Pharmacophore Analyses of SARS-CoV-2 Active Main Protease Inhibitors Using Pharmacophore Query and Docking Study

Muhammet Karaman [a,b*]

[a]Department of Molecular Biology and Genetics, Faculty of Arts and Science, Kilis 7 Aralik University, Kilis, Turkey

[b]Advanced Technology Application and Research Center (ATACR), Kilis 7 Aralik University, Kilis, Turkey

***Correspondence:**

Muhammet Karaman Ph.D.

Department of Molecular Biology and Genetics, Faculty of Arts and Science,

Kilis 7 Aralik University, 79000 Kilis, Turkey.

Email: mhmmtkaraman@gmail.com

Abstract

The coronavirus disease (COVID-19) pandemic is the most important current problem in the world. Many researchers have focused on approved drugs or new drug candidates to combat the pandemic. Structural and nonstructural proteins of SARS-CoV-2 have been detected as targets for prevention of host cell infection or blockade of vital function. The main protease that plays an essential role in the virus life cycle is the optimal target. To design new inhibitors against the enzyme, the catalytic active site and substrate-binding site should be well analyzed. In this study, we generated a pharmacophore model using the cocrystallized pose of an active SARS-CoV-2 main protease inhibitor. According to the model, the inhibitor inhibits the enzyme via three hydrogen bond donors, two hydrogen bond acceptors and two aromatic ring interactions. Moreover, we docked reported active inhibitors of the main protease into the catalytic active site and detected matches between their pharmacophore models. The results showed that two close hydrogen acceptor/donor atom pairs and an aromatic ring are essential for enzyme inhibition.

Keywords: SARS-CoV-2, main protease, pharmacophore model, docking

1. Introduction

Novel coronavirus, which leads to a lung disorder, appeared in the Wuhan state of China in November 2019 and spread worldwide. The novel coronavirus has led to 4,529,027 infections and 307,565 deaths [1]. The genome of the novel coronavirus shares a highly similar gene sequence (approximately 82%) with other SARS coronaviruses [2]. For this reason, it is named as SARS-CoV-2. The 30 kb genome of SARS-CoV-2 expresses proteins, including spike (S), envelope (E), membrane (M), nucleocapsid (N) and nonstructural proteins. The coronavirus genome contains many open reading frames (ORFs) coding two-thirds of the genome. The first (ORF1a/b) of these directly translates two large polyproteins known as pp1a and pp1ab. These polyproteins are processed into 16 nonstructural proteins by the main protease (chymotrypsin-like protease), and subgenomic RNAs coding the structural proteins are produced from these nonstructural proteins [3]. Therefore, the main protease is a vital part of the coronavirus life cycle. Viral replication can be stopped by inhibiting enzyme activity. Inhibitors should be nontoxic to humans due to different cleavage specificity between the main proteases of human and coronavirus.

Many researchers have focused on new inhibitor design and identification of effective ready-to-use drugs for inhibition of the main protease. Several studies have demonstrated that α-ketoamides are effective inhibitors against the main protease [4, 5]. Numerous studies have also discovered that some approved drugs may inhibit enzyme activity [6-8]. However, the pharmacophore site of the enzyme is still not clear. As pharmacophore features are important for structural-drug design, we elucidated pharmacophore features using active main protease inhibitors.

2. Results and Discussion

Many researchers have focused on efforts to overcome the coronavirus (COVID-19) pandemic. For this purpose, researchers have focused on inhibition of SARS-CoV-2 proteins by using FDA-approved drugs or synthesizing novel compounds. The main protease of SARS-CoV-2 is one of the most important targets for these studies. Knowledge about the pharmacophore site is essential to understand the inhibitory mechanism of the approved drug and to design a novel inhibitor against the main protease. For these reasons, we generated a pharmacophore model of the main protease using a crystal structure cocrystallized with an inhibitor, and we determined the effectiveness by *in vitro* experiments. To generate a pharmacophore model, it is important to have a crystal structure of the enzyme in complex with a potent inhibitor. The present model consisted of three hydrogen bond donors, two hydrogen bond acceptors and two aromatic ring features (Figure 1 left-top). His41 and Cys145 residues are responsible for the catalytic activity of the enzyme [5]. Benzyl moiety contributes to enzyme inhibition by interacting with the His41 residue through an aromatic interaction. Therefore, the aromatic plane feature (R1) is the most important part of the pharmacophore model, and the distance between features was calculated (Figure 1 left-bottom). The distance between the A2 and D3 features represented interactions between the Glu166 residue and atoms of 2-dihydropyridin and carbamate moieties, and this distance was calculated to be 2.3 Å. These interactions are also important for enzyme inhibition because two promoters create dimers by interacting with their N-finger via the Glu166 residue, and dimerization generates an active enzyme [9]. The distance between the A1 and D2 features represented interactions between the Hie163 and Phe140 residues and atoms of oxopyrrolidin moiety, and this distance was calculated to be 2.6 Å. The closeness between features provides structural interactions with neighbor residues, and this distance was approximately 6 Å. The pharmacophore model was validated using pharmacophore models of best-posed inhibitors, and the validation results are presented in Table 1. The validation results indicated that all active

ligands had a good fitness score and matched well with features. The merged pharmacophore models are shown in Supplementary Figure 1.

