





Proposition of a new allosteric binding site for potential SARS-CoV-2 3CL protease inhibitors by utilizing molecular dynamics simulations and ensemble docking

Jurica Novak^{a*} , Hrvoje Rimac^{b*} , Shivananda Kandagalla^a , Prateek Pathak^a , Vladislav Naumovich^a,
Maria Grishina^a  and Vladimir Potemkin^a 

^aLaboratory of Computational Modeling of Drugs, Higher Medical and Biological School, South Ural State University, Chelyabinsk, Russia;

^bDepartment of Medicinal Chemistry, Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia

Communicated by Ramaswamy H. Sarma

ABSTRACT

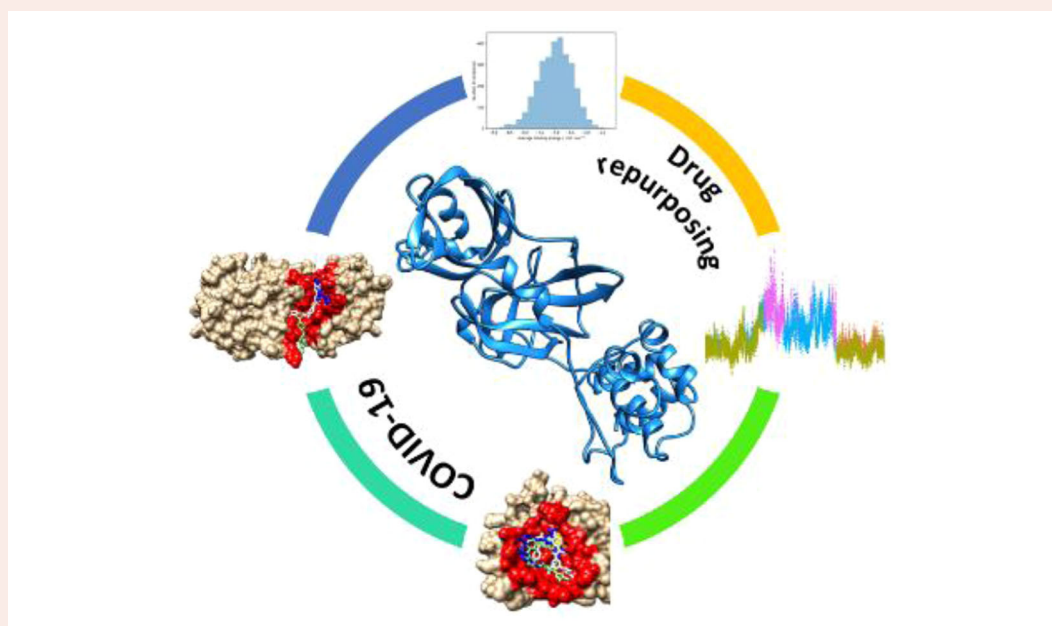
The SARS-CoV-2 3CL protease (3CLpro) shows a high similarity with 3CL proteases of other beta-coronaviruses, such as SARS and MERS. It is the main enzyme involved in generating various non-structural proteins that are important for viral replication and is one of the most important proteins responsible for SARS-CoV-2 virulence. In this study, we have conducted an ensemble docking of molecules from the DrugBank database using both the crystallographic structure of the SARS-CoV-2 3CLpro, as well as five conformations obtained after performing a cluster analysis of a 300 ns molecular dynamics (MD) simulation. This procedure elucidated the inappropriateness of the active site for non-covalent inhibitors, but it has also shown that there exists an additional, more favorable, allosteric binding site, which could be a better target for non-covalent inhibitors, as it could prevent dimerization and activation of SARS-CoV-2 3CLpro. Two such examples are radotinib and nilotinib, tyrosine kinase inhibitors already in use for treatment of leukemia and which binding to the newly found allosteric binding site was also confirmed using MD simulations.

ARTICLE HISTORY





Received 26 January 2021
Accepted 5 May 2021

KEYWORDS

3CL protease; molecular dynamics; SARS-CoV-2; radotinib; nilotinib



Abbreviations: 3CLpro: 3-chymotrypsin-like protease; COVID-19: Coronavirus disease 2019; MD: molecular dynamics; MERS-CoV: Middle East respiratory syndrome coronavirus; MM/GBSA: Molecular Mechanics/Generalized Born Surface Area; PLpro: Papain-like protease; RMSD: Root-mean-square deviation; RNA: ribonucleic acid; SARS-CoV: Severe acute respiratory syndrome coronavirus; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2

CONTACT Hrvoje Rimac  hrvoje.rimac@pharma.unizg.hr  Department of Medicinal Chemistry, Faculty of Pharmacy and Biochemistry, University of Zagreb, Ante Kovacica 1, Zagreb, 10000 Croatia; Shivananda Kandagalla  kandagallas@susu.ru  Laboratory of Computational Modeling of Drugs, Higher Medical and Biological School, South Ural State University, Tchaikovsky str. 20-A, Chelyabinsk, 454008 Russia

*Equally contributing authors.

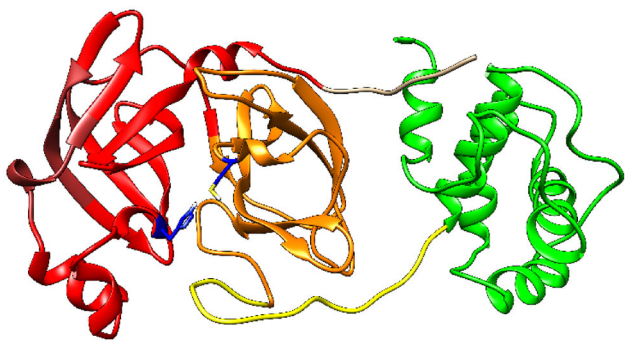


Figure 1. Structure of SARS-CoV-2 3CLpro (PDB ID: 6LU7). The *N*-finger (residues 1–7) is depicted in tan, domain I (residues 8–101) in red, domain II (residues 102–184) in orange, the loop region (residues 185–200) in yellow, domain III (residues 201–306) in green, and the conserved His 41 and Cys 145 in dark blue.

Introduction

Coronaviruses are a class of single-stranded positive-sense RNA viruses with a large viral RNA genome (Y. Chen et al., 2020). The SARS-CoV-2 is classified as a beta-coronavirus and has a highly similar genomic organization as other beta-coronaviruses (<80% nucleotide identity and 89.10% nucleotide similarity with SARS-CoV genes) (Wu et al., 2020; Zhu et al., 2020). A recent genome annotation of SARS-CoV-2 identified 14 orfs (open reading frames i.e. continuous stretches of codons that have the ability of being translated) encoding for 27 proteins. Typically, beta-coronaviruses produce a polypeptide (~800 kDa) upon transcription of the genome. This polypeptide is proteolytically cleaved by a papain-like protease (PLpro) and the 3-chymotrypsin-like protease (3CLpro). The 3CLpro cleaves the polyprotein at 11 distinct sites to generate various non-structural proteins that are important for the viral replication (Anand et al., 2003). Thus, this main protease is required for the maturation of coronaviruses and is vital for the viral life cycle, making it an attractive target for anti-SARS-CoV-2 inhibitors.

Further sequence comparison of the 3CLpro protein with its closest homologs shows that the SARS-CoV-2 3CLpro shares a sequence identity of 99.02% with bat SARS-like coronaviruses. It also shows a sequence identity with SARS-CoV (96%), MERS-CoV (87%), human-CoV (90%) and bovine-CoV (90%) homologs. All these reports mean that the 3CLpro is a highly conserved enzyme and a good target for anti-viral drugs (Xu et al., 2020; Zhou et al., 2020; Zhu et al., 2020).

Recent availability of the SARS-CoV-2 3CLpro crystal structure (PDB ID: 6LU7) confirms its high structural similarity with the SARS-CoV 3CLpro (PDB ID: 1UJ1) (Yang et al., 2003). The 3CLpro protomers of both SARS-CoV-2 and SARS-CoV contain three distinct domains i.e. domain I, domain II, and domain III, which are together comprised of nine α -helices and 13 β -strands (Figure 1). Similar to the other corona proteases, domains I (residues 8–101) and II (residues 102–184) include an antiparallel β -sheet structure with 13 β -strands, which resemble trypsin-like serine proteases structures. Domain III (the C-terminal domain, residues 201–306) consists of five α -helices and is connected to domain II by a long loop (residues 185–200). The main substrate binding site is formed by a cleft between domains I and II, and has a catalytic dyad

composed of conserved residues His 41 and Cys 145. Domains I and II are collectively referred to as the *N*-terminal domain. Amino acid residues 1–7 in the *N*-terminus form the *N*-finger which plays a significant role in the dimerization and the formation of the active 3CLpro dimer (Sang et al., 2020).

Through enzyme activity measurements and molecular dynamics simulations, H. Chen et al. (2006) determined that a SARS 3CLpro monomer is unable to establish a normal enzymatic activity and that only one protomer in the homodimer is active. Additionally, Shi and Song (Shi & Song, 2006) identified four regions associated with 3CLpro dimerization: residues 1–7 from the *N*-terminus forming the *N*-finger (1), the Asn 214 residue (2), the region around residues Glu 288 – Asp 289 – Glu 290 in a close contact with the *N*-finger (3), and the C-terminus' last helix region around residues Arg 298 – Gln 299 (4). Due to their high similarity, this behavior is also expected for SARS-CoV-2.

