Report #1

Interactions of drug-set-1 and drug-set-2 with 3CL protease

(work in progress**)**

K. A. Krzyśko[1], Ł. Charzewski[1], P. Chyży[1], M. Kolinski[2], B. Lesyng[1]

[1]Laboratory for Molecular Design and Bioinformatics, Department of Biophysics, Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland

[2]Bioinformatics Laboratory, Mossakowski Medical Research Centre, Polish Academy of Sciences, Pawinskiego 5, 02-106 Warsaw, Poland

1. An overview of the project

The COVID-19 pandemic originated in China and it has quickly spread over all continents affecting most countries in the world. Due to the easy transmission of the 2019-nCOV virus (coronavirus) from human to human, and due to the lack of a vaccine and/or specific antiviral drugs, the pandemic has been developing rapidly. Much of the focus for developing antiviral treatments today still lies on small molecule drugs that block the molecular machinery of a specific virus. Medical practice shows that some antiviral, antimalarial or antiparasitic drugs such as *favipiravir*, *chloroquine*, *ivermectin* and a number of others, show a weakening effect on this viral infection. However, it should be emphasized that understanding the activity of these drugs directed to coronavirus infection is still very poor. Meanwhile, multiscale modelling and bioinformatics methods, including molecular design methods, can significantly contribute to the understanding of such activity and to creating a more effective frontline against this infection. It is also the cheapest way to positively contribute to the anti-pandemic war.

One should note that structures of a number of SARS-CoV-2 proteins and glycoproteins and the ACE2 enzyme, which are important for the course of the virus infection process, have already been isolated and their structures have been determined using diffraction techniques (structures available in Protein Data Bank). It gives the possibility to check with molecular modelling and bioinformatics methods whether the existing drugs can interact specifically with these molecular objects by blocking their biological activity. We have already checked a number of anti-malarial and anti-AIDS drugs. In parallel, we have been performing a high-throughput screening of chemical compounds from DrugBank (13500 approved drugs and discovery-phase compounds, plus 5200 proteins) as well as from the NCI Open Database (containing over 250000 compounds), and have been selecting some compounds that meet the conditions necessary to block the activity of viral proteins. The structures of the selected compounds can be further modified to achieve optimal interactions with the given SARS-CoV-2 molecular targets and/o human proteins involved in the infection process. This includes, 3CL-protease, PL-protease,

Nsp15 endoribonuclease, spike (S1 & S2), Nsp3 ADP ribose phosphatase, Nsp9 RNA binding protein, N-terminal RNA binding protein of nucleocapsid, subunits of nsp16/nsp10 complex, replicase (RNA-dependent RNA polymerase (RdRp, nsp12) in complex with nsp7 and nsp8) and others whose structures may appear soon.

The simplified pipeline technological process that we have been implementing can be illustrated in the scheme below:

[high-performance screening]-> [docking using a multi-scale approach]-> [selection of the most optimal chemical compounds (including approved drugs)]-> [their possible modification (including, in particular boron substitution)]-> [detailed analysis of interactions of the most promising inhibitors]-> [extending the study of their specificity, to classify them as potential pro-drugs]-> [publishing reports on the project's web site, open to all research groups and/or pharmaceutical/biotechnological companies]. Regarding boron derivatives, these can form stable chemical bonds with hydroxyl groups of molecular targets. At the same time, these derivatives typically show few side effects.

What distinguishes our approach from other approaches of the same type is that we consider it very important to use a multi-scale approach, applying among others excellent coarse-grained CABS-dock platform, through conventional modelling environments (e.g. MOE, NAMD/VMD), up to modelling at the microscopic quantum level (e.g. SCIGRESS, quantum DFT or *ab initio*). The latter are especially entitled to predict the formation of effective chemical bonds of inhibitors with their targets.

The cooperation of the Laboratory of Bioinformatics MMRC PAS with the Laboratory for Molecular Design and Bioinformatics at the Faculty of Physics of the University of Warsaw can prove very effective in all of the above-mentioned research methods.