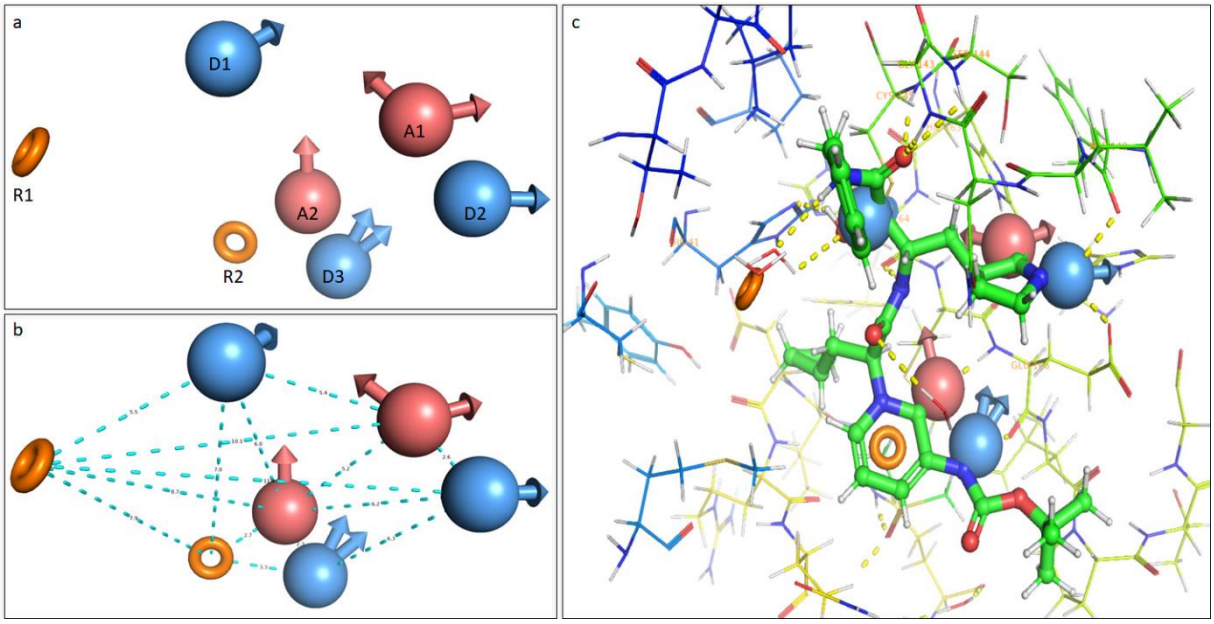

Figure 1. The pharmacophore model generated a cocrystallized pose of a potent main protease inhibitor (EC_{50} =4-5 μM, PDB code:6Y2F). Hydrogen bond acceptors are shown as red vectored spheres, and hydrogen bond donors are shown as blue vectored spheres. Aromatic planes are shown as orange circles **(a)**. The distances between the pharmacophore features were calculated. Distances are shown as cyan dashed lines **(b)**. Cocrystallized ligand and the pharmacophore features in the catalytic active site. Hydrogen bonds are shown as yellow dashed lines **(c)**.

Table 1. Pharmacophore model validation results.

Active inhibitors	Matches	RMSD (Å)
11a	6 of 7	0.477
11b	3 of 7	1.224
11r	4 of 7	0.655
13a	6 of 7	0.359
13b	4 of 7	1.424
14b	6 of 7	0.328
N3	3 of 7	0.329
Atazanavir	No matching	-

In addition, the pharmacophore features were analyzed with docked active inhibitors. For this purpose, we docked all active inhibitors into the catalytic active site of the enzyme by centering residue 166. The binding affinities of the active inhibitors were analyzed based on the Glide score and binding score. The pose with the highest Glide score in a negative direction was selected as the best pose. The docking scores and reported EC_{50} values are presented in Table

2. The scores indicated that all active ligands exhibited high and similar binding affinity against the enzyme. Furthermore, the scores were compatible with *in vitro* experiment results [5, 10].

Table 2. Binding affinity scores of the compounds according to the Glide score and ΔG binding score. The table has also included previously reported *in vitro* results.

