond contacts, Isavuconazonium makes a salt bridge electrostatic contact with a  $\text{Cl}^-$ . Overall, more H-bond contact is made during the MD simulation in comparison with the IFD simulation. The root-mean-square deviation between the structures when they are superimposed on apo- $\text{M}^{\text{pro}}$  is 1.22-1.66 Å, which shows that the structures were conserved throughout the 50 ns simulation time. In other words, the binding of the ligand did not induce any major conformational perturbations in the protease.

We examined the dynamics between the catalytic His 41 and Cys 145 side chains by measuring the distance between the HE1 of His 41 and HG of Cys 145 along the 50 ns MD simulation (Figure 8). The results show that the distance between HE1 of His 41 and HG of Cys 145 was consistent throughout the simulation time in apo- $\text{M}^{\text{pro}}$ . However, there was an initial perturbation in the distance between these two atoms in the  $\text{M}^{\text{pro}}$ -ligand complexes within the first 25 ns of the simulation. The atomic distance convergence after 25 ns was maintained at a stable distance in the last 25 ns.

To further explore the effect of ligand binding on the conformational states of  $\text{M}^{\text{pro}}$  in a principal component (PC) phase space, PCA was carried out for all  $\text{C}\alpha$  atoms in the five simulated systems as a function of the 50 ns trajectory. Every single mode in the apo- and complexed-  $\text{M}^{\text{pro}}$  systems is related to an eigenvalue, which is congruent with the degree of fluctuation along a mode, and it contains different contributions of the  $\text{C}\alpha$  atoms. The eigenvalues were related to large intensive motions initially, followed by a rapid decline, and finally, reached more localised fluctuations (Figure S8). The top 20 PCs for apo- $\text{M}^{\text{pro}}$ , Lig13b- $\text{M}^{\text{pro}}$ ,  $\alpha$ -KI- $\text{M}^{\text{pro}}$ , Pentagastrin- $\text{M}^{\text{pro}}$  and Isavuconazonium- $\text{M}^{\text{pro}}$  systems described 77.8, 74.9, 74.1, 76.4 and 80.1% of the total variance during the 50 ns of the trajectories, respectively. Table 3 summarises the contributions of the first two principal components (PCs) to the variance in  $\text{C}\alpha$  atomic fluctuations, relative to the remaining PCs. For instance, the contributions of the first and second PCs to the variance in  $\text{C}\alpha$  atomic fluctuations in Lig13b- $\text{M}^{\text{pro}}$  system were 24.3% and 12.9%, respectively, whereas the contribution of the remaining PCs to this variance was less than 9%. Therefore, PC1 and PC2 was enough to provide a useful description of the atomic fluctuation in the sampled conformations, while still retaining most of the variance in the original distribution. Therefore, the first and second PCs were used to

analyse the transitions in conformations of the simulated systems by deconvoluting their trajectories onto a two-dimensional subspace (Figure 9). Therefore, this projection of the distribution onto the subspace described by the two largest PCs (PC1 and PC2) resulted in a lower dimensional representation of the fluctuation of the C $\alpha$  atoms in three-dimensional spaces. The continuous colour scale, which changed from blue to white and then red, indicates that there were intermittent transitions between these conformations. The results show that the conformational behaviour of the simulated systems exhibited an observable difference between the blue and red conformations as projected along the direction of PC1 and PC2. The scattered dots (Figure 9) were dispersed for the liganded systems, whereas the un-liganded systems showed a slightly different distribution of the scattered dots.

## Discussion

Among the 112 crystal structures of SARS-Cov-2 polypeptides deposited in the PDB, SARS-CoV-2 M<sup>pro</sup> (6Y2G) may be the most attractive target for computational modelling studies because 6Y2G is an M<sup>pro</sup>-inhibitor (Lig13b) complex. In addition, an empirical study has shown the inhibitor (an  $\alpha$ -ketoamide inhibitor derivative) to be effective against the virus. Using 6Y2G as a template, we aimed to generate theoretical extrapolations on plausible inhibitors of SARS-CoV-2 M<sup>pro</sup> capable of being introduced as experimental drugs for the treatment of COVID-19. We have utilised accessible databases, including PubChem and DrugBank, to search for analogs of 11 antiretrovirals and ketoamide inhibitors that can serve as (i) a template for rational drug design and (ii) FDA-approved drugs that can be introduced immediately to treat the disease and (iii) a proof of a concept of using any known drug to treat this viral pandemic. Molecular modelling has, in the past, proved to be a viable option in accelerating drug discovery and complements experimental studies in drug discovery<sup>30-32</sup>. Our hypothesis is linked to the fact that a crystal structure of a protein provides a wealth of opportunity for a computational approach to drug design<sup>33,34</sup>. In addition, a druggable target must also be critical for the biochemistry of the target pathogen-in this case, the SARS-CoV-2.

To cast our net as wide as possible in searching for this best-fit molecule, we extracted over 5000 molecules, generated over 33000 conformers (isomers) and critically assed more that15 initial hits using induced-fit ligand docking. We tested the stability of the top-scoring (using Lig13b as the benchmark) molecules with a 50 ns molecular dynamics simulation. We used

the Glide Emodel parameter implemented in Maestro v12 to score and rank our ligands. Our HTVS protocol in the Maestro modelling algorithm is designed to rationalise and filter the number of ligands that were carried over to the second step, which is IFD followed by free binding energy calculation using MM/GBSA and MD simulation. We did not apply Lipinski's Rule of Five<sup>35</sup> penalties because most of these molecules are peptidomimetics, which are often larger than 500 g/mol and may not pass the absorption parameter of this rule due to their molecular weights<sup>36</sup>. However, we excluded molecules with reactive functional groups and ensured that the molecules conformed with the physiological conditions in terms of the pKa at pH ~7.2. This is because ligands with reactive functional groups may form covalent bonds with M<sup>pro</sup>, therefore resulting in a false positive interaction<sup>37-39</sup>. The Extended Protocol (XP) docking algorithm (Maestro) was not used because we intended to extend the modelling to molecular dynamics simulation. The results from HTVS showed that most eight classes (out of 11) showed some form of biological activity towards the SARS-CoV-2 M<sup>pro</sup>. All the ligands preferred to bind at the active site (as indicated in 6Y2G) because no other site in the SARS-CoV-2 M<sup>pro</sup> attracted any of the over 33000 conformers used in the HTVS.

Although the coordinates of 6Y2G indicate that M<sup>pro</sup> is a homodimer with two active sites per subunit; for this study, we used a single subunit because we observed that it is only Ser 1 in subunit A that is within possible non-bonding interaction distance (3.9-4.1 Å) with the ligand in subunit B (and vice versa). There was no initial indication that Ser 1 makes any interaction with the active site, which largely depends on the nature of the ligand. From the architecture of the M<sup>pro</sup> active site, one is tempted to predict that several water molecules will be involved in binding of any ligand at this site. The structure of 6Y2G illustrates that two water molecules form either direct or indirect interactions with the Lig13b.

Our IFD studies showed that some of the known antiretrovirals and their analogs may interact with M<sup>pro</sup> based on docking scores and Emodel energies. However, Remdesivir, which is proposed as a potential COVID-19 drug, had binding energy values more than -316 kJ/mol, which is higher when compared with Pentagastrin, P2-P3 ketoamide derivative ( $\alpha$ -KI), and Isavuconazonium. These drugs are FDA-approved and may be optional experimental drugs for the treatment of SARS-CoV2. Studies by other groups have suggested the use or re-purposing of FDA-approved drugs for the treatment of COVID-19<sup>29,40</sup>. All three drugs showed lower Glide Emodel binding energies in comparison with the benchmark Lig13b. The ligands are mostly stabilised by hydrogen bond interactions followed by van der Waals interactions. Isavuconazonium interactions with SARS-CoV-2 M<sup>pro</sup> is stabilised by  $\pi$ - $\pi^*$  stacking interaction

between His 41 and the imidazole moiety of the ligand (in addition to H-bond and van der Waals interactions). His 41 is believed to be the catalytic histidine in this protease<sup>21</sup>. The presence of catalytic histidine residues in proteases is well-documented<sup>41-44</sup>.

One remarkable feature of the IFD algorithm is the ability to predict a realistic interaction that is comparable with experimental models<sup>45-47</sup>. A comparison of the modelled interaction between M<sup>pro</sup> and Lig13b with the experimental interaction with the experimental interaction (6Y2G) shows that the RMSD between the experimental pose and the theoretical pose is 0.27 Å. Therefore, we have strong reasons to accept the results generated in this study as “close to reality” as possible and allow rational inferences to be made from this study. All the top-scoring ligands bind within 4 Å of the catalytic Cys 145, which could be a target for covalent inhibition<sup>48,49</sup> of the SARS-CoV-2 M<sup>pro</sup>.