2. Research methodology

Multiscale molecular modeling and bioinformatics methods, from coarse-grained simulations, through classical molecular dynamics, to microscopic quantum mechanics, are being applied. Docking of peptides are carried out using CABS-dock applying Monte Carlo simulation with replica-exchange schemes for sampling the conformational space of bound molecules. CABS-dock uses a unique simulation approach, allowing unrestrained docking of fully flexible peptide (or a small protein and/or a protein fragment) ligands to protein receptors that can significantly modify their tertiary structures upon the ligand binding. The DEDAL environment is applied for algorithmic comparison of protein structures. The MOE platform is used for microscopic molecular mechanics (MM) and molecular dynamics (MD) simulations, and SCIGRESS allows for fast quantum mechanical calculations (QM). In turn NAMD/VMD molecular modeling platform is applied for large scale MM/MD simulations, including quantum-classical QCMD simulations, where the MOPAC package is used as the generator of the quantum potential energy function. The DrugBank and Open NCI Database of chemicals are used, along with the high-throughput analysis environment provided with the MOE platform.

3. First results

Systematic screening of known antiviral, antimalarial and antiparasitic drugs against viral proteins or glycoproteins 2019-nCOV has been already initiated. 3CL protease was chosen as the first molecular target. The MOE environment was used for docking as well as for the refinement of structures of the complexes using molecular MM and MD methods.

The following anti-malarian drugs (**drug-set-1**) have been docked against the 3CL protease:
Clindamycine, Lumefantrine, Quinine, Mefloquine, Dihydroqinghaosu (artenimol, dihydroartemisinin, DHA), Chloroquine, Doxycycline, Hydroxychloroquine, Pyrimethamine, Piperaquine, Amodiaquine, Artesunate, Atovaquone , Proguanil, Primaquine, Sulfadoxine, Artemether and:
the following anti-HIV drugs (**drug-set-2**) have been docked against the 3CL protease:
Brecanavir, Saquinavir, Nelfinavir, Darunavir, Lopinavir, Remdesivir, Amprenavir, Ritonavir, Indinavir and Tipranavir.

