Original article:

UNREVEALING SEQUENCE AND STRUCTURAL FEATURES OF NOVEL CORONAVIRUS USING *IN SILICO* APPROACHES: THE MAIN PROTEASE AS MOLECULAR TARGET

Joseph Thomas Ortega[1], Maria Luisa Serrano[2], Flor Helene Pujol[3], Hector Rafael Rangel[3]*

[1] Department of Pharmacology and Cleveland Center for Membrane and Structural Biology, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA

[2] Unidad de Química Medicinal, Facultad de Farmacia, Universidad Central de Venezuela, Caracas, Venezuela

[3] Laboratorio de Virología Molecular, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela

* **Corresponding author:** Hector Rafael Rangel, Laboratorio de Virología Molecular, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela; E-mail: hrangel2006@gmail.com

ABSTRACT

Direct-acting antivirals are effective tools to control viral infections. SARS-CoV-2 is a coronavirus associated with the epidemiological outbreak in late 2019. Previous reports showed that HIV-1 protease inhibitors could block SARS-CoV main protease. Based on that and using an *in silico* approach, we evaluated SARS-CoV-2 main protease as a target for HIV-1 protease inhibitors to reveal the structural features related to their antiviral effect. Our results showed that several HIV inhibitors such as lopinavir, ritonavir, and saquinavir produce strong interaction with the active site of SARS-CoV-2 main protease. Furthermore, broad library protease inhibitors obtained from PubChem and ZINC (www.zinc.docking.org) were evaluated. Our analysis revealed 20 compounds that could be clustered into three groups based on their chemical features. Then, these structures could serve as leading compounds to develop a series of derivatives optimizing their activity against SARS-CoV-2 and other coronaviruses. Altogether, the results presented in this work contribute to gain a deep understanding of the molecular pharmacology of SARS-CoV-2 treatment and validate the use of protease inhibitors against SARS-CoV-2.

Keywords: Coronavirus, SARS-CoV-2, protease, treatment, HIV

INTRODUCTION

At present there is not specific treatment for coronavirus infection. However, prior outbreaks have helped to decipher the main pharmacological targets in these viruses. The therapies have been focused on targeting protease, helicase, polymerase, and using immunomodulators such as interferons and corticosteroids (Zumla et al., 2016). Ribavi-rin alone or combined with IFN has been the most common therapeutic intervention in patients with SARS and MERS (Khalid et al., 2015). Also, protease inhibitors have been used *in vitro* and *in vivo* to block coronavirus replication. Lopinavir (LPV) and ritonavir (RTV) are protease inhibitors currently used in HIV therapy that could block SARS-CoV and MERS-CoV main proteases (Savarino 2005). The combination of lopinavir or ri-

tonavir with ribavirin was associated with improvement in clinical outcome, compared with ribavirin alone, in SARS-CoV-infected patients (Lai, 2005). During the MERS-CoV outbreak, the Food and Drugs Administration approved the use of ritonavir/lopinavir, based mainly on data obtained from *in vitro* studies (Sheahan et al 2020). Altogether, these data support the assumption that some protease inhibitors may have an antiviral effect by blocking coronavirus main protease. However, like other RNA viruses, the main challenge associated with antiviral therapy is the selection of resistant variants. Mechanisms of generation of diversity in coronavirus are related to a moderate error rate of the polymerase (with proof-reading capacity) and homologous or heterologous recombination, factors that lead to antigenic drift and shift, similar to those described for Influenza viruses (Menachery et al., 2017). Thus, viral replication will produce a diverse population of genome variants having different fitness profiles. These variants could be associated with the development of drug resistance (Yin and Wunderink, 2018; Pruijssers and Denison, 2019). The ongoing efforts toward discovering efficient drugs to prevent and treat SARS-CoV-2 infection should include the prior structural and pharmacological knowledge gained with the other coronavirus outbreaks. Based on that, in this work was established a comparative theoretical study to rationalize the potential use of protease inhibitors as a treatment against SARS-CoV-2 infections.

MATERIALS AND METHODS

Sequence analysis

Protein sequences of the main protease were individually retrieved from GenBank (accession numbers are shown in the phylogenetic tree, Figure 1) for SARS-CoV, SARS-CoV-2 and several Bat-CoV from the genus Betacoronavirus.