A large amount of useable and inferable information can be obtained from a MD simulation study. MD simulation gives us first-hand information on the stability of the ligand within the ligand-binding site and how the binding of these ligands impact on the conformational dynamics of the receptor, especially when compared with apo-M<sup>pro</sup> (unliganded M<sup>pro</sup>). In order to validate our study, we used the Desmond molecular dynamics engine to generate trajectories of M<sup>pro</sup> in a complex with Lig13b, Pentagastrin,  $\alpha$ -KI, and Isavuconazonium. Although it is advisable to test all the ligands listed in Table 1, the computational power required is demanding and this precluded us from performing extensive MD simulations. Instead, we focussed on three FDA-approved molecules which showed promise when compared with the benchmark molecule, Lig13b. Overall, all four ligands were stable within the active site throughout the 50 ns simulation period. It would have also been worthwhile extending the time scale of the simulation to microseconds. However, the residence time of a drug within a receptor is determined by several factors including the effect of the drug on the conformational dynamics of the protein receptor<sup>50,51</sup>. Nonetheless, the summary of the trajectories showed that the systems were stable after the initial time frames. But it appears that due to the structure of each ligand, the conformational dynamics of M<sup>pro</sup> varied in comparison with apo-M<sup>pro</sup>, as suggested by PCA. The binding of the ligand did not induce a global perturbation on M<sup>pro</sup>, but rather a local perturbation. This is because there was no formation or deformation of secondary structure elements within the 50 ns time frame.

Judging from the best representative structure from an ensemble of structures obtained after performing MD using the Trajectory Cluster analysis algorithm implemented in Desmond, all

four ligands formed water bridges, maintained H-bonds with the side chains, as well as close contact with the catalytic His 41 and Cys 145 residues. Therefore, our top candidate drugs may be a competitive inhibitor of M<sup>pro</sup>. Our results further affirm that the binding of the ligand affects the dynamics between these two residues and each ligand exerted a specific landscape of dynamics between these two residues. Cys 145 and His 41 may both be involved in the binding and stabilisation of the ligands within the M<sup>pro</sup> catalytic site<sup>21</sup> because these two catalytic residues are within 4 Å of the ligands in the binding pocket. This substantial piece of information may be a molecular insight into the mode of inhibition of this enzyme.

In conclusion, we have used a computational approach which includes HTVS, IFD, MM/GBSA free binding energy calculations and MD simulation to study potential drug candidates for COVID-19. We also used PCA to deconvolute the dynamics between the C $\alpha$  atoms as a function of ligand binding. Zhang et al.<sup>21</sup> laid the foundation of this study by depositing the coordinates of the atomic structure of SARS-CoV-2 M<sup>pro</sup> (6Y2G) in complex with a derivative of an  $\alpha$ -ketoamide inhibitor (Lig13b). Using this molecule, we searched publicly available small molecule databases for potential analogs that we employed in our computational studies. With Lig13b as the molecular gauge for inhibition of SARS-CoV-2 M<sup>pro</sup>, we could confidently infer that Isavuconazonium,  $\alpha$ -KI and Pentagastrin are potential drug candidates that could comparatively exert similar effects on M<sup>pro</sup> as Lig13b. Information on these three drugs, including pharmacokinetics, mode of action, toxicity and doses can be extracted from the DrugBank and PubChem databases. At this stage it is not known whether there are other compounds in these databases that can outperform these three drugs. Detailed biochemical, biophysical, cell culture and high-throughput screening studies are needed to further interrogate the potential effectiveness of the molecules, Isavuconazonium,  $\alpha$ -KI and Pentagastrin, identified in this study.

## **Materials and Methods**

### *Computer hardware*

Two high-performance desktop servers equipped with a 16 CPU Intel<sup>®</sup> core i7<sup>™</sup> 5960x Extreme Edition (3.3 GHz, 20 M cache 16x cores), Nvidia GTX 750Ti graphics card, 32 GB DDR4-2133 MHz memory on an MSI X99 motherboard and 264 GB RAM were used for this study. The Window OS desktop (containing the Schrödinger Maestro v12) was used for HTVS and ligand docking, while the Linux server (containing the 2019 GPU-enabled Schrödinger

Desmond version) was used for molecular dynamics simulation studies. All post-dynamic analysis was done on a 64-bit Windows Lenovo Legion Y540 laptop with the following configuration: Intel 9th Generation Intel Core i7-9750H, 4.5 GHz processor, 12 MB cache, 6x cores, 12x threads, 64 GB DDR4 RAM, 2 TB ultra-fast SSD, 1 TB hard drive and Nvidia series 20 GeForce RTX 2060 6 GB GDDR6 graphics card (GPU).

#### *Ligand and receptor extraction and preparation*

The  $\alpha$ -ketoamide derivative (Lig13b) by Zhang et al.<sup>21</sup> and eleven FDS-approved antiretrovirals (Table S1) were submitted to the PubChem and DrugBank databases for analog search. Similar compounds were extracted in a structured data file (SDF) format and submitted to the LigPrep module implemented in Maestro v12 for ligand preparation, which involves energy minimisation using OPLS 2005 force field. The algorithm was set to generate possible states of the molecules at pH  $7.0\pm 2$ , while accurately predicting the  $pK_a$  of these states at the set pH using the Epik module of the algorithm. The ligands were also desalted and possible tautomeric states ( $\sim 32$  tautomers/ligand) were further generated at pH  $7.0\pm 2$ . Additionally, specific chiral centres were retained (for molecules with multiple chiral centres), while other chiral centres were varied during the ligand preparation to return chemically sensible structures. These generated molecules were saved as a compressed Maestro file. The atomic coordinate for the SARS-CoV-2 M<sup>Pro</sup> (PDB ID 6Y2G) was extracted from the RCSB-PDB database and submitted to the Protein Preparation Wizard module implemented in Maestro. The entire structure was energy-minimised by assignment of accurate protonation state at physiological pH and hydrogen atoms were added to the crystal structure using the default parameters. The stereochemistry of the side chains was checked to ensure that no major perturbations were induced while preparing the structure.

#### *High-throughput virtual screening*

A grid file of the receptor was prepared using Maestro for the HTVS. More than 33000 molecules were prepared using the LigPrep algorithm and were submitted to the high-throughput virtual screening (HTVS) module implemented in Maestro. Three steps of the virtual screening workflow were used, beginning with the HTVS, the standard protocol (SP) and finally the extend protocol. The option for MM/GBSA was not applied at this step. The Lipinski ADME filtering was not applied, but the QikProp filtering was applied during the HTVS. The ligand docking step in the HTVS performed initial docking of the entire set of more

than 33000 molecules and 10% of the HTVS-docked ligands were further subjected to SP-docking protocol. This rigorous and systematic process generated docked potential hits that were scored using Glide docking scores.

#### *Induced-fit ligand docking*

Top scoring ligands in each class of drug were extracted and re-submitted to the induced-fit docking (IFD) module implemented in the Maestro v12 algorithm, which employs a mixed molecular docking and dynamic protocol. Briefly, the standard IFD protocol was applied to the selected (centroid) amino acid side chains (19-29, 38-54, 85, 114-119, 126, 136-147, 161-175, 181, 185-193) in an implicit solvent model using the OPLS\_2005 force field. H-bond and metal ion constraints were applied to both the initial and re-docking stages. Ring conformational sampling with a 2.5 kcal/mol energy barrier, as well as a non-planar conformation penalty on amide bonds was applied to the IFD protocol. The scaling for both receptor and ligand was set at 0.5 with a maximum of 20 allowable poses per ligand. Residues within 5 Å of the docked ligand were further refined using Prime Refinement algorithm implemented in Maestro v12. Prime energy was used to rank the refined protein-ligand complexes. The receptor structures within 30 kcal/mol of the minimum energy structure were submitted for a final round of Glide docking and scoring. Each ligand was re-docked into every single refined low-energy receptor structure in the subsequent second docking step using the default Glide XP settings.

