

Virtual screening of an FDA approved drugs database on two COVID-19 coronavirus proteins

Alessandro Contini^{1*}

¹Università degli Studi di Milano, Dipartimento di Scienze Farmaceutiche, Sezione di Chimica Generale e Organica “A. Marchesini”, Via Venezian, 21 20133 Milano, Italy

*To whom correspondence should be addressed. E-mail: alessandro.contini@unimi.it

Abstract. The infection by the 2019-nCoV coronavirus (COVID-19) is a world-wide emergency. The crystal structure of a protein essential for virus replication has been filed in the Protein Data Bank recently. Additionally, homology models of 24 COVID-19 proteins were made available by the Zhang group. In this paper, we present results deriving from the virtual screening of a database of more than 3000 FDA approved drugs on two distinct targets. Results showed that some of the known protease inhibitors currently used in HIV infections might be helpful for the therapy of COVID-19 also.

Keywords: COVID-19, coronavirus, protease, virtual screening, Nwat-MMGBSA

Introduction

The infection by the COVID-19 coronavirus originated in the Hubei Province of China in late 2019 and, today,[1] infections spread out to 25 countries, including China. The total number of confirmed COVID-19 infections are 44730, plus 16067 suspect cases, and the number of deaths reached 1114. The scientific community is working for a vaccine, that will probably be available in 18 months.[2] However, in the meanwhile, finding a therapy to treat infections and reduce death cases is necessary. The scientific community immediately responded to this call, and different results were made available.[3] The crystal structure of the COVID-19 main protease in complex with a peptidomimetic inhibitor (PDB code 6LU7) was made available recently.[4] Additionally, the Zhang group, developer of the popular homology-modelling software I-TASSER,[5] made available 24 3D structural models [6] of proteins in the COVID-19 genome.[7]

Aiming to give our contribution, we designed a Virtual Screening campaign of currently FDA approved drugs.[8,9] This choice was driven by the need to provide a therapy quickly, since the development of a new drug might take years to complete. Two separate targets, the COVID-19 main protease 6LU7 and the Proteinase 3CL-PRO (QHD43415), were considered in our screening. The former is the only crystal structure currently available, to our knowledge, while the latter was chosen due to the high reliability of the model (Estimate TM-score=0.96). Moreover, being both proteases, they share the mechanism with many antiviral drugs currently used in the therapy of HIV infection, thus giving additional chance to find an effective drug quickly.

Methods

Receptors preparation.

Receptor models were prepared using the MOE2019 software,[10] with the following protocol: *6LU7*: all water molecules were deleted, then the covalently bound peptidomimetic ligand was unbound from Cys145 and the α,β double bond of the ligand, that behave as a Michael acceptor, was restored. The Structure Preparation module of MOE was used to correct PDB inconsistencies and to assign the protonation state at pH=7.0. The default Amber10EHT forcefield, coupled to the Born solvation model was assigned to the system. The ligand was then minimized, keeping the receptor constrained. Then, receptor was minimized by applying backbone restraints and keeping the ligand constrained. Finally, the complex was minimized in two separate steps, first by keeping backbone restraints, second by removing all restraints. The receptor and the ligand were then saved for future use.

4MDS: the crystal structure of SARS-CoV 3CL-PRO proteinase, that is a close homologue of COVID-19 3CL-PRO proteinase, in complex with a carboxamide inhibitor was also downloaded and used as a reference.[11] The system was prepared for calculations as follow: the PDB was corrected and protonated at pH=7.0 using MOE as stated above. The ligand was minimized, keeping the receptor constrained, using the MMFF94x forcefield coupled with the Born solvation model. The receptor was then minimized, keeping the ligand constrained, using Amber10EHT+Born. Finally, the complex was minimized in two steps, as described above.

QHD43415: the I-TASSER model QHD43415_5 [6] was superposed to 4MDS (prepared as previously described) in order to define the precise binding site. Since we observed that the 4MDS ligand also fitted QHD43415_5, the ligand was transferred, and the complex was prepared as described for 4MDS. The resulting structure was used for docking.

Database preparation.