Drug repurposing (or drug repositioning) strategy includes various data-driven and experimental procedures for the identification of new uses for approved or investigational drugs that are outside the scope of the original medicinal indication (Ashburn & Thor, 2004; Pushpakom et al., 2019; Vanhaelen, 2019). The emergence of the SARS-CoV-2 virus motivated both science and pharmaceutical communities to speed up drug discovery against SARS-CoV-2 by applying drug repurposing approaches (Harrison, 2020). According to clinicaltrials.gov (<https://clinicaltrials.gov/ct2/home>), on March 30, 2021, 5193 studies were underway across the world, many of them trying to repurpose currently available anti-viral drugs. One of these clinical trials (registration number: ChiCTR2000029603) is testing the suitability of HIV-1 protease inhibitors, namely ASC09/ritonavir and lopinavir/ritonavir cocktails. In the initial epicenter of the epidemic, in Wuhan, China, another clinical trial (registration number: ChiCTR2000029541) is running, testing the appropriateness of darunavir/cobicistat and lopinavir/ritonavir combined with thymosin α 1. Based on pre-clinical studies in SARS-CoV and MERS-CoV infections, two clinical trials (identifiers: NCT04252664 and NCT04257656) are testing potential of remdesivir drug as a potential antiviral therapy for COVID-2019.

In computational studies, 3CLpro seems to be a very attractive target. Nukoolkarn et al. (Nukoolkarn et al., 2008) performed molecular dynamics simulations for the SARS-CoV 3CLpro free enzyme and its complexes with lopinavir and ritonavir. They discovered that complex intermolecular interactions when the inhibitors are bound to the proteinase active site result in enzyme's flap closing. Several different groups have also performed virtual screening studies targeting this protein by using various ligand libraries and using both non-covalent and covalent docking methods, as well as molecular dynamics (MD) simulations (Alamri et al., 2020; Bhardwaj et al., 2020, 2021; Elmezayen et al., 2021; Gyebi et al., 2020; Koulgi et al., 2020). Li et al. (Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening, unpublished data), aside from identifying drugs with high binding capacity to the SARS-CoV main protease

(prulifloxacin, bictegravir, nelfinavir, and tegobuvir), identified three binding sites: binding site of a natural substrate, a site in the groove between the two monomers and a small pocket near the C-terminus. Additionally, Novak et al. performed virtual screening of molecules from the Natural Product Atlas and molecular dynamics simulations of the most potent 3CLpro inhibitor, demonstrating a similar binding potential for a variety of inhibitors for the catalytic and the allosteric groove binding sites (Novak et al., 2021). An alternative way of choosing a virtual screening target is to identify the most important virus-host interactions and try to find ligands which could bind to their interface and disrupt these interactions (Gollapalli et al., 2020). To summarize the current progress, Dotolo et al. published an extensive review on computational drug-repurposing for SARS-CoV-2, with its advantages and drawbacks (Dotolo et al., 2021).

The recent availability of the SARS-CoV-2 3CLpro 3D structure provides a deep insight into the viral life cycle and aids the screening of anti-COVID-19 drugs. In this concern, the present work identified potential drug candidates against 3CLpro protease through virtual screening, but more importantly it showed a presence of an additional druggable site on the enzyme. Screening for compounds which bind to this binding site could identify drugs which would be missed if only binding to the active site would be performed.

Materials and methods

Molecular dynamics (MD) simulations of the free enzyme

The SARS-CoV-2 3CLpro structure was obtained from RCSB Protein Data Bank (code 6LU7). After removing the ligand and water molecules and visual inspection of amino acid residues, the protein was prepared for MD simulations to obtain its different conformations. The AMBER ff14SB force field was used, and the protein was solvated in a truncated octahedral box of TIP3P water molecules spanning a 12 Å thick buffer. The 3CLpro was neutralized by Na⁺ ions and submitted to geometry optimization in the AMBER16 program, employing periodic boundary conditions in all directions (Case et al., 2016). For the first 1500 cycles the complex was restrained and only water molecules were optimized, after which another 2500 cycles of optimization followed, where both water molecules and complex were unrestrained. Optimized systems were gradually heated from 0 K to 300 K and equilibrated during 30 ps using NVT conditions, followed by a productive and unconstrained MD simulations of 300 ns employing a time step of 2 fs at a constant pressure (1 atm) and temperature (300 K). The latter was held constant using the Langevin thermostat with a collision frequency of 1 ps⁻¹. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm (Ryckaert et al., 1977), while the long-range electrostatic interactions were calculated employing the Particle Mesh Ewald method (Darden et al., 1993). The non-bonded interactions were truncated at 11.0 Å. After the simulation, the *k*-means cluster analysis was performed using CPPTRAJ (Roe & Cheatham, 2013). Enzyme structures were clustered into five groups based on RMSD in backbone

atoms of amino acids located within 5.5 Å of the Cys 145 residue (Thr 24, Thr 25, Thr 26, Leu 27, His 41, Met 49, Tyr 54, Phe 140, Leu 141, Asn 142, Gly 143, Ser 144, Cys 145, His 163, His 164, Met 165, Glu 166, Leu 167, Pro 168, His 172, Asp 187, Arg 188, Gln 189, Thr 190, Ala 191, and Gln 192), with the maximal number of iterations set to 500, randomized initial set of points used and sieving set to 10. After clustering, frames closest to the centroids of each cluster were identified as five different conformations and were used as receptors in the second round of docking.

Molecular docking

First, a database of drugs was obtained from DrugBank (Release Version 5.1.5, <https://www.drugbank.ca/>) After cross-checking the collected drug molecules for redundancies, an sdf file containing 8756 potential ligands was procured. All structures were then converted to the pdb format using the Open Babel 3.0.0. program (O'Boyle et al., 2011) and were prepared for non-covalent docking using the AutoDockTools 4 (Morris et al., 2009) program's script `prepare_ligand4.py` and saved in the pdbqt format. Before docking, the crystallographic inhibitor molecule was removed, hydrogens were added where necessary, all Lys, Arg, His, and Cys side chains were protonated, all Asp and Glu side chains were deprotonated, and the amino and carboxy termini were charged. This molecule was then saved in the pdbqt format. All dockings were performed using AutoDock Vina (Handoko et al., 2012) locally on 6 personal computers with 8 Intel® Core™ i7-6700K CPU @ 4.00 GHz, 32 GB RAM, and the 64-bit Windows 10 Pro operating system. Docking of the crystallographic ligand (code name N3) was performed to assess the suitability of the docking procedure. The first round of ligand docking was performed on the protein in the crystallographic conformation and docking was centered at the Cys 145 residue, with coordinates -11.4, 12.8, 70.1 and the size of the box was 20 × 30 × 20 Å. The number of runs was set to 100 and the exhaustiveness to 20. All ligands with a binding energy lower than -7.0 kcal/mol proceeded to the second round of docking. These ligands were then docked to five different conformations of the SARS-CoV-2 3CLpro, which were obtained using MD simulations (described in the previous paragraph), using the same parameters as for the first round of docking. For these ligands, a weighted binding constant across all five receptor conformations was calculated using proportion of time the receptor was in each conformation (Equation (1)):

$$K_w = \sum_{i=1}^n K_i \times p_i \quad (1)$$

K_w is the ligand's weighted non-covalent binding constant, K_i is the ligand's non-covalent binding constant for a given protein conformation i , p_i is the proportion of time the protein is in that conformation, and n is the number of different conformations (in this case 5). For the best 10 ligands obtained in this way a final, third, round of docking was performed on all 5 protein conformations, but this time the exhaustiveness was set to 100, to confirm if the best binding

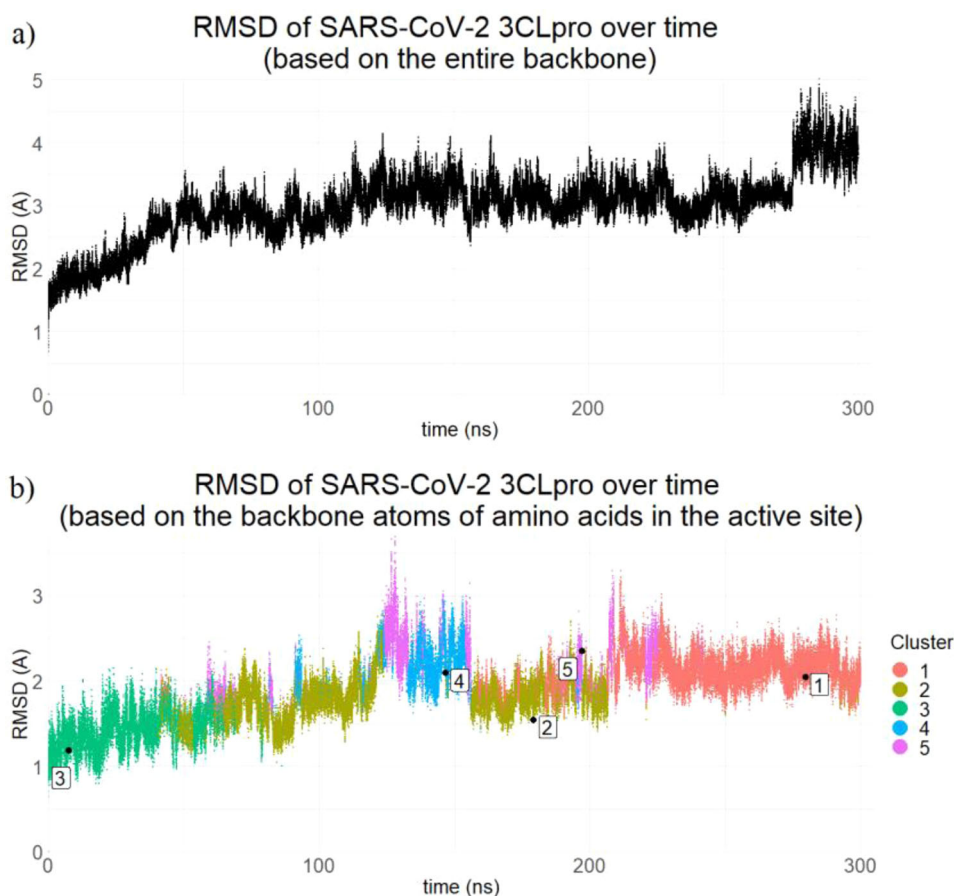


Figure 2. Root-Mean-Square Deviation (RMSD) of the SARS-CoV-2 3CLpro (PDB entry 6LU7) over time, in relation to the first frame: a) based on the whole backbone, b) based on the backbone atoms of amino acids located within 5.5 Å of the Cys 145 residue with position of the representative conformations shown.

poses were truly found. Additionally, blind docking was also performed for these best 10 ligands, with the exhaustiveness set to 100, to possibly locate an alternative binding site outside the active pocket.