#### *Binding free energy calculations*

Binding free energy ( $\Delta G_{\text{bind}}$ ) calculations were carried out using the molecular mechanics/generalised Born solvent area (MM/GBSA) method<sup>52-54</sup> implemented in Amber 18 in order to gain more insight into the binding of the ligands to M<sup>PrO</sup>. Briefly, the free energy of binding of Lig13b,  $\alpha$ -KI, Pentagastrin and Isavuconazonium to M<sup>PrO</sup> were calculated by averaging 2000 snapshots of the simulated complexes (from 20 ns molecular dynamics simulation).  $\Delta G_{\text{bind}}$  of ligands at the M<sup>PrO</sup> active site was calculated using (1)

$$\Delta G_{\text{bind}} = \Delta G_{\text{RL}} - (\Delta G_{\text{R}} + \Delta G_{\text{L}}) \quad (1)$$

where  $\Delta G_{\text{RL}}$ ,  $\Delta G_{\text{R}}$ , and  $\Delta G_{\text{L}}$  represent the free energies of complex, receptor, and the ligand, respectively. The free energy ( $G$ ) of each state was calculated using the following equations:

$$\Delta G_{\text{bind}} = \Delta E_{\text{gas}} + \Delta G_{\text{sol}} - T\Delta S \quad (2)$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{vdW}} + E_{\text{ele}} \quad (3)$$

$$G_{\text{sol}} = G_{\text{GB}} + G_{\text{SA}} \quad (4)$$

$$G_{\text{SA}} = \gamma \text{SASA} \quad (5)$$

The FF14SB force field terms were used to estimate the gas phase energy ( $E_{\text{gas}}$ ), which is the sum of the internal energy ( $E_{\text{int}}$ ); Coulomb energy ( $E_{\text{ele}}$ ) and the van der Waals energies ( $E_{\text{vdw}}$ ). The energy contribution from the polar states ( $G_{\text{GB}}$ ) and non-polar states ( $G_{\text{SA}}$ ) were employed to evaluate the solvation free energy ( $\Delta G_{\text{sol}}$ ). The solvent accessible surface area (SASA in  $\text{\AA}^2$ ) was also used to derive the non-polar solvation energy ( $G_{\text{SA}}$ ) using a water probe radius of 1.4  $\text{\AA}$ , while the contribution from polar solvation ( $G_{\text{GB}}$ ) was determined by solving the Generalised Born equation, where the total entropy of the solute and temperature is represented by  $S$  and  $T$ , respectively. To obtain the contribution of each residue to the total binding free energy profile between the peptidomimetics and  $\text{M}^{\text{pro}}$ , per-residue free energy decomposition was carried out at the atomic level for imperative residues using the MM/GBSA method in Amber 18.

### *Molecular dynamic simulation*

Molecular dynamics simulation was carried out using GPU-enabled Desmond<sup>55-57</sup> engine implemented in Maestro v12. The complex corresponding to the top-scoring pose for each ligand or the un-complexed (Apo) protein was saved as a PDB file and submitted to the Linux (Ubuntu) computer for the Desmond high-performance molecular dynamics simulations studies. This study has two main phases; namely, system building (solvation and ionisation) and production. The System Builder module implemented in the Desmond algorithm was used to solvate the system using the TIP3P explicit solvent model with the OPLS\_2005 force field. The model was placed in an orthorhombic water box (distance from the box face to the outermost protein/ligand atom = 10  $\text{\AA}$ , box angle  $\alpha = \beta = \gamma = 90^\circ$ ). The box volume was minimised, and counter ions added to neutralise the system, making sure the ions are placed at least 20  $\text{\AA}$  from each ligand. The system was physiologically conditioned by adding 0.15 M NaCl into the solvent box. After the solvation and ionisation phase in the explicit solvent model was completed, the system was submitted to the molecular dynamics production phase. This phase of MD simulation is divided into eight distinct stages with specified parameters. The first seven stages involve the equilibration phase and is composed of short simulation steps. Step 8 is a final, long simulation stage. A total of 50 ns production stage was carried out. In the first stage, the type and parameters of the solvated system were detected. In stage 2, a 100 ps simulation was carried out using Brownian Dynamics under NVT conditions at 10 K, while

placing restraints on solute heavy atoms. Stage 3 involved a 12 ps simulation under NVT conditions at 10 K with restraints on heavy atoms. Stages 4, 6 and 7 (the pocket solvation at stage 5 was omitted) employed short simulation steps (12, 12 and 24 ps, respectively) under NPT conditions (at 10 K and restraints on heavy atoms for stages 4 and 6). No restraints were placed on heavy atoms at stage 7. The final production stage at constant temperature (300 K) was carried out at stage 8, for 50 ns.

### *Principal Component Analysis*

In order to further study the effect of ligand binding on dynamics of the C $\alpha$  atoms, we deconvoluted the C $\alpha$  RMSD using principal component analysis (PCA). This method enabled the interpretation of the relationship between different C $\alpha$  conformations of M<sup>Pro</sup> across the five simulated systems, which were sampled during the trajectory as implemented in the Bio3D R-Statistical package for comparative analysis of protein structures<sup>58</sup>. Briefly, the Desmond trajectory (.dtr) files were imported in VMD<sup>59</sup> and re-written as CHARMM/NAMD trajectory (.dcd) files. The dcd files together with the pdb file of the minimised structure were imported into RStudio and C $\alpha$  atoms for trajectory frames were superimposed to form a list containing numeric indices in  $x$ ,  $y$  and  $z$  coordinates used for the PCA analysis. Using the *pca.xyz* function implemented in the Bio3D package for R statistical analysis, a lower dimensional representation of the structural dataset of the simulated systems were obtained by projecting the minimised structure and snapshots from MD trajectories into the sub-space defined by the largest principal component (PC), which describes the largest C $\alpha$  atoms variance between the minimised structure and the MD trajectories.

## **Acknowledgements**

This work is supported by the University of the Witwatersrand, the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation (grant 64788 to I.A.). A.E.I. acknowledges the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation for supporting his post-doctoral fellowship studies. O.J.A. acknowledges the DELTAS Africa Initiative-SSACAB and the GlaxoSmithKline (GSK) Grant No. 107754/Z/15/Z-DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme for supporting her doctoral studies. The authors will like to acknowledge the Centre for High Performance Computing (CHPC), Cape Town, South Africa for providing access to Schrödinger Maestro v12 molecular modelling suite. YS acknowledges the South African Medical Research Council (SA MRC) under a Self-Initiated Research Grant and the South African National Research Foundation (NRF) via the Competitive Programme for Rated Researchers (CSR170428229183). The views and opinions expressed are those of the authors and do not necessarily represent the official views of the SA MRC.

## **Author contributions**

I.A conceived the study, performed the parts computational studies and wrote the manuscript. A.E.I performed MM/GBSA and per-residue energy decomposition analysis, O.J.A performed and interpreted the PCA, DCCM and other statistical analysis using the R statistical environment. Y.S. co-conceived the study. M.F assisted in setting up the Desmond molecular dynamics simulation engine and analysed the results. All authors reviewed the manuscript.

## **Competing interests**

The authors declare no competing interests.

## **Additional information**

Supplementary information accompanying this article can be found at <https://...>