Molecular docking

The coordinates for SARS-CoV and SARS-CoV-2 main proteases were obtained from the protein data bank, PDB code 1UJ1/2GX4 and 6LU7 respectively. Also, HIV-1 protease bounded to lopinavir under PDB code 1MIU and Bat HKU4 coronavirus PDB code 2YNB were evaluated. The PDB files to be used under further computational analysis were optimized by removing co-crystallized molecules and all crystallographic water molecules. Hydrogens were added and partial charges were assigned to all atoms. The obtained PDB files for each protein were further submitted to restrained molecular mechanics refinement. All molecular dynamics simulations described in this study were performed with NAMD 2.12 (Phillips et al., 2005), Vega ZZ 3.1.0.21 (Pedretti et al., 2004; Vanommeslaeghe et al., 2010) as described in Ortega et al. (2019). Following, structural analysis of the binding pocket was developed by using CASTp 3.0 software using the http://sts.bioe.uic.edu/castp/ server. The ligand-binding pocket located in the catalytic site was obtained manually and then verified by a priori docking approach with lopinavir by using the Achilles Blind Docking server (Sánchez-Linares et al., 2012). The 3D structure of each inhibitor was obtained from PubChem. Also, public libraries for protease inhibitors were obtained from PubChem and ZINC databases. Molecular docking was performed with VINA/VegaZZ 3.1.0.21 and 30 runs conducted for each compound. The results were prioritized according to the predicted binding energy in kcal/mol. The results obtained from the docking simulation were visualized with the Biovia Discovery Studio Visualizer 17.2.0 software.

ADME compound characterization

A comprehensive analysis of physico-chemical descriptors, as well as ADME parameters, pharmacokinetic properties, drug-like nature, and medicinal chemistry for the top 5 compounds obtained from the library virtually screened, was carried out by using SWISSADME tools. These tools were asset through the website at http://www.swissadme.ch.

RESULTS

Homology sequence analysis of the main protease of SARS-CoVs and related Bat-CoVs

Phylogenetic analysis of the main protease protein sequences of SARS-CoVs and Bat-CoVs is shown in Figure 1. The results are in agreement with recent reports of an independent introduction of SARS-CoV-2 from a Bat-CoV, different from the spillover which led to the introduction of SARS-CoV, being the Bat-CoV of *Rhinolophus affinis* the probable ancestor of this new virus (Wong et al., 2020). The sequences of the whole main protease of this Bat-CoV and of SARS-CoV-2 share 99.3 % identity, while the MERS-CoV protease was only 50.6% identical (Figure 1). The main protease sequence is highly conserved among all the SARS-like coronaviruses from human and related CoVs from other animals (Figure 1).
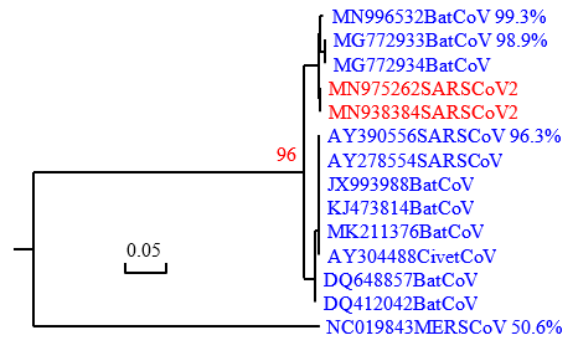

Figure 1: Phylogenetic analysis of SARS-CoV-2 and other coronaviruses main protease protein. Phylogenetic tree constructed with Poisson correction and 100 bootstrap replicas. The sequences are named with their accession number. The percent homology with SARS-CoV-2 protease protein is shown for some proteins.
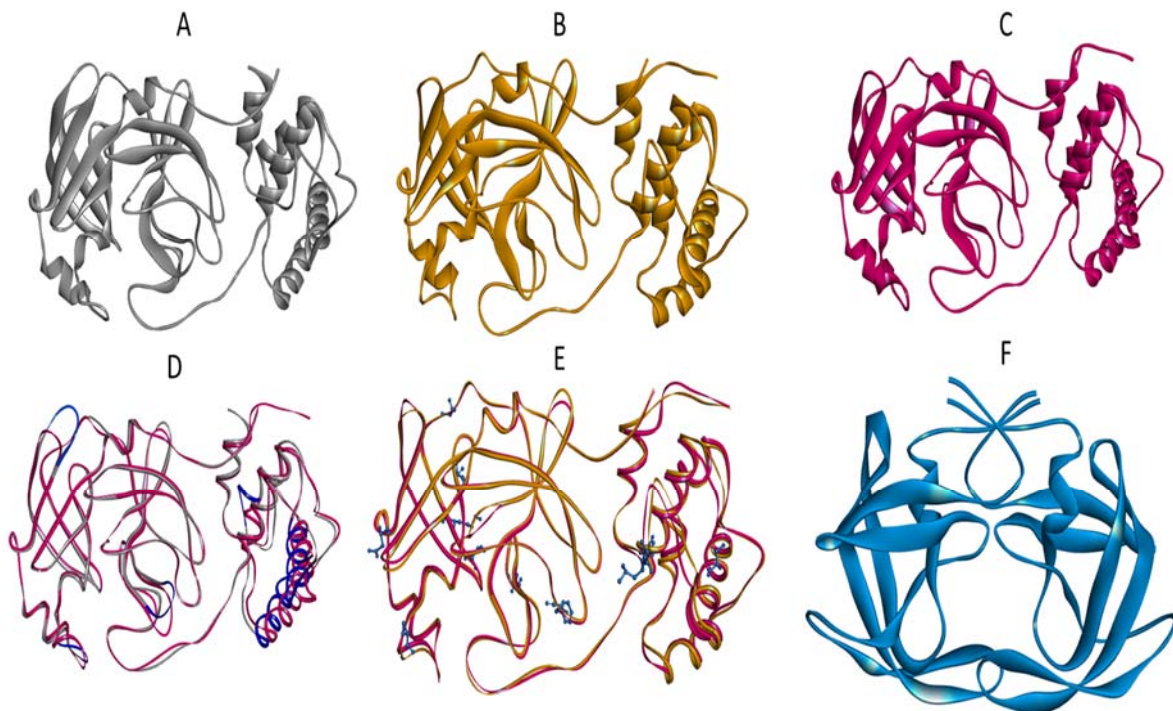