An additional verification of the alternative binding site was confirmed using the DoGSiteScorer program (Volkamer et al., 2012) using the default settings available on the Proteins plus server (<https://proteins.plus/>).

Molecular dynamics (MD) simulations of the enzyme-ligand complexes

Complexes of the best 10 ligands docked to the allosteric groove binding site of the SARS-CoV-2 3CLpro in the cluster 1 and cluster 2 representative conformations were subjected to 150 ns MD simulations (a total of 20 simulations of 150 ns in duration). The systems were prepared as described in the "Molecular dynamics (MD) simulations of the free enzyme" section, with only difference being the temperature of the simulations, which was set to 310 K, and the number of Na⁺ and Cl⁻ ions (Na⁺ and Cl⁻ ions were added according to Machado and Pantano (2020) which were set to achieve a neutral environment with salt concentration of 0.15 M). This was done to better simulate the conditions inside the human body.

The binding energy, ΔG_{bind} , of the simulated complexes was calculated using the MM/GBSA (Molecular Mechanics/Generalized Born Surface Area) protocol (Genheden & Ryde, 2015; Hou et al., 2011), available as a part of AmberTools16 (Case et al., 2016). MM/GBSA is a method for the calculation of ΔG_{bind} from snapshots of MD trajectory (Ferenczy, 2015) with an estimated standard error of 1–3 kcal/mol (Genheden & Ryde, 2015). ΔG_{bind} is calculated in the following manner:

$$\Delta G_{\text{bind}} = \langle G_{\text{complex}} \rangle - \langle G_{\text{protein}} \rangle - \langle G_{\text{ligand}} \rangle \quad (2)$$

where the symbol $\langle \rangle$ represents the average value over 100 snapshots collected from a 30 ns part of the corresponding MD trajectories. For all enzyme-ligand complexes, the whole trajectory was divided into 5 parts of 30 ns length and ΔG_{bind} was calculated for all 5 parts of the simulations and reported as mean \pm standard deviation. The calculated MM/GBSA binding free energies were decomposed into specific residue contribution on a per-residue basis according to established procedures. This protocol calculates the contributions to ΔG_{bind} arising from each amino acid side chains and identifies the nature of the energy change in terms of interaction and solvation energies, or entropic contributions (Gohlke et al., 2003; Rastelli et al., 2010). In this case, the entropic term was not calculated.

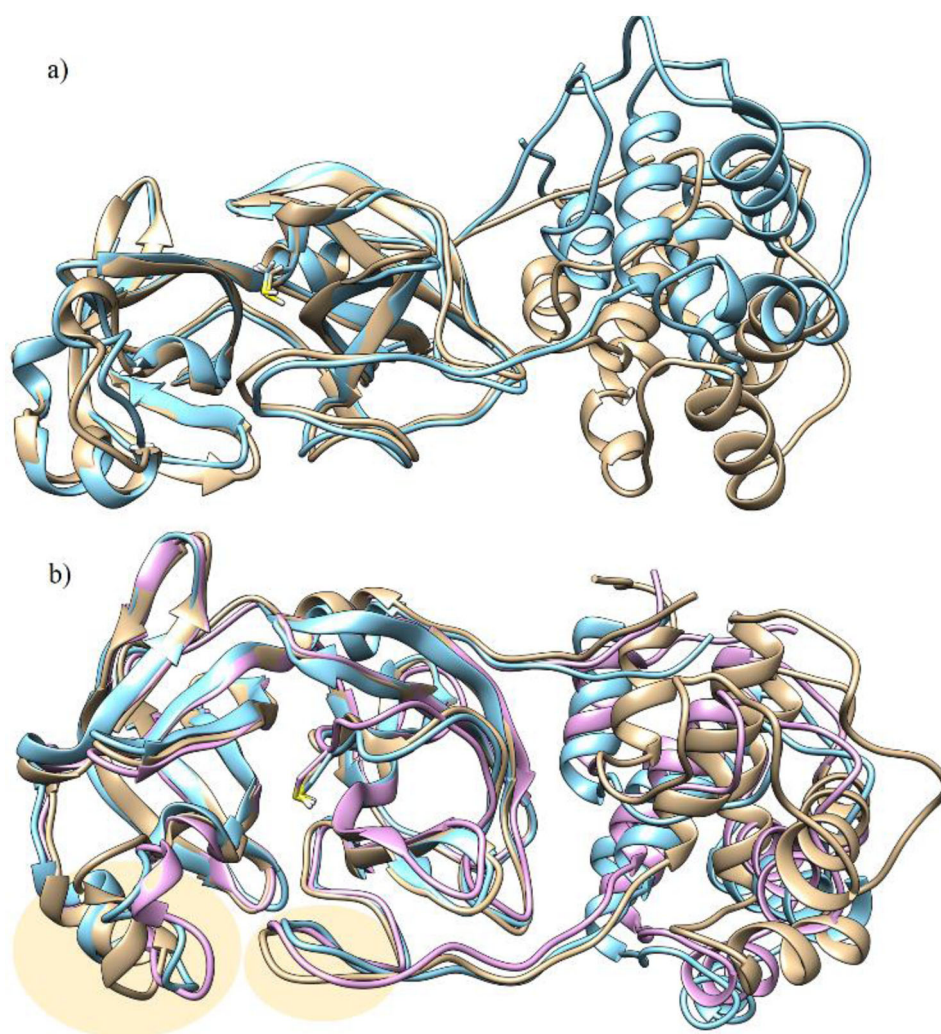


Figure 3. Overlay of SARS-CoV-2 3CLpro (PDB entry 6LU7) with depicted Cys 145 residue. (a) After 250 (brown) and 300 ns (light blue) of MD simulation, (b) representative conformations of clusters 1 (brown), 2 (light blue), and 3 (pink). The main conformational differences are present outside of the binding site (highlighted).

Table 1. Cluster occupancy of SARS-CoV-2 3CLpro through time (300 ns).

| Cluster # | Cluster population (%) | Average distance from the cluster centroid | Cluster standard deviation |
|-----------|------------------------|--|----------------------------|
| 1 | 32.6 | 1.119 | 0.256 |
| 2 | 31.9 | 1.474 | 0.300 |
| 3 | 18.6 | 1.375 | 0.299 |
| 4 | 8.5 | 1.441 | 0.298 |
| 5 | 8.3 | 1.504 | 0.306 |

Results and discussion

Molecular dynamics (MD) simulations of the free enzyme

In Figure 2(a) changes in RMSD of the entire backbone compared to the first frame are displayed. It is immediately noticeable that in the last 25 ns, a significant conformational change occurs, depicted in Figure 3(a). A highly flexible loop, connecting domain III and the anti-parallel cluster of five α -helices of domain II, is responsible for this conformational change. According to our simulation, the secondary structure of all three domains is conserved in both conformations, with slight changes in the unstructured loops, connecting

the secondary structure motifs. Clustering of the SARS-CoV-2 3CLpro trajectory based on RMSD of the active pocket (amino acid residues within 5.5 Å of Cys 145) is shown in Figure 2(b), while the three representative conformations of the three most populated clusters are depicted in Figure 3(b). Comparing the RMSD's obtained in these two ways (Figure 2(a,b)) and by visual inspection of the generated conformations, it was concluded that the greatest conformational change occurs outside of the active site.

Even though the clustering was based on the amino acid residues in the active site (Table 1), the main conformational differences still occurred in the domain III. Obviously, the structural changes in the domain III are reflected on the

structure of the active site itself. The only noticeable conformational differences in the active site are found between residues Cys 44 – Leu 67 (a region in the domain I with 3 small α -helices) and Val 186 – Gln 192 (in the loop region) (Figure 3(b)). The same is also true for clusters 4 and 5 (data not shown).

Docking studies

Since ligands in the DrugBank database are compounds which are approved or undergoing clinical trials, they are usually small molecules. The mean and the highest molecular weights of the tested compounds were 331.3 Da and 1268.9 Da, respectively. The mean binding energy of potential inhibitors to the crystal structure of the SARS-CoV-2 3CLpro according to the present study is -6.4 ± 1.3 kcal/mol (Figure 4). The arbitrary cut-off binding energy was -7.0 kcal/mol and 3056 out of 8756 ligands ($\sim 35\%$) passed the first screening and proceeded to the second round. Docking energy values of all ligands can be found in Supporting Information.