## Figures and Tables

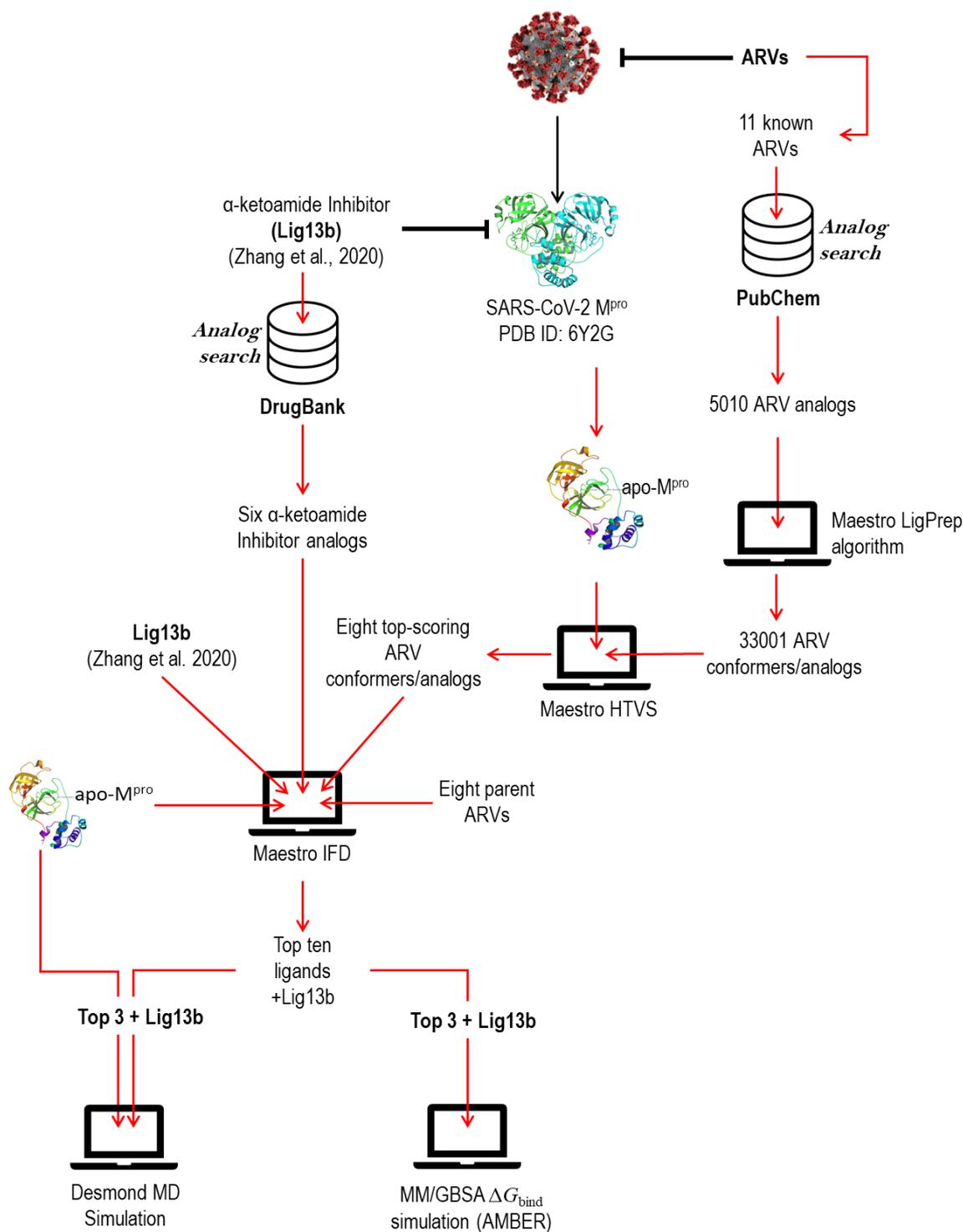


Figure 1. Conceptual framework of the study. The 11 known ARVs are (i) HIV PIs: Atazanavir, Darunavir, Fosamprenavir, Indinavir, Lopinavir, Ritonavir, Saquinavir and Tipranavir; (ii) NNRTIs: Delavirdine and Nevirapine; (iii) A broad spectrum antiviral: Remdesivir. A list of the top scoring conformers is detailed in Table S1. Six analogs of Lig13b<sup>21</sup> are Isavuconazonium, P2-P3 Ketoamide derivative ( $\alpha$ -KI), Pentagastrin, Bromocriptine, Cefotolozane and Cobicistat (see Table 1).

Table 1: Ligands that bound to M<sup>pro</sup> with overall Emodel energy lower than the benchmark ligand (Lig13b). Complexes of ligands in bold were taken further for MD simulation studies.

Ligand	PubChem CID	Glide Score (kJ/mol)	Glide Emodel (kJ/mol)
<b>Pentagastrin</b>	<b>9853654</b>	<b>-51.66</b>	<b>-700.54</b>
<b><math>\alpha</math>-KI</b>	<b>10129151</b>	<b>-46.10</b>	<b>-575.20</b>
<b>Isavuconazonium</b>	<b>6918606</b>	<b>-42.75</b>	<b>-507.17</b>
Bromocriptine	31101	-46.98	-470.11
Ceftolozane	53234134	-40.76	-459.58
Cobicistat	25151504	-41.30	-451.16
Delavirdine	5625	-40.28	-412.68
Atazanavir-analog	118983632	-37.05	-405.64
Fosamprenavir-analog	68322161	-36.02	-384.99
Nevirapine	71750992	-37.48	-383.47
<b>Lig13b</b>	<b>N/A</b>	<b>-32.13</b>	<b>-316.24</b>

Table 2: Binding-free energy and individual energy components calculated using MM/GBSA method for M<sup>pro</sup>-complex with Lig13b, Isavuconazonium,  $\alpha$ -KI and Pentagastrin

M <sup>pro</sup> -complex	Energy components (kJ/mol)				
	$\Delta E_{vdW}$	$\Delta E_{ele}$	$\Delta E_{gas}$	$\Delta E_{sol}$	$\Delta G_{bind}$
Lig13b	-155.66	-60.71	-216.37	98.68	-117.68
Isavuconazonium	-226.35	-1094.19	-1320.54	1129.43	-191.11
$\alpha$ -KI	-207.43	-78.39	-285.81	98.45	-187.36
Pentagastrin	-220.06	-196.45	-416.52	270.76	-145.76

Table 3: Projections of the distribution of the simulated systems onto the subspace defined by the largest principal components

Simulated systems	Variance in C $\alpha$ atomic fluctuations (%)		
	PC1	PC2	PC3-PC20
apo-M <sup>pro</sup>	23.7	15.1	<8.7
Lig13b-M <sup>pro</sup>	24.3	12.9	<9.0
$\alpha$ -KI-M <sup>pro</sup>	21.4	10.4	<9.0
Pentagastrin-M <sup>pro</sup>	21.7	13.0	<9.9
Isavuconazonium-M <sup>pro</sup>	34.0	13.5	<6.5

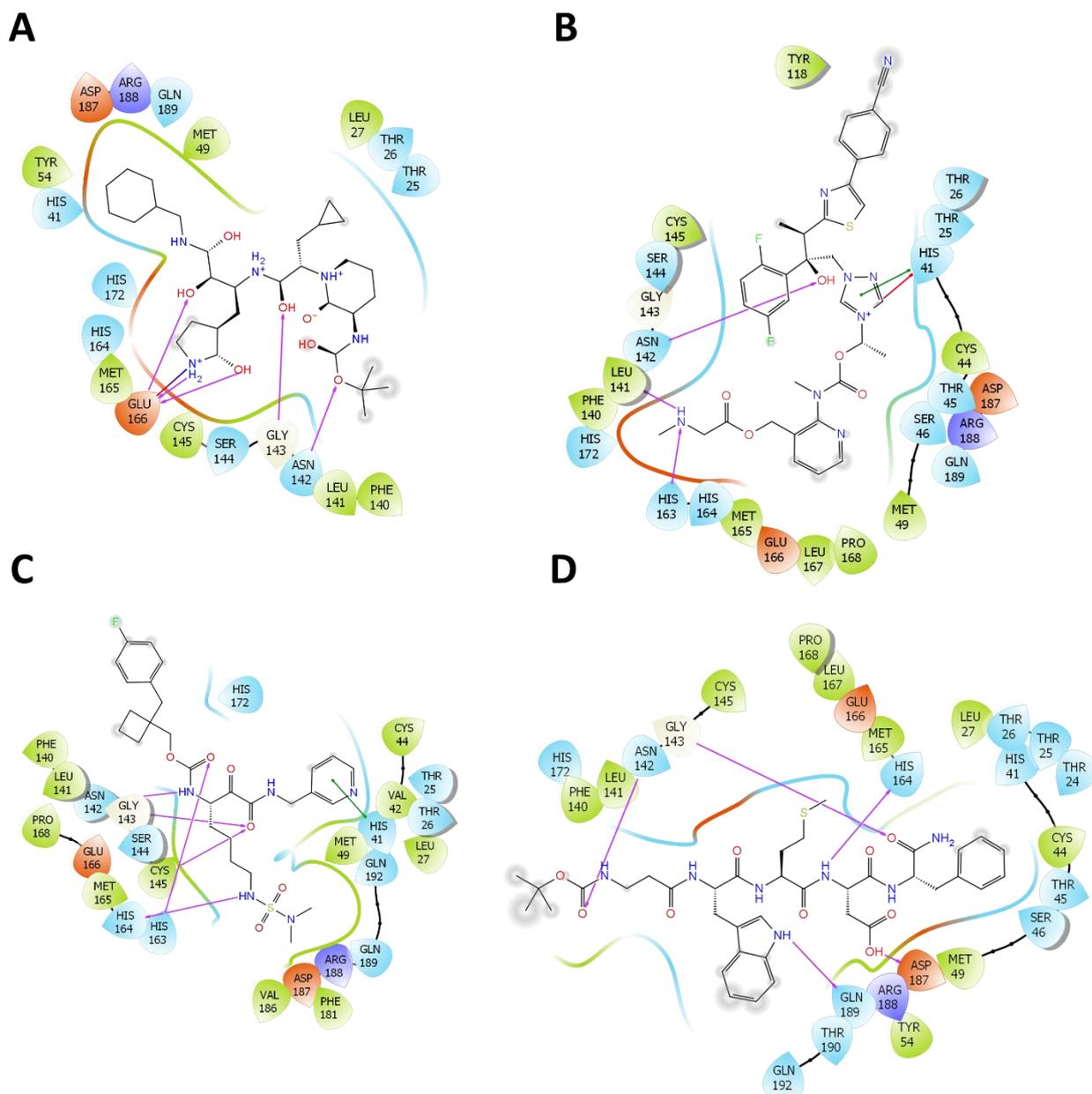


Figure 2: 2D interaction plot of top poses of (A) Lig13b, (B) Isavuconazonium (C)  $\alpha$ -KI and (D) Pentagastrin in complex with M<sup>Pro</sup> from induced-fit ligand docking. The residues are within 4 Å from the ligand. Amino acid side chains are represented as follows: green is non-polar; blue is polar; violet is positively charged; orange is negatively charged. Purple lines represent H-bond, blue-red line represent salt bridge interactions and grey shades represent exposure to solvent. The image was generated using Maestro 2D interaction diagram implemented in Schrödinger Maestro v12.