Figure 2: Coronavirus main proteases. Graphical representation for each protease was achieved in Biovia Discovery Studio Visualizer by using the coordinates obtained from the Protein Data Bank. Structure for **A**) Bat-CoV (2YNB), **B**) SARS-CoV (2GX4), **C**) SARS-CoV-2 (6LU7) are shown. Also, a comparison by molecular overlapping between these structures is shown in **D**. The main residues and changes in SARS-CoV-2 compared to SARS-CoV are shown in **E**. The HIV-1 protease (1MIU) is shown in **F**.

Structural analysis of the main protease

The SARS-CoV and SARS-CoV-2 proteins exhibit 96 % sequence identity to each other (Figure 1). Figure 2 shows the 3D representation for Bat-CoV, SARS-CoV, and SARS-CoV-2 main proteases and indeed, very similar structures were observed, particularly when comparing SARS-CoV proteases, although very different to the one of HIV-1. Interestingly the SARS-CoV-2 protease shares 21/29 amino acids involved in the drug interaction with SARS-CoV protease, while this later exhibits only 21 amino acids involved in this interaction (Figure 3). SARS-CoV-2 protease has an active site with a surface accessible to solvent of 356Å and SARS-CoV only 256 Å (Figure 3).
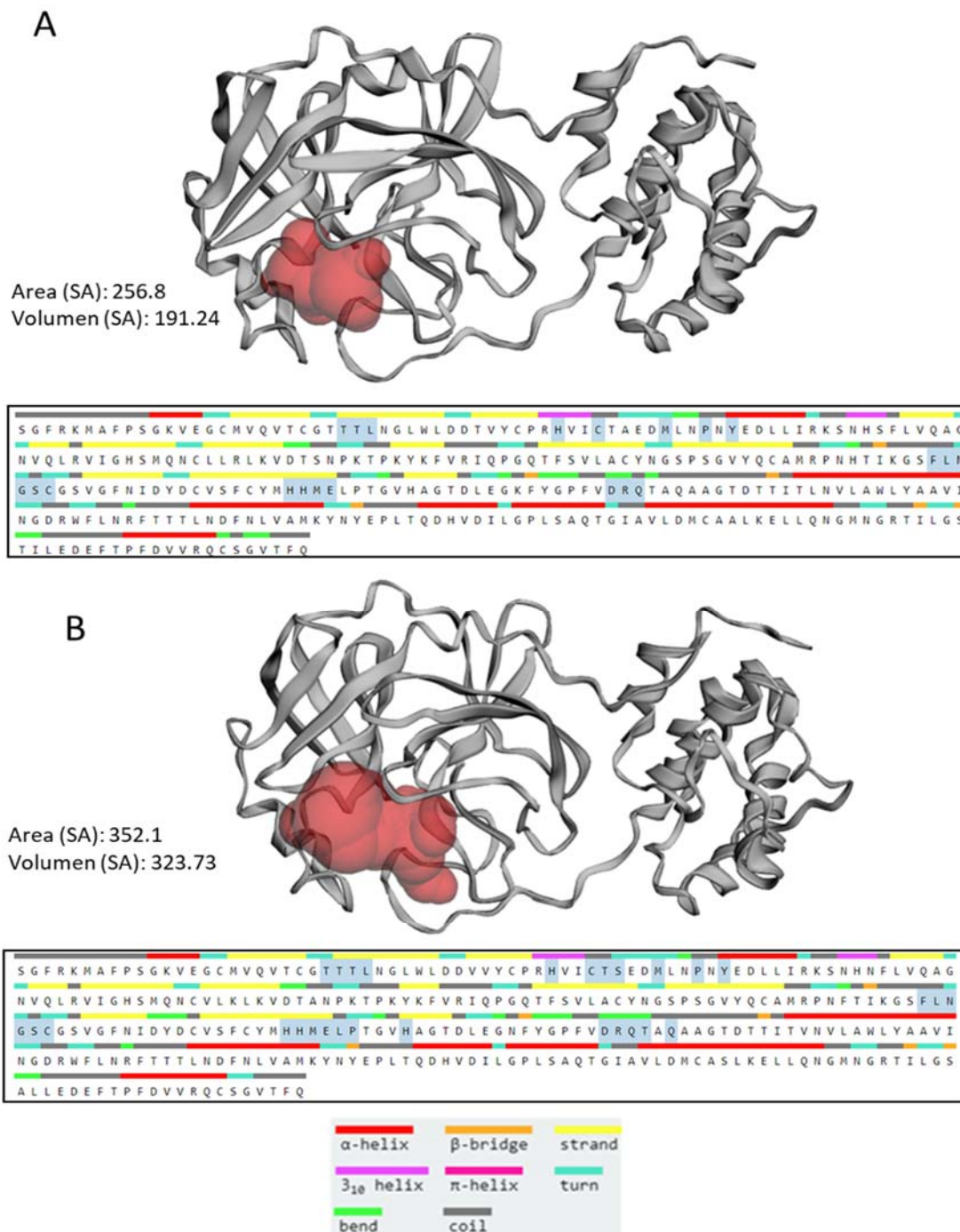