Active Inhibitors	Glide Score (kcal/mol)	ΔG Binding Score (kcal/mol)	*In Vitro* Results (μM)
11a	-10.954	-114.025	0.053 ± 0.005 (IC_{50}) 0.41 ± 0.08 (EC_{50}) [4]
11b	-11.631	-119.171	0.040 ± 0.002 (IC_{50}) 0.33 ± 0.09 (EC_{50}) [4]
11r	-10.080	-105.179	0.18 ± 0.02 (IC_{50}) [5]
13a	-12.753	-134.788	2.39 ± 0.63 (IC_{50}) [5]
13b	-9.331	-98.161	4 – 5 (EC_{50}) [5]
14b	-10.610	-106.473	Weak the inhibitory potency [5]
N3	-9.305	-110.492	16.77 ± 1.70 (EC_{50}) [10]
Atazanavir	-5.207	-62.460	2.0 ± 0.12 (EC_{50}) [11]

We then analyzed the binding mode and pharmacophore matches of the best-posed inhibitors. As shown in Figure 1, all inhibitors exhibited similar interaction with active site residues. The best pose of inhibitor **11a** had nine hydrogen bonds with catalytic active site residues (Figure 2a). The nitrogen atoms of γ-lactam, propane-2-yl amino and indole moieties donated hydrogen to Phe140, Hie164 and Glu166. The oxygen atoms of γ-lactam and carboxamide moieties accepted hydrogen from Hie163 and Glu166. These interactions matched with the pharmacophore feature as shown in Figure 3a. Other interactions of the best pose showed overlapping interaction between the inhibitor and main protease, and they may stabilize the inhibitor into the catalytic active site as reported by Dia et al. [4]. The cyclohexyl moiety was surrounded by His41, Cys44, Met49, Try54, Asp187, and Arg188. The hydrophobic interactions matched with the aromatic plane feature, which formed hydrophobic interactions between the inhibitor and residues. Another best-scored inhibitor, **11b**, was constructed of nine hydrogen bonds with catalytic active site residues both directly and indirectly (Figure 2b). The

nitrogen atoms of γ-lactam and indole moieties donated hydrogen to the same residues. However, the nitrogen atom of the propane-2-yl amino moiety donated hydrogen to water in catalytic active sites. The oxygen atoms of γ-lactam and carboxamide moieties of inhibitor **11b** accepted hydrogen to Hie163 and Gln189. Similar to inhibitor **11a**, the interactions of inhibitor **11b** matched with the pharmacophore feature. Changing the fluorophenyl moiety of cyclohexyl resulted in near interactions between the inhibitor and key residues (Figure 3b).