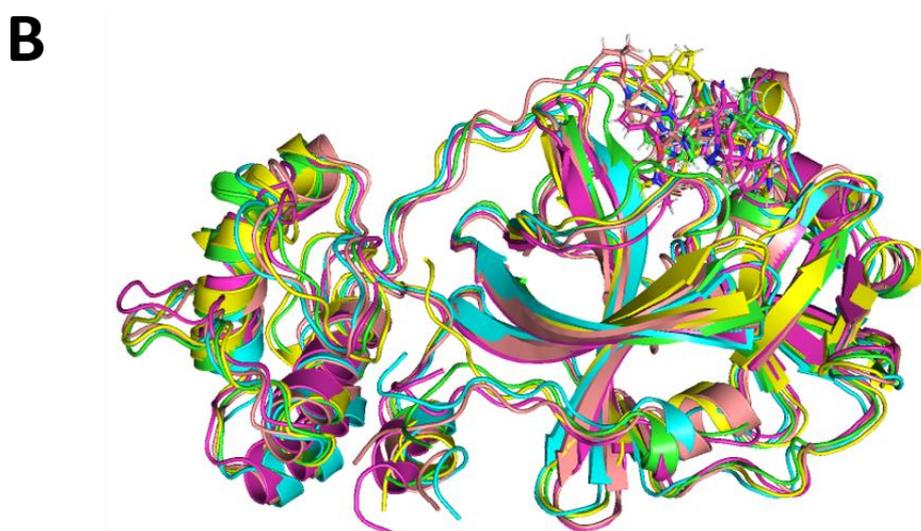
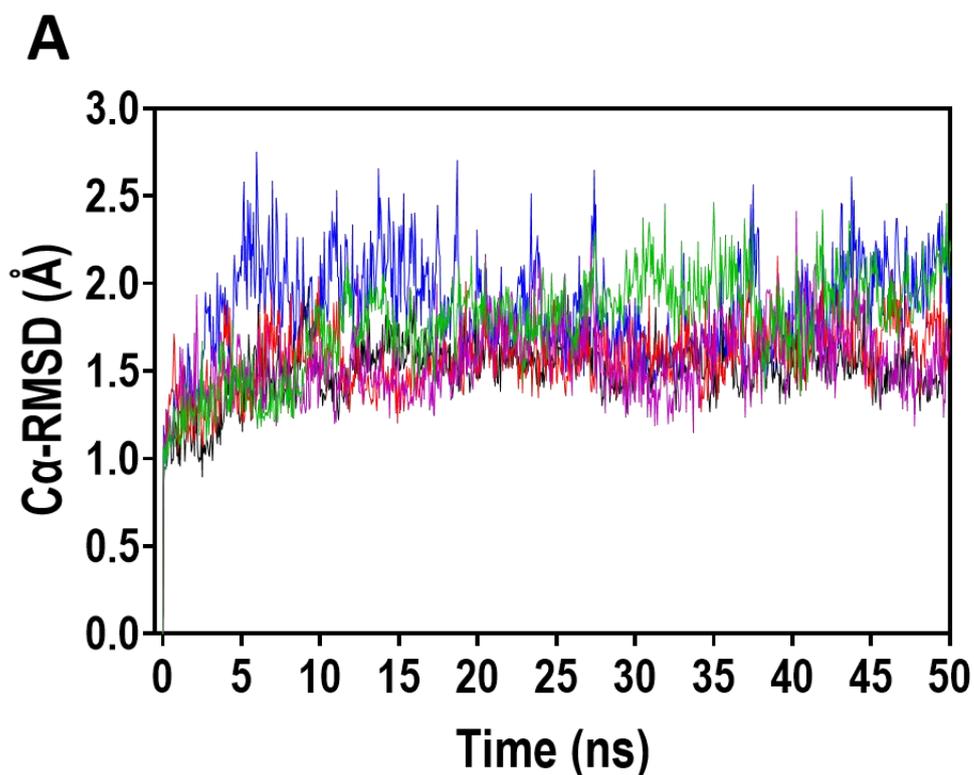


Figure 3: (A) The root-mean-square deviation (RMSD) of the  $C\alpha$  atoms as a function of 50 ns simulation time for the apo- $M^{pro}$  (black line), Lig13b- $M^{pro}$  complex (purple line),  $\alpha$ -KI- $M^{pro}$  complex (green line), Pentagastrin- $M^{pro}$  complex (red line) and Isavuconazonium- $M^{pro}$  complex (blue line). The image was generated using GraphPad Prism. (B) Superimposed images of representative ribbon structures of apo- $M^{pro}$  (cyan), Lig13b- $M^{pro}$  complex (green),  $\alpha$ -KI- $M^{pro}$  complex (yellow), Pentagastrin- $M^{pro}$  complex (brown) and Isavuconazonium- $M^{pro}$  complex (purple) taken from a trajectory cluster of the most abundant structures. The image was generated using PyMol. The ligands are shown in stick representations.

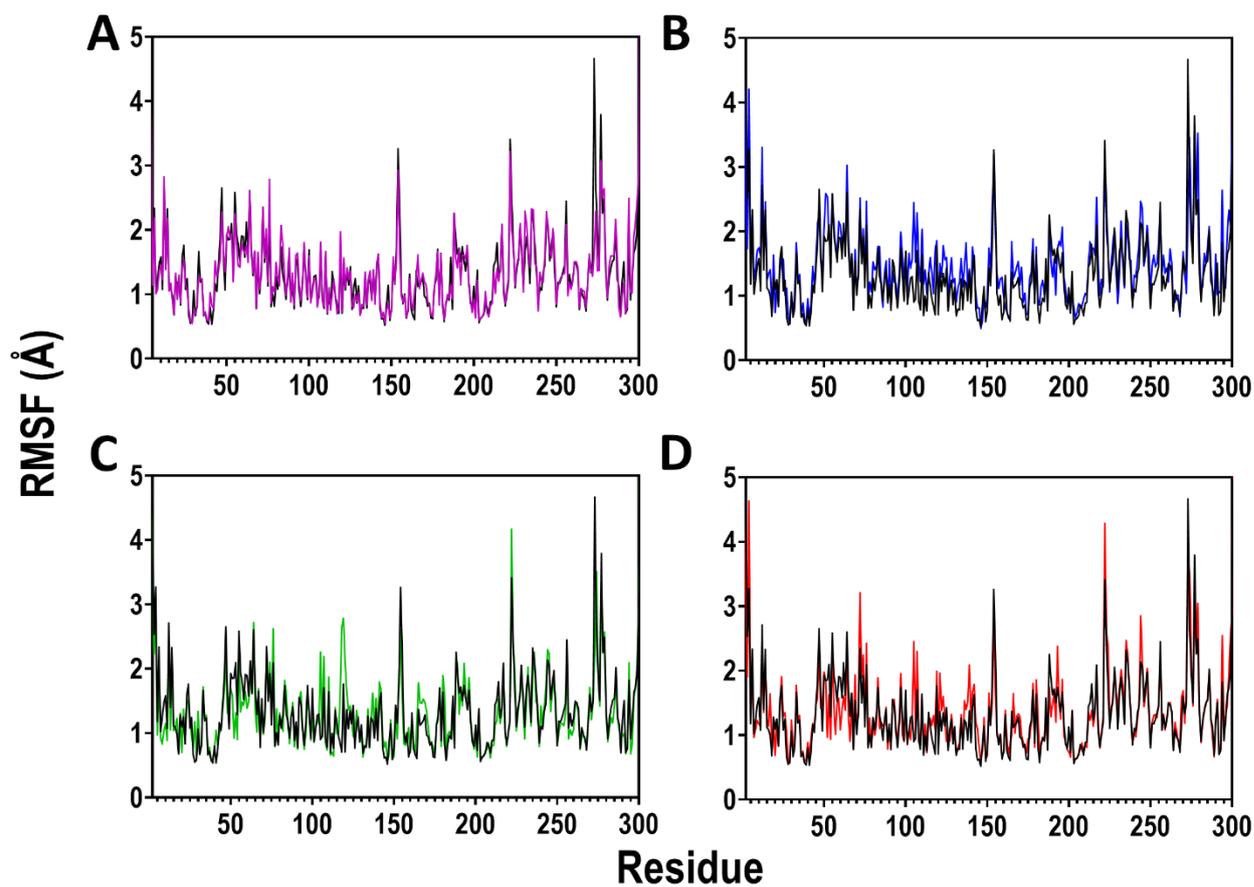


Figure 4: The root-mean-square fluctuation (RMSF) per residue over 50 ns simulation time for (A) apo-M<sup>pro</sup> (black line) in comparison with Lig13b-M<sup>pro</sup> complex (purple line), (B) apo-M<sup>pro</sup> (black line) comparison with Isavuconazonium-M<sup>pro</sup> complex (blue line), (C) apo-M<sup>pro</sup> (black line) comparison with  $\alpha$ -KI-M<sup>pro</sup> complex (green line) and (D) apo-M<sup>pro</sup> (black line) comparison with Pentagastrin-M<sup>pro</sup> complex (red line). The image was generated using GraphPad Prism.