Koulgi et al. demonstrated that docking results obtained for protein structures extracted from MD simulations are better compared to docking results from docking to the crystal structure (Koulgi et al., 2020). For this reason, the selected ligands were then non-covalently docked to the five different protein conformations extracted from the MD simulation and their weighted binding constants were calculated according to Equation (1). (docking energies of the second round of docking can be found in Supporting Information). Best 10 performing ligands additionally underwent a third round of non-covalent docking with an increased exhaustiveness to confirm the best binding pose. The results for the most favorable binding poses of the top 10 ligands to all five protein conformations showed that the binding constants of the same ligand to different protein conformations do not differ significantly (Table 2). This is due to fact that the shape of the active site does not change significantly throughout the MD simulation (Figure 3). However, as can be seen in Figure 5, the active site is very shallow, wide, and located on the surface of the protein, which results in many ligands binding here in different conformations with similar affinities. This is further confirmed by the docking of the crystallographic inhibitor. The best (non-covalent) docking binding energy of the N3 inhibitor was found to be -7.3 kcal/mol, which was better than $\sim 71\%$ of all docked ligands. However, even though the α,β -unsaturated carbonyl group of the docked inhibitor was also located in the vicinity of the Cys 145 residue, the conformation of the rest of the molecule differs significantly from the crystallographic conformation (RMSD of the crystallographic and the redocked ligand is 6.194 Å) (Figure 6). This lack of a small, buried active site makes the active site a bad target for ligands that form only non-covalent bonds, due to their easy displacement by water or other ligands. The shape of the 3CLpro active site is a direct result of its role as a protease, which has to enable an easy access to proteins destined to be processed. However, since its catalytic mechanism is known and well described (Anand et al.,

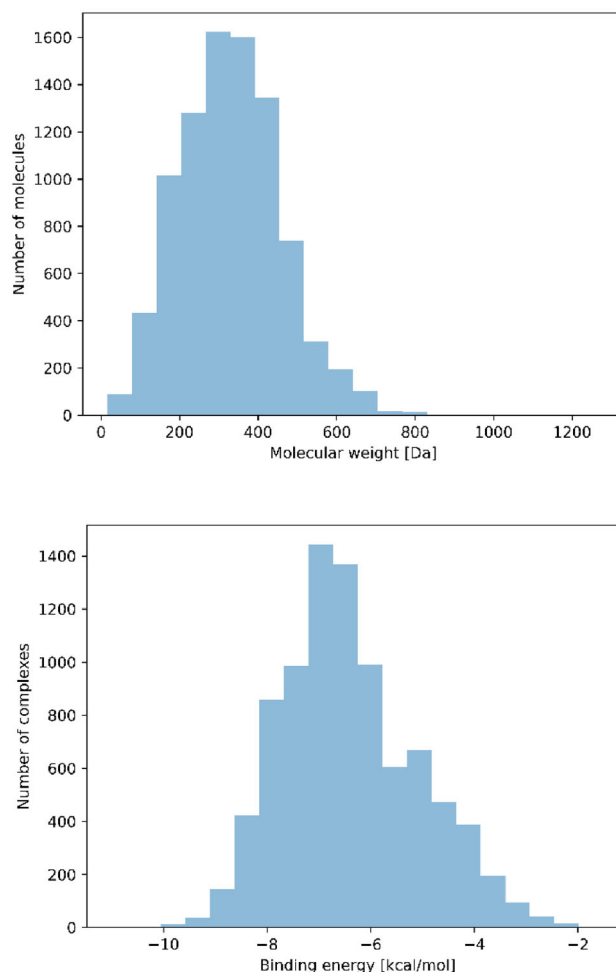
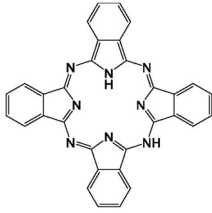
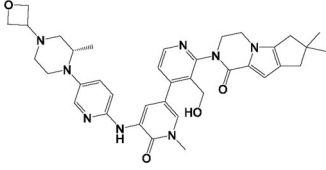
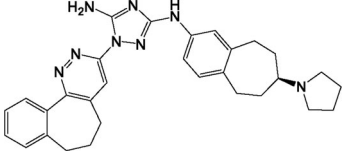
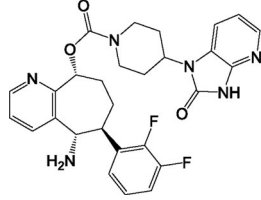
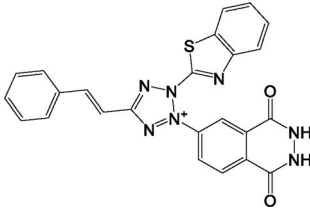
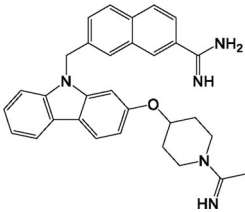
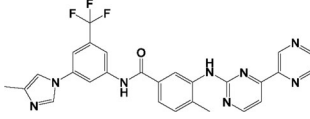
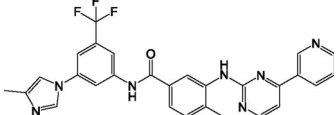
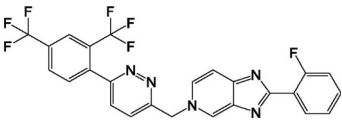


Figure 4. Distribution of molecular weights of ligands in the DrugBank database (top) and of binding energies of virtual screening to SARS-CoV-2 3CLpro (bottom) (PDB entry 6LU7).

2003; Li et al., 2016; Nukoolkarn et al., 2008), it is a good target for covalent inhibitors, which can take advantage of this and bind to the key amino acid residue (Cys 145).

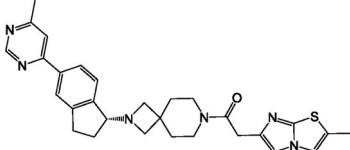
Due to these limitations of the active site, we investigated a possibility of an alternative, allosteric binding site, which could impede the function of the SARS-CoV-2 3CLpro. Therefore, a blind docking study for the 10 best inhibitors was also performed in the search of a more favorable non-covalent binding site. In the screening, the ligands have shown preferred binding in the groove between domains II and III, and in most cases with higher affinities than for the active site. The groove of the cluster 1 representative conformation (groove 1) is delimited by residues Arg 4, Lys 5, Ala 7, Ser 10, Gln 127, Lys 137, Gly 138, Ser 139, Glu 166, Gly 170, Trp 207, Ser 284, and Gln 306 (Table 3 and Figure 7(a)). For all other clusters, due to the conformational changes and large amplitude motion of the domain III, this groove 1 is non-existent. However, another groove (groove 2) forms, where ligands bind to the side opposite of the Cys 145 residue, delimited by amino acid residues Met 6, Val 104, Gly 109, Gln 110, Gln 127, Asn 151, Tyr 154, Asn 203, Asp 245, His 246, Thr 292, Phe 294, and Arg 298 (Table 3 and Figure 7(b)). By analyzing the obtained clusters using the Davies-Bouldin index (DBI), pseudo-F statistic (pSF), and the ratio of

Table 2. Non-covalent docking binding energies (in kcal/mol) of 10 hit molecules in the active site of SARS-CoV-2 3CLpro. The weighted binding constant (K_w) was calculated using Equation (1).

| Ligand name | Structural formula | K_1 | K_2 | K_3 | K_4 | K_5 | K_w | Drug status |
|----------------------|---|-------|-------|-------|-------|-------|-------|---|
| Phthalocyanine |  | -8.9 | -8.6 | -9.7 | -9.2 | -10.1 | -9.1 | Investigational for the treatment of actinic keratosis, Bowen's disease, skin cancer, and stage I or stage II mycosis fungoides |
| Fenebrutinib |  | -9.4 | -8.9 | -9.2 | -8.2 | -8.5 | -9.0 | Investigational for the treatment of non-Hodgkin's lymphoma and chronic lymphocytic leukemia |
| R-428 |  | -9.2 | -8.7 | -8.7 | -9.0 | -8.9 | -8.9 | Investigational for the treatment of non-small cell lung cancer |
| Rimegepant |  | -9.2 | -8.4 | -9.2 | -8.5 | -8.7 | -8.8 | Approved for the treatment of acute migraine headache |
| DB01897 ^a |  | -8.9 | -8.9 | -8.8 | -8.9 | -8.4 | -8.8 | Experimental as a hematopoietic prostaglandin D synthase inhibitor |
| Zk-806450 |  | -9.1 | -8.5 | -9.3 | -8.7 | -8.2 | -8.8 | Experimental as a serine-type endopeptidase inhibitor |
| Radotinib |  | -8.7 | -9.0 | -8.9 | -8.7 | -8.3 | -8.8 | Investigational for the treatment of myelogenous, chronic, BCR-ABL positive leukemia |
| Nilotinib |  | -8.7 | -8.9 | -8.9 | -8.8 | -8.4 | -8.8 | Investigational for the treatment of chronic myelogenous leukemia (CML) |
| Tegobuvir |  | -8.8 | -9.1 | -8.3 | -8.8 | -8.3 | -8.8 | Investigational for the treatment of chronic hepatitis C |

(continued)

Table 2. Continued.

| Ligand name | Structural formula | K_1 | K_2 | K_3 | K_4 | K_5 | K_w | Drug status |
|-------------|---|-------|-------|-------|-------|-------|-------|--|
| PF-5190457 |  | -8.5 | -9 | -8.9 | -8.4 | -8.7 | -8.7 | Investigational as a ghrelin receptor antagonist |

³2-(2f-benzothiazolyl)-5-styryl-3-(4f-phthalhydrazidyl)tetrazolium chloride.