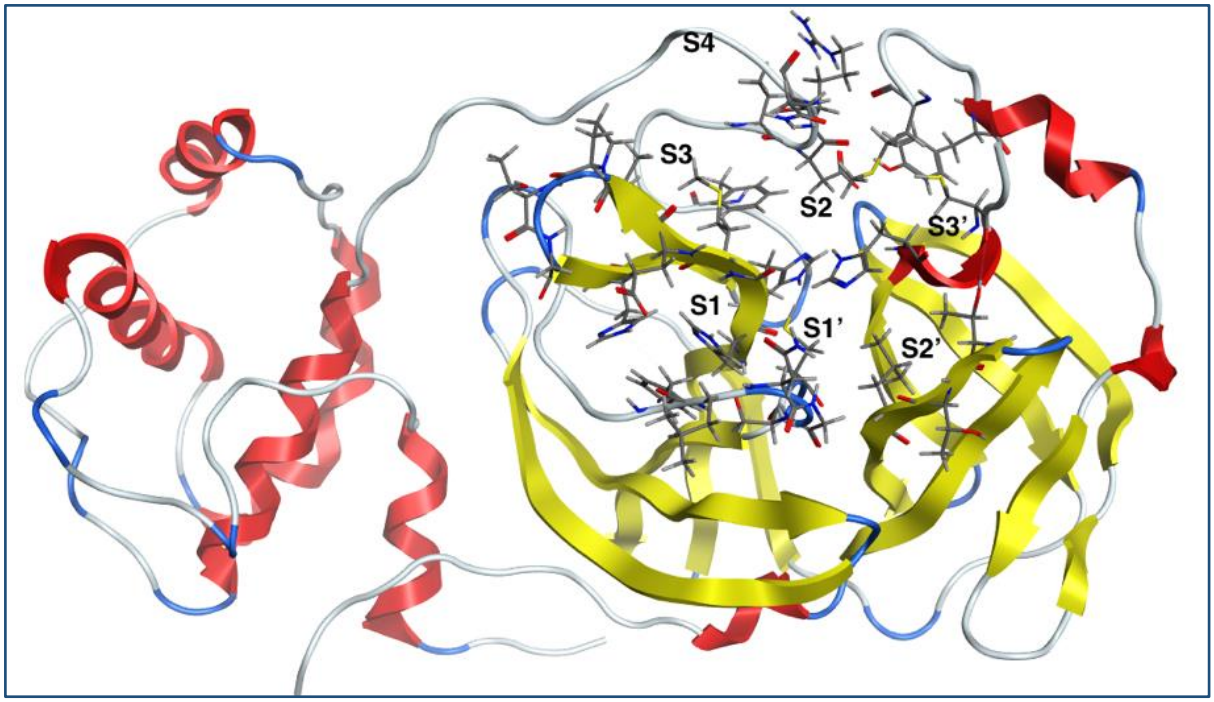


Fig. 1. Overall view of the 3CL protease structure with the identified binding subdomains.