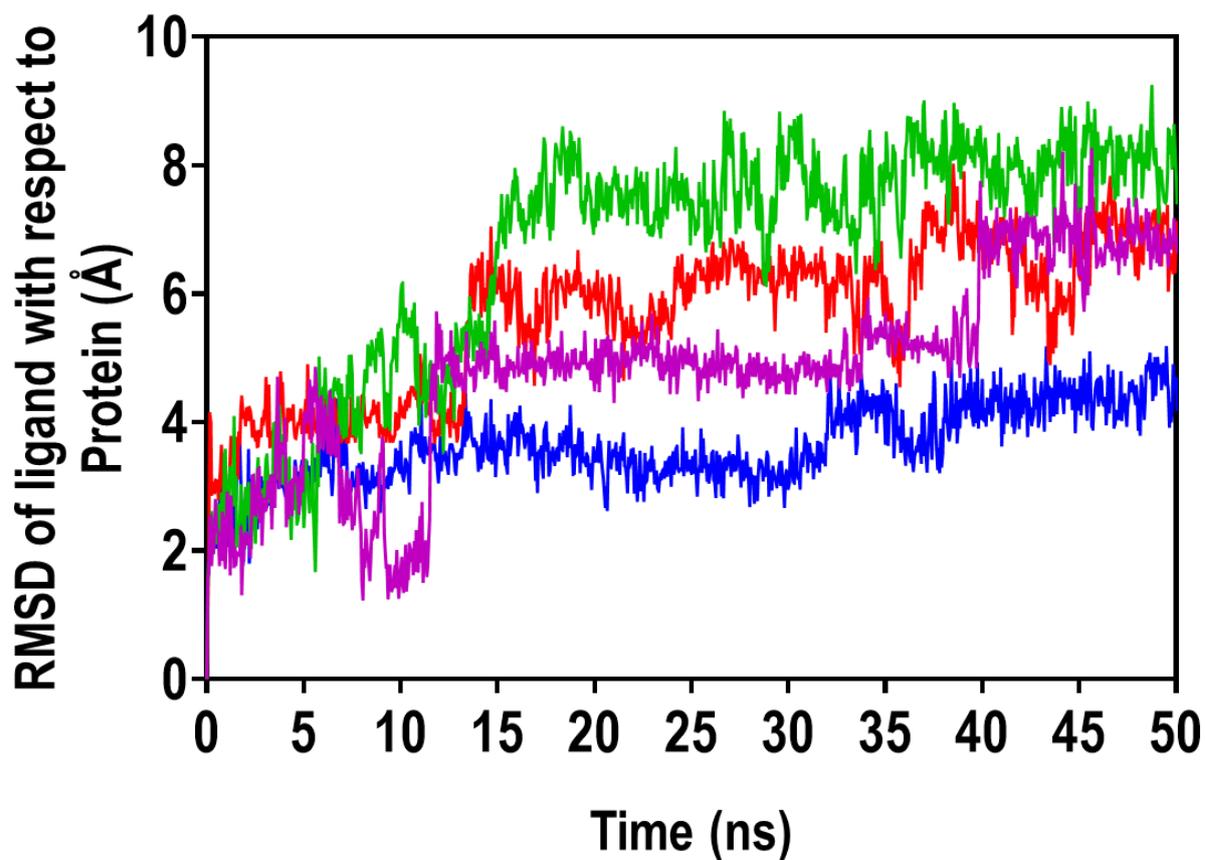


Figure 5: The root-mean-square deviation (RMSD) of the ligands with respect to the receptor ( $M^{pro}$ ) as a function of 50 ns simulation time for Lig13b- $M^{pro}$  complex (purple line),  $\alpha$ -KI- $M^{pro}$  complex (green line), Pentagastrin- $M^{pro}$  complex (red line) and Isavuconazonium- $M^{pro}$  complex (blue line). The image was generated using GraphPad Prism.

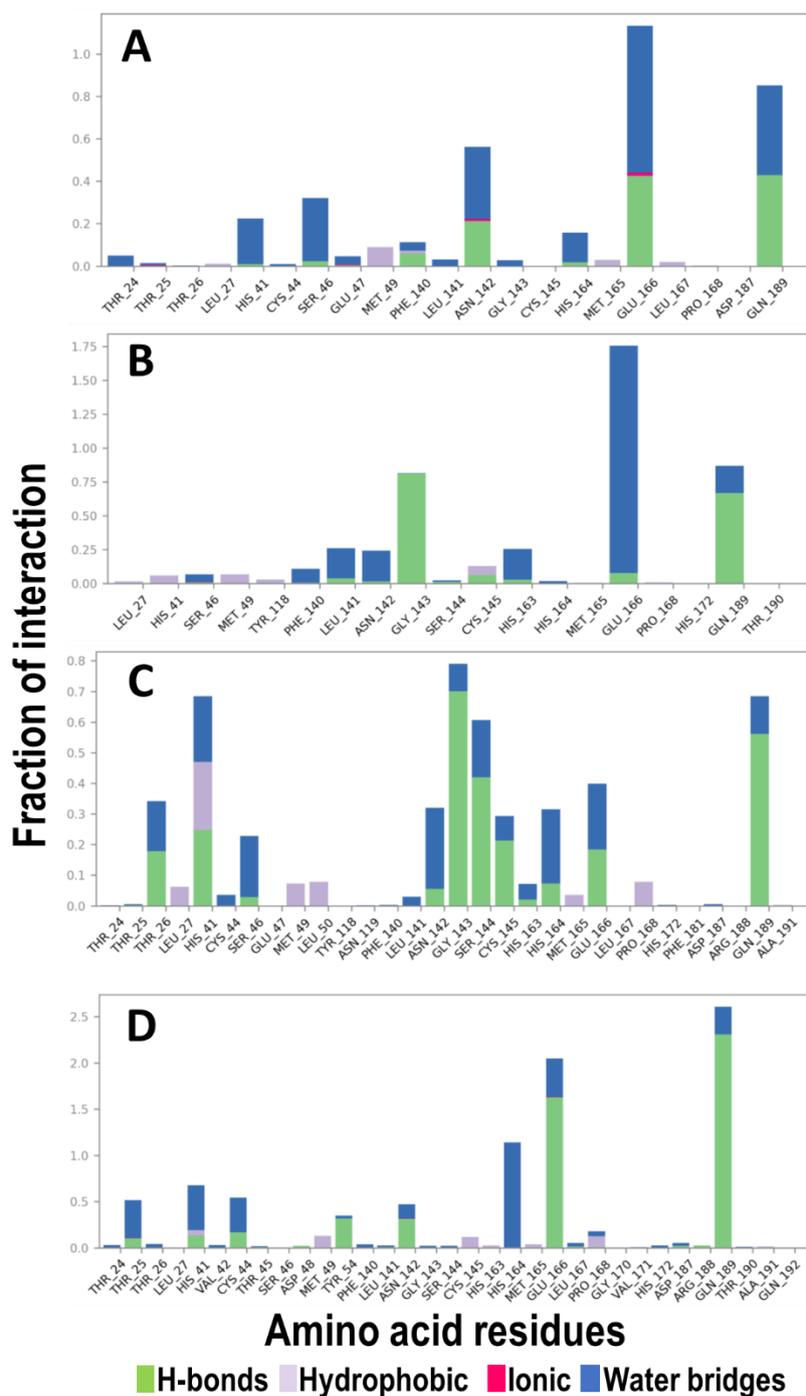


Figure 6: Stacked bar charts of side chains interactions and the types of interaction with (A) Lig13b, (B) Isavuconazonium, (C)  $\alpha$ -KI and (D) Pentagastrin monitored throughout the 50 ns simulation. Green bars represent a fraction of time for H-bond (categorised into backbone acceptor; backbone donor; side-chain acceptor; side-chain donor). Hydrophobic interactions are categorised into  $\pi$ -cation;  $\pi$ - $\pi^*$ ; and other, non-specific interactions. The stacked bar charts are normalised throughout the 50 ns trajectory. Therefore, the fraction of interaction can be construed to be percentage time, which could be more than 100%.