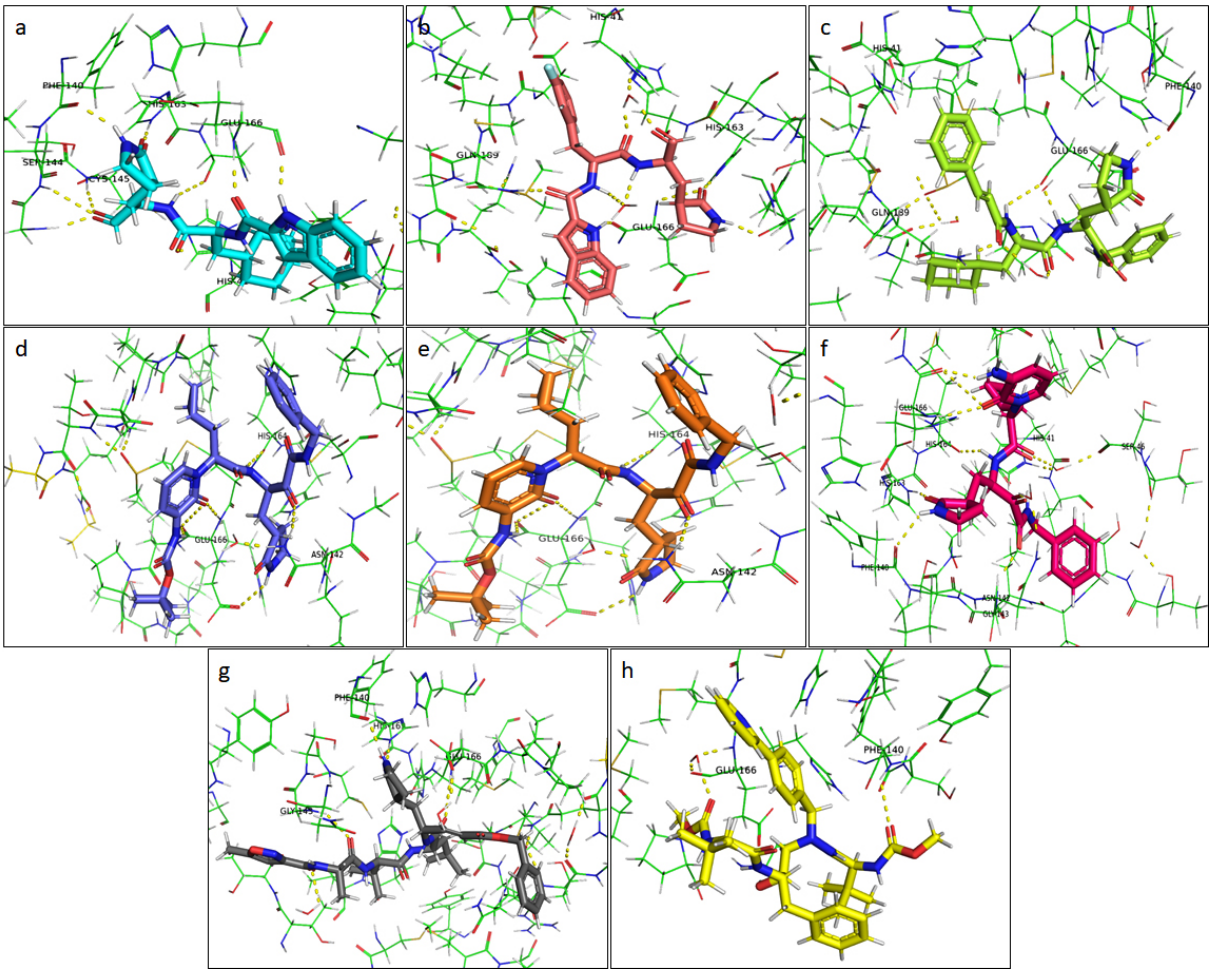

Figure 2. 3D binding mode of best-posed main protease inhibitors. Hydrogen bonds are represented as yellow dashed lines, and aromatic hydrogen bonds are represented as turquoise dashed lines.

Inhibitor **11r,** which is highly effective against MERS-CoV and SARS-CoV [12], is accepted as a broad-spectrum inhibitor of coronavirus due to effect against SARS-CoV-2 main protease [5]. Inhibitor **11r** includes cinnamamido and cyclohexylpropanamido moieties instead of γ-lactam and indole moieties, thereby differing from **11a**. The differences caused a π-π stacking

interaction between cinnamamido and His41, and it matched with the aromatic plane feature (Figure 3c). On the other side, the nitrogen atom of the cinnamamido moiety donated hydrogen to Gln189, and its oxygen atom accepted hydrogen from Glu166. In addition, the nitrogen atom of the γ-lactam moiety donated hydrogen to Phe140 (Figure 2c). These interactions matched with the D3, A2 and D2 features. To improve solubility and decrease binding to plasma proteins of the compounds, inhibitor **13a** was obtained by modifying inhibitor **11r**. The modification resulted in many interactions with key residues in the catalytic active sites (Figure 2d). Therefore, inhibitor **13a** exhibited the highest binding affinity among the active inhibitors, and it also matched two hydrogen bond acceptor and two hydrogen bond donor features (Figure 3d). As shown in Figure 2d, the cyclopropyl moiety was too small to stabilize the inhibitor in the catalytic active site by interacting with hydrophobic residues, resulting in less activity compared to **11r** inhibitors [5].