Figure 3: Coronavirus main proteases. Topological analysis of the catalytic site for SARS-CoV (**A**) and SARS-CoV-2 (**B**) main proteases. Conformational residues are highlighted in blue.

I*n silico* data demonstrated that the different protease inhibitors, used against HIV-1, could interact with the active site of SARS-CoV-2 protease producing an interaction with a binding energy lower than -6.9 Kcal/mol. However, the compound that produced the strongest interaction with the active site was saquinavir, with a binding energy of -9.6 Kcal/mol (Table 1). Binding energy of SARS-CoV-2 main protease to Saquinavir (SQV) and LPV were slightly higher, but similar, to the one of HIV-1 protease (Table 1). The interaction of LPV, SQV and RTV with the SARS-CoV-2 main protease is shown in Figure 4, the different amino-acids involved in the interaction with all these drugs are also shown.

In order to contribute with further studies related to developing more effective drugs, in this work was evaluated a broad library of protease inhibitors available in the ZINC database (over 100 compounds) and PubChem (over 200 compounds). After molecular docking, compounds were prioritized based on binding energies. The top 20 compounds were clustered using a hierarchical matrix and represented into a heat map (Figure 5). These compounds could be clustered into three big groups based on their chemical structures. Also, the binding energies and chemical motives for each compound are shown in Table 2. Interestingly, in the output of this virtual screening at least 5 compounds, representatives of these chemical clusters with binding energy around -8.2 Kcal/mol, were obtained. These scaffolds could be used as leading compounds to fur-

ther chemical optimization to develop more potent inhibitors. The structure and main ADME parameters for each inhibitor are shown in Figure 6, in general the 5 compounds shown, have physicochemical parameters under the expected values for a good drug.

DISCUSSION

Due to the rapid spread of SARS-CoV-2, affecting more than 70 countries, therapeutic alternatives are urgently required. Viral main protease plays a pivotal role in coronavirus replication. This enzyme is responsible for the cleavage of the polyprotein, producing functional proteins that will be packed into the virion. The molecular study of SARS-CoV-2 protease showed that this virus has high homology with SARS-CoV protease. Previous data suggest the use of protease inhibitors as potential inhibitors of SARS-CoV protease, in special some of those used for HIV-1 therapy. HIV-1 protease is an aspartyl protease, while in coronavirus it is a cysteine protease. However, the inhibition by protease inhibitors could be driven by a similar mechanism. HIV-1 is one of the best-studied models in antiviral research, their targets are well characterized and its inhibition by drugs has been broadly studied. Protease inhibitors play a main role in HIV-1 therapy. Furthermore, these compounds have shown inhibitory effects over proteases from pathogens such as Leishmania and malaria (Valdivieso et al., 2010; Sonoiki et al., 2016; de Wilde et al., 2014). Thus, the use of protease inhibitors

Table 1: Molecular docking of Protease inhibitors used against HIV-1 over SARS-CoVs main proteases

Protease inhibitor	Binding energy (Kcal/mol) for SARS CoV-2 protease	Binding energy (Kcal/mol) for SARS CoV protease	Binding energy (Kcal/mol) for HIV-1 protease
Saquinavir	-9.6	-8.1	-9.7
Lopinavir	-9.1	-8.4	-11.4
Tipranavir	-8.7	-7.7	-10.3
Darunavir	-8.2	-6.7	-10.9
Amprenavir	-7.6	-7.3	-9.2
Atazanavir	-7.2	-7	-9.8
Ritonavir	-6.9	-7.1	-9.4