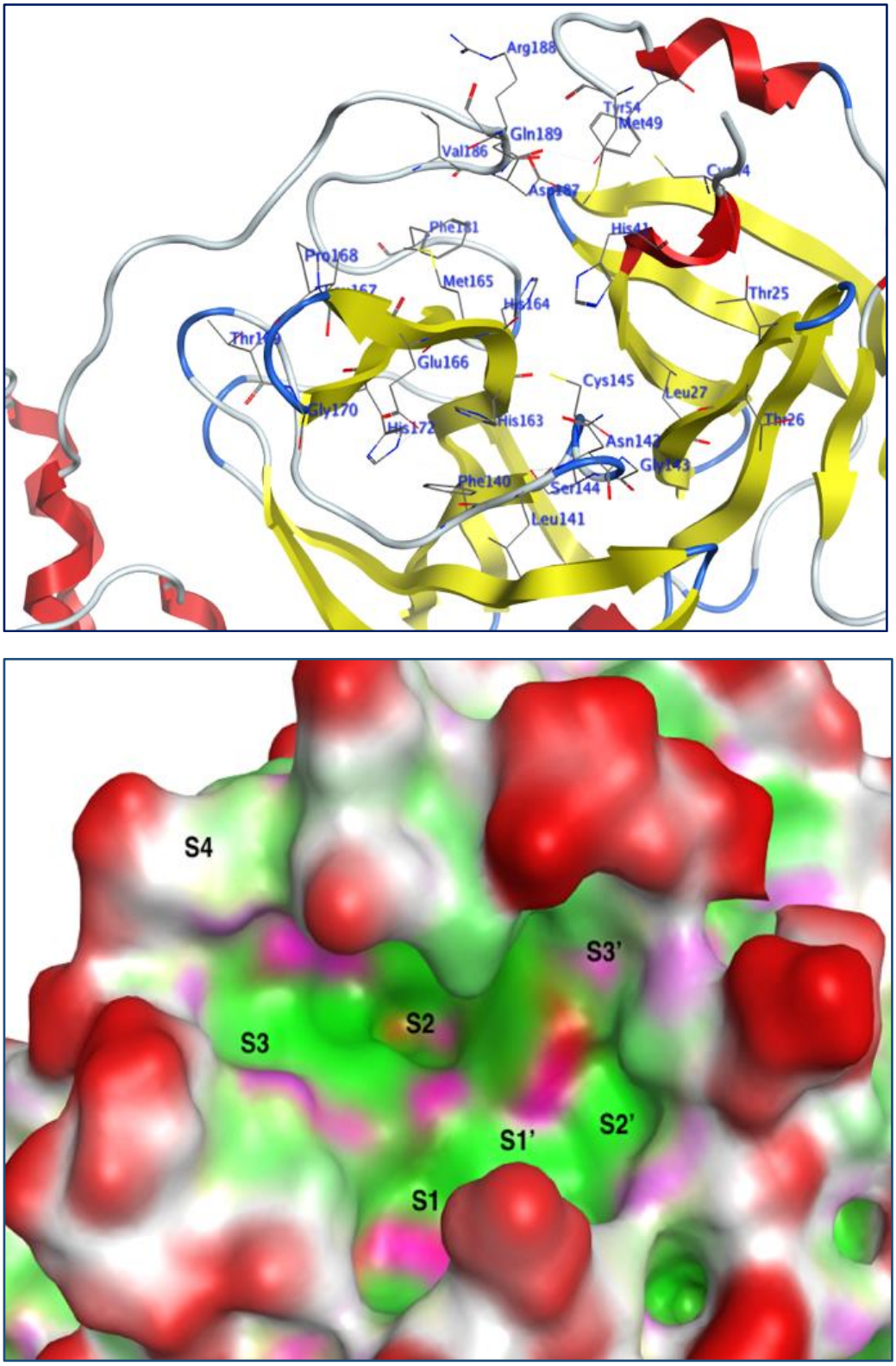
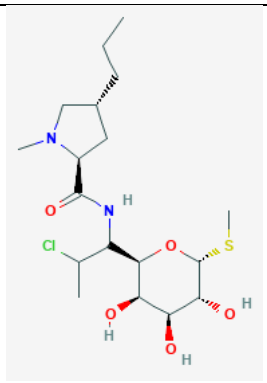
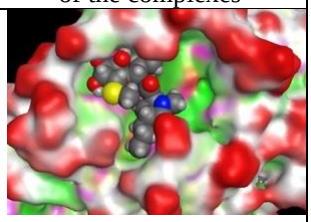
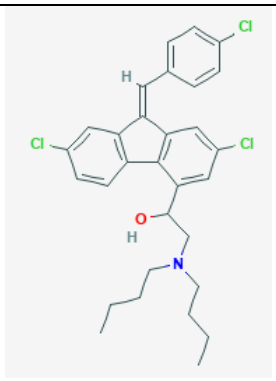
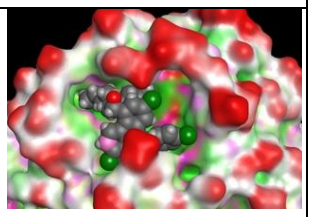
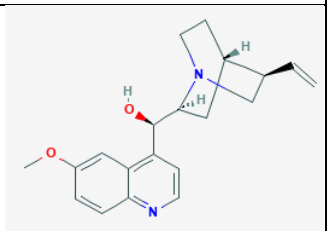
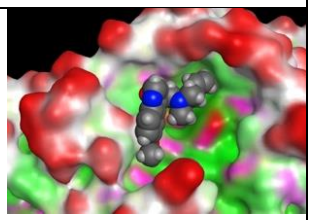
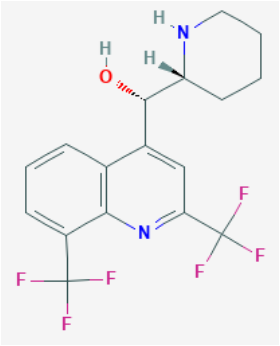
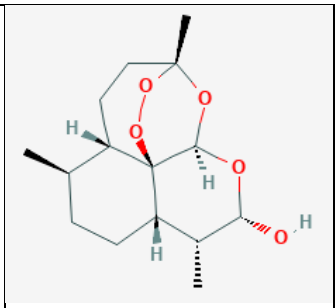
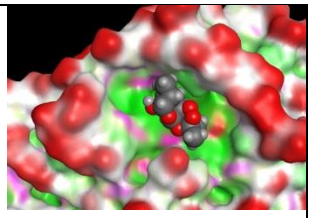
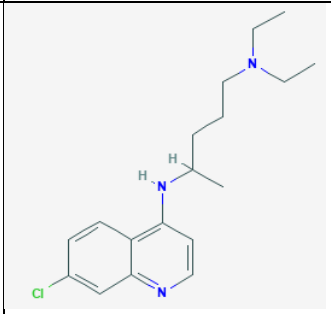
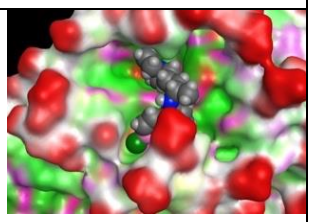
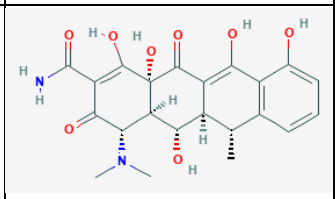
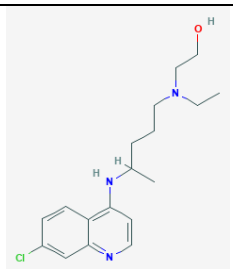
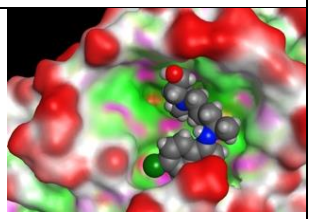