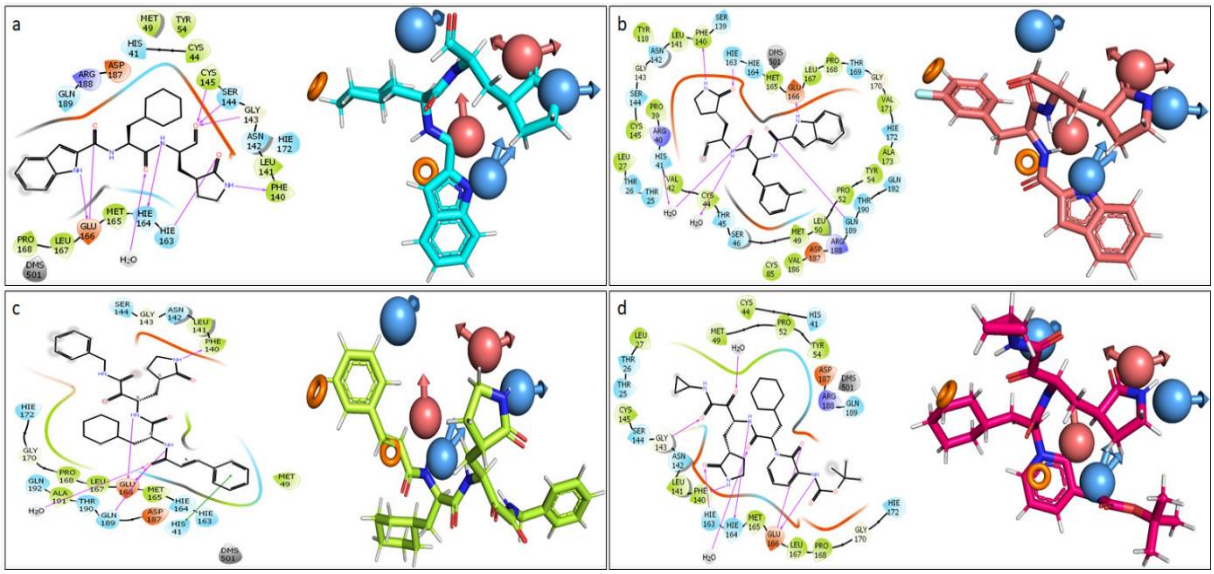

Figure 3. Binding modes and pharmacophore feature match of the best-posed active inhibitors, including **11a** (a), **11b** (b), **11r** (c) and **13a** (d). Hydrogen bond is represented as purple arrow. π-π sticking interaction is represented as green line. Hydrogen bond acceptors are shown as red vectored spheres, and hydrogen bond donors are shown as blue vectored spheres. Aromatic planes are shown as orange circles.

Inhibitor **13b** was designed from **13a** by switching the cyclohexyl and cyclopropyl moieties. The small changes hit the target from twelve due to the perfect fit of the inhibitor in the catalytic

active site because the cyclohexyl moiety contributed stabilization of the inhibitor in the catalytic active site by surrounding His41, Val42, Cys44 and Met49. The oxygen atom of 3,4-dioxo and dihydropyridin moieties accepted hydrogen from the Cys145 and Glu166 residues. The nitrogen atoms of γ-lactam, butan-2-yl amino and *tert*-butyl carbamate moieties donated hydrogen to Hie164 and Glu166 (Figure 2e). Therefore, **13b** perfectly matched another aromatic ring and acceptor/donor hydrogen bond feature as shown in Figure 4a. Another α-ketoamide inhibitor, **14b**, exhibited similar interaction with inhibitor **13b** via the dihydropyridin moiety. The γ-lactam moiety formed hydrogen bonds with Phe140 and Hie163. Moreover, the 3,4-dioxo moiety accepted hydrogen from Asn142 and Gly143, and the butan-2-yl amino moiety formed a hydrogen bond with His41 and Hie164 as shown in Figure 2f. These interactions matched all pharmacophore features (Figure 4b). However, **14b** did not contain a Boc group, which caused it to differ from **13a** and **13b** because the Boc group forces some residues to move outward due to interaction between the Boc group and Pro168. Removing the Boc group from **13a** and **13b** led to stronger interaction with active site residues of **14b**. The present results were compatible with previously reported *in vitro* results [5]. Removal of the Boc group also led to less inhibitory potency compared to other α-ketoamide inhibitors because the resulting compound cannot cross the cellular membrane. Finally, inhibitor N3 exhibited slightly lower binding affinity than other active inhibitors because it did not accurately bind in the catalytic active site due to its large structure. As shown in Figure 2g, the inhibitor formed hydrogen bonds with Phe140, Gly143, Hie163 and Glu166. The interactions matched with two hydrogen acceptor features and a hydrogen donor feature (A1, A2 and D2) as shown in Figure 4c. The present *in silico* results were verified by *in vitro* results. The FDA-approved drug, Atazanavir, exhibited week inhibitor potency against the SARS-CoV-2 main protease. Atazanavir formed only two hydrogen bonds with Phe140 and Glu166 as a hydrogen acceptor, and interactions between the methyl ester and Glu166 were indirect interactions (Figure 2h).

The interacting residues were similar to those residues reported by Fintelman-Rodrigues et al. [11]. However, Atazanavir did not match any generated pharmacophore features, and it had weak binding affinity compared to the other main protease inhibitors as shown in Table 1. However, the present *in silico* results did not match the *in vitro* studies reported by Fintelman-Rodrigues et al., but they were confirmed by *in vitro* studies reported by Jeon et al. [13]. Moreover, the present *in silico* results were compatible with other *in silico* studies [14-16].