in other viral species could be supported by these non-specific protease inhibitions. Previous reports showed the inhibition of SARS-CoV protease by protease inhibitors used in HIV-1 therapy, being an interesting pharmacological target to treat SARS-CoV-2. Previously, scientific reports showed that some HIV-1 protease inhibitors could decrease SARS-CoV replication (Savarino, 2005). Nevertheless, not all of the protease inhibitors showed the same efficacy against this virus (Jenwitheesuk and Samudrala, 2003). Based on that, in this work, we established an *in silico* approach to validate the inhibition of SARS-CoV-2 protease by HIV-1 protease inhibitor. The results showed that the main protease inhibitors used in HIV-1 therapy could produce a decrease in the protease activity in coronavirus enzyme. The compound with the highest binding affinity was saquinavir. Preliminary reports are showing that patients infected with SARS-CoV-2 and treated with lopinavir could achieve viral clearance (Lim et al., 2020). Our study suggests that drugs such as SQV and LPV have shown slightly higher, but comparative binding energies against the SARS-CoV-2, than HIV-1 protease. Thus, based on their safety profiles, SQV and LPV could be used as therapy and/or pre-exposure prophylaxis to reduce new infections by SARS-CoV-2. Additionally, the data obtained from prior outbreaks, the clinical results obtained by using protease inhibitors in patients infected with SARS-CoV-2, and the results shown here, support the use of protease inhibitors to treat SARS-CoV-2.
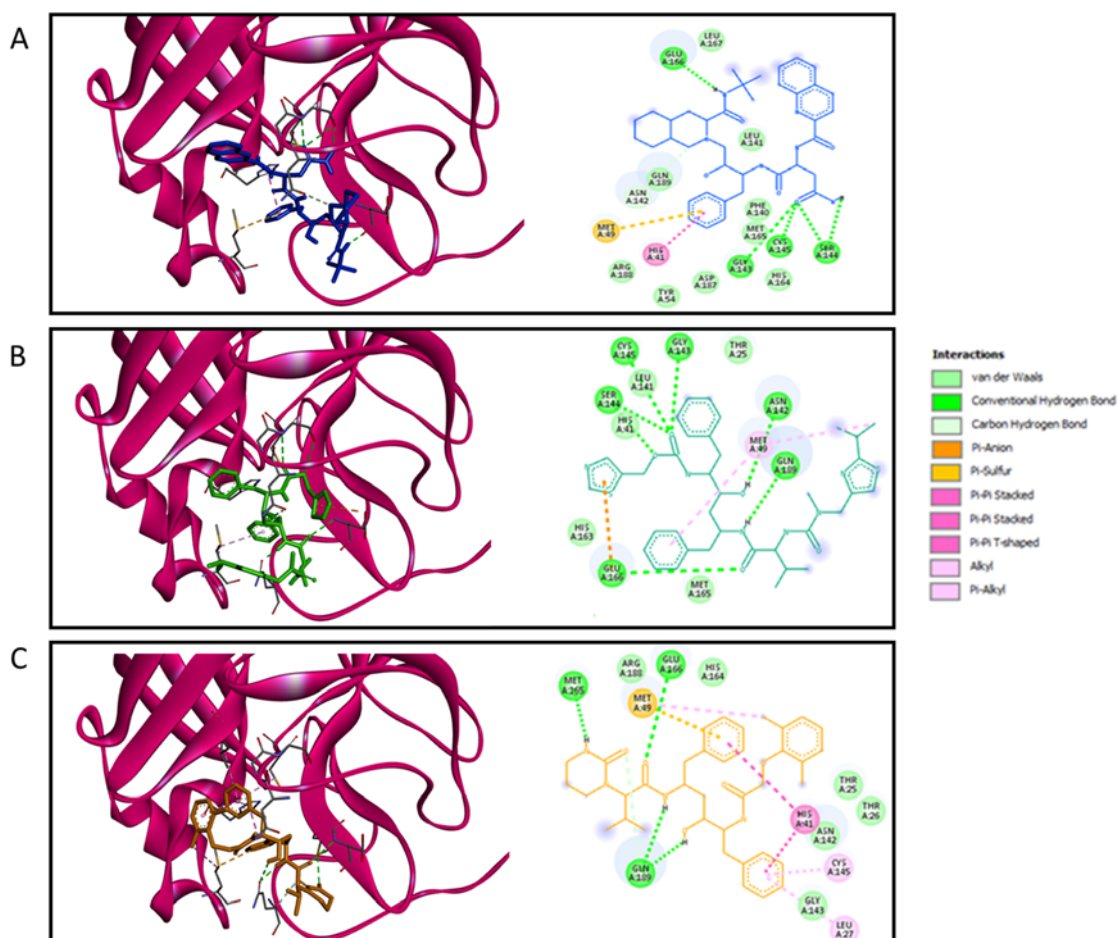