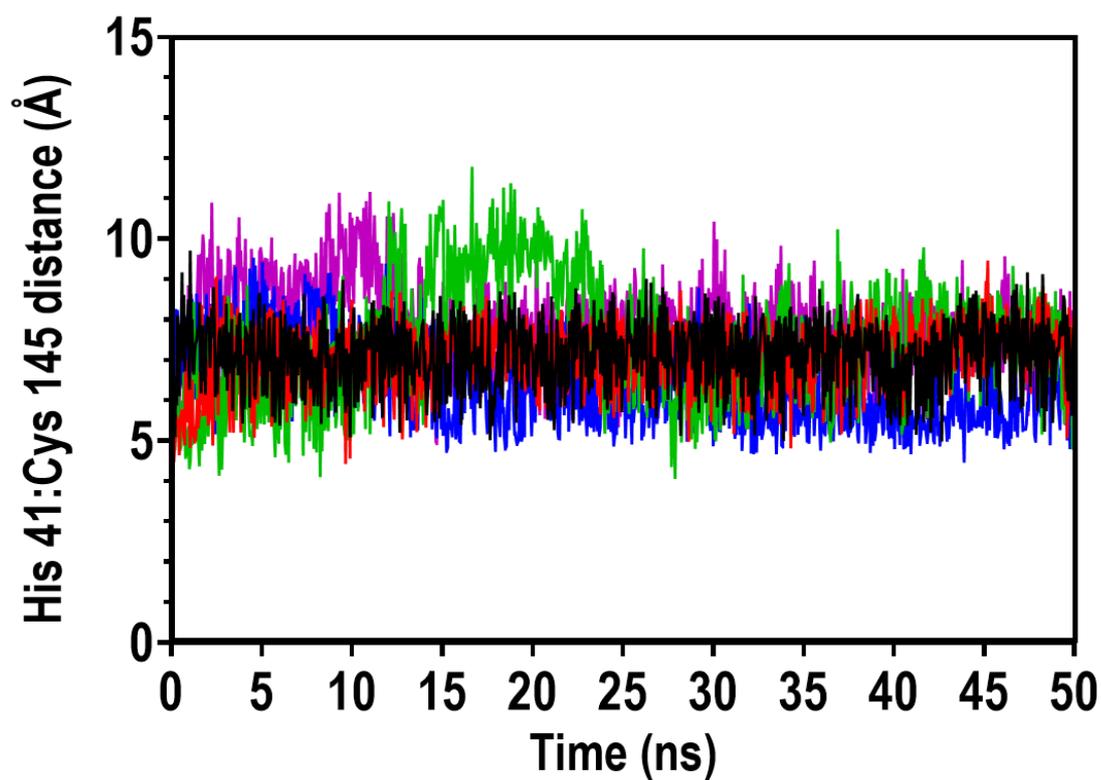


Figure 8: Atomic distance between His 41 and Cys 145 as a function of 50 ns simulation time for the apo-M<sup>pro</sup> (black line), Lig13b-M<sup>pro</sup> complex (purple line), α-KI-M<sup>pro</sup> complex (green line), Pentagastrin-M<sup>pro</sup> complex (red line) and Isavuconazonium-M<sup>pro</sup> complex (blue line). The image was generated using GraphPad Prism.

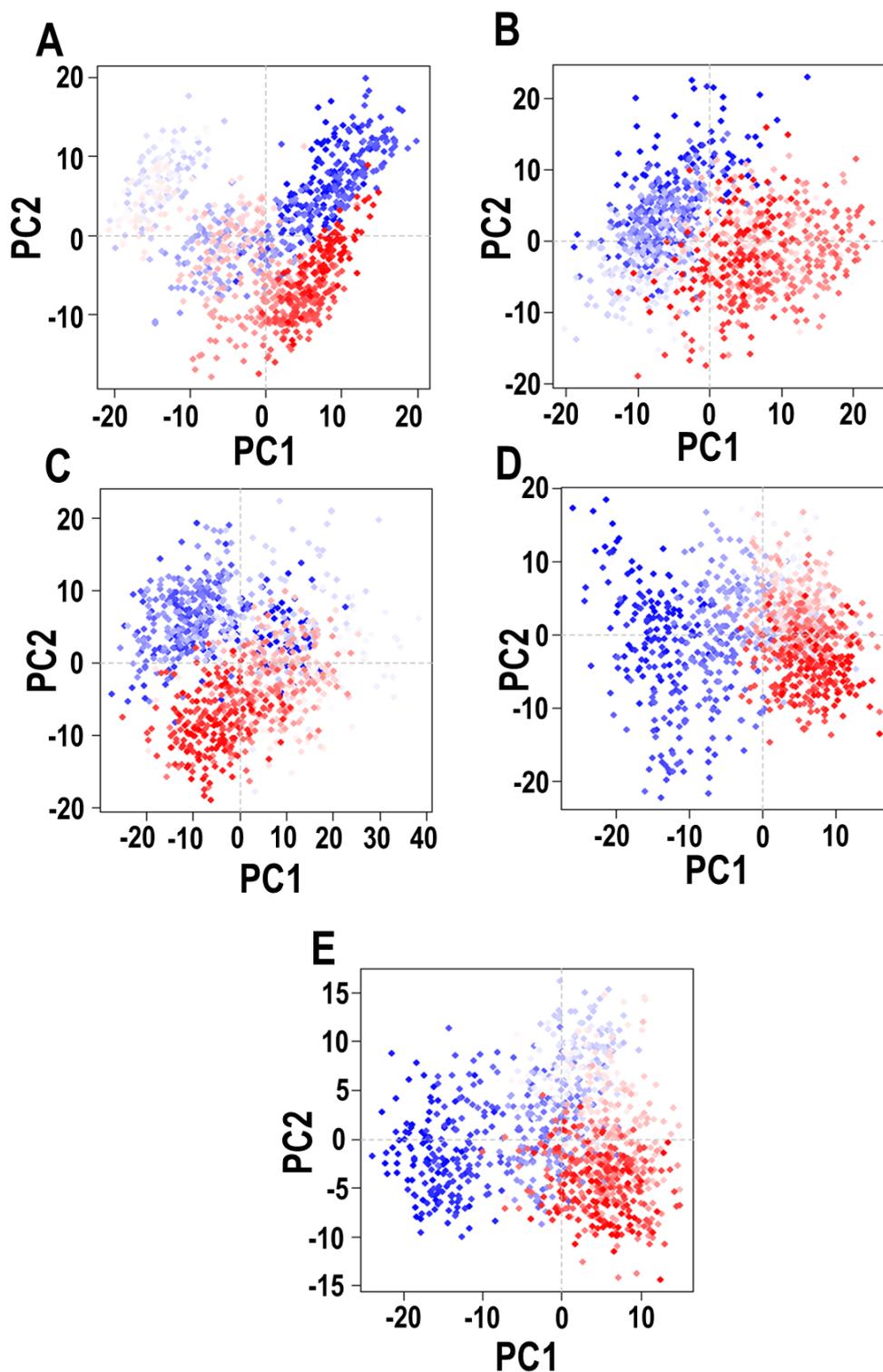


Figure 9: Projections of trajectories onto the subspace by the first two principal components (PC1 and PC2) for (A) apo-M<sup>pro</sup>, (B) Lig13b-M<sup>pro</sup> complex, (C) Isavuconazonium-M<sup>pro</sup>, (D)  $\alpha$ -KI-M<sup>pro</sup> complex and (E) Pentagastrin-M<sup>pro</sup> complex simulated systems. The two distinct conformations along the trajectory are coloured blue (the start) and red (the end). Between these two conformations are the intermediate states (coloured from the lower shade of blue to grey, white and lower shade of red).