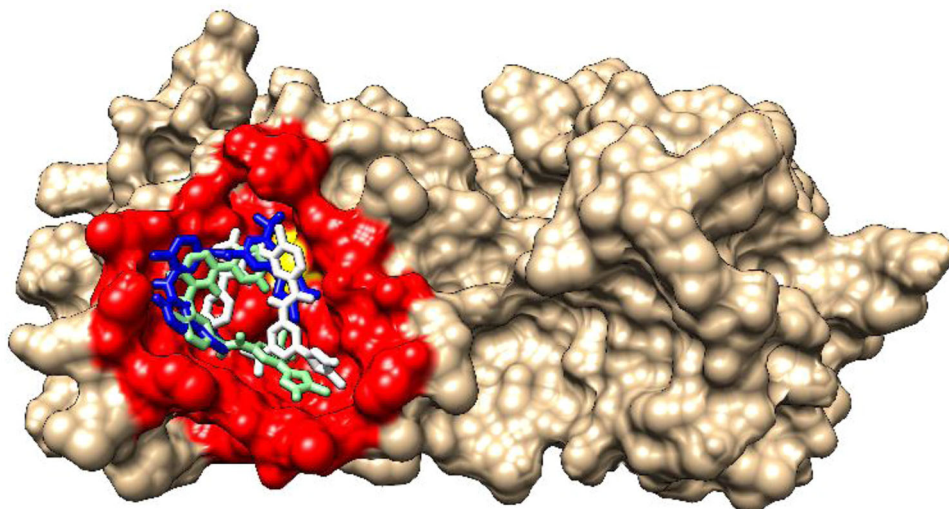


Figure 5. Active site of the SARS-CoV-2 3CLpro (PDB entry 6LU7) (red) and Cys 145 depicted in yellow with the best three conformations of nilotinib (conformations with non-covalent binding energies of -8.7 (blue), -8.7 (light green), -8.5 kcal/mol (white)).

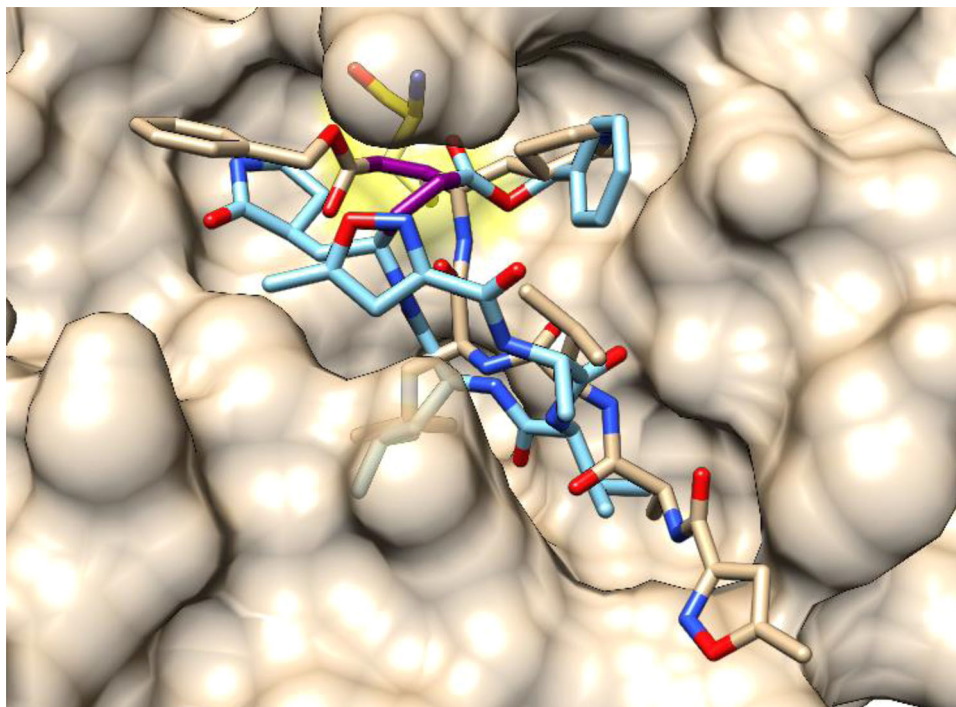


Figure 6. Comparison of the crystallographic (tan) and the best docked N3 inhibitor (light blue) with depicted α,β -unsaturated carbonyl group (purple) and the Cys 145 residue (yellow). Location of the α,β -unsaturated carbonyl group is approximately the same even though the conformation of the rest of the molecule differs significantly.

sum of squares regression and sum of squares error (SSR/SST) (Table SI 1 and SI 2), it was determined that there are, in fact, two main conformations of the SARS-CoV-2 3CLpro,

represented by clusters 1 and 2 (backbone RMSD of these clusters is 8.761 \AA), with the backbone RMSD between representative conformations of cluster 2 and cluster 3, 4, and 5

Table 3. Docking binding energies (in kcal/mol) of the 10 hit molecules in the groove between the domains II and III of SARS-CoV-2 3CLpro. The weighted non-covalent binding constant (K_w) was calculated using Equation (1).

| Ligand name | K_1 | K_2 | K_3 | K_4 | K_5 | K_w |
|----------------------|-------|-------|-------|-------|-------|-------|
| Phthalocyanine | -11.0 | -10.8 | -11.4 | -10.0 | -11.2 | -10.9 |
| R - 428 | -10.3 | -9.6 | -10.2 | -9.7 | -10.7 | -10.0 |
| Zk - 806450 | -10.3 | -9.2 | -10.3 | -9.8 | -10.1 | -9.9 |
| Nilotinib | -9.6 | -9.6 | -9.8 | -10.2 | -10.6 | -9.8 |
| Fenebrutinib | -10.4 | -8.8 | -9.4 | -9.3 | -9.4 | -9.5 |
| DB01897 ^a | -9.8 | -8.5 | -9.8 | -9.0 | -10.3 | -9.3 |
| Radotinib | -9.4 | -8.7 | -9.6 | -10 | -10.5 | -9.3 |
| Tegobuvir | -9.4 | -8.9 | -9.3 | -9.8 | -10.3 | -9.3 |
| PF - 5190457 | -8.9 | -8.6 | -8.8 | -8.4 | -10.4 | -8.9 |
| rimegepant | -8.4 | -8.8 | -9.6 | -8.6 | -9.1 | -8.8 |

^a2-(2f-benzothiazolyl)-5-styryl-3-(4f-phthalhydrazidyl)tetrazolium chloride.

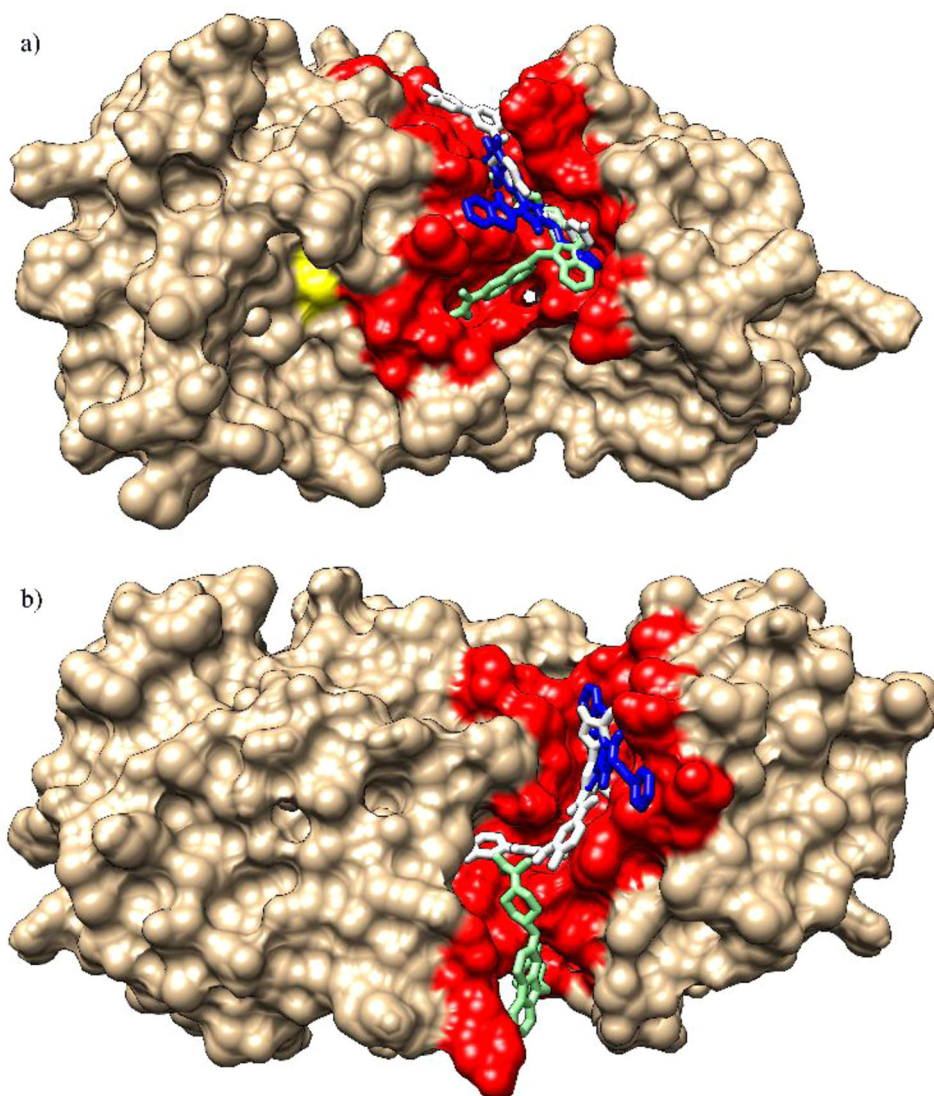


Figure 7. The SARS-CoV-2 3CLpro (PDB entry 6LU7) with depicted Cys 145 (yellow) and bound ligands DB01897 (blue), Zk-806450 (light green), and nilotinib (white) in grooves of the two highest populated conformations. (a) Conformation 1, groove 1, (b) conformation 2, groove 2.

being very low (2.264, 1.406, and 1.178 Å, respectively). We compared the conformations of clusters 1 and 2 to the crystallographic structures of the PDB entries 6LU7 and 6Y2G (Zhang et al., 2020). The analysis revealed that the conformation of cluster 2 is the most similar to the crystallographic conformations (with RMSD of 2.149 Å in the case of 6LU7 and 1.588 Å in the case of 6Y2G). The similarity between

cluster 1 and crystallographic conformations is lower (with RMSD of 7.378 Å and 8.172 Å for 6LU7 and 6Y2G, respectively), while the difference between the 6LU7 and 6Y2G structures is 0.721 Å.