Figure 4: Molecular docking of HIV protease inhibitors on SARS-CoV-2 main protease. The molecular docking output represented as lower binding energy frame is shown for each inhibitor. 3D (left) and 2D (right) representations showing the main interaction between the inhibitor and the receptor are displayed. **A**) Saquinavir, **B**) Ritonavir, and **C**) Lopinavir
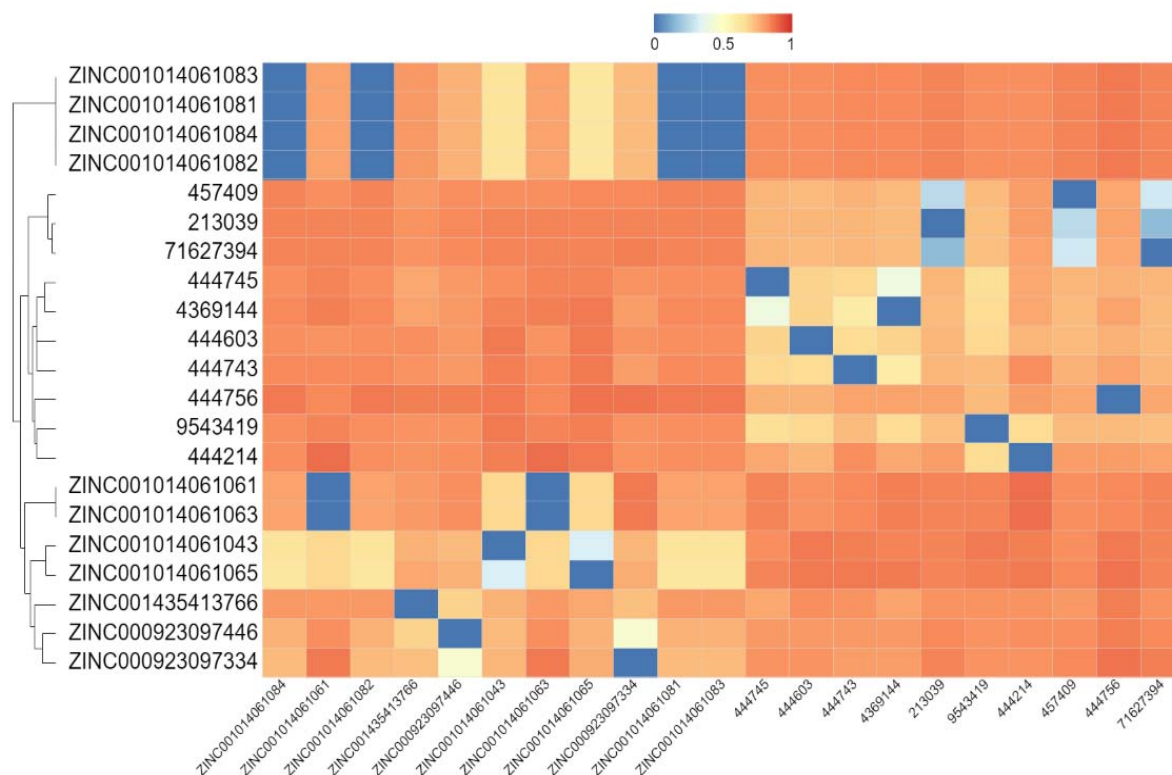
Figure 5: Hierarchical clustering of the top 20 best protease inhibitors obtained from public libraries (ZINC and PubChem) that could block SARS-CoV-2 protease. The compounds were clustered by chemical structure based in the Tanimoto coefficient where 0: the same structure and 1: no chemical relationship.

In addition, to contribute to further development of antiviral drugs against SARS-CoV-2, in this work experimental compounds against the SARS-CoV-2 protease were evaluated by virtual screening. Some compounds showed comparable binding energy values to the SARS-CoV-2 protease than to those observed for SQV, LPV and RTV. ADME (for Absorption, Distribution, Metabolism, and Excretion) parameters estimation for new drugs, significantly helps to reduce the pharmacokinetics-related failure in clinical phases. These results evidence that these compounds could be suitable for further medicinal chemistry optimization to produce the second generation of inhibitors more specific and potent against SARS-CoV-2.

Other antivirals have been evaluated in SARS-CoV-2 as the polymerase inhibitor remdesivir, a nucleotide analog (currently in clinical trials against Ebola virus and SARS-CoV-2), alone or in combination with chloroquine, an inhibitor of lysosome acidification, with interesting results (Wang et al., 2020; de Wit et al., 2020). Therefore, proposing a combinatory therapy against SARS-CoV-2 could be a feasible approach. The starting therapeutic scheme could include remdesivir, with the disadvantage of its intravenous administration, and HIV-1 protease inhibitors such as Lopinavir or Saquinavir. Nevertheless, more pre-clinical and clinical data are required to support the development of new therapies against SARS-CoV-2.

Table 2: Binding energy (BE) values of the best 20 compounds selected as potential inhibitors of SARS-CoV-2 protease. Compound structures were obtained from ZINC database (Bold) and PubChem (italics), lowest BE compound of each group are shown in red.

