

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro

Alamri, MA; ul Qamar M, Tahir; Mirza, MU; Bhadane, R; Alqahtani, SM; Muneer, I; Froeyen, M; Salo-Ahen, Outi

Published in:
Journal of Biomolecular Structure and Dynamics

DOI:
[10.1080/07391102.2020.1782768](https://doi.org/10.1080/07391102.2020.1782768)

Published: 01/01/2020

Document Version
Accepted author manuscript

Document License
Publisher rights policy

[Link to publication](#)

Please cite the original version:
Alamri, MA., ul Qamar M, T., Mirza, MU., Bhadane, R., Alqahtani, SM., Muneer, I., Froeyen, M., & Salo-Ahen, O. (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*, -. <https://doi.org/10.1080/07391102.2020.1782768>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Biomolecular Structure and Dynamics on the 24th June 2020, available online: <http://www.tandfonline.com/10.1080/07391102.2020.1782768>.

Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CL^{pro}

Mubarak A. Alamri ^{a,†}, Muhammad Tahir ul Qamar ^{b,†,*}, Muhammad Usman Mirza ^{c,†}, Rajendra Bhadane ^{d,e}, Safar M. Alqahtani ^a, Iqra Muneer ^f, Matheus Froeyen ^c, Outi M. H. Salo-Ahen ^{d,e,*}

^a Department of Pharmaceutical Chemistry, College of Pharmacy, Prince Sattam Bin Abdulaziz University, P.O. Box 11323, Alkarj, Saudi Arabia

^b College of Life Science and Technology, Guangxi University, Nanning 530004, P. R. China

^c Department of Pharmaceutical and Pharmacological Sciences, Rega Institute for Medical Research, Medicinal Chemistry, University of Leuven, B-3000 Leuven, Belgium

^d Structural Bioinformatics Laboratory, Faculty of Science and Engineering, Biochemistry, Åbo Akademi University, FI-20520 Turku, Finland

^e Pharmaceutical Sciences Laboratory, Faculty of Science and Engineering, Pharmacy, Åbo Akademi University, FI-20520 Turku, Finland

^f School of Life Sciences, University of Science and Technology of China, Hefei 230026, P. R. China

[†]These authors contributed equally in this study

*Corresponding authors: Muhammad Tahir ul Qamar - m.tahirulqamar@hotmail.com; Outi

M. H. Salo-Ahen - outi.salo-ahen@abo.fi

Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CL^{pro}

Abstract: The SARS-CoV-2 was confirmed to cause the global pandemic of coronavirus disease 2019 (COVID-19). The 3-chymotrypsin-like protease (3CL^{pro}), an essential enzyme for viral replication, is a valid target to combat SARS-CoV and MERS-CoV. In this work, we present a structure-based study to identify potential covalent inhibitors containing a variety of chemical warheads. The targeted Asinex Focused Covalent (AFCL) library was screened based on different reaction types and potential covalent inhibitors were identified. In addition, we screened FDA-approved protease inhibitors to find candidates to be repurposed against SARS-CoV-2 3CL^{pro}. A number of compounds with significant covalent docking scores were identified. These compounds were able to establish a covalent bond (C-S) with the reactive thiol group of Cys145 and to form favorable interactions with residues lining the substrate-binding site. Moreover, paritaprevir and simeprevir from FDA-approved protease inhibitors were identified as potential inhibitors of SARS-CoV-2 3CL^{pro}. The mechanism and dynamic stability of binding between the identified compounds and SARS-CoV-2 3CL^{pro} were characterized by molecular dynamics (MD) simulations. The identified compounds are potential inhibitors worthy of further development as COVID-19 drugs. Importantly, the identified FDA-approved anti-hepatitis-C virus (HCV) drugs paritaprevir and simeprevir could be ready for clinical trials to treat infected patients and help curb COVID-19.

Keywords: SARS-CoV-2; COVID-19; 3CL protease; Covalent inhibitors; Paritaprevir; Simeprevir; Molecular dynamics simulation

Abbreviations: 3CL^{pro}: 3-chymotrypsin-like cysteine protease; COVID-19: coronavirus disease 2019; MD: molecular dynamics; MERS-CoV: Middle East respiratory syndrome coronavirus; PDB: Protein Data Bank; RMSD: root-mean-square deviation; RMSF: root-mean-square fluctuation; SARS-CoV: severe acute respiratory syndrome coronavirus; WHO: World Health Organization

1. Introduction

In the last two decades, several pathogens have spilled over and caused outbreaks. Among them, emergence and reemergence of coronavirus-related epidemics have widely spread fatal respiratory illnesses (N. Zhu et al., 2020). Coronaviruses are enveloped RNA viruses that are distributed broadly among humans, other mammals, and birds, and cause respiratory, enteric, hepatic, and neurologic diseases (Weiss & Leibowitz, 2011). Six coronavirus strains are known to cause human diseases. Four strains (229E, OC43, NL63, and HKU1) are prevalent and typically cause common cold symptoms in immunocompetent individuals. The two other strains (severe acute respiratory syndrome coronavirus, SARS-CoV, and Middle East respiratory syndrome coronavirus, MERS-CoV) are zoonotic in origin and have been linked to fatal illnesses (Cui et al., 2019; Su et al., 2016). Recently, severe respiratory illness outbreak was reported. Follow-up studies found that this illness is linked with a new highly infectious strain of coronavirus SARS-CoV-2, and term it as coronavirus disease 2019 (COVID-19) (F. Wu et al., 2020; Zhou et al., 2020; N. Zhu et al., 2020). COVID-19 has quickly turned into a global pandemic, affected millions of people, and claimed tens of thousands of human lives worldwide. Currently, there is no approved drug or vaccine to combat COVID-19 (Ahmed et al., 2020; Tahir ul Qamar, Rehman, et al., 2020). Several antiviral drugs are in clinical trials (R.-f. Zhu et al., 2020) and only remdesivir appeared to be the most promising (Cao et al., 2020) which is being tested in patients with severe disease (Cao et al., 2020; Hillaker et al., 2020).

Later investigations have revealed that the SARS-CoV-2 belongs to the beta-coronavirus family and it is closely related to SARS-CoV (Tahir ul Qamar, Alqahtani, et al., 2020; F. Wu et al., 2020; Xu et al., 2020; Zhou et al., 2020). Similar to other beta-corona-viruses, SARS-CoV-2 produces an 800-kDa polypeptide upon transcription of its genome (N. Zhu et al., 2020). This polypeptide is proteolytically cleaved to generate various proteins including Spike (S) protein, envelope (E) protein, membrane (M) protein and nucleocapsid (N) protein (Xu et al., 2020; N. Zhu et al., 2020). S protein is involved in virus-cell receptor binding, while the other structural proteins E, M and N are crucial for virion assembly (C. Wu et al., 2020; Zumla et al., 2016). The proteolytic processing is mediated by papain-like protease (PL^{pro}) and 3-chymotrypsin-like protease (3CL^{pro}). 3CL^{pro} (also termed as the main protease, M^{pro}) cleaves the polyprotein at 11 distinct sites to generate many of the non-structural proteins, including RNA-dependent RNA polymerase (RdRp) and helicase (Hel), which are important in viral transcription and replication (Zumla et al., 2016). Thus, this main protease plays a critical role

in transcription and replication of the virus (Shirato et al., 2013; Zumla et al., 2016). Structure-based activity studies and various high-throughput studies have identified distinct inhibitors of SARS-CoV and MERS-CoV main proteases (Anand et al., 2003; Bacha et al., 2004; Kumar et al., 2017; Ryu et al., 2010). Therefore, it is crucial to identify novel inhibitors of SARS-CoV-2 main protease to control COVID-19 (Farag et al., 2020; Haider et al., 2020; Jin et al., 2020; Wang, 2020). On the other hand, main protease is highly conserved across coronaviruses. Therefore, it is a potential target for identification of compounds that could have broad spectrum anti-viral activity (McInnes, 2007; Mirza & Froeyen, 2020).

Covalent inhibition of therapeutic protein targets is becoming popular. Nearly 30% of the marketed drugs in the last two decades target enzymes by covalent binding mechanism (De Cesco et al., 2017; Lagoutte et al., 2017). Covalent inhibitors form a chemical bond between an electrophilic group of the ligand and a nucleophilic residue (e.g. Cys) on the target protein (Mah et al., 2014). The mechanism of action of covalent ligands is known to have two-steps. The mechanism initiates by the formation of a non-covalent complex essential to orient the electrophilic group in favor of bond formation with the nucleophilic residue of the target protein. The second step involves the formation of the covalent bond (Mah et al., 2014; Mukherjee & Grimster, 2018). The nature of the electrophile dictates if covalent inhibition is reversible or irreversible. Irreversible inhibition caused by covalent bond formation results in a remarkably different profile as compared to the reversible counterpart (De Cesco et al., 2017).

Cysteine moieties are present in a range of diverse proteins such as proteases, oxidoreductases, and kinases and thus, take part in a variety of functions (Go et al., 2015). Cysteine thiol is a highly reactive moiety due to its high electron density and polarizability and can therefore, be targeted with less reactive ligands allowing to minimize potential side effects (Flanagan et al., 2014; Poole, 2015; Schultz et al., 2006). Hence, cysteine-targeted electrophiles can be employed to understand their effect on a wide range of proteins. For example, cysteine proteases including Cathepsin B, K, and S have been targeted by covalent inhibitors (vinyl sulfones, epoxides, isothiazolones, and ketoamides, as well as other chemical groups) (Toledo Warshaviak et al., 2014). Hepatitis C virus (HCV) protease is also targeted covalently by the ketoamide groups of boceprevir (Rotella, 2013) and telaprevir (Kwong et al., 2011). In this work, given the importance of cysteine targeting covalent inhibitors, we aimed to identify cysteine thiol targeting warheads (reactive groups in compounds). Specifically, this study is focused on targeting Cys145 of the catalytic dyad of SARS-CoV-2 main protease.

Traditional methods for identification of inhibitors are expensive and time-consuming (Hou & Xu, 2004; Kolb & Sharpless, 2003; Mirza et al., 2019; Tahir ul Qamar et al., 2016; C. Wu et al., 2020; Xu, 2006). Therefore, the use of *in silico* techniques for identification of potential inhibitors has gained importance in recent years (Ece, 2020; Er et al., 2018; Mirza & Ikram, 2016; Mirza et al., 2019; Tahir ul Qamar et al., 2017; Tahir ul Qamar et al., 2019; C. Wu et al., 2020). The available small molecule databases can be utilized for structure and ligand-based virtual screening to identify novel scaffolds that could serve as hits or leads to be optimized for enhanced activity (McInnes, 2007; Mumtaz et al., 2017). In this contribution, combined virtual screening approaches were used to identify potential covalent inhibitors of SARS-CoV-2 M^{pro} from specific databases consisting of a variety of chemical warheads. In addition, FDA approved protease inhibitors were screened for the purpose of drug repurposing. Molecular dynamics (MD) simulations were utilized to investigate the binding mode of the potential inhibitors at the active site of SARS-CoV-2 main protease. The here discovered virtual hits could serve as a starting point to develop future drug candidates. Furthermore, the FDA approved protease inhibitors identified in this study warrant further *in vitro* evaluations.

2. Materials and Method

2.1. Sequence and Structural Alignment Analysis

A multiple sequence and structure alignment analysis was carried out to find out the evolutionarily conserved functional residues among SARS-CoV-2, SARS-CoV and MERS-CoV that could be further targeted for the discovery of drug hits. Sequences and 3D structures of SARS-CoV-2 (PDB ID 6LU7) (Jin et al., 2020), SARS-CoV (PDB ID 2A5I) (Lee et al., 2005) and MERS-CoV (PDB ID 5WKK) (Kankanamalage et al., 2018) main proteases were retrieved from the Protein Data Bank (PDB) (Berman et al., 2003). The 3CL^{pro} sequences were aligned using Clustal Omega (Sievers et al., 2011). To ensure broad-spectrum relevance of these protein targets, the conserved functional residues within their active sites were analyzed through structural alignment as well. Structural alignment/superposition analysis was carried out using the Pymol Molecular Graphics System v1.3 program (Seeliger & de Groot, 2010), while the interactive protein/ligand snapshots were generated with the UCSF Chimera v1.14 (Pettersen et al., 2004).