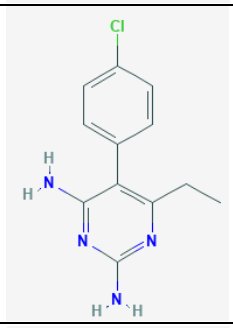
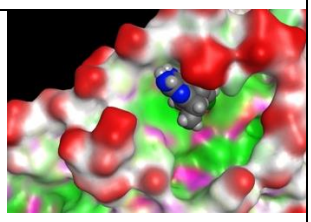
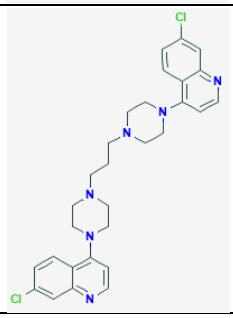
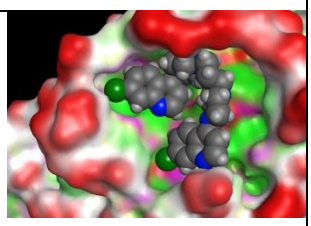
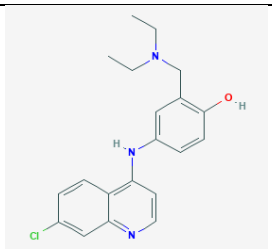
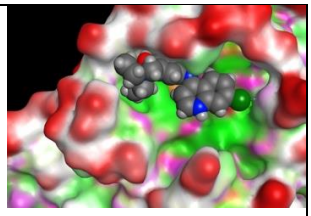
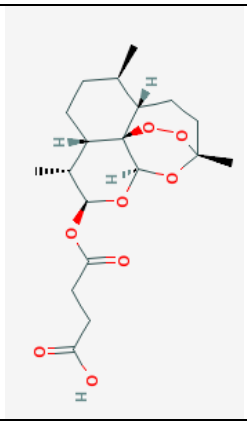
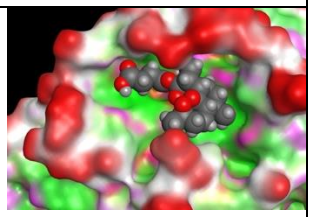
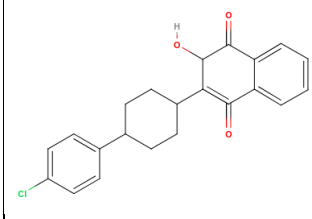
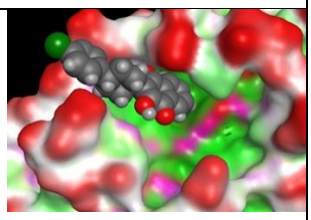
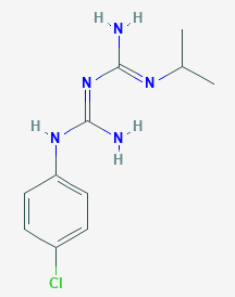
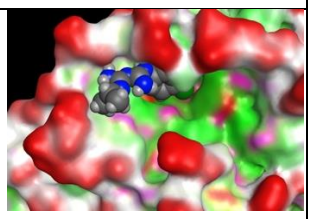
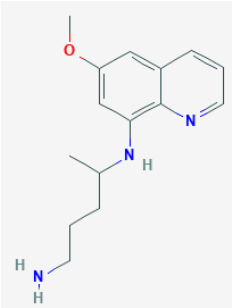
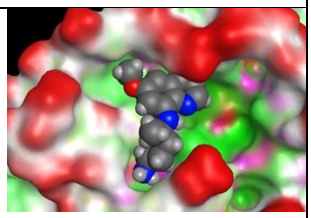