Compound ID	BE (kcal/mol)	Molecular formula	MW	F	S	Br	Cl	RNH2	R2NH	R3N	ROH	RCOR	RCOOR	ROR	RINGS	AROMATIC
213039	-8,1	C27H37N3O7S	547,6636	0	1	0	0	1	1	0	1	0	1	2	4	2
4369144	-8,2	C30H34N6O3	526,6294	0	0	0	0	0	1	1	0	0	0	0	5	3
444214	-8,1	C28H31F5N4O4	582,5622	5	0	0	0	0	3	1	1	0	0	0	4	3
444603	-8,7	C25H34N6O5S	530,6397	0	1	0	0	0	1	2	1	0	0	0	4	2
444743	-8,3	C23H29BrN6O5S	581,4826	0	1	1	0	0	1	1	0	0	0	0	4	2
444745	-9,3	C31H35N7O3	553,6547	0	0	0	0	0	2	1	0	0	0	0	6	4
444756	-8	C37H55ClN8O5	727,3362	0	0	0	1	0	2	4	2	0	0	0	5	2
457409	-8	C28H36N2O9S	576,6584	0	1	0	0	0	1	0	1	0	1	4	5	2
71627394	-7,9	C28H37N3O8S	575,6737	0	1	0	0	0	1	0	1	0	1	2	4	2
9543419	-8,1	C30H39N7O6	593,674	0	0	0	0	0	2	1	1	0	0	0	2	1
ZINC001014061065	-7,6	C15H17N5O4	331,33	0	0	0	0	0	1	1	1	0	0	0	3	2
ZINC001014061063	-7,6	C18H28N4O3	348,44	0	0	0	0	0	1	1	0	0	0	0	3	1
ZINC001014061081	-8,7	C16H17N7O2	339,35	0	0	0	0	0	1	1	0	0	0	0	4	3
ZINC000923097446	-7,6	C17H21N5O3	343,38	0	0	0	0	0	1	2	0	1	0	1	3	2
ZINC001014061043	-7,6	C18H22N4O3	342,39	0	0	0	0	0	1	1	0	0	0	0	3	2
ZINC001014061084	-8,1	C16H17N7O2	339,35	0	0	0	0	0	1	1	0	0	0	0	4	3
ZINC001014061083	-8,3	C16H17N7O2	339,35	0	0	0	0	0	1	1	0	0	0	0	4	3
ZINC000923097334	-7,7	C17H18N4O2	310,35	0	0	0	0	0	1	1	0	1	0	0	3	3
ZINC001014061082	-7,7	C16H17N7O2	339,35	0	0	0	0	0	1	1	0	0	0	0	4	3
ZINC001435413766	-7,7	C20H24FN3O2	357,42	1	0	0	0	0	0	1	0	0	1	0	4	2
ZINC001014061061	-7,8	C18H28N4O3	348,44	0	0	0	0	0	1	1	0	0	0	0	3	1

F= fluoride, S= sulphurous, Br= bromide, Cl=chloride, RNH2= primary amine, R2NH= secondary amine, R3N= tertiary amine, ROH= alcohol, RCOR= ketone, RCOOR= ester, ROR= ether
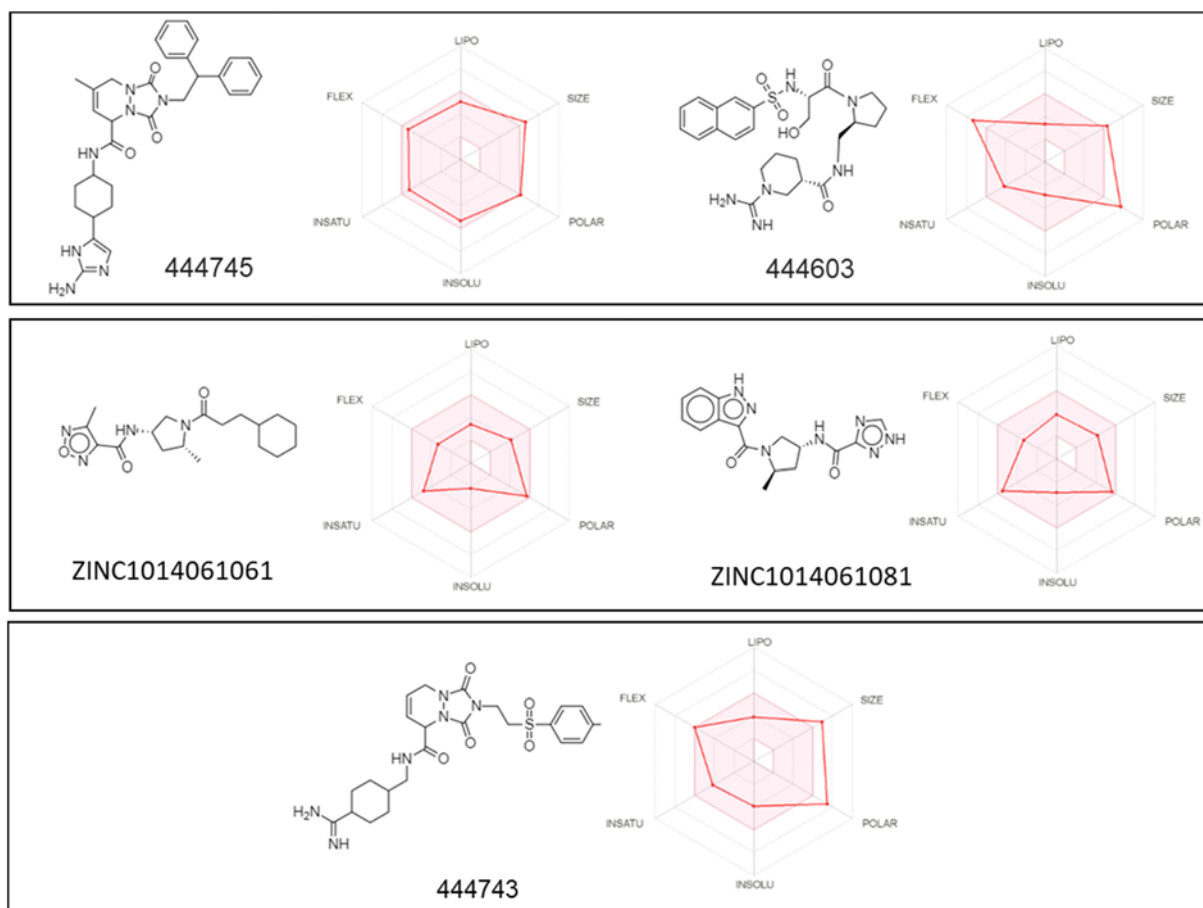
Figure 6: Chemical structure and ADME parameters of top five protease inhibitors obtained from public libraries (ZINC and PubChem) that could block SARS-CoV-2 protease. The colored zone is the suitable physicochemical space for oral bioavailability. LIPO (Lipophility): -0.7 <XLOGP3< +5.0. SIZE: 150 g/mol <MW <500 g/mol. POLAR (Polarity): 20Å2 <TPSA< 130Å2. INSOLU (Insolubility): 0< LogS (ESOL) <6. INSATU (Insaturation) 0.25< fraction Csp3 <1. FLEX (Flexibility): 0< Number of rotatable bonds <9