In both cases, the groove is located in the vicinity of amino acids crucial for the dimerization process. As it was determined by Lim et al. (2014), a single Arg298Ala mutation

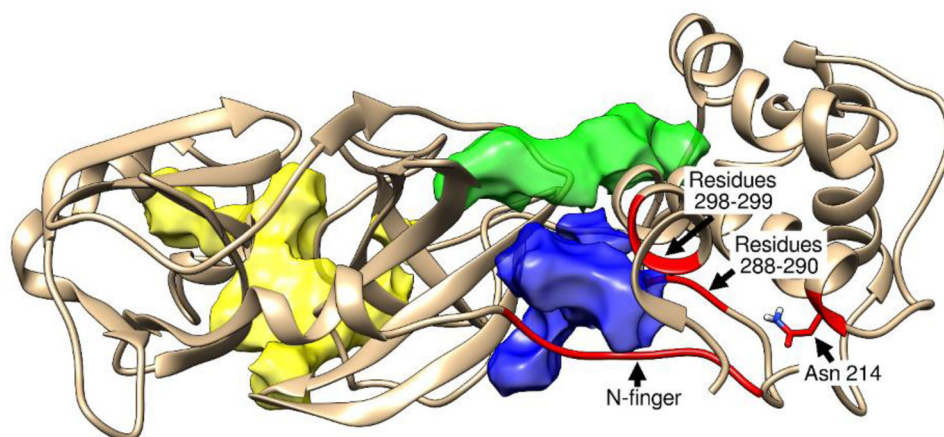


Figure 8. Top three 3CLpro binding pockets identified by DoGSiteScorer (Volkamer et al., 2012): the active site (yellow), the first groove site (blue), and the second groove site (green), with the key amino acid residues for dimerization marked in red.

completely stops the dimerization, resulting in an inactive monomeric form of the enzyme. Since the role of the *N*-finger in correct dimerization is also well known (Wei et al., 2006; Zhong et al., 2008), binding of ligands between domains II and III (most notably, the *N*-finger and the Arg 298) could allosterically inhibit the activity of SARS-CoV-2 3CLpro by preventing it to dimerize into the active form (Yang et al., 2003; Zhong et al., 2008). The location of this binding site corresponds to the possible binding site found by Li et al. (Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening, unpublished data), where a possibly druggable binding site was found in the groove between the two monomers.

The existence of this binding site was also confirmed by the DoGSiteScorer (Volkamer et al., 2012) score (Figure 8, Table 4), showing that the potential groove binding site (divided here into two subpockets), represents a good drug target, with the larger subpocket protruding inside the protein. This subpocket is also in the direct vicinity of all the four key elements (the *N*-finger, the Asn 214 residue, residues Glu 288 – Asp 289 – Glu 290, and the residues Arg 298 – Gln 299) essential for the dimerization and activation of the 3CLpro. Additionally, all hit molecules have a higher binding potential to the groove pocket compared to the catalytic pocket (except rimegepant, whose affinities for both sites are similar).

Molecular dynamics (MD) simulations of the enzyme-ligand complexes

The best 10 docked ligands docked to the groove of the cluster 1 and cluster 2 representative conformations (examples in Figure 7) were subjected to 150 ns MD simulations to test the stability of such complexes. A total of 20 simulations was performed and the results are shown in Table 5.

As can be seen, all the tested ligands bind to the allosteric groove binding site of the cluster 2 representative conformation and remain bound for the entire duration of the simulation. However, only four ligands (radotinib, nilotinib, R-428, and Zk-806450) remain in the groove binding site of the

Table 4. Binding site properties calculated using DoGSiteScorer (Volkamer et al., 2012).

| Position | Volume (Å ³) | Surface (Å ²) | Drug Score |
|-------------------------|--------------------------|---------------------------|------------|
| Active site | 1053.5 | 1600.2 | 0.80 |
| First groove subpocket | 613.5 | 1047.9 | 0.78 |
| Second groove subpocket | 507.6 | 644.5 | 0.86 |

Table 5. ΔG_{bind} energies of the 10 best ligands bound to the allosteric groove binding site of the cluster 1 and cluster 2 representative conformations. The energies are expressed in kcal/mol as mean \pm standard deviation.

| Ligand name | ΔG_{bind} (conformation 1) | ΔG_{bind} (conformation 2) |
|----------------|---|---|
| Radotinib | -39.06 ± 1.21 | -25.72 ± 2.20 |
| Nilotinib | -26.92 ± 2.55 | -34.75 ± 0.92 |
| R-428 | -26.55 ± 2.10 | -27.07 ± 1.75 |
| Zk-806450 | -26.04 ± 4.40 | -36.43 ± 4.21 |
| Fenebrutinib | / | -32.48 ± 3.17 |
| PF-5190457 | / | -30.75 ± 3.05 |
| DB01897 | / | -30.33 ± 1.12 |
| Tegobuvir | / | -25.67 ± 1.64 |
| Rimegepant | / | -24.64 ± 2.92 |
| Phthalocyanine | / | -24.62 ± 3.56 |

Table 6. Contributions of the most important amino acid residues for binding of radotinib and nilotinib, with differing amino acid residues bolded.

| Conformation 1 | | Conformation 2 | | | | | |
|----------------|--------------------------|----------------|--------------------------|-----------|--------------------------|---------|-------|
| Radotinib | Nilotinib | Radotinib | Nilotinib | Radotinib | Nilotinib | | |
| Residue | ΔG_{bind} | Residue | ΔG_{bind} | Residue | ΔG_{bind} | | |
| Arg 4 | -2.86 | Arg 4 | -3.73 | Phe 294 | -2.36 | Gln 110 | -2.41 |
| Lys 5 | -2.58 | Lys 5 | -3.51 | Ile 249 | -1.94 | Gln 107 | -2.36 |
| Tyr 126 | -2.51 | Met 6 | -1.07 | Gln 107 | -1.68 | Ile 249 | -2.28 |
| Val 125 | -1.73 | Glu 290 | -1.07 | Gln 110 | -1.47 | His 246 | -2.08 |
| Met 6 | -1.49 | Tyr 126 | -0.75 | His 246 | -1.19 | Pro 108 | -1.53 |
| Ala 7 | -1.47 | Phe 291 | -0.59 | Pro 108 | -1.08 | Thr 292 | -1.17 |
| Glu 290 | -1.19 | Glu 288 | -0.57 | Val 202 | -0.56 | Val 202 | -0.99 |
| Gln 127 | -1.18 | Phe 3 | -0.55 | Thr 292 | -0.54 | Thr 243 | -0.90 |
| Pro 9 | -1.12 | Trp 207 | -0.39 | Thr 243 | -0.51 | Phe 294 | -0.78 |

cluster 1 representative conformation. Complete MM/GBSA data, with contributions of individual amino acid residues to the ΔG_{bind} , as well as complex conformations every 30 ns for all complexes can be found in the [Supplementary Information](#).

Radotinib and nilotinib were found to be the most stable ligands with the smallest conformational changes in the complexes, which is also reflected in their small ΔG_{bind}