## References

- 1 Gorbalenya, A. E. Severe acute respiratory syndrome-related coronavirus–The species and its viruses, a statement of the Coronavirus Study Group. *BioRxiv* (2020).
- 2 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature* **579**, 270-273 (2020).
- 3 Ayittey, F. K., Dzuovor, C., Ayittey, M. K., Chiwero, N. B. & Habib, A. Updates on Wuhan 2019 novel coronavirus epidemic. *Journal of Medical Virology* **92**, 403 (2020).
- 4 Sohrabi, C. *et al.* World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery* (2020).
- 5 Jeffs, B. A clinical guide to viral haemorrhagic fevers: Ebola, Marburg and Lassa. *Tropical doctor* **36**, 1-4 (2006).
- 6 Jarynowski, A., Wójta-Kempa, M., Płatek, D. & Czopek, K. Attempt to understand public health relevant social dimensions of COVID-19 outbreak in Poland. *Available at SSRN 3570609* (2020).
- 7 Colson, P., Rolain, J.-M., Lagier, J.-C., Brouqui, P. & Raoult, D. Chloroquine and hydroxychloroquine as available weapons to fight COVID-19. *Int J Antimicrob Agents* **105932** (2020).
- 8 Cortegiani, A., Ingoglia, G., Ippolito, M., Giarratano, A. & Einav, S. A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. *Journal of critical care* (2020).
- 9 Delvecchio, R. *et al.* Chloroquine, an endocytosis blocking agent, inhibits Zika virus infection in different cell models. *Viruses* **8**, 322 (2016).
- 10 Devaux, C. A., Rolain, J.-M., Colson, P. & Raoult, D. New insights on the antiviral effects of chloroquine against coronavirus: what to expect for COVID-19? *International Journal of Antimicrobial Agents*, 105938 (2020).
- 11 Gao, J., Tian, Z. & Yang, X. Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. *Bioscience trends* (2020).
- 12 Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell research* **30**, 269-271 (2020).
- 13 Te Velthuis, A. J. *et al.* Zn<sup>2+</sup> inhibits coronavirus and arterivirus RNA polymerase activity in vitro and zinc ionophores block the replication of these viruses in cell culture. *PLoS pathogens* **6** (2010).
- 14 Al-Tawfiq, J. A., Al-Homoud, A. H. & Memish, Z. A. Remdesivir as a possible therapeutic option for the COVID-19. *Travel Med Infect Dis* **101615** (2020).
- 15 Gordon, C. J., Tchesnokov, E. P., Feng, J. Y., Porter, D. P. & Götte, M. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *Journal of Biological Chemistry* **295**, 4773-4779 (2020).
- 16 Molina, J. M. *et al.* No evidence of rapid antiviral clearance or clinical benefit with the combination of hydroxychloroquine and azithromycin in patients with severe COVID-19 infection. *Med Mal Infect*, 30085-30088 (2020).
- 17 Batool, M., Ahmad, B. & Choi, S. A structure-based drug discovery paradigm. *International journal of molecular sciences* **20**, 2783 (2019).
- 18 Chang, Y.-C. *et al.* Potential therapeutic agents for COVID-19 based on the analysis of protease and RNA polymerase docking. *Preprints* **2020020242**, doi:doi:10.20944/preprints202002.0242.v2 (2020).
- 19 Kumar, Y. & Singh, H. In silico identification and docking-based drug repurposing against the main protease of SARS-CoV-2, causative agent of COVID-19. *ChemRxiv* (2020).
- 20 Exscientia. Celgene and Exscientia enter 3-year AI drug discovery collaboration focused on accelerating drug discovery in oncology and autoimmunity. (2019).
- 21 Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* (2020).

- 22 Dunn, B. M., Goodenow, M. M., Gustchina, A. & Wlodawer, A. Retroviral proteases. *Genome biology* **3**, reviews3006. 3001 (2002).
- 23 Konvalinka, J., Kräusslich, H.-G. & Müller, B. Retroviral proteases and their roles in virion maturation. *Virology* **479**, 403-417 (2015).
- 24 Kräusslich, H.-G. & Wimmer, E. Viral proteinases. *Annual review of biochemistry* **57**, 701-754 (1988).
- 25 Wishart, D. S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research* **36**, D901-D906 (2008).
- 26 Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research* **34**, D668-D672 (2006).
- 27 Kim, S. *et al.* PubChem substance and compound databases. *Nucleic acids research* **44**, D1202-D1213 (2016).
- 28 Li, Q., Cheng, T., Wang, Y. & Bryant, S. H. PubChem as a public resource for drug discovery. *Drug discovery today* **15**, 1052-1057 (2010).
- 29 Kumar, S., Zhi, K., Mukherji, A. & Gerth, K. Repurposing Antiviral Protease Inhibitors Using Extracellular Vesicles for Potential Therapy of COVID-19. *Viruses* **12**, 486 (2020).
- 30 Bajorath, J. Rational drug discovery revisited: interfacing experimental programs with bio- and chemo-informatics. *Drug discovery today* **6**, 989-995 (2001).
- 31 Braiuca, P., Ebert, C., Basso, A., Linda, P. & Gardossi, L. Computational methods to rationalize experimental strategies in biocatalysis. *TRENDS in Biotechnology* **24**, 419-425 (2006).
- 32 Yu, H. & Adedoyin, A. ADME-Tox in drug discovery: integration of experimental and computational technologies. *Drug discovery today* **8**, 852-861 (2003).
- 33 Blundell, T. L. Structure-based drug design. *Nature* **384**, 23 (1996).
- 34 Blundell, T. L., Jhoti, H. & Abell, C. High-throughput crystallography for lead discovery in drug design. *Nature Reviews Drug Discovery* **1**, 45-54 (2002).
- 35 Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technologies* **1**, 337-341 (2004).
- 36 Amidon, G. L. & Lee, H. Absorption of peptide and peptidomimetic drugs. *Annual review of pharmacology and toxicology* **34**, 321-341 (1994).
- 37 Dahlin, J. L., Baell, J. & Walters, M. A. in *Assay Guidance Manual [Internet]* (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2015).
- 38 Thorne, N., Auld, D. S. & Inglese, J. Apparent activity in high-throughput screening: origins of compound-dependent assay interference. *Current opinion in chemical biology* **14**, 315-324 (2010).
- 39 Rishton, G. M. Reactive compounds and in vitro false positives in HTS. *Drug discovery today* **2**, 382-384 (1997).
- 40 Kandeel, M. & Al-Nazawi, M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life sciences*, 117627 (2020).
- 41 Wang, J., Loveland, A. N., Kattenhorn, L. M., Ploegh, H. L. & Gibson, W. High-molecular-weight protein (pUL48) of human cytomegalovirus is a competent deubiquitinating protease: mutant viruses altered in its active-site cysteine or histidine are viable. *Journal of virology* **80**, 6003-6012 (2006).
- 42 Husain, S. & Lowe, G. Evidence for histidine in the active site of papain. *Biochemical Journal* **108**, 855-859 (1968).
- 43 Mladenovic, M., Schirmeister, T., Thiel, S., Thiel, W. & Engels, B. The importance of the active site histidine for the activity of epoxide- or aziridine-based inhibitors of cysteine proteases. *ChemMedChem: Chemistry Enabling Drug Discovery* **2**, 120-128 (2007).
- 44 Banerjee, R. *et al.* Four plasmepsins are active in the Plasmodium falciparum food vacuole, including a protease with an active-site histidine. *Proceedings of the National Academy of Sciences* **99**, 990-995 (2002).

- 45 Sherman, W., Beard, H. S. & Farid, R. Use of an induced fit receptor structure in virtual screening. *Chemical biology & drug design* **67**, 83-84 (2006).
- 46 Koska, J. r. *et al.* Fully automated molecular mechanics based induced fit protein– ligand docking method. *Journal of chemical information and modeling* **48**, 1965-1973 (2008).
- 47 Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *Journal of medicinal chemistry* **49**, 534-553 (2006).
- 48 Matsumoto, K. *et al.* Structural basis of inhibition of cysteine proteases by E-64 and its derivatives. *Peptide Science* **51**, 99-107 (1999).
- 49 Roush, W. R. *et al.* Vinyl sulfonate esters and vinyl sulfonamides: potent, irreversible inhibitors of cysteine proteases. *Journal of the American Chemical Society* **120**, 10994-10995 (1998).
- 50 Copeland, R. A. The drug–target residence time model: a 10-year retrospective. *Nature Reviews Drug Discovery* **15**, 87 (2016).
- 51 Copeland, R. A. Conformational adaptation in drug–target interactions and residence time. *Future medicinal chemistry* **3**, 1491-1501 (2011).
- 52 Kollman, P. A. *et al.* Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research* **33**, 889-897 (2000).
- 53 Massova, I. & Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in drug discovery and design* **18**, 113-135 (2000).
- 54 Tsui, V. & Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers: Original Research on Biomolecules* **56**, 275-291 (2000).
- 55 Bergdorf, M., Kim, E. T., Rendleman, C. A. & Shaw, D. E. Desmond/GPU Performance as of November 2014. *DE Shaw Research Technical Report DESRES/TR—2014-01* (2014).
- 56 Chow, E. *et al.* Desmond performance on a cluster of multicore processors. *DE Shaw Research Technical Report DESRES/TR--2008-01* (2008).
- 57 Release, S. 1: Desmond molecular dynamics system, version 3.7. *DE Shaw Research, New York, NY, Maestro-Desmond Interoperability Tools, version 3* (2014).
- 58 Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. & Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695-2696 (2006).
- 59 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33-38 (1996).