2.2. Chemical libraries preparation

Two chemical libraries were used in this study, the commercially available Asinex Focused Covalent (AFCL) library that consists of 1000 molecules (<http://www.asinex.com/>) and the FDA-approved protease inhibitor library that includes in total 116 anti-viral compounds. The structures of these 116 inhibitors were downloaded individually from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>). The AFCL database was used to identify potential compounds bearing reactive warheads, while the FDA-approved protease inhibitor database was used to search for existing anti-viral compounds that could be repurposed against SARS-CoV-2 main protease.

2.3. Covalent docking-based virtual screening protocol

Since the active site of SARS-CoV-2 main protease contains a catalytic cysteine, it is possible to target it with covalently binding compounds. For example, in the crystal structure of SARS-CoV-2 main protease (PDB ID 6LU7), the N3 peptide (a designed Michael acceptor inhibitor) (Yang et al., 2005) is covalently attached to the S_γ atom of C145. The Michael addition reaction has occurred with the C_β of the N3 vinyl group. Thus, we carried out a covalent docking protocol to screen the AFCL library, using Schrödinger's Covalent Docking (CovDock) tool (Zhu et al., 2014) as implemented in the Maestro molecular modelling package (release 2019-4; Schrödinger, LLC, New York, NY, 2019). In addition to Michael addition, we also explored the other reaction types implemented in CovDock, such as nucleophilic addition to double bond or triple bond.

Protein Preparation Wizard (Sastry et al., 2013) of Maestro was used for preparing the target structure of SARS-CoV-2 main protease (PDB ID 6LU7). The missing hydrogen atoms were added, and the hydrogen bond network was optimized with PROPKA at pH 7.0. The protonation states of histidine residues at the protease active site were optimized interactively. The catalytic His41 and His164 were set as δ -protonated (HID) while His163 was set as ϵ protonated (HIE) and His172 as double protonated (HIP). In the active conformation of the substrate-binding site, His172, rather than His163 has been proposed to have a salt bridge interaction with Glu166 (Tan et al., 2005), which is consistent with the target's active crystal conformation. The covalently bound N3 peptide inhibitor was removed from the protease active site and its structure was remodeled to the unreacted state and used in the validation docking employing the Michael addition reaction. Moreover, all water molecules were removed, and a restrained minimization was carried out using the newly optimized OPLS3e

force field (Roos et al., 2019) and the convergence criteria of 0.3 Å root-mean-square deviation (RMSD) for all heavy atoms. The AFCL library was processed with the LigPrep tool of Maestro (release 2019-4). The missing hydrogen atoms were added and different tautomeric states for each ligand were generated at pH of 7.0 ± 2.0 with Epik (Shelley et al., 2007). Maximum of two alternative stereoisomers per ligand were generated whilst retaining the specified chiralities. Finally, the OPLS3e force field was used to generate optimized low-energy 3D conformers of the ligands.

The center of the docking site was defined at the centroid of the catalytic Cys145 residue and the box size was set for ligands with a length of ≤ 20 Å. For the N3 peptide redocking, the box centroid was defined by four residues: Cys145, His163, Pro168 and Gln189 (the grid center xyz coordinates were either [-14.142582, 19.133875, 64.827842] or [-13.805531, 12.077630, 68.251552], depending on in which order the residues were picked). At first, all the library compounds were docked using the Michael addition reaction and the fast-virtual screening mode of CovDock to investigate the feasibility of forming a covalent interaction with Cys145. All compounds that successfully formed a covalent complex with the protease (as well as the N3 peptide) were then docked using the more thorough pose prediction mode. The cut-off to retain poses for further refinement was set at the docking score of 2.5 kcal/mol and maximum number of 100 poses (200 or 300 for the N3 peptide). The poses were then ranked using the Prime (Jacobson et al., 2004) MM-GBSA score in the OPLS3e force field. The Prime/MM-GBSA (molecular mechanics–generalized Born surface area) calculation employs the variable dielectric generalized Born solvation model (VSGB 2.1) (Li et al., 2011). The MM-GBSA binding free energy calculation does not take into account covalent binding and thus, the CovDock tool treats the compounds as non-covalent (capped) binders during the calculation. The above steps were then repeated for other reaction types implemented in CovDock (nucleophilic addition to a double and a triple bond, nucleophilic substitution, aryl and nitrile activated conjugate addition to alkyne). The top-ranking compounds were selected for further validation by MD simulations.

2.4. Virtual screening of FDA-approved protease inhibitors

Molecular (non-covalent) docking-based virtual screening was performed to search for potential inhibitors from the FDA-approved protease inhibitors library using AutoDock Vina v1.1.2 (Trott & Olson, 2010) in PyRx 0.8 virtual screening tool (Dallakyan & Olson, 2015). Initially, the compounds were imported into the Open Babel tool (O'Boyle et al., 2011)

implemented in PyRx 0.8 for energy minimization using the MMFF94 force field (Halgren, 1996a, 1996b). The same program was utilized to convert the compounds from the SDF format to the AutoDock Vina PDBQT format. The screening was carried out against the active site of SARS-CoV-2 (PDB ID 6LU7) (Jin et al., 2020) in which the grid box was defined to cover the catalytic dyad residues His41 and Cys145 and other essential residues within the binding pocket. The protein was kept rigid and multiple docking poses were generated for each compound. After screening, the docked poses of the compounds were ranked on the basis of the docking score (kcal/mol) generated by AutoDock Vina (Trott & Olson, 2010). The top compounds with docking scores lower than -8.0 kcal/mol were selected as hits for further analyses.

The molecular interaction of the candidate compounds was studied by re-docking each compound individually against the SARS-CoV-2 3CL^{pro} structure using AutoDock Vina v1.1.2 program (Trott & Olson, 2010). The protein structure (PDB ID 6LU7) in monomer form was prepared using Discovery Studio Visualizer v2.5 (Biovia, 2017) by removing the N3 peptide ligand and crystallographic water molecules. The Gasteiger partial atomic charges were assigned and polar hydrogen atoms were added to the protein using AutoDockTools (Huey & Morris, 2008). The same program was used to generate the PDBQT file of the protein. The docking grid of size 24 Å x 22 Å x 26 Å (X, Y, Z) was centered using the following xyz coordinates: [-1.549, 2.454, 7.117] to cover the active site. The molecular interactions and binding modes of the top-ranked docked poses were visually inspected using the Discovery Studio Visualizer v2.5 and Pymol Molecular Graphics System v1.3 (Seeliger & de Groot, 2010). In furtherance of validation of the docking parameters, the co-crystallized inhibitor N3 was re-docked non-covalently at same active site of SARS-CoV-2 3CL^{pro} to assess the ability of the AutoDock Vina docking parameters to reproduce the bioactive conformation. The inhibitory constant (pK_i) was calculated based on the docking energy score using the following equation [22-24]:

$$pK_i = 10^{[\text{Binding energy score} / 1.366]}$$

2.5. Molecular dynamics simulations

In order to investigate the binding mode of the most promising hit compounds inside the active site of SARS-CoV-2 3CL^{pro}, molecular dynamics (MD) simulations of each complex were performed for a period of 50 ns. Complex stability and interaction profile were elucidated from the simulation trajectories. All simulations were performed with the AMBER 18 simulation

package (Case et al.). The same MD simulation protocol was implemented as described previously (Ikram et al., 2019; Jabbar et al., 2018; Mirza et al., 2016), but the length of the production run was increased to 50 ns. Moreover, the parametrization was performed by creating a new molecular topology file for each compound, starting from ligand and ending in covalently bound Cys145. The tleap program of AMBER was used to generate the topology and coordinate files of the complexes. The Antechamber package of AmberTools was utilized and parameters for the ligands (using AM1–BCC charge definitions) were generated from the AMBER force field (GAFF) (Wang et al., 2004). The charges of each simulation system were neutralized by adding counter ions around the ligand-protease complex that was centered in a dodecahedral TIP3P (Jorgensen et al., 1983) water box with a 10-Å distance between the solute and the box edge. The covalent bonds were constrained using the SHAKE algorithm to maintain constant bond length (Ryckaert et al., 1977). After a stepwise minimization, the system was heated and equilibrated. Finally, a production run of 50 ns was performed at 300 K and 1 bar pressure. The time step was set to 2 fs and the trajectory snapshots were saved every 2 ps and analyzed using the CPPTRAJ program (Roe & Cheatham Iii, 2013) of AMBER.

2.6. MM-GBSA calculations using AMBER

The binding free energies (ΔG_{tol}) of SARS-CoV-2 3CL^{pro} complexed with the most promising hit compounds were calculated using the MM-GBSA method, implemented in AMBER 18. Briefly, 5000 snapshots were generated for each system from the last 20 ns stable trajectories with an interval of 2 ps. The total binding free energy is calculated as a sum of the molecular mechanics binding energy (ΔE_{MM}) and solvation free energy (ΔG_{sol}) as given below:

$$\Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{p}} + \Delta G_{\text{np}}$$

$$\Delta G_{\text{tol}} = \Delta E_{\text{MM}} + \Delta G_{\text{sol}}$$

Where, ΔE_{MM} is further divided into internal energy (ΔE_{int}), electrostatic energy (ΔE_{ele}), and van der Waals energy (ΔE_{vdw}), and the total solvation free energy (ΔG_{sol}) is contributed by the sum of polar (ΔG_{p}) and non-polar (ΔG_{np}) components. The MM-GBSA approach is well illustrated in binding free energy calculations (Hou et al., 2011) for antiviral inhibitors (Srivastava & Sastry, 2012; Tan et al., 2006).

3. Results and discussion

3.1. Structural analysis of SARS-CoV-2 main protease

The sequence alignment showed that the SARS-CoV-2 3CL^{pro} is 96.08% and 51.82% identical to SARS-CoV (PDB: 2A5I) and MERS-CoV (PDB: 5WKK) main proteases, respectively (see **Figure S1**). The sequence alignment also revealed that the catalytic dyad residues His41 and Cys145 of SARS-CoV-2 main protease, are conserved among SARS-CoV-2, SARS-CoV and MERS-CoV. Furthermore, the structural alignment/superposition of the main proteases of all 3 coronaviruses revealed conserved catalytic dyad residues His41 and Cys145, inlaid at the same position in the active site with an average RMSD of 0.12 Å (**Figure 1**). These findings were consistent with a recent study reported by Martin Stoermer (Stoermer, 2020). Therefore, the covalent docking-based virtual screening was performed against the reactive nucleophilic Cys145 to identify compounds having reactive electrophilic moieties, while the molecular (non-covalent) docking-based virtual screening was performed against the catalytic dyad residues to identify potential FDA-approved protease inhibitors for drug repurposing.

Superimposition of several co-crystallized SARS-CoV 3CL^{pro} structures bound with various inhibitors (N1, PDB ID 1WOF; I2, PDB ID 2D2D; N3, PDB ID 2AMQ; N9, PDB ID 2AMD) and recently deposited SARS-CoV-2 3CL^{pro} complexed with N3 (PDB ID 6LU7), revealed similar positions of subsites S1, S2, and S4 near the catalytic dyad that are crucial for substrate recognition (Chang et al., 2007; Yang et al., 2005). In SARS-CoV-2 protease, the side chains of Phe140, His163, Glu166 and His172 constitute the S1 site. The side chains of Pro52 and Tyr54, alongside the alkyl and methylene moieties of the side chains of Asp187 and Glu47, constitute a deep hydrophobic S2 subsite. Met165, Leu167, Phe185 and Gln192 are involved in forming S4 subsite.