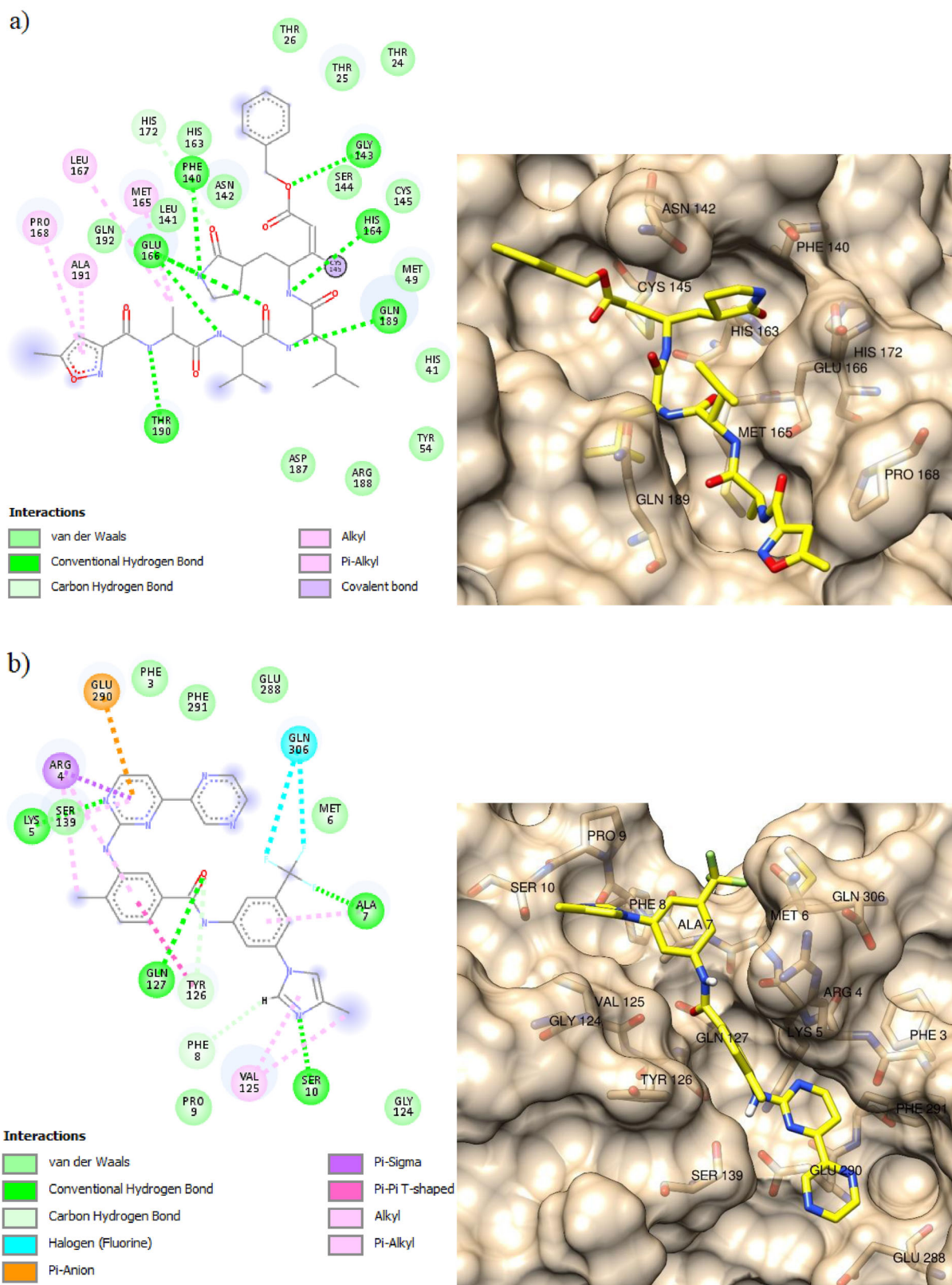


Figure 9. Comparison of the 3CLpro active site with bound N3 ligand (a) with radotinib bound in the allosteric groove binding site in the representative conformation of the cluster 1 (b), and radotinib bound in the allosteric groove binding site in the representative conformation of the cluster 2 (c). Left-hand side of the figures shows the most important interactions shown in 2D, recognized by Discovery Studio 2021 and the right-hand side shows amino acids located less than 4 Å from the bound ligand (in yellow).

standard deviations. The most important amino acid residues in their binding, as well as their contributions are shown in Table 6.

In binding to the cluster 1 representative conformation, radotinib and nilotinib differ in only four out of ten most important amino acid residues. As already mentioned, Phe 3,

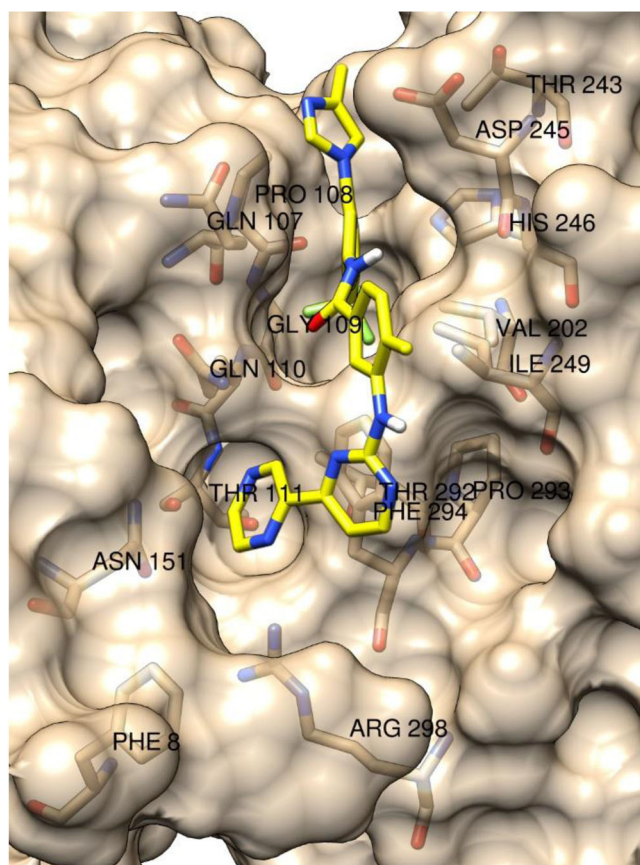
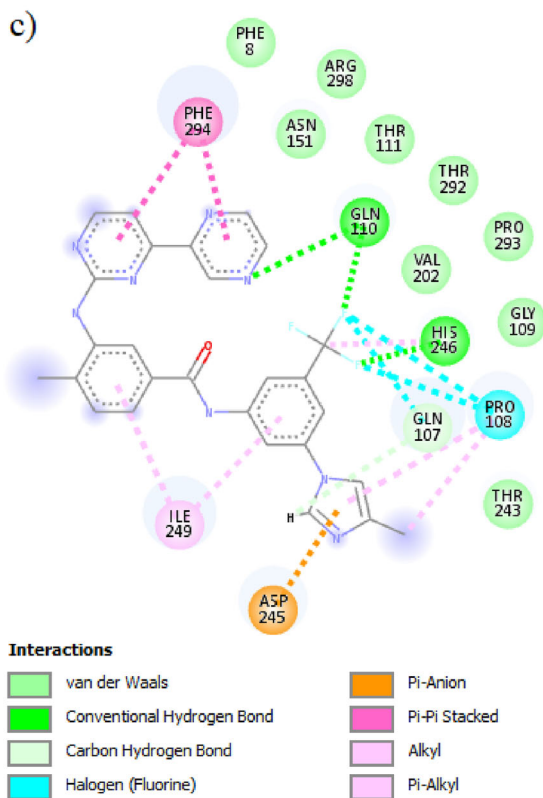


Figure 9. (Continued).

Arg 4, Lys 5, Met 6, and Ala 7 amino acid residues are a part of the *N*-finger, which is crucial for SARS-CoV-2 3CLpro dimerization and activation. Additionally, both radotinib and nilotinib also interact with and Glu 290 and nilotinib also with Glu 288, which are also very important for the 3CLpro activation. In the case of the cluster 2 representative conformation, radotinib and nilotinib share all ten most important amino acid residues.

Figure 9 shows interactions of the crystallographic N3 ligand bound to the active site of the SARS-CoV-2 3CLpro and radotinib interactions with the allosteric groove after 150 ns MD simulations for both conformations 1 and 2. The figure was generated using Discovery Studio 2021 (BIOVIA, Dassault Systèmes, Discovery Studio v21.1.020298, San Diego: Dassault Systèmes, 2021). Since the N3 ligand is a covalently bound peptidomimetic, in addition to a few alkyl interactions, peptide $-C(=O)N-$ groups form many hydrogen bonds (Figure 9(a)). On the other hand, radotinib has only one such group (which forms a hydrogen bond in both conformations) but establishes additional π -interactions (Figure 9(b,c)), which are not present in the case of the N3 ligand. Combining this knowledge with the fact that the most important amino acid residues for radotinib binding possess hydrogen donor and acceptors groups, it can be said that the potential of this binding site is not fully exploited. This is also visible when comparing radotinib and nilotinib poses (Figure SI 1) and their binding energies for different 3CLpro conformations (Table 5). Radotinib binds much stronger than nilotinib to the 3CLpro groove site in conformation 1, while the opposite

is true for 3CLpro groove site in conformation 2. In Figure SI 1a, it can be seen that nilotinib is protruding outside of the groove site, while radotinib (Figure 9(b)) is tightly bound. As for the conformation 2, nilotinib is located deep inside the groove pocket (Figure SI 1 b), while radotinib is located more on the surface of the groove (Figure 9(c)). In general, the groove in the conformation 2 is deeper and better defined than that in the conformation 1, and better suited for ligand binding. This is also visible from the MM/GBSA data (Table 5), where all ten ligands remained stable during the entire 150 MD simulation when bound to the enzyme in the conformation 2, while only four remained stable in conformation 1. Additionally, all ligands which remained stable in both enzyme conformations (except radotinib) bound more strongly to the enzyme in the conformation 2. With a significant difference in radotinib and nilotinib binding energies to the groove site of the two 3CLpro conformations, there still exists a great potential to find a groove site inhibitor which would bind to both conformations even stronger than the tested compounds.

In short, the groove represents a possible alternative inhibition target to inhibition of the active site. While binding of the ten tested ligands is better in the enzyme conformation 2 (except for radotinib) due to its larger depth, ligands that also remain bound in the conformation 1 interact more significantly with amino acids important for 3CLpro dimerization and activation. Additionally, considering the shallowness of the active site and the lack of discrimination between ligands' binding energies when binding to it, this

study suggests that binding of ligands inside the groove (as opposed to the active site) might even be more favorable for non-covalent inhibitors. Therefore, we propose considering the groove between domains II and III as a target binding site in future 3CLpro inhibition studies and in drug repurposing.

Conclusion

In this article an ensemble virtual screening study of DrugBank library was conducted using both the crystallographic conformation of the SARS-CoV-2 3CLpro, as well as its five representative conformations which were obtained after a 300 ns MD simulation and *k*-means clustering. The study showed that the 3CLpro active site is not the best target for non-covalent inhibitors due to its shallowness and wideness. However, we propose targeting the additional, allosteric binding site, located in the groove between domains II and III. This groove is partially made of amino acids in the *N*-finger region (residues 1–7) and the crucial Arg 298 residue, which are involved in the SARS-CoV-2 3CLpro dimerization and activation processes. MD simulations of the 10 best ligands demonstrated the viability of the groove site as a potential anti-SARS-CoV-2 target, with radotinib and nilotinib, tyrosine kinase inhibitors already in use for treatment of leukemia, as potential allosteric 3CLpro inhibitors. Therefore, binding of ligands to this site could inhibit protein dimerization, and consequently, SARS-CoV-2 3CLpro activation.

Acknowledgements

The authors also acknowledge University of Zagreb, University Computing Centre (SRCE) for granting computational time on the Isabella cluster.

Disclosure statement

The authors declare no conflict of interest.

Funding

This study was funded by RFBR, DST, CNPq, SAMRC, project number 20-53-80002.