Two separate databases were downloaded from the corresponding sources [8,9] and merged. The database was then checked for redundancy and redundant molecules, identified by CAS number, were removed. The database was then processed by MOE in order to build the 3D structure and to minimize the geometry of each molecule. The Wash function of the MOE database tool was used with the MMFF94x+Born force field, requesting the dominant protonation state at pH=7.0 and preserving existing chirality. The final database, consisting of 3118 unique molecules, was saved in SDF format.

Virtual Screening.

The Virtual Screening (VS) was done accordingly to our recently developed protocol.[12] The protocol is applied using a set of scripts (available for download as Supplementary Material within ref. [12]) that do the following steps automatically: 1. Preparation of the screening library. 2. Docking of all molecules (and corresponding tautomers, protonation states and ring conformers, if requested) using PLANTS [13]. 3. Analysis of results 4. Parameterization of docked ligands selected for rescoring. 5. Molecular dynamics of complexes selected for rescoring, using Amber [14]. Rescoring using the Nwat-MMGBSA method.[12,15] All screenings were performed by using the ligand database prepared as stated before (no tautomers or alternative protonation states). The following parameters were set for the screenings:

6LU7: center="-10.2858 12.3088 69.3271", rad=16, speed=1, sfunct=chemplp

QHD43415: center="-15.124 15.0521 -24.6152", rad=14, speed=1, sfunct=chemplp

Nwat-MMGBSA rescoring was requested for the top 1% compounds (about 30 molecules for each target). Rescoring consist in performing a short MD simulation (about 2.5 ns, including 1.5 ns of equilibration and 1 ns of production), followed by calculation of binding energy by MMGBSA[16] using the default protocol (Nwat=0) or selecting a certain number of explicit waters to be included in the calculation, accordingly to the Nwat-MMGBSA method (in this case, rescoring was repeated using Nwat=10, 20, 30, 60 and 100).

The same protocol was also applied to *4MDS*, since the binding energy computed for the 4MDS crystallographic ligand was used as a reference. Analogously, the binding energy of the 6LU7 ligand (whose covalent bond was broken as described above), was computed as a reference.

Results and discussion.

The results of the VS campaign conducted on COVID-19 main protease (PDB code 6LU7) and on COVID-19 3CL-PRO proteinase are summarized in Tables 1 and 2, respectively. Results of Nwat-MMGBSA rescoring, using 30 explicit waters, are the only reported, since Nwat=30 was considered a reasonable value in previous publications. [12,15]

Table 1. Results of the VS campaign on the crystal structure of COVID-19 main protease (PDB code 6LU7). Top 1% compounds selected from the docking of 3118 FDA approved drugs and rescored by Nwat-MMGBSA (Nwat=30) are shown. Compounds ranked better than the reference^a are highlighted in bold.

<i>Drug Name</i>	<i>Dock score</i>	<i>Nwat-MMGBSA^{b,c}</i>
Angiotensin II human Acetate	-124.4	-120.3±10.1
GHRP-2	-132.6	-106.0±8.3
Indinavir	-122.4	-86.5±5.7
Cobicistat (GS-9350)	-124.3	-72.8±8.3
Montelukast	-112.8	-54.9±6.8
Octenidine Dihydrochloride	-114.0	-50.8±4.9
Tyloxapol	-112.3	-42.5±6.5
Salvianolic acid B	-124.4	-41.1±11.0
Travoprost	-114.9	-35.6±6.1
Monomethyl auristatin E (MMAE)	-110.2	-67.2±7.8
Nafarelin Acetate	-155.4	F
Leuprorelin Acetate	-145.6	F
Leuprorelin Acetate	-140.9	F
Somatostatin Acetate	-139.6	F
Icatibant Acetate	-139.3	F
Nystatin (Fungicidin)	-132.7	F
Goserelin Acetate	-128.8	F
Alarelin Acetate	-128.5	F
Gonadorelin Acetate	-125.5	F
Amphotericin B	-124.7	F
Carfilzomib (PR-171)	-122.4	F
Thymopentin	-121.2	F
Lentinan	-116.4	F
Ritonavir	-116.1	F
NAD+	-114.1	F
Octreotide Acetate	-113.6	F
Colistin Sulfate	-113.5	F
Cangrelor Tetrasodium	-111.6	F
Oxytocin (Syntocinon)	-111.4	F
Flucytosine (Ancobon)	-111.3	F
Echinacoside	-110.3	F

a. The 6LU7 crystallographic ligand of 2019-nCoV main protease (6LU7) was used as the reference. Docking and Nwat-MMGBSA scores are -132.7 and -70.6±8.0 kcal/mol, respectively.