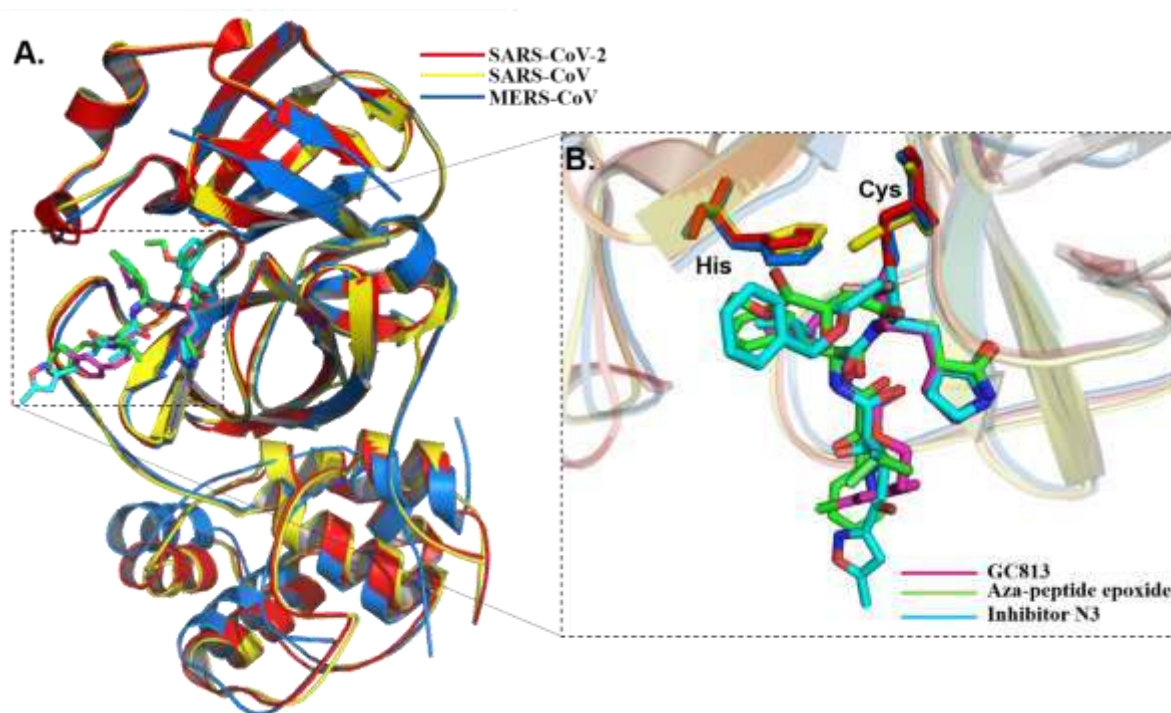


Figure 1: (A) Ribbon representation of the superimposed SARS-CoV-2 3CL^{pro} (red) (PDB ID 6LU7) bound to inhibitor N3 (cyan sticks), SARS-CoV 3CL^{pro} (yellow) bound to an aza-peptide epoxide inhibitor (green sticks) (PDB: 2A5I) and MERS-CoV 3CL^{pro} (blue) bound to GC813 (magenta). (B) Zoom-in to the main protease active site. The catalytic dyad His41, Cys145 within SARS-CoV-2, SARS-CoV and MERS-CoV main proteases is shown in red, yellow, and blue sticks, respectively.

3.2. Identification and evaluation of covalent inhibitors

3.2.1. Validation docking of the N3 peptide inhibitor

As an essential validation step before virtual screening, the co-crystallized N3 peptide was re-docked by employing the Michael addition reaction. All the docked poses of the N3 peptide successfully formed a covalent bond between the C β of the vinyl group and the thiol group of the catalytic cysteine. The best re-docked pose of the N3 peptide (using the first centroid xyz coordinates and maximum number of 300 refined poses, see the Methods) resembles the original crystal pose closely. Especially the deeply buried valine side chain and the other residues in the middle are well positioned, while the side chains at the ends of the peptide are somewhat shifted from the original pose (see **Figure 2A**). As a reference, the MM-GBSA score for this pose was -79.80 kcal/mol and the covalent docking affinity score was -9.551 kcal/mol. On the other hand, selecting only maximum 200 poses for further refinement in the docking process did not yield as good docking poses in the end.

The second centroid coordinate option gave also reasonably highly scored poses (MM-GBSA energy varying between -67 and -90 and the docking score between -7.434 and -9.994), but none of the poses was so closely resembling the crystal pose as the best pose from the docking run with the slightly different centroid coordinates. Interestingly, when we also docked the SRSSS isomer of the N3 peptide (which is SSSSS) (**Figure 2B**), the MM-GBSA score for the best pose (from maximum number of 200 retained poses and from 5 final output poses) was as low as -98.97 kcal/mol and the covalent docking affinity score was -8.361 kcal/mol. The best pose also resembled the original crystal pose very closely apart from the 5-methylisoxazol ring containing side chain that had more space to move at the outer edge of the active site (**Figure S2**). Moreover, we also assumed that different stereochemistry at this chiral center (SSSSS to SRSSS) could be utilized in drug design, particularly peptide inhibitor design for improved selectivity. In this regard, N3 already showed quite a good binding affinity. During the MD simulation, the RMSD of the N3 re-docked pose remained comparable to that of the co-crystallized pose. After small fluctuations, where the peptide tried to adopt a more favorable conformation, both poses displayed stability inside the substrate-binding pocket for the period of the last ~10 ns (**Figure 2C**). This clearly validates the covalent docking procedure with regard to the true binding mode reproducibility.

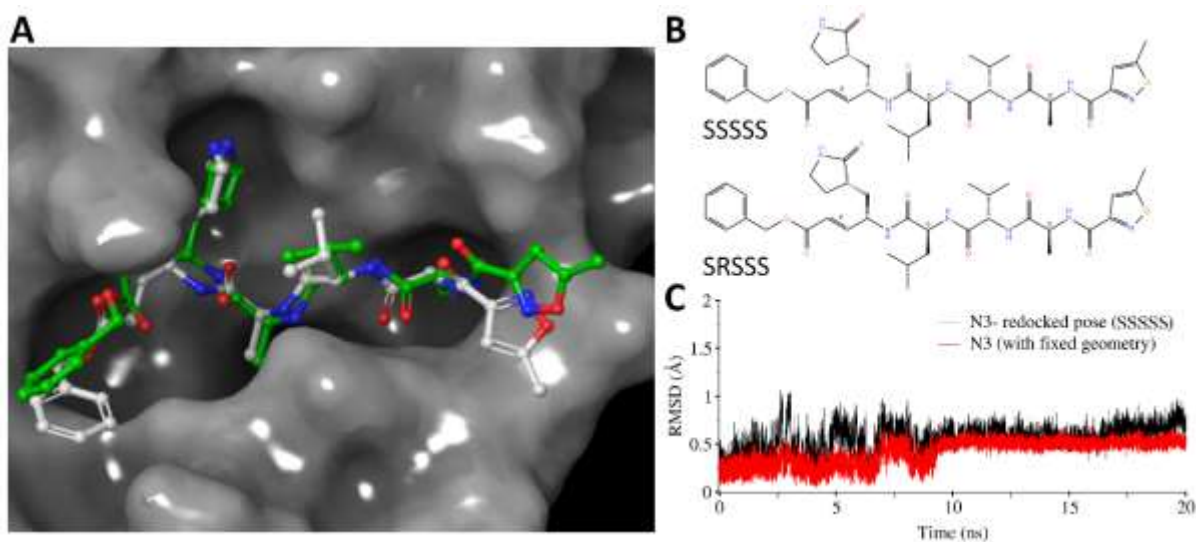


Figure 2: (A) Covalently re-docked N3 peptide (white carbons, ball and stick presentation) in the active site of the SARS-CoV-2 3CL^{pro} (PDB ID 6LU7, grey surface presentation). The original crystal pose of the peptide is shown as ball and stick presentation with green carbon atoms. Atom color code: red – oxygen; blue – nitrogen. Hydrogen atoms are omitted for clarity. (B) Two stereoisomers of N3 peptide (see the chiral center at the carbon next to the reactive

vinyl group). (C) RMSD of C α atoms of the N3 peptide (re-docked N3 in black, co-crystallized N3 in red) over a period of 20 ns.

3.2.2. Covalent docking-based virtual screening

Previous studies have demonstrated Cys145 a key residue in the active site of SARS-CoV 3CL^{pro}, which makes it an important target for covalent inhibitors (Pillaiyar et al., 2016; Pillaiyar et al., 2020; Tang et al., 2020). The virtual screening process includes two steps: (i) the selection of the candidate ligand with an appropriate pose with its reactive group in close proximity to Cys145; (ii) and a virtual chemical reaction between the reactive groups, leading to the formation of a stable covalent bond (Pillaiyar et al., 2020; Tang et al., 2020). Ligand poses within the distance cut-offs (reacting pair of atoms within 5 Å) are kept, and a covalent bond (S-C) is formed according to the reaction type. For the identification of potential covalent inhibitors from the AFCL library containing a variety of chemical warheads, six possible covalent binding reactions were used: Michael addition, nucleophilic addition to a double and a triple bond, nucleophilic substitution, and aryl and nitrile activated conjugate addition to alkyne. As an integral part of the covalent docking protocol, the free energies of binding were calculated with the Prime/MM-GBSA method for all the docked poses as described in the Methods. The compounds having the lowest Prime/MM-GBSA and/or CovDock scores were considered for further molecular inspection; thus, 19 compounds from the results of the Michael addition screening, together with 23 and 2 compounds from two other reaction types: nucleophilic addition to double bond and nitrile activated conjugate addition to alkyne respectively (details in SI, **Figures S3–S7**). After careful inspection of the structures and poses of these compounds, only the best 3 inhibitor candidates, compounds (cmp) **51**, **78**, and **223** (one from each reaction type, ranking according to the CovDock score) were selected for a detailed interaction analysis and MD simulations to analyze the binding mode and the stability of the interactions inside the enzyme active site.

3.2.3. Molecular insights of identified potential covalent inhibitors

The docking complexes of the three candidate inhibitors were investigated for the molecular interactions governed by the covalent bond formation. The structures of the candidate inhibitors, their molecular interactions within the substrate-binding pocket and the corresponding reaction sites between the chemical warheads of the candidates and the catalytic Cys145 are displayed in **Figure 3**. Cmp **223** formed a covalent bond between the electrophilic β -C atom of pyrrolidine moiety with the reactive nucleophilic thiol group of Cys145 (1.83 Å

C-S distance), and the oxygen atom of its terminal furan ring established a H-bond with the side chain nitrogen atom (NE) of His163 (2.47 Å) located at S1 subsite. At S1 subsite, the furan moiety also had van der Waals (vdW) interactions with Glu166 and Phe140, while the other hydrophobic contacts include Met165 and His41 at S4 subsite.