ORCID

Jurica Novak  <http://orcid.org/0000-0001-5744-6677>
 Hrvoje Rimac  <http://orcid.org/0000-0001-7232-6489>
 Shivananda Kandagalla  <http://orcid.org/0000-0003-4540-3500>
 Prateek Pathak  <http://orcid.org/0000-0002-6160-9755>
 Maria Grishina  <http://orcid.org/0000-0002-2573-4831>
 Vladimir Potemkin  <http://orcid.org/0000-0002-5244-8718>

References

Alamri, M. A., Tahir Ul Qamar, M., Mirza, M. U., Bhadane, R., Alqahtani, S. M., Muneer, I., Froeyen, M., & Salo-Ahen, O. M. H. (2020). Pharmacoinformatics and molecular dynamics simulation studies

reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro. *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–13. <https://doi.org/10.1080/07391102.2020.1782768>

Anand, K., Ziebuhr, J., Wadhvani, P., Mesters, J. R., & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-SARS drugs. *Science (New York, N.Y.)*, 300(5626), 1763–1767. <https://doi.org/10.1126/science.1085658>

Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: Identifying and developing new uses for existing drugs. *Nature Reviews. Drug Discovery*, 3(8), 673–683. <https://doi.org/10.1038/nrd1468>

Bhardwaj, V. K., Singh, R., Das, P., & Purohit, R. (2021). Evaluation of acridinedione analogs as potential SARS-CoV-2 main protease inhibitors and their comparison with repurposed anti-viral drugs. *Computers in Biology and Medicine*, 128, 104117. <https://doi.org/10.1016/j.compbiomed.2020.104117>

Bhardwaj, V. K., Singh, R., Sharma, J., Rajendran, V., Purohit, R., & Kumar, S. (2020). Identification of bioactive molecules from tea plant as SARS-CoV-2 main protease inhibitors. *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–10. <https://doi.org/10.1080/07391102.2020.1766572>

Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., C., Lin, Luchko, T., ... Kollman, P. A. (2016). *Amber 2016*. University of California.

Chen, H., Wei, P., Huang, C., Tan, L., Liu, Y., & Lai, L. (2006). Only one promoter is active in the dimer of SARS 3C-like proteinase. *The Journal of Biological Chemistry*, 281(20), 13894–13898. <https://doi.org/10.1074/jbc.M510745200>

Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*, 92(4), 418–423. <https://doi.org/10.1002/jmv.25681>

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N -log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12), 10089–10092. <https://doi.org/10.1063/1.464397>

Dotolo, S., Marabotti, A., Facchiano, A., & Tagliaferri, R. (2021). A review on drug repurposing applicable to COVID-19. *Briefings in Bioinformatics*, 22(2), 716–726. <https://doi.org/10.1093/bib/bbaa288>

Elmezayen, A. D., Al-Obaidi, A., Şahin, A. T., & Yelekcı, K. (2021). Drug repurposing for coronavirus (COVID-19): In silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *Journal of Biomolecular Structure & Dynamics*, 39(8), 2913–2980. <https://doi.org/10.1080/07391102.2020.1758791>

Ferency, G. G. (2015). Computation of drug-binding thermodynamics. In G. M. Keserü & D. C. Swinney (Eds.), *Thermodynamics and kinetics of drug binding* (pp. 37–61). Wiley-VCH Verlag GmbH & Co. <https://doi.org/10.1002/9783527673025.ch3>

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 10(5), 449–461. <https://doi.org/10.1517/17460441.2015.1032936>

Gohlke, H., Kiel, C., & Case, D. A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *Journal of Molecular Biology*, 330(4), 891–913. [https://doi.org/10.1016/S0022-2836\(03\)00610-7](https://doi.org/10.1016/S0022-2836(03)00610-7)

Gollapalli, P., Sharath, B. S., Rimac, H., Patil, P., Nalilu, S. K., Kandagalla, S., & Shetty, P. (2020). Pathway enrichment analysis of virus-host interactome and prioritization of novel compounds targeting the spike glycoprotein receptor binding domain-human angiotensin-converting enzyme 2 interface to combat SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, 1–14. <https://doi.org/10.1080/07391102.2020.1841681>

Gyebi, G. A., Ogunro, O. B., Adegunloye, A. P., Ogunyemi, O. M., & Afolabi, S. O. (2020). Potential inhibitors of coronavirus 3-chymotrypsin-like protease (3CLpro): An in silico screening of alkaloids and terpenoids from African medicinal plants. *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–13. <https://doi.org/10.1080/07391102.2020.1764868>

- Handoko, S. D., Ouyang, X., Su, C. T. T., Kwok, C. K., & Ong, Y. S. (2012). QuickVina: Accelerating AutoDock Vina using gradient-based heuristics for global optimization. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(5), 1266–1272. <https://doi.org/10.1109/TCBB.2012.82>
- Harrison, C. (2020). Coronavirus puts drug repurposing on the fast track. *Nature Biotechnology*, 38(4), 379–381. <https://doi.org/10.1038/d41587-020-00003-1>
- Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 51(1), 69–82. <https://doi.org/10.1021/ci100275a>
- Koulgi, S., Jani, V., Uppuladinne, M., Sonavane, U., Nath, A. K., Darbari, H., & Joshi, R. (2020). Drug repurposing studies targeting SARS-CoV-2: An ensemble docking approach on drug target 3C-like protease (3CLpro). *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–21. <https://doi.org/10.1080/07391102.2020.1792344>
- Li, C., Teng, X., Qi, Y., Tang, B., Shi, H., Ma, X., & Lai, L. (2016). Conformational flexibility of a short loop near the active site of the SARS-3CLpro is essential to maintain catalytic activity. *Scientific Reports*, 6(January), 20918. <https://doi.org/10.1038/srep20918>
- Lim, L., Shi, J., Mu, Y., & Song, J. (2014). Dynamically-driven enhancement of the catalytic machinery of the SARS 3C-like protease by the S284-T285-I286/A mutations on the extra domain. *PLoS One*, 9(7), e101941. <https://doi.org/10.1371/journal.pone.0101941>
- Machado, M. R., & Pantano, S. (2020). Split the charge difference in two! A rule of thumb for adding proper amounts of ions in MD simulations. *Journal of Chemical Theory and Computation*, 16(3), 1367–1372. <https://doi.org/10.1021/acs.jctc.9b00953>
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16), 2785–2791. <https://doi.org/10.1002/jcc.21256>
- Novak, J., Rimac, H., Kandagalla, S., Grishina, M. A., & Potemkin, V. A. (2021). Can natural products stop the SARS-CoV-2 virus? A docking and molecular dynamics study of a natural product database. *Future Medicinal Chemistry*, 13(4), 363–378. <https://doi.org/10.4155/fmc-2020-0248>
- Nukoolkarn, V., Lee, V. S., Malaisree, M., Aruksakulwong, O., & Hannongbua, S. (2008). Molecular dynamic simulations analysis of ritonavir and lopinavir as SARS-CoV 3CLpro inhibitors. *Journal of Theoretical Biology*, 254(4), 861–867. <https://doi.org/10.1016/j.jtbi.2008.07.030>
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1), 33. <https://doi.org/10.1186/1758-2946-3-33>
- Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D., & Pirmohamed, M. (2019). Drug repurposing: Progress, challenges and recommendations. *Nature Reviews. Drug Discovery*, 18(1), 41–58. <https://doi.org/10.1038/nrd.2018.168>
- Rastelli, G., Rio, A. D., Degliesposti, G., & Sgobba, M. (2010). Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *Journal of Computational Chemistry*, 31(4), 797–810. <https://doi.org/10.1002/jcc.21372>
- Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9(7), 3084–3095. <https://doi.org/10.1021/ct400341p>
- Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3), 327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5) [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5)
- Sang, P., Tian, S. H., Meng, Z. H., & Yang, L. Q. (2020). Anti-HIV drug repurposing against SARS-CoV-2. *RSC Advances*, 10(27), 15775–15783. <https://doi.org/10.1039/D0RA01899F>
- Shi, J., & Song, J. (2006). The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *The FEBS Journal*, 273(5), 1035–1045. <https://doi.org/10.1111/j.1742-4658.2006.05130.x>
- Vanhaelen, Q. (2019). *Computational methods for drug repurposing* (Vol. 1903). Springer New York. <https://doi.org/10.1007/978-1-4939-8955-3>
- Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., & Rarey, M. (2012). Combining global and local measures for structure-based druggability predictions. *Journal of Chemical Information and Modeling*, 52(2), 360–372. <https://doi.org/10.1021/ci200454v>
- Wei, P., Fan, K., Chen, H., Ma, L., Huang, C., Tan, L., Xi, D., Li, C., Liu, Y., Cao, A., & Lai, L. (2006). The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochemical and Biophysical Research Communications*, 339(3), 865–872. <https://doi.org/10.1016/j.bbrc.2005.11.102>
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265–269. <https://doi.org/10.1038/s41586-020-2008-3>
- Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., & Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Science China. Life Sciences*, 63(3), 457–460. <https://doi.org/10.1007/s11427-020-1637-5>
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R., & Rao, Z. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23), 13190–13195. <https://doi.org/10.1073/pnas.1835675100>
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science (New York, N.Y.)*, 368(6489), 409–412. <https://doi.org/10.1126/science.abb3405>
- Zhong, N., Zhang, S., Zou, P., Chen, J., Kang, X., Li, Z., Liang, C., Jin, C., & Xia, B. (2008). Without its N-finger, the main protease of severe acute respiratory syndrome coronavirus can form a novel dimer through its C-terminal domain. *Journal of Virology*, 82(9), 4227–4234. <https://doi.org/10.1128/JVI.02612-07>
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 382(8), 727–733. <https://doi.org/10.1056/NEJMoa2001017>