Fig. 2. The 3CL protease binding cavity. *Top*: Amino acids forming the binding cavity. His41-Cys145 form the catalytic dye. *Bottom*: Molecular surface of the binding cavity. S and S' denote subdomains of the cavity. Amino acids upstream of the cleavage site, bind to the S4-S1 subdomains. Amino acids located after the cleavage site, bind to the S1'-S3' subdomains.

Table 1. Interaction free energies of antimalarial drugs with 3CL coronavirus proteinase. Drugs are ordered in ascending order of the energy. The more negative energy, the stronger the interaction. For identification of the binding sites, see Fig.1and Fig. 2.

	Drug names	Chemical formulae	ΔG kcal/mol	Binding sites	3D structure of the complexes
1	*Clindamycine*		-18.49	S1 S2 S3 S1’	
2	*Lumefantrine*		-16.22	S1 S2 S3 S1’ S2’	
3	*Quinine*		-13.95	(S1) S2 (S3)	
4	*Mefloquine*				Work in progress

5	*Dihydroqinghaosu* (*artenimol dihydroartemisinin, DHA*)		-12.37	S2 S1’	
6	*Chloroquine*		-12.04	S1 S2 S1’	
7	*Doxycycline*				Work in progress
8	*Hydroxychloroquine*		-9.07	S1 S2 S1’ S2’	
9	*Pyrimethamine*		-7.87	S2	
10	*Piperaquine*		-7.85	(S1) S2 S3 S1’	

11	*Amodiaquine*		-7.13	S2, S3, (S3')	
12	*Artesunate*		-6.75	S2, (S3) S1' S2'	
13	*Atovaquone*		-6.50	S2 S3	
14	*Proguanil*		-6.37	S2 S3	
15	*Primaquine*		-5.86	S1 S2 (S3)	