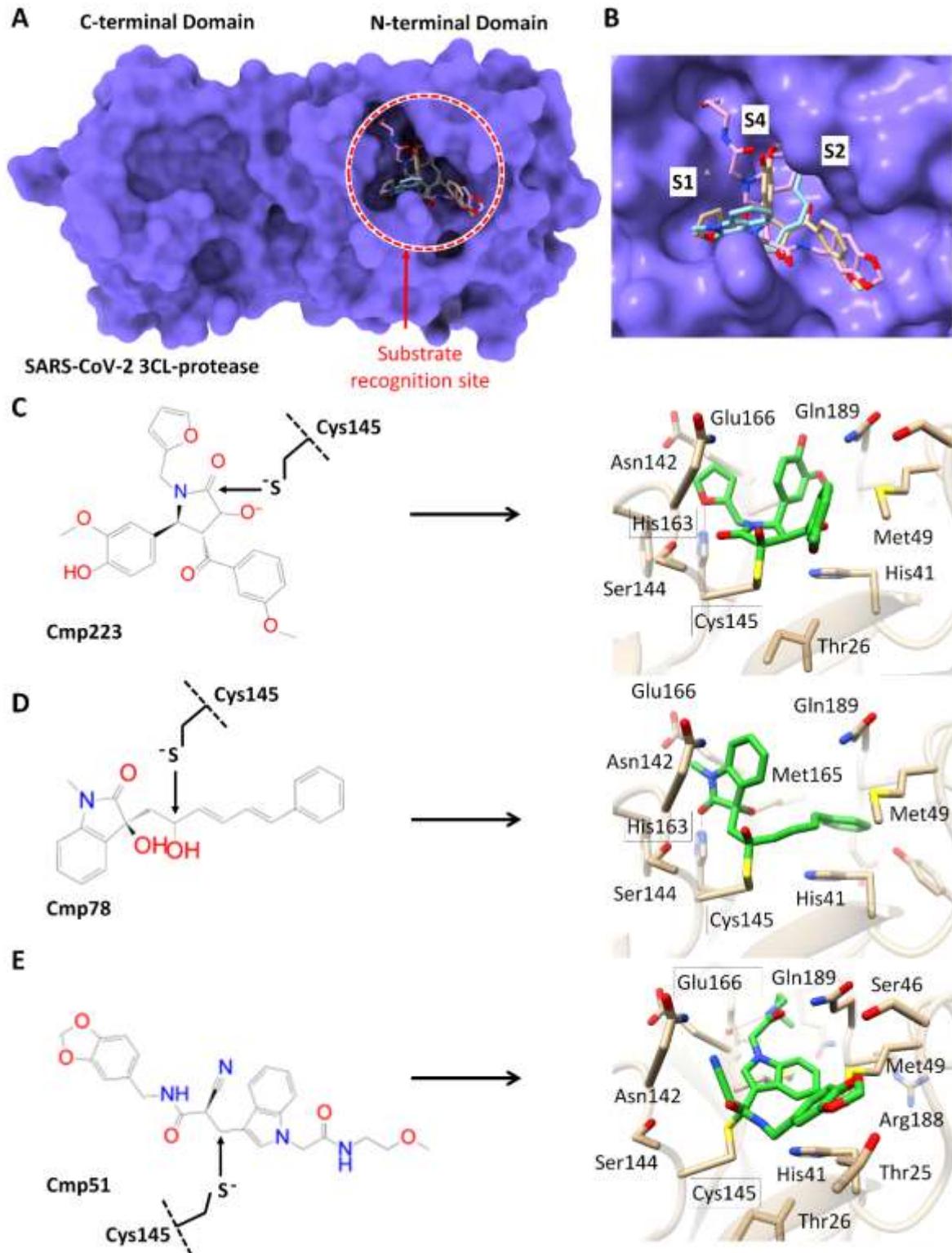


Figure 3: (A) Molecular surface representation of SARS-CoV-2 3CL^{pro} with the covalently docked hit compounds. (B) Zoomed-in SARS-CoV-2 3CL^{pro} substrate-binding pocket labelled with subsites S1, S2 and S4. (C-E) Representation of the chemical reaction of the reactive thiol group of Cys145 with the reactive nucleophilic group of the hit compounds, and the corresponding covalently docked poses (green sticks) inside the substrate-binding site of SARS-CoV-2 3CL^{pro} (white ribbon presentation, ligand-interacting amino acids are shown in sticks). Atom color code: carbon – green/white; nitrogen – blue; oxygen – red; sulfur – yellow. Hydrogen atoms are omitted for clarity.

Cmp **78** was ranked the best compound from the ‘nucleophilic addition to double bond’ screen according to the CovDock score and established a covalent bond between the β -C to the indoline moiety and the thiol group of Cys145 (1.82 Å, C-S). It also formed a H-bond between the carbonyl carbon of indoline and His163 (2.65 Å) of S1 subsite. The structural analysis revealed that the indoline moiety of **78** formed hydrophobic interactions with the S1 residues and its terminal benzene extended in the direction of the S2 subsite, where it formed π - π stacking interaction with the catalytic His41. Cmp **51** was identified as a potential covalent inhibitor from the third reaction type, nitrile activated conjugate addition to alkyne (also among the best in the Michael addition screen, according to the CovDock score). Cmp **51** established a covalent bond (1.9 Å C-S distance) with its electrophilic β -C atom present between the indole moiety and the nitrile. The indoline moiety of **51** formed hydrophobic contacts and the side chain nitrogen atom established a H-bond with Glu166 (2.51 Å) located at S1 subsite.

3.2.4. Protein stability analysis through MD simulations

MD simulations are considered to be a reliable approach in investigating the underlying dynamic consequences of protein-ligand complexes (Ahmed et al., 2018; Durdagi et al., 2018; Durrant & McCammon, 2011; Liu et al., 2017; Muneer et al., 2019; Zhou et al., 2017). In order to analyze the conformational flexibility and stability of the complexes, the RMSD of the backbone atoms of the SARS-CoV-2 main protease with reference to the initial structures of the complexes and the apo (without ligand) conformation of the main protease was plotted for the 50-ns production simulations. Overall, the main protease remained more stable in the presence of the bound inhibitors when compared to its apo conformation that displayed greater fluctuation in its backbone. As shown in **Figure 4A**, the RMSD of all three complexes increased gradually in the beginning of the MD simulation. After about 20 ns, the RMSD values converged to 1-2 Å. The initial RMSD increase in the complexes was expected due to the

flexible loop connected to the helical domain III. In general, the inhibitor interaction with the protein decreased the overall protein flexibility. Among these complexes, the M^{pro}/cmp78 complex displayed the lowest averaged RMSD value (0.5 Å). This is possibly due to the compound's best interaction profile with an additional π - π stacking interaction with the catalytic His41. However, a more detailed analysis of the flexibility of the protein backbone was possible from the per-residue root-mean-square fluctuation (RMSF). **Figure 4B** demonstrates the RMSF for the backbone atoms of each residue in M^{pro}. Overall, the fluctuations showed a similar trend with highest fluctuations in the loop regions. The highest amplitude of motion occurred in the long flexible loop of 16 residues (residues 184-199), which retained its freedom of movement and triggered dynamic effects in the helical domain III (residues 200-306). On the other hand, a decrease in flexibility was observed in the substrate-binding region, which revealed the influence of inhibitor interactions on the residues located in the substrate-binding pocket.

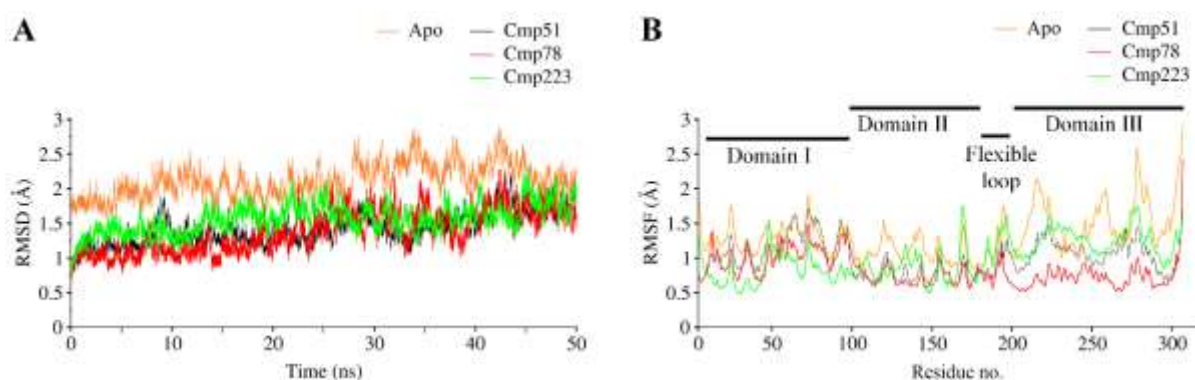


Figure 4: Molecular dynamics simulations of SARS-CoV-2 3CL^{pro}. **(A)** RMSD and **(B)** RMSF of both unbound (apo) and ligand-bound SARS-CoV-2 3CL^{pro}. The domains of SARS-CoV-2 3CL^{pro} are displayed in **(B)**.

3.2.5. Binding free energy calculations with AMBER

The molecular mechanics energies combined with the generalized Born and surface area continuum solvation (MM-GBSA) method is a well-established approach to estimate the free energy of the binding of small ligands to biological macromolecules. The approach is typically based on MD simulations of protein-ligand complexes. It has been applied to numerous systems with varying success (Chen et al., 2016; Durrani et al., 2019; Parveen et al., 2019; Sirin et al., 2014) and better covered in various reviews (Chen et al., 2016; Gohlke & Case, 2004; Gohlke et al., 2003). The AMBER MM-GBSA energetic components for each complex are tabulated in **Table 1**. Since the RMSD analysis of all complexes showed backbone stability

especially in the last 10 ns, a total of 5000 snapshots were generated and the binding free energies were calculated from the time period of 40-50 ns. For all three complexes, the contributions of van der Waals (ΔE_{vdw}), electrostatic (ΔE_{ele}), and non-polar solvation energy (ΔG_{np}) were favorable for protein-ligand interactions. All inhibitors mainly contributed to the binding energy through van der Waals interactions (< -48 kcal/mol), followed by electrostatic interactions (< -10 kcal/mol). Cmp **78** displayed the most favorable total binding free energy ($\Delta G_{\text{tot}} = -60.05$ kcal/mol) as compared to the other two, which was likely due to the additional π - π stacking interaction with the catalytic His41. Apart from the molecular mechanics contributions, the solvation energy also plays a crucial role in small ligand-protein systems and the accuracy of these calculations has been extensively studied (Genheden et al., 2010; Gohlke & Case, 2004; Gohlke et al., 2003). For all these simulation systems, the non-polar solvation energy contributed least to the total binding free energy, which is evident from the surface area accessible to the solvent as the ligands are barely exposed, while the polar solvation energy was unfavorable in all three complexes.

Table 1: Molecular mechanics generalized born surface area (MM-GBSA) binding free energy calculation of the identified covalent inhibitor candidates against SARS-CoV-2 3CL^{pro}.

Hit compounds	ΔE_{vdw}	ΔE_{ele}	ΔE_{MM}	ΔG_{p}	ΔG_{np}	ΔG_{sol}	ΔG_{tot}
Cmp223	-51.26	-12.01	-63.27	22.3	-5.77	16.53	-46.74
Cmp78	-67.45	-14.69	-82.14	28.98	-6.89	22.09	-60.05
Cmp51	-48.47	-10.52	-58.99	15.72	-6.69	9.03	-49.96

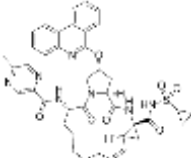
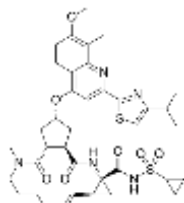
Note: ΔG_{tot} represents total binding free energy, which is the sum of molecular mechanics energy (ΔE_{MM}) and solvation free energy (ΔG_{sol}). Both ΔE_{MM} and ΔG_{sol} are further divided into internal energy (ΔE_{int}), electrostatic energy (ΔE_{ele}), and van der Waals (ΔE_{vdw}) energy in the gas phase, and polar (ΔG_{p}) and non-polar (ΔG_{np}) contributions to the solvation free energy. The units of these energy parameters are calculated in kcal/mol.

3.3. Repurposing of FDA-approved protease inhibitors

In order to identify possible repurposable SARS-CoV-2 3CL^{pro} inhibitors from the FDA-approved protease inhibitors, a library of 116 FDA-approved protease inhibitors was virtually screened against SARS-CoV-2 main protease. Initially, the docking protocol was validated by non-covalent re-docking of the co-crystallized ligand, inhibitor N3, into the active site of SARS-CoV-2 3CL^{pro}. As shown in **Figure S8**, the top-docked pose adapted a binding mode within the active site similar to that observed for the co-crystallized ligand with an RMSD value of 0.760 Å, indicating the robustness of our docking protocol. Next, molecular docking was carried out as described in the Methods, and the drugs were ranked based on their binding

energies (kcal/mol) and potential interaction with active site residues. The screening of the database revealed two potential protease inhibitors namely, paritaprevir and simeprevir with the lowest binding energy scores of -8.8 and -8.78 kcal/mol, respectively (**Table 2**).

Table 2: Chemical structures, binding energy scores, predicted inhibitory constant (pKi) and molecular interactions of identified FDA-approved protease inhibitors.

Name	Chemical Structure	Binding Energy Score (kcal/mol)	pKi (μM)	Hydrogen bond	Hydrophobic	MM-GBSA value (Amber)
Paritaprevir		-8.8	0.36	Gly143 and Cys145	Thr45, Met49, Met165 and Pro168	-47.15
Simeprevir		-8.78	0.37	Gly143 and Gln189	Pro168	-51.84

The predicted inhibitory constants (pKi) against SARS-CoV-2 3CL^{pro} based on the AutoDock Vina docking scores for paritaprevir and simeprevir were 0.36 μM and 0.37 μM , respectively. Interestingly, paritaprevir and simeprevir are FDA-approved anti-hepatitis C virus (HCV) drugs that act by targeting the NS3/4A serine protease of HCV (McConachie et al., 2016). To our knowledge, there are no clinical data available regarding the use of these two compounds to treat SARS-CoV-2, MERS-CoV or SARS-CoV. Both compounds adopted poses near the catalytic dyad, the sulfonamide group posing near Cys145 (**Figure 5**). Paritaprevir was found to form hydrogen bonds with Gly143 and Cys145, while simeprevir established hydrogen bonds with Gly143 and Gln189.