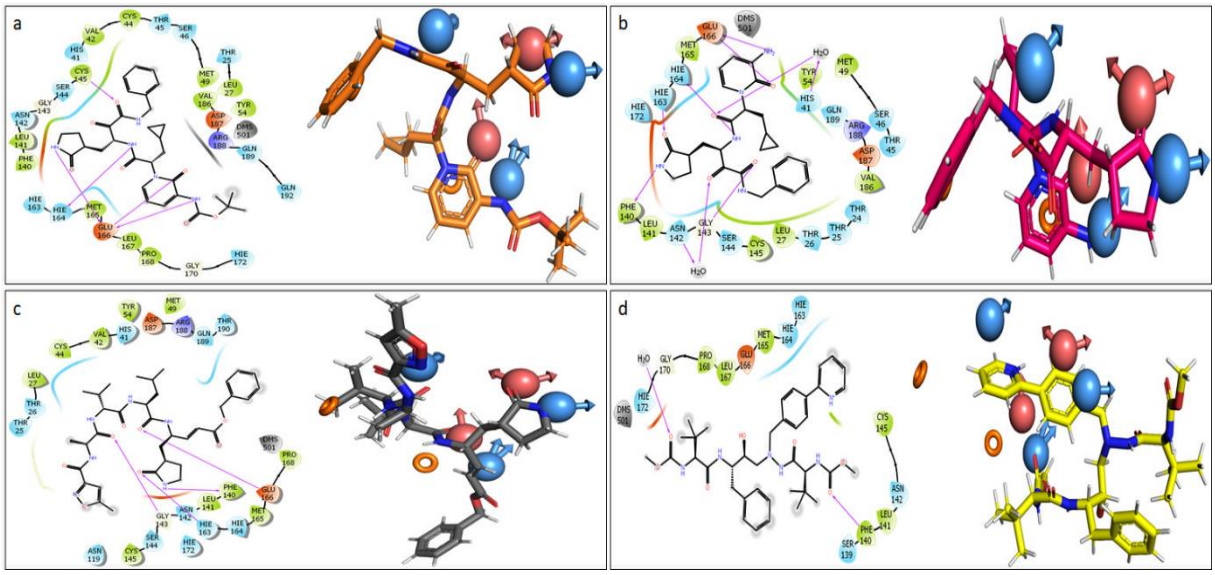

Figure 4. Binding modes and pharmacophore feature match of the best-posed active inhibitors, **13b** (a), **14b** (b), **N3** (c) and **Atazanavir** (d). Hydrogen bond is represented as purple arrow. π-π sticking interaction is represented as green line. Hydrogen bond acceptors are shown as red vectored spheres, and hydrogen bond donors are shown as blue vectored spheres. Aromatic planes are shown as orange circles.

In silico studies have shown that compounds that interact with His41, Phe140, Asn142, Cys145, His163 and Glu166 through hydrogen bonds exhibit potent inhibitory effects against the SARS-CoV-2 main protease. The distance between hydrogen acceptor/donor atoms that interact with Phe140 and His163 should be approximately 2.6 Å. The distance between other hydrogen acceptor/donor atom pairs should be approximately 2.3 Å. The distance is shorter than that of other atom pairs due to interacting atoms of one residue, namely, Glu166. In addition, it is important that compounds include an aromatic ring that is located 8-11 Å. The aromatic ring contributes to enzyme inhibition by interacting with His41 and other hydrophobic residues.

3. Conclusion

The approved drugs are an excellent starting point to overcome the coronavirus (COVID-19) pandemic because they are safe and ready for use in clinical treatment. However, the location of the approved drugs in the catalytic active site is important for evaluating fitness. In addition to evaluating approved drugs, novel drug design is necessary for blocking the coronavirus life cycle. Thus, we analyzed the inhibition mechanism of active inhibitors on the main protease and generated a pharmacophore model. The computational studies provide information about the key residues, pharmacophore features and distance between features. The features can be used as a valuable directory to synthesize main protease inhibitor candidates with optimal geometry and chemical functionality necessary to fit the model features.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that influenced the work reported in this paper.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

4. Materials and Methods

4.1. Protein and Ligand Preparation

X-ray crystal structures of SARS-CoV-2 main protease (PDB Code: 6Y2E and 6Y2F) were downloaded from the RCSB protein data bank. The structures were prepared using the protein prep wizard tool of Maestro [17]. The protein preparation workflow was performed in several steps, including bond order assignment, charge assignment, hydrogen atom addition, missing

side chain filling, amino acid ionization and energy minimization according to previously reported studies [18, 19]. Ligand files were downloaded from PubChem, UCSF DUD and Zinc Database, and they were prepared using the Ligprep tool of Maestro [20]. The 3D structures of ligands were opened in Maestro, and they were prepared by obtaining correct molecular geometries and protonation state at pH 7.0 ± 2.0.

4.2. Induced-Fit Docking Study

The effectiveness of the main protease inhibitors was determined by *in vitro* experiments and docked into the main protease using the induced-fit docking method [21]. Induced-fit docking was performed in three steps, including centroid generation, side-chain trimming and refinement of ligand pose, according to previously reported studies [22].