REFERENCES

de Wilde AH, Jochmans D, Posthuma CC, Zevenhoven-Dobbe JC, van Nieuwkoop S, Bestebroer TM, et al. Screening of an FDA-approved compound library identifies four small-molecule inhibitors of Middle East respiratory syndrome coronavirus replication in cell culture. Antimicrob Agents Chemother. 2014;58: 4875–84.

de Wit E, Feldmann F, Cronin J, Jordan R, Okumura A, Thomas T, et al. Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. Proc Natl Acad Sci U S A. 2020;13:epub ahead of print.

Jenwitheesuk E, Samudrala R. Identifying inhibitors of the SARS coronavirus proteinase. Bioorg Med Chem Lett. 2003;13:3989-92.

Khalid M, Al Rabiah F, Khan B, Al Mobeireek A, Butt TS, Al Mutairy E. Ribavirin and interferon-α2b as primary and preventive treatment for Middle East respiratory syndrome coronavirus: a preliminary report of two cases. Antivir Ther. 2015;20:87–91.

Lai ST. Treatment of severe acute respiratory syndrome. Eur J Clin Microbiol Infect Dis. 2005;24: 583–91.

Lim J, Jeon S, Shin HY, Kim MJ, Seong YM, Lee WJ, et al. Case of the index patient who caused tertiary transmission of COVID-19 infection in Korea: the application of Lopinavir/Ritonavir for the treatment of COVID-19 infected pneumonia monitored by quantitative RT-PCR. J Korean Med Sci. 2020;35(6): e79.

Menachery VD, Graham RL, Baric RS. Jumping species-a mechanism for coronavirus persistence and survival. Curr Opin Virol. 2017;23:1-7.

Ortega JT, Serrano ML, Suárez AI, Baptista J, Pujol FH, Cavallaro LV, et al. Antiviral activity of flavonoids present in aerial parts of Marcetia taxifolia against Hepatitis B virus, Poliovirus, and Herpes Simplex Virus in vitro. EXCLI J. 2019;18:1037-48.

Pedretti A, Villa L, Vistoli G. VEGA - An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. J Comput Aided Mol Des.2004;18:167–73.

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26:1781–1802.

Pruijssers AJ, Denison MR. Nucleoside analogues for the treatment of coronavirus infections. Curr Opin Virol. 2019;35:57-62.

Sánchez-Linares I, Pérez-Sánchez H, Cecilia JM, García JM. High-throughput parallel blind virtual screening using BINDSURF. BMC Bioinformatics. 2012;13 (Suppl 14):S13.

Savarino A. Expanding the frontiers of existing antiviral drugs: possible effects of HIV-1 protease inhibitors against SARS and avian influenza. J Clin Virol. 2005;34:170-8.

Sheahan TP, Sims AC, Leist SR, Schäfer A, Won J, Brown AJ, et al. Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. Nat Commun. 2020;11(1):222.

Sonoiki E, Nsanzabana C, Legac J, Sindhe KM, DeRisi J, Rosenthal PJ. Altered Plasmodium falciparum sensitivity to the antiretroviral protease inhibitor lopinavir associated with polymorphisms in pfmdr1. Antimicrob Agents Chemother. 2017;61(1):01949-16.

Valdivieso E, Rangel A, Moreno J, Saugar JM, Cañavate C, Alvar J, et al. Effects of HIV aspartylproteinase inhibitors on Leishmania sp. Exp Parasitol. 2010;126:557-63.

Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010; 1:671-90.

Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020;30:269–271.

Wong MC, Javornik Creegen SJ, Ajami NJ, Petrosino JF. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. bioRxiv. 2020; https://doi.org/10.1101/2020.02.07.939207.

Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. Respirology. 2018;23:130-7.

Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY. Coronaviruses - drug discovery and therapeutic options. Nat Rev Drug Discov. 2016;15:327-47.