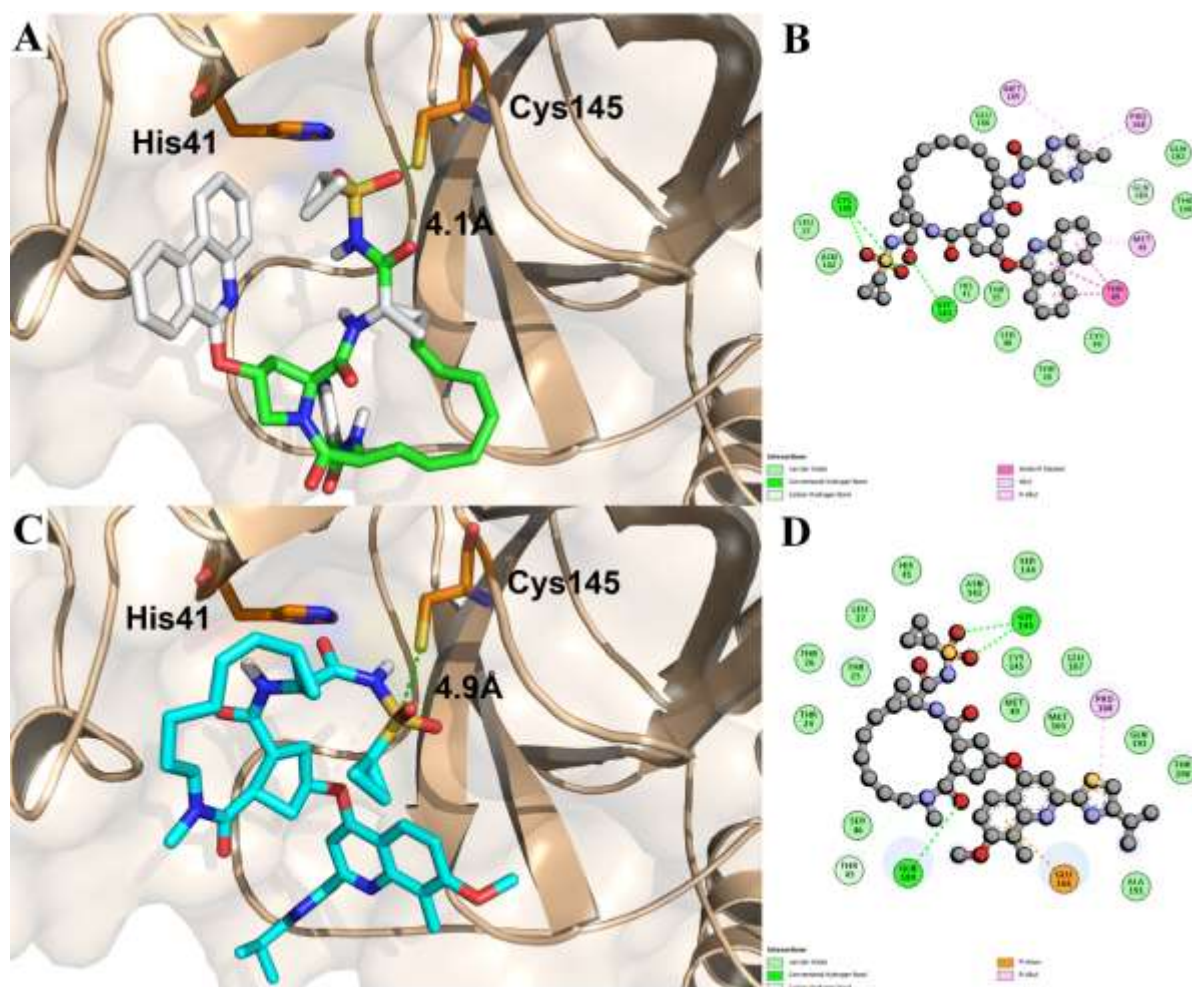


Figure 5: 3D binding mode and 2D molecular interactions of identified candidate compounds within the active site of SARS-CoV-2 3CL^{pro} (PDB ID 6LU7); (A-B) paritaprevir; (C-D) simeprevir. The catalytic dyad His41, Cys145 within the active site is shown in orange sticks.

The stability of the docked complexes was investigated with the MD simulations using the same protocol as for the covalent hit compound complexes (**Figure 6**). Overall, the complexes remained stable. The average RMSD values for the last 40 ns for paritaprevir and simeprevir were 3.2 and 3.5 Å, while the average RMSF values were 1.2 and 1.5 Å, respectively. Moreover, the total binding free energy was calculated from the last 20 ns period and revealed favorable ΔG_{tot} for paritaprevir (-47.15 kcal/mol) and simeprevir (-51.84 kcal/mol).

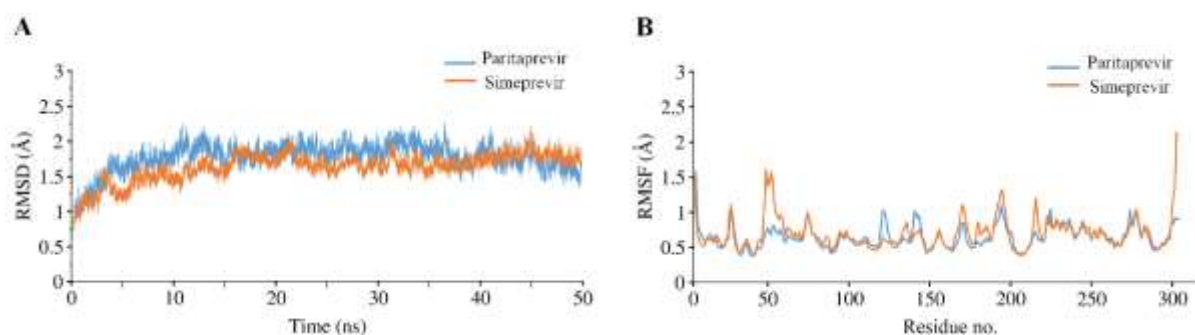


Figure 6: Molecular dynamics simulations of SARS-CoV-2 3CL^{pro}. (A) RMSD and (B) RMSF of SARS-CoV-2 3CL^{pro} with bound paritaprevir and simeprevir.

4. Conclusion

In the current study, pharmacoinformatics and MD approaches were utilized to identify potential covalent inhibitors of SARS-CoV-2 3CL^{pro} for the treatment of COVID-19. Several covalent inhibitors, together with FDA-approved paritaprevir and simeprevir, were identified as potential inhibitors of SARS-CoV-2 3CL^{pro} enzyme. The binding affinity, mechanism, and stability of binding of these compounds to SARS-CoV-2 3CL^{pro} were investigated by molecular docking and MD simulations. The potential warheads identified in this study could serve as a guideline to design covalent inhibitors targeting the catalytic Cys145. Moreover, the FDA approved anti-HCV drugs, paritaprevir and simeprevir may also play a key role in expediting the drug discovery process and could be tested in clinical trials as a treatment for COVID-19.

Acknowledgment

This work was supported by the Postdoctoral Project from Guangxi University. Authors would like to thank Guangxi University, Prince Sattam Bin Abdulaziz University, University of Leuven and Åbo Akademi University for providing the necessary tools to conduct this research. The Sigrid Jusélius Foundation, Biocenter Finland Bioinformatics and Drug Discovery and Chemical Biology networks, CSC IT Center for Science, Joe, Pentti and Tor Borg Memorial Fund and Prof. Mark Johnson and Dr. Jukka Lehtonen are gratefully acknowledged for the excellent computational infrastructure at the Åbo Akademi University.

Conflict of interest

Authors have no conflict of interest to declare.

References

- Ahmed, B., Ali Ashfaq, U., & Usman Mirza, M. (2018). Medicinal plant phytochemicals and their inhibitory activities against pancreatic lipase: molecular docking combined with molecular dynamics simulation approach. *Natural product research*, 32(10), 1123-1129.
- Ahmed, T., Noman, M., Almatroudi, A., Shahid, M., Khurshid, M., Tariq, F., ul Qamar, M. T., Yu, R., & Li, B. (2020). Coronavirus Disease 2019 Associated Pneumonia in China: Current Status and Future Prospects. *Preprints*, 2020020358. <https://doi.org/10.20944/preprints202002.0358.v3>
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*, 300(5626), 1763-1767.
- Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A., & Freire, E. (2004). Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry*, 43(17), 4906-4912.
- Berman, H. M., Bourne, P. E., Westbrook, J., & Zardecki, C. (2003). The protein data bank. In *Protein Structure* (pp. 394-410). CRC Press.
- Biovia, D. S. (2017). *Discovery studio modeling environment*. In Biovia.
- Cao, Y.-c., Deng, Q.-x., & Dai, S.-x. (2020). Remdesivir for severe acute respiratory syndrome coronavirus 2 causing COVID-19: An evaluation of the evidence. *Travel Medicine and Infectious Disease*, 101647.
- Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham Iii, T. E., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghoreishi, D., & Gilson, M. K. AMBER 2018; 2018. *University of California, San Francisco*.
- Chang, H.-P., Chou, C.-Y., & Chang, G.-G. (2007). Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride. *Biophysical Journal*, 92(4), 1374-1383.
- Chen, F., Liu, H., Sun, H., Pan, P., Li, Y., Li, D., & Hou, T. (2016). Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking. *Physical Chemistry Chemical Physics*, 18(32), 22129-22139.
- Cui, J., Li, F., & Shi, Z.-L. J. N. r. M. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3), 181-192.
- Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. In *Chemical biology* (pp. 243-250). Springer.
- De Cesco, S., Kurian, J., Dufresne, C., Mittermaier, A. K., & Moitessier, N. (2017). Covalent inhibitors design and discovery. *European journal of medicinal chemistry*, 138, 96-114.
- Durdagi, S., Tahir ul Qamar, M., Salmas, R. E., Tariq, Q., Anwar, F., & Ashfaq, U. A. (2018). Investigating the molecular mechanism of staphylococcal DNA gyrase inhibitors: A combined ligand-based and structure-based resources pipeline. *Journal of Molecular Graphics and Modelling*, 85, 122-129.

- Durrani, F. G., Gul, R., Mirza, M. U., Kaderbhai, N. N., Froeyen, M., & Saleem, M. (2019). Mutagenesis of DsbAss is Crucial for the Signal Recognition Particle Mechanism in *Escherichia coli*: Insights from Molecular Dynamics Simulations. *Biomolecules*, 9(4), 133.
- Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, 9(1), 71.
- Ece, A. (2020). Towards more effective acetylcholinesterase inhibitors: a comprehensive modelling study based on human acetylcholinesterase protein–drug complex. *Journal of Biomolecular Structure and Dynamics*, 38(2), 565-572.
- Er, M., Abounakhla, A. M., Tahtaci, H., Bawah, A. H., Çınaroğlu, S. S., Onaran, A., & Ece, A. (2018). An integrated approach towards the development of novel antifungal agents containing thiadiazole: synthesis and a combined similarity search, homology modelling, molecular dynamics and molecular docking study. *Chemistry Central Journal*, 12(1), 121.
- Farag, A., Wang, P., Ahmed, M., & Sadek, H. (2020). Identification of FDA Approved Drugs Targeting COVID-19 Virus by Structure-Based Drug Repositioning (Version 4). *chemrxiv*. <https://doi.org/10.26434/chemrxiv.12003930.v4>
- Flanagan, M. E., Abramite, J. A., Anderson, D. P., Aulabaugh, A., Dahal, U. P., Gilbert, A. M., Li, C., Montgomery, J., Oppenheimer, S. R., & Ryder, T. (2014). Chemical and computational methods for the characterization of covalent reactive groups for the prospective design of irreversible inhibitors. *Journal of medicinal chemistry*, 57(23), 10072-10079.
- Genheden, S., Luchko, T., Gusarov, S., Kovalenko, A., & Ryde, U. (2010). An MM/3D-RISM approach for ligand binding affinities. *The Journal of Physical Chemistry B*, 114(25), 8505-8516.
- Go, Y.-M., Chandler, J. D., & Jones, D. (2015). The cysteine proteome. *Free Radical Biology & Medicine*, 84, 227-245.
- Gohlke, H., & Case, D. A. (2004). Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf. *Journal of computational chemistry*, 25(2), 238-250.
- 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.
- Haider, Z., Subhani, M. M., Farooq, M. A., Ishaq, M., Khalid, M., Khan, R. S. A., & Niazi, A. K. (2020). In Silico discovery of novel inhibitors against main protease (Mpro) of SARS-CoV-2 using pharmacophore and molecular docking based virtual screening from ZINC database. *Preprints*. <https://www.preprints.org/manuscript/202002.0431/v1>
- Halgren, T. A. (1996a). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, 17(5-6), 490-519.
- Halgren, T. A. (1996b). Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *Journal of Computational Chemistry*, 17(5-6), 520-552.