4.3. Pharmacophore Generation

The energy-optimized pharmacophore (e-pharmacophore) model was generated using the phase tool of Maestro [23] according to the proposed method by Ece [24]. The x-ray crystal structures of the main protease were complexed with a potent inhibitor (O6K-13b) for E-pharmacophore generation. The pharmacophore model was represented by hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic site (H) and aromatic ring (R) features. Additionally, other pharmacophore models were generated using pharmacophore models of best-scored active ligands, and the first model was validated using these models.

References

[1] Who, Coronavirus disease (COVID-2019) situation reports, in, 2020.

[2] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C.L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R.D. Jiang, M.Q. Liu, Y. Chen, X.R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F.X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature, 579 (2020) 270-273.

[3] P.S. Masters, The Molecular Biology of Coronaviruses, Advances in Virus Research, 66 (2006) 193-292.

[4] W. Dai, B. Zhang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.K. Zhang, Y. Xu, H. Yang, H. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, Science, (2020).

[5] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors, Science, 368 (2020) 409-412.

[6] M. Kandeel, M. Al-Nazawi, Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease, Life Sci, 251 (2020) 117627.

[7] A.T. Ton, F. Gentile, M. Hsing, F. Ban, A. Cherkasov, Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds, Mol Inform, (2020).

[8] A. Kumar, S. C.S, A. J, D. Francis, S. C, Drug Repurposing Against SARS-CoV-2 Using E-Pharmacophore Based Virtual Screening and Molecular Docking with Main Protease as the Target, Chem Rxiv, (2020).

[9] K. Anand, G.J. Palm, J.R. Mesters, S.G. Siddell, J. Ziebuhr, R. Hilgenfeld, Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain, EMBO J, 21 (2002) 3213-3224.

[10] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of M(pro) from COVID-19 virus and discovery of its inhibitors, Nature, (2020).

[11] N. Fintelman-Rodrigues, C.Q. Sacramento, C.R. Lima, F.S.d. Silva, M.M. André C. Ferreira, C.S.d. Freitas, V.C. Soares, S.d.S.G. Dias, J.R. Temerozo, M. Miranda, A.R. Matos, F.A. Bozza, N. Carels, C.R. Alves, M.M. Siqueira, P.T. Bozza, T.M.L. Souza, Atazanavir inhibits SARS-CoV-2 replication and pro-inflammatory cytokine production, BioRxiv, (2020).

[12] L. Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E.J. Snijder, H. Liu, R. Hilgenfeld, alpha-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment, J Med Chem, 63 (2020) 4562-4578.

[13] S. Jeon, M. Ko, J. Lee, I. Choi, S.Y. Byun, S. Park, D. Shum, S. Kim, Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs, BioRxiv, (2020).

[14] V.K. Bhardwaj, R. Singh, J. Sharma, V. Rajendran, R. Purohit, S. Kumar, Identification of bioactive molecules from Tea plant as SARS-CoV-2 main protease inhibitors, J Biomol Struct Dyn, (2020) 1-13.

[15] S.T. Ngo, N. Quynh, A. Pham, L. Le, D.-H. Pham, V. Vu, Computational Determination of Potential Inhibitors of SARS-CoV-2 Main Protease, ChemRxiv, (2020).

[16] C. Mendoza-Martinez, A. Rodriguez-Lezama, Identification of Potential Inhibitors of SARS-CoV-2 Main Protease via a Rapid In-Silico Drug Repurposing Approach, ChemRxiv, (2020).

[17] G.M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, J Comput Aided Mol Des, 27 (2013) 221-234.

[18] C. Bayrak, P. Taslimi, H.S. Karaman, I. Gulcin, A. Menzek, The first synthesis, carbonic anhydrase inhibition and anticholinergic activities of some bromophenol derivatives with S including natural products, Bioorg Chem, 85 (2019) 128-139.

[19] B. Yiğit, R. Kaya, P. Taslimi, Y. Işık, M. Karaman, M. Yiğit, İ. Özdemir, İlhamiGulçin, Imidazolinium chloride salts bearing wingtip groups: Synthesis, molecular docking and metabolic enzymes inhibition, Journal of Molecular Structure, 1179 (2019) 709-718.

[20] L. Schrödinger, LigPrep, in, Schrödinger, LLC, New York, NY, 2020.

[21] W. Sherman, T. Day, M.P. Jacobson, R.A. Friesner, R. Farid, Novel procedure for modeling ligand/receptor induced fit effects, J Med Chem, 49 (2006) 534-553.