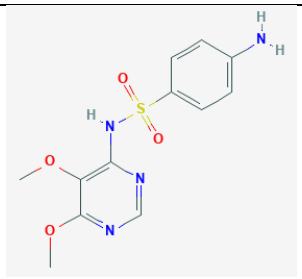
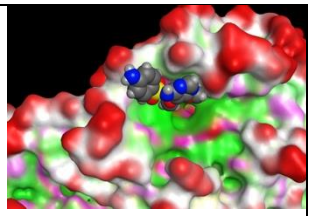
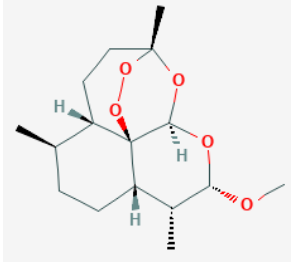
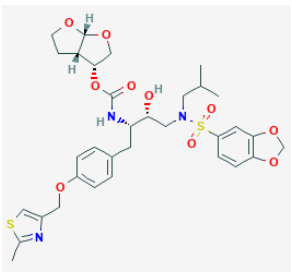
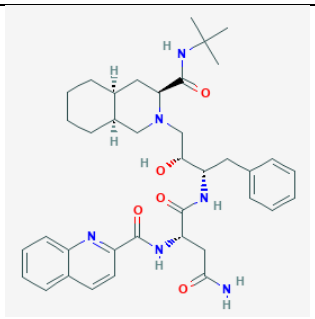
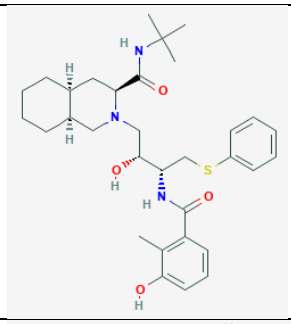
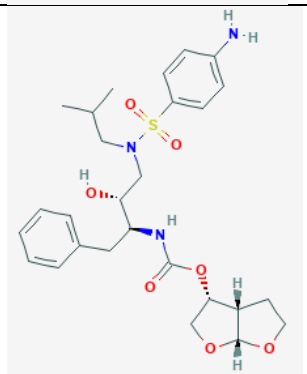
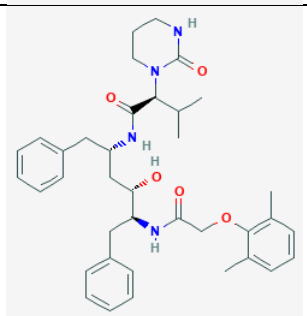
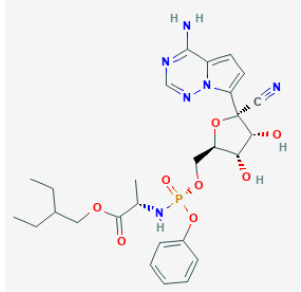
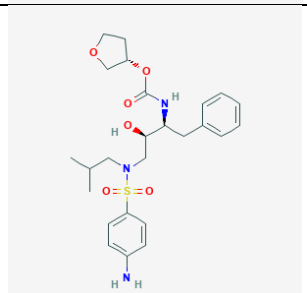
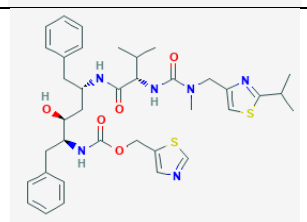
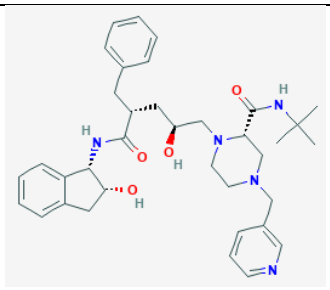
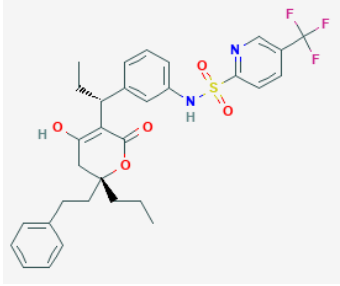
16	*Sulfadoxine*		-3.56	(S3)	
17	*Artemether*				Work in progress

Table 2. Interaction free energies of anti-AIDS drugs with 3CL coronavirus proteinase. Drugs are ordered in ascending order of the energy. The more negative energy, the stronger the interaction (energy results are subject to a refinement procedure). For identification of the binding sites, see Fig.1and Fig. 2

	Drug names	Chemical formulae	ΔG kcal/mol	Binding sites	3D structure of the complexes
1	*Brecanavir*		-24.6	(S1) S2 S3 S1' S2'	
2	*Saquinavir*		-21.00	S1 S2 (S3') S1	

3	*Nelfinavir*		-17.95	(S1) S2 (S1') S2'	
4	*Darunavir*		-5.89	S1 S2 S3	
5	*Lopinavir*		-2.97	S1 S2 (S3)	
6	*Remdesivir*		being computed		
7	*Amprenavir*		1.54	S2 S3' (S1', S2')	
8	*Ritonavir*		1.58	S1 S2 S2' (S1')	

9	*Indinavir*		5.18	S2 S3 S3’ (S2’)	
10	*Tipranavir*		5.74	(S1) S2 (S1’)	