- Hillaker, E., Belfer, J. J., Bondici, A., Murad, H., & Dumkow, L. E. (2020). Delayed Initiation of Remdesivir in a COVID-19 Positive Patient. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*. <https://doi.org/10.1002/phar.2403>
- 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 Modeling*, 51(1), 69-82.
- Hou, T., & Xu, X. (2004). Recent development and application of virtual screening in drug discovery: an overview. *Current Pharmaceutical Design*, 10(9), 1011-1033.
- Huey, R., & Morris, G. M. (2008). Using AutoDock 4 with AutoDocktools: a tutorial. *The Scripps Research Institute, USA*, 54-56.
- Ikram, N., Mirza, M. U., Vanmeert, M., Froeyen, M., Salo-Ahen, O. M. H., Tahir, M., Qazi, A., & Ahmad, S. (2019). Inhibition of Oncogenic Kinases: An In Vitro Validated Computational Approach Identified Potential Multi-Target Anticancer Compounds. *Biomolecules*, 9(4), 124.
- Jabbar, B., Rafique, S., Salo-Ahen, O. M. H., Ali, A., Munir, M., Idrees, M., Mirza, M. U., Vanmeert, M., Shah, S. Z., & Jabbar, I. (2018). Antigenic Peptide Prediction From E6 and E7 Oncoproteins of HPV Types 16 and 18 for Therapeutic Vaccine Design Using Immunoinformatics and MD Simulation Analysis. *Frontiers in immunology*, 9, 3000.
- Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., & Friesner, R. A. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics*, 55(2), 351-367.
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., & Duan, Y. (2020). Structure-based drug design, virtual screening and high-throughput screening rapidly identify antiviral leads targeting COVID-19. *bioRxiv*. <https://doi.org/10.1101/2020.02.26.964882>
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., & Yang, H. (2020). Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature*. <https://doi.org/10.1038/s41586-020-2223-y>
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2), 926-935.
- Kankanamalage, A. C. G., Kim, Y., Damalanka, V. C., Rathnayake, A. D., Fehr, A. R., Mehzabeen, N., Battaile, K. P., Lovell, S., Lushington, G. H., & Perlman, S. (2018). Structure-guided design of potent and permeable inhibitors of MERS coronavirus 3CL protease that utilize a piperidine moiety as a novel design element. *European Journal of Medicinal Chemistry*, 150, 334-346.
- Kolb, H. C., & Sharpless, K. B. (2003). The growing impact of click chemistry on drug discovery. *Drug Discovery Today*, 8(24), 1128-1137.
- Kumar, V., Shin, J. S., Shie, J.-J., Ku, K. B., Kim, C., Go, Y. Y., Huang, K.-F., Kim, M., & Liang, P.-H. (2017). Identification and evaluation of potent Middle East respiratory

- syndrome coronavirus (MERS-CoV) 3CLPro inhibitors. *Antiviral Research*, 141, 101-106.
- Kwong, A. D., Kauffman, R. S., Hurter, P., & Mueller, P. (2011). Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus. *Nature Biotechnology*, 29(11), 993.
- Lagoutte, R., Patouret, R., & Winssinger, N. (2017). Covalent inhibitors: an opportunity for rational target selectivity. *Current Opinion in Chemical Biology*, 39, 54-63.
- Lee, T.-W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C., Eltis, L. D., & James, M. N. G. (2005). Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *Journal of Molecular Biology*, 353(5), 1137-1151.
- Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., & Friesner, R. A. (2011). The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins: Structure, Function, and Bioinformatics*, 79(10), 2794-2812.
- Liu, K., Watanabe, E., & Kokubo, H. (2017). Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. *Journal of computer-aided molecular design*, 31(2), 201-211.
- Mah, R., Thomas, J. R., & Shafer, C. M. (2014). Drug discovery considerations in the development of covalent inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 24(1), 33-39.
- McConachie, S. M., Wilhelm, S. M., & Kale-Pradhan, P. B. (2016). New direct-acting antivirals in hepatitis C therapy: a review of sofosbuvir, ledipasvir, daclatasvir, simeprevir, paritaprevir, ombitasvir and dasabuvir. *Expert Review of Clinical Pharmacology*, 9(2), 287-302.
- McInnes, C. (2007). Virtual screening strategies in drug discovery. *Current Opinion in Chemical Biology*, 11(5), 494-502.
- Mirza, M. U., & Froeyen, M. (2020). Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. *Journal of Pharmaceutical Analysis*. <https://doi.org/10.1016/j.jpha.2020.04.008>
- Mirza, M. U., & Ikram, N. (2016). Integrated computational approach for virtual hit identification against ebola viral proteins VP35 and VP40. *International Journal of Molecular Sciences*, 17(11), 1748.
- Mirza, M. U., Rafique, S., Ali, A., Munir, M., Ikram, N., Manan, A., Salo-Ahen, O. M. H., & Idrees, M. (2016). Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Scientific Reports*, 6, 37313.
- Mirza, M. U., Vanmeert, M., Ali, A., Iman, K., Froeyen, M., & Idrees, M. (2019). Perspectives towards antiviral drug discovery against Ebola virus. *Journal of Medical Virology*, 91(12), 2029-2048.
- Mukherjee, H., & Grimster, N. P. (2018). Beyond cysteine: recent developments in the area of targeted covalent inhibition. *Current Opinion in Chemical Biology*, 44, 30-38.

- Mumtaz, A., Ashfaq, U. A., Tahir ul Qamar, M., Anwar, F., Gulzar, F., Ali, M. A., Saari, N., & Pervez, M. T. (2017). MPD3: a useful medicinal plants database for drug designing. *Natural Product Research*, 31(11), 1228-1236.
- Muneer, I., Tahir ul Qamar, M., Tusleem, K., Abdul Rauf, S., Hussain, H. M., & Siddiqi, A. R. (2019). Discovery of selective inhibitors for cyclic AMP response element-binding protein: a combined ligand and structure-based resources pipeline. *Anti-cancer Drugs*, 30(4), 363-373.
- 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.
- Parveen, A., Khan, S. A., Mirza, M. U., Bashir, H., Arshad, F., Iqbal, M., Ahmad, W., Wahab, A., Fiaz, A., & Naz, S. (2019). Deleterious Variants in WNT10A, EDAR, and EDA Causing Isolated and Syndromic Tooth Agenesis: A Structural Perspective from Molecular Dynamics Simulations. *International Journal of Molecular Sciences*, 20(21), 5282.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605-1612.
- Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y., & Jung, S.-H. (2016). An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *Journal of Medicinal Chemistry*, 59(14), 6595-6628.
- Pillaiyar, T., Meenakshisundaram, S., & Manickam, M. (2020). Recent discovery and development of inhibitors targeting coronaviruses. *Drug Discovery Today*. <https://doi.org/10.1016/j.drudis.2020.01.015>
- Poole, L. B. (2015). The basics of thiols and cysteines in redox biology and chemistry. *Free Radical Biology and Medicine*, 80, 148-157.
- Roe, D. R., & Cheatham Iii, 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.
- Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J. M., Lu, C., Dahlgren, M. K., Mondal, S., Chen, W., & Wang, L. (2019). OPLS3e: Extending force field coverage for drug-like small molecules. *Journal of Chemical Theory and Computation*, 15(3), 1863-1874.
- Rotella, D. P. (2013). The discovery and development of boceprevir. *Expert Opinion on Drug Discovery*, 8(11), 1439-1447.
- 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.
- Ryu, Y. B., Jeong, H. J., Kim, J. H., Kim, Y. M., Park, J.-Y., Kim, D., Nguyen, T. T. H., Park, S.-J., Chang, J. S., & Park, K. H. (2010). Biflavonoids from *Torreya nucifera* displaying SARS-CoV 3CLpro inhibition. *Bioorganic & Medicinal Chemistry*, 18(22), 7940-7947.

- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-aided Molecular Design*, 27(3), 221-234.
- Schultz, T. W., Carlson, R. E., Cronin, M. T. D., Hermens, J. L. M., Johnson, R., O'Brien, P. J., Roberts, D. W., Siraki, A., Wallace, K. B., & Veith, G. D. (2006). A conceptual framework for predicting the toxicity of reactive chemicals: modeling soft electrophilicity. *SAR and QSAR in Environmental Research*, 17(4), 413-428.
- Seeliger, D., & de Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of Computer-aided Molecular Design*, 24(5), 417-422.
- Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., & Uchimaya, M. (2007). Epik: a software program for pK_a prediction and protonation state generation for drug-like molecules. *Journal of Computer-aided Molecular Design*, 21(12), 681-691.
- Shirato, K., Kawase, M., & Matsuyama, S. (2013). Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. *Journal of Virology*, 87(23), 12552-12561.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., & Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1).
- Sirin, S., Kumar, R., Martinez, C., Karmilowicz, M. J., Ghosh, P., Abramov, Y. A., Martin, V., & Sherman, W. (2014). A computational approach to enzyme design: predicting ω -aminotransferase catalytic activity using docking and MM-GBSA scoring. *Journal of Chemical Information and Modeling*, 54(8), 2334-2346.
- Srivastava, H. K., & Sastry, G. N. (2012). Molecular dynamics investigation on a series of HIV protease inhibitors: assessing the performance of MM-PBSA and MM-GBSA approaches. *Journal of Chemical Information Modeling*, 52(11), 3088-3098.
- Stoermer, M. (2020). Homology Models of Coronavirus 2019-nCoV 3CLpro Protease. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.11637294.v3>
- Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C., Zhou, J., Liu, W., Bi, Y., & Gao, G. F. J. T. i. m. (2016). Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends in Microbiology*, 24(6), 490-502.
- Tahir ul Qamar, M., Alqahtani, S. M., Alamri, M. A., & Chen, L.-L. (2020). Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *Journal of pharmaceutical analysis*. <https://doi.org/10.1016/j.jpha.2020.03.009>
- Tahir ul Qamar, M., Ashfaq, U. A., Tusleem, K., Mumtaz, A., Tariq, Q., Goheer, A., & Ahmed, B. (2017). In-silico identification and evaluation of plant flavonoids as dengue NS2B/NS3 protease inhibitors using molecular docking and simulation approach. *Pakistan Journal of Pharmaceutical Sciences*, 30(6), 2119-2137.
- Tahir ul Qamar, M., Kiran, S., Ashfaq, U. A., Javed, M. R., Anwar, F., & Ali, M. A. (2016). Discovery of novel dengue NS2B/NS3 protease inhibitors using pharmacophore modeling and molecular docking based virtual screening of the zinc database. *International Journal of Pharmacology*, 12, 621-632.

- Tahir ul Qamar, M., Maryam, A., Muneer, I., Xing, F., Ashfaq, U. A., Khan, F. A., Anwar, F., Geesi, M. H., Khalid, R. R., & Rauf, S. A. J. S. r. (2019). Computational screening of medicinal plant phytochemicals to discover potent pan-serotype inhibitors against dengue virus. *Scientific Reports*, 9(1), 1-16.
- Tahir ul Qamar, M., Rehman, A., Ashfaq, U. A., Awan, M. Q., Fatima, I., Shahid, F., & Chen, L.-L. (2020). Designing of a next generation multiepitope based vaccine (MEV) against SARS-COV-2: Immunoinformatics and *in silico* approaches. *bioRxiv*. <https://doi.org/10.1101/2020.02.28.970343>
- Tan, J., Verschueren, K. H., Anand, K., Shen, J., Yang, M., Xu, Y., Rao, Z., Bigalke, J., Heisen, B., & Mesters, J. R. (2005). pH-dependent conformational flexibility of the SARS-CoV main proteinase (Mpro) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *Journal of Molecular Biology*, 354(1), 25-40.
- Tan, J. J., Zu Chen, W., & Wang, C. X. (2006). Investigating interactions between HIV-1 gp41 and inhibitors by molecular dynamics simulation and MM–PBSA/GBSA calculations. *Journal of Molecular Structure: THEOCHEM*, 766(2-3), 77-82.
- Tang, B., He, F., Liu, D., Fang, M., Wu, Z., & Xu, D. (2020). AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *bioRxiv*. <https://doi.org/10.1101/2020.03.03.972133>
- Toledo Warshaviak, D., Golan, G., Borrelli, K. W., Zhu, K., & Kalid, O. (2014). Structure-based virtual screening approach for discovery of covalently bound ligands. *Journal of Chemical Information and Modeling*, 54(7), 1941-1950.
- Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455-461.
- Wang, J. (2020). Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) Through Computational Drug Repurposing Study. *Journal of Chemical Information and Modeling*. <https://doi.org/10.1021/acs.jcim.0c00179>
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9), 1157-1174.
- Weiss, S. R., & Leibowitz, J. L. (2011). Coronavirus pathogenesis. In *Advances in virus research* (Vol. 81, pp. 85-164). Elsevier.
- Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M., & Li, X. (2020). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharmaceutica Sinica B*. <https://doi.org/10.1016/j.apsb.2020.02.008>
- 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.
- Xu, X. (2006). New concepts and approaches for drug discovery based on traditional Chinese medicine. *Drug Discovery Today: Technologies*, 3(3), 247-253.

- Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., & Hao, P. J. S. C. L. S. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci.* <https://doi.org/10.1007/s11427-020-1637-5>
- Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., & Ziebuhr, J. (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biology*, 3(10).
- Zhou, P., Silverstein, K. A., Ramaraj, T., Guhlin, J., Denny, R., Liu, J., Farmer, A. D., Steele, K. P., Stupar, R. M., Miller, J. R., Tiffin, P., Mudge, J., & Young, N. D. (2017). Exploring structural variation and gene family architecture with De Novo assemblies of 15 Medicago genomes. *BMC Genomics*, 18(1), 261. <https://doi.org/10.1186/s12864-017-3654-1>
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., & Huang, C.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270-273.
- Zhu, K., Borrelli, K. W., Greenwood, J. R., Day, T., Abel, R., Farid, R. S., & Harder, E. (2014). Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *Journal of Chemical Information and Modeling*, 54(7), 1932-1940.
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., & Lu, R. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*. <https://doi.org/10.1056/NEJMoa2001017>
- Zhu, R.-f., Gao, R.-l., Robert, S.-H., Gao, J.-p., Yang, S.-g., & Zhu, C. (2020). Systematic Review of the Registered Clinical Trials of Coronavirus Diseases 2019 (COVID-19). *medRxiv*. <https://doi.org/10.1101/2020.03.01.20029611>
- Zumla, A., Chan, J. F., Azhar, E. I., Hui, D. S., & Yuen, K.-Y. (2016). Coronaviruses—drug discovery and therapeutic options. *Nature reviews Drug discovery*, 15(5), 327.

Supplementary Information

Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CL^{pro}

Mubarak A. Alamri ^{a,†}, Muhammad Tahir ul Qamar ^{b,†,*}, Muhammad Usman Mirza ^{c,†}, Rajendra Bhadane ^{d,e}, Safar M. Alqahtani ^a, Iqra Muneer ^f, Matheus Froeyen ^c, Outi M. H. Salo-Ahen ^{d,e,*}

^a Department of Pharmaceutical Chemistry, College of Pharmacy, Prince Sattam Bin Abdulaziz University, P.O. Box 11323, Alkarj, Saudi Arabia

^b College of Life Science and Technology, Guangxi University, Nanning 530004, P. R. China

^c Department of Pharmaceutical and Pharmacological Sciences, Rega Institute for Medical Research, Medicinal Chemistry, University of Leuven, B-3000 Leuven, Belgium

^d Structural Bioinformatics Laboratory, Faculty of Science and Engineering, Biochemistry, Åbo Akademi University, FI-20520 Turku, Finland

^e Pharmaceutical Sciences Laboratory, Faculty of Science and Engineering, Pharmacy, Åbo Akademi University, FI-20520 Turku, Finland

^f School of Life Sciences, University of Science and Technology of China, Hefei 230026, P. R. China

†These authors contributed equally in this study

*Corresponding authors: Muhammad Tahir ul Qamar - m.tahirulqamar@hotmail.com; Outi outi.salo-ahen@abo.fi



Figure S1: Multiple sequence alignment of SARS-CoV-2 3CL^{pro} (PDB ID: 6LU7) against 3CL^{pro} of SARS-CoV (PDB: 2A5I) and MERS-CoV (PDB: 5WKK) created with Clustal Omega. The conserved catalytic dyad residues His41 and Cys145 are highlighted with red color stars.

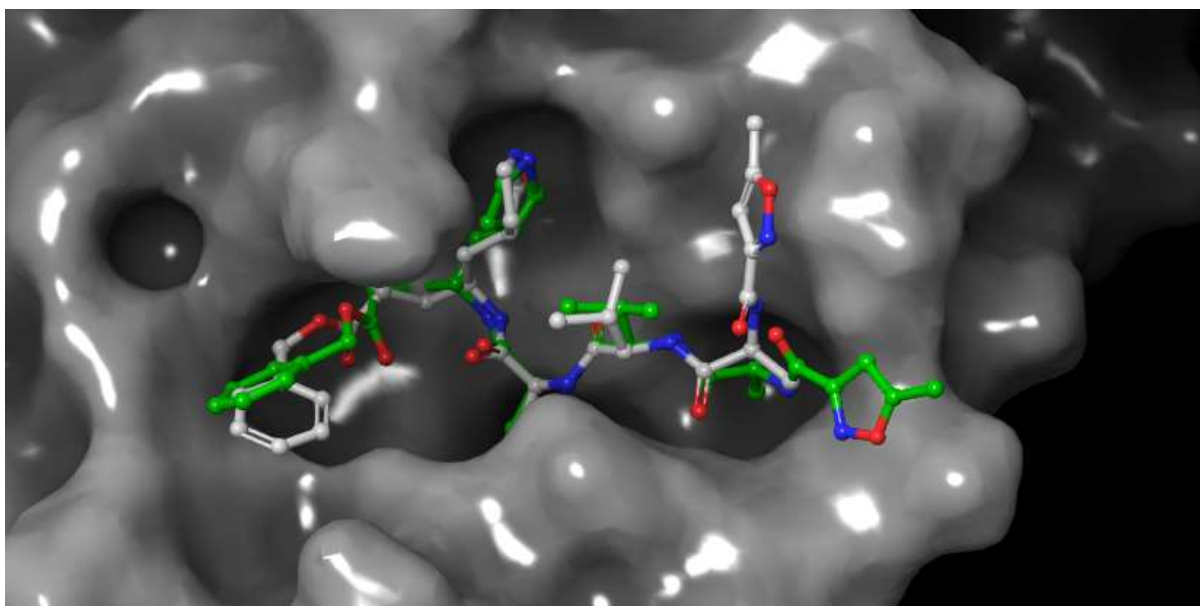


Figure S2. Covalently docked SRSSS-isomer of the N3 peptide (white carbons, ball and stick presentation) in the active site of the SARS-CoV-2 main protease (PDB ID 6LU7, grey surface presentation). The original crystal pose of the peptide is shown as ball and stick presentation with green carbon atoms. Atom colour code: red – oxygen; blue – nitrogen. Hydrogen atoms are omitted for clarity.

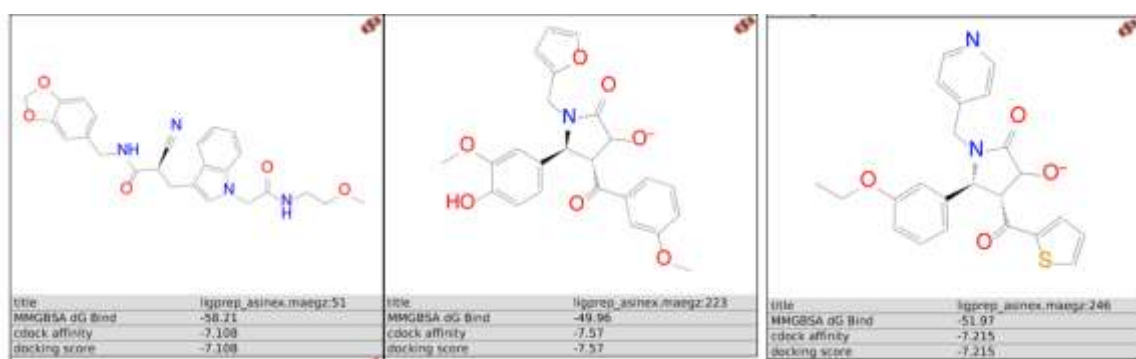


Figure S3: Compounds identified by the covalent-docking-based virtual screening using Michael addition reaction, ranked by CovDock score.

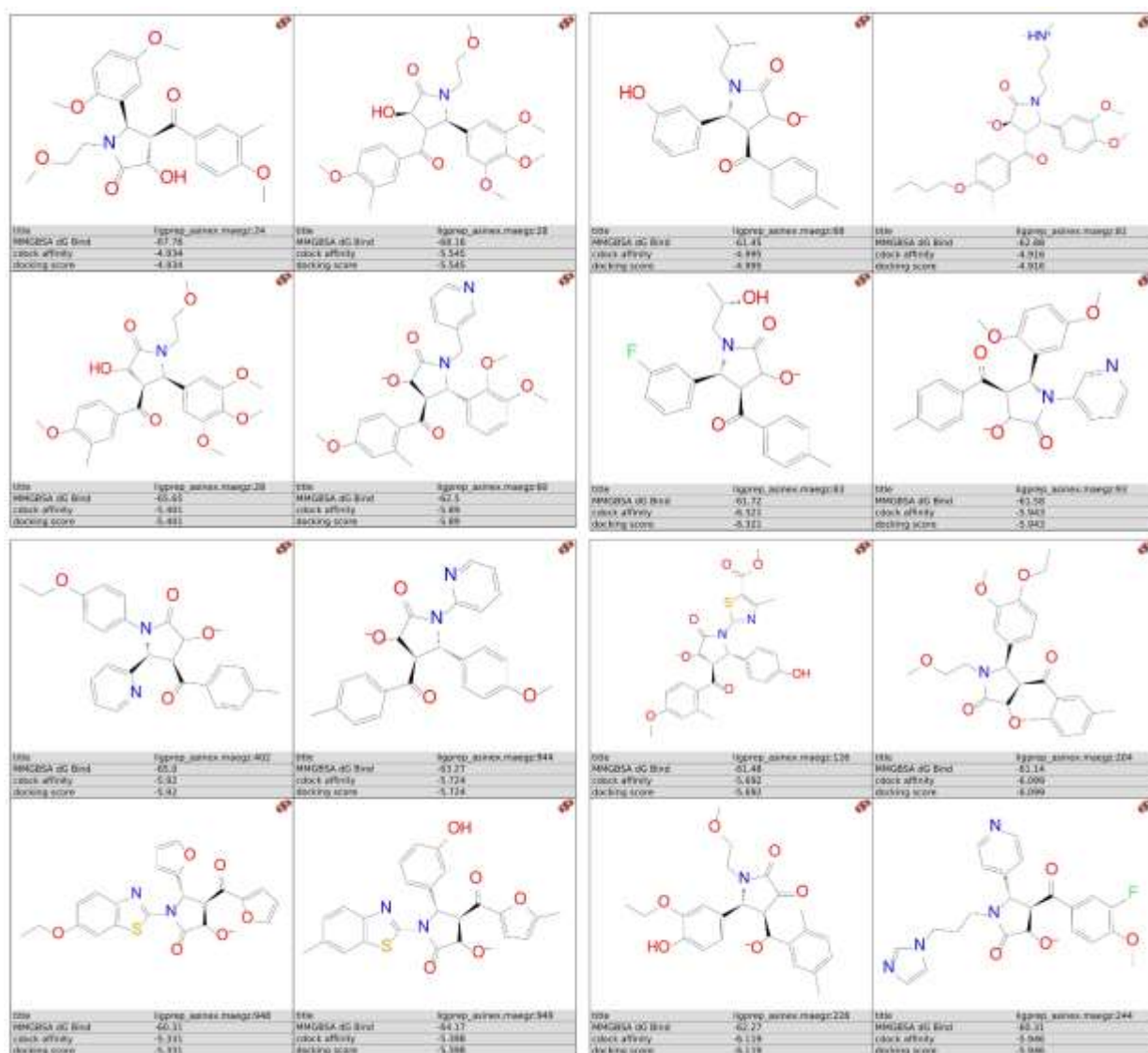


Figure S4: Compounds identified by the covalent-docking-based virtual screening using Michael addition reaction, ranked according to the Prime-MMGBSA binding energies.