[22] F. Türkan, M.H. Calimli, G.S. Kanberoğlu, M. Karaman, Inhibition effects of isoproterenol, chlorpromazine, carbamazepine, tamoxifen drugs on Glutathione S- transferase, Cholinesterases enzymes and molecular docking studies, Journal of Biomolecular Structure and Dynamics, (2020).

[23] S.L. Dixon, A.M. Smondyrev, S.N. Rao, PHASE: a novel approach to pharmacophore modeling and 3D database searching, Chem Biol Drug Des, 67 (2006) 370-372.

[24] A. Ece, Towards more effective acetylcholinesterase inhibitors: a comprehensive modelling study based on human acetylcholinesterase protein-drug complex, J Biomol Struct Dyn, 38 (2020) 565-572.

Supplementary Figure

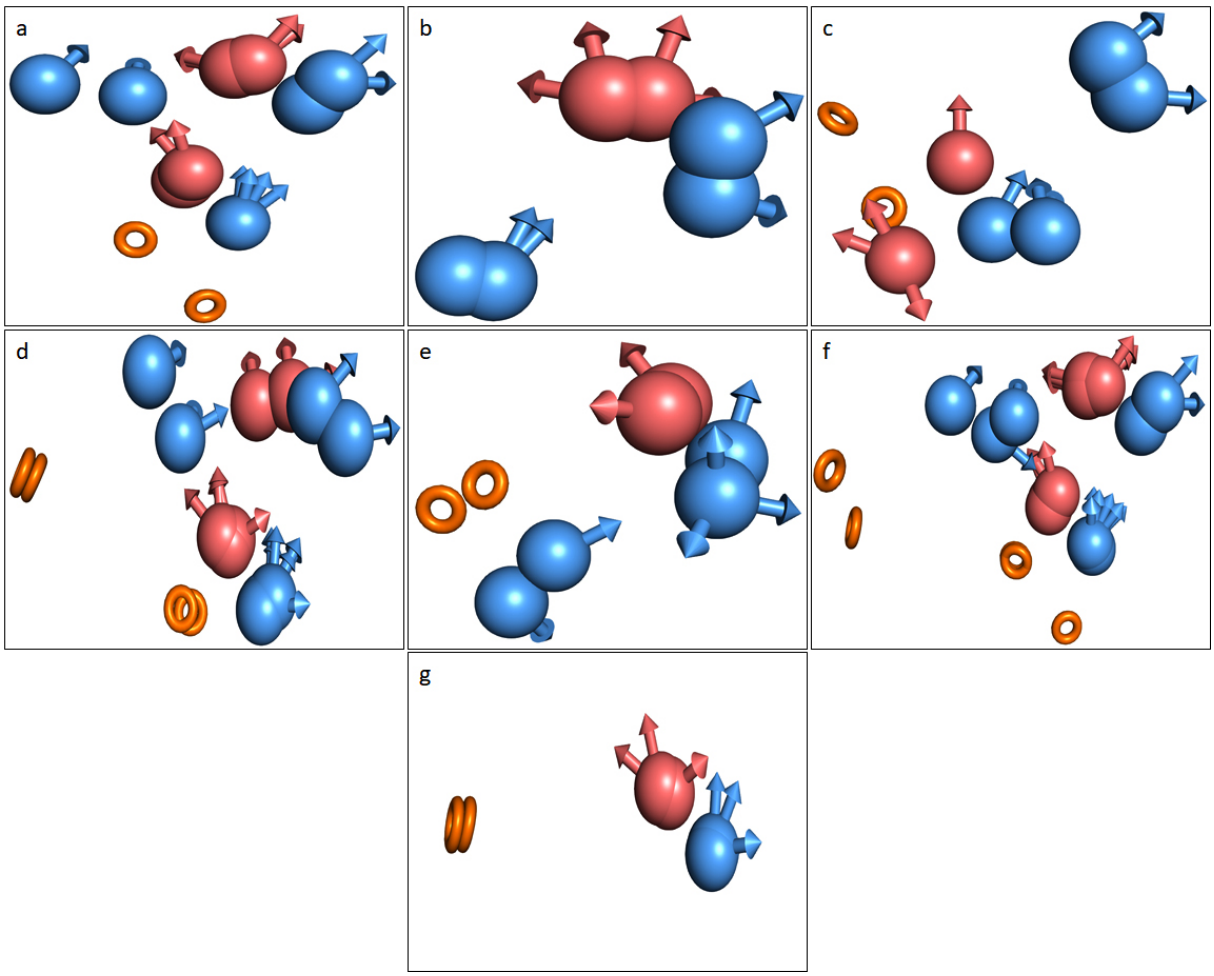

Figure S1. Merged pharmacophore models. a) **11a**, b)**11b**, c)**11r**, d) **13a**, e)**13b**, f)**14b**, and g)**N3**.