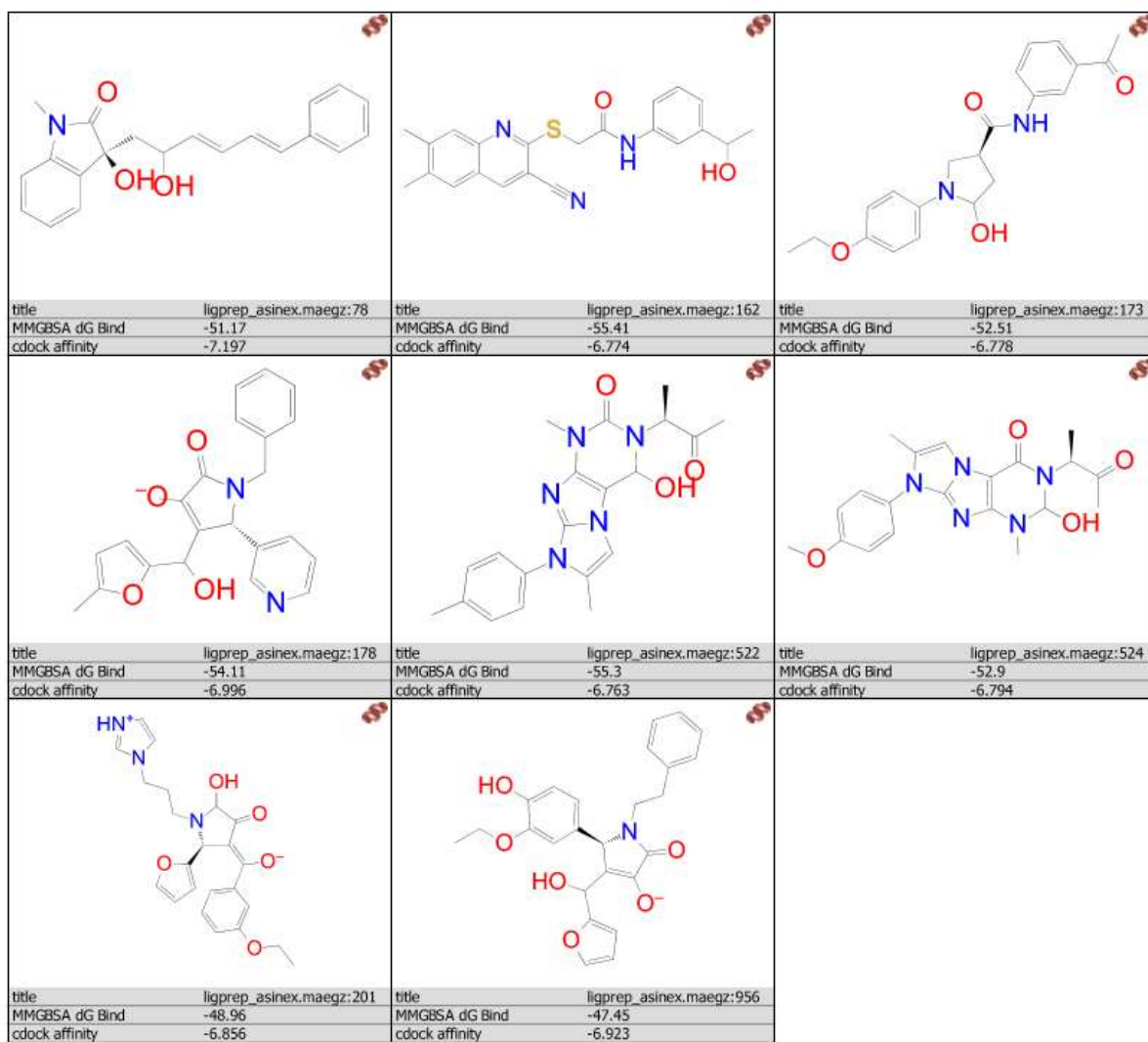


Figure S5: Compounds identified by the covalent-docking-based virtual screening using nucleophilic addition to a double bond, ranked by CovDock score.

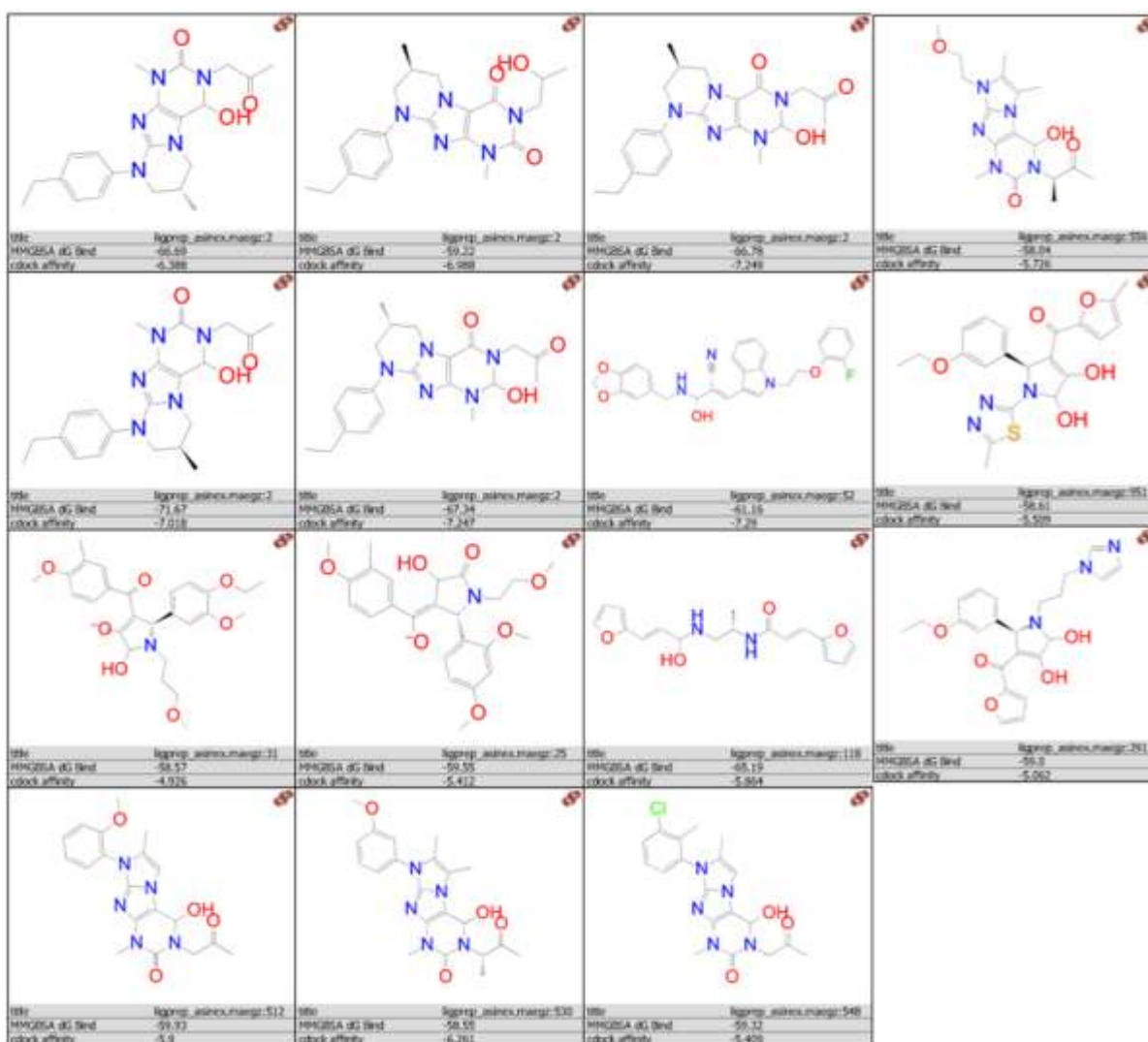


Figure S6: Compounds identified by covalent-docking-based virtual screening using nucleophilic addition to a double bond, re-ranked according to the Prime-MMGBSA binding energies.

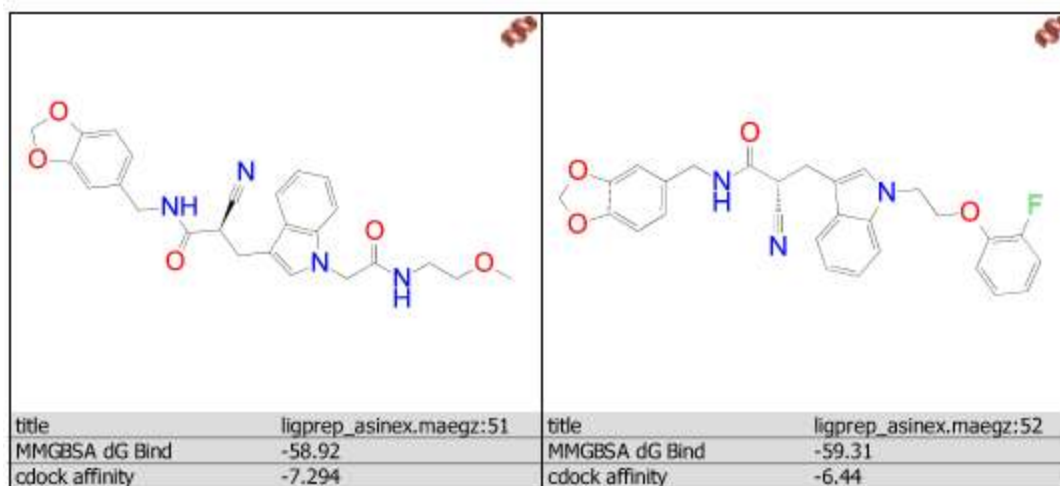


Figure S7: Compounds identified by the covalent-docking-based virtual screening using nitrile activated conjugate addition to alkyne, ranked by CovDock score.

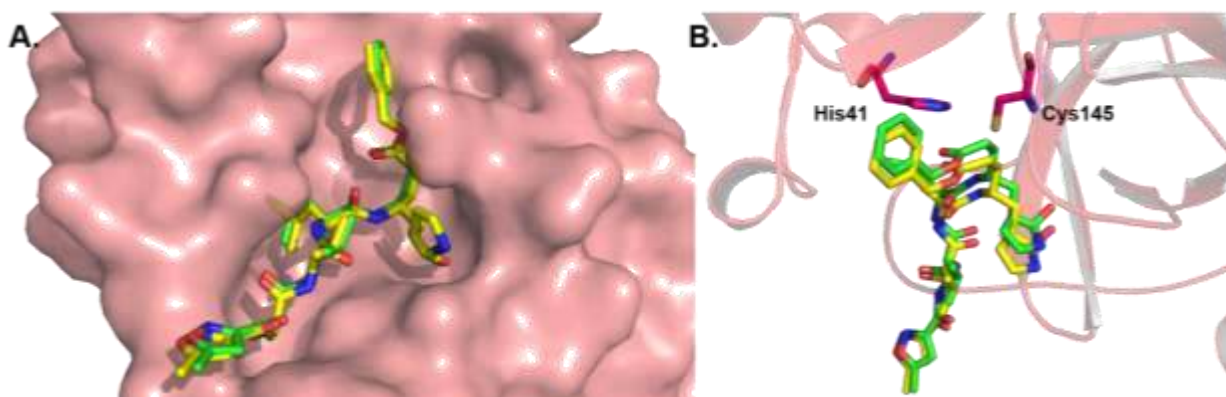


Figure S8: AutoDock Vina non-covalent docking validation. Superimposition of co-crystallized (green sticks) and best-docked pose (yellow sticks) of inhibitor N3 in the active site of SARS-CoV-2 3CL^{pro} (PDB ID: 6LU7, pink molecular surface (**A**) and pink ribbon representation (**B**)). Catalytic dyad His41 and Cys145 within the active site are shown as pink sticks. Atom colour code: carbon – green/yellow/pink; oxygen – red; nitrogen – blue; sulphur – yellow. Hydrogen atoms are omitted for clarity.