b. Nwat-MMGBSA rescoring was done considering 30 explicit water molecules around the ligand (Nwat=30).

c. "F" means that the molecular dynamic simulation failed to converge, possibly due to poor parameterization or unacceptable steric clashes in the docking pose.

Table 2. Results of the VS campaign on the homology model of COVID-19 Proteinase 3CL-PRO (QHD43415).^a Top 1% compounds selected from the docking of 3118 FDA approved drugs and rescored by Nwat-MMGBSA (Nwat=30) are shown. Compounds that ranked better than the reference^b are highlighted in bold.

<i>Drug Name</i>	<i>Dock score</i>	<i>Nwat-MMGBSA^{c,d}</i>
Caspofungin Acetate	-108.3	-97.9±12.4
Lopinavir (ABT-378)	-106.5	-89.9±5.9
Atazanavir	-109.9	-86.0±7.0
GHRP-2	-116.7	-79.2±11.1
Indinavir	-105.4	-78.6±6.5
Angiotensin II human Acetate	-125.7	-75.7±9.2
Ritonavir	-112.3	-58.3±7.8
Salvianolic acid B	-116.0	-51.0±7.7
Elbasvir	-110.2	-48.0±7.7
Montelukast Sodium	-110.1	-47.5±6.9
Cobicistat (GS-9350)	-119.3	-45.4±11.6
Tyloxapol	-105.5	-41.1±8.3
Salmeterol Xinafoate	-105.3	-37.9±7.3
Penfluridol	-106.2	-31.8±9.6
Gonadorelin Acetate	-131.4	F
Leuprorelin Acetate	-127.9	F
Nafarelin Acetate	-127.9	F
Leuprorelin Acetate	-124.1	F
Goserelin Acetate	-117.8	F
Bacitracin Zinc	-117.0	F
Amphotericin B	-115.2	F
Alarelin Acetate	-114.6	F
Deferoxamine mesylate	-113.7	F
Nystatin (Fungicidin)	-112.8	F
Octreotide Acetate	-108.0	F
Carfilzomib (PR-171)	-107.7	F
Terlipressin Acetate	-107.3	F
Somatostatin Acetate	-106.5	F
Flucytosine (Ancobon)	-106.4	F
Lypressin Acetate	-106.3	F
Sennoside A	-105.6	F

a. The homology model of 2019-nCoV 3CL-PRO was done and made available by the Zhang group at <https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/2019-nCov/>

b. The 4MDS crystallographic ligand in complex with SARS-CoV 3CLpro, a close homologue of 2019-nCoV 3CL-PRO, was used to compute reference scorings. Docking and Nwat-MMGBSA scores are -96.4 and -59.8±5.3 kcal/mol, respectively.

c. Nwat-MMGBSA rescoring was done considering 30 explicit water molecules around the ligand (Nwat=30).

d. “F” means that the molecular dynamic simulation failed to converge, possibly due to poor parameterization or unacceptable steric clashes in the docking pose.

As it can be observed, the protease inhibitor Indinavir,[16] currently used to treat HIV infections, has been selected by both targets. Additionally, the other protease inhibitors Lopinavir [17] and Atazanavir [18] are also top ranked for 3CL-PRO. Cobicistat, another drug that is approved for the treatment of HIV infection, has been selected as a potential COVID-19 main protease inhibitor, even if its main mechanism is claimed to be the inhibition of CYP3A.[19]

Angiotensin II and GHRP-2 are also selected in both screenings. However, it should be considered that both Angiotensin II and GHRP-2 are peptides, and so they are likely to be substrates for proteases. Even if they might bind the target, they would be rapidly cleaved by the target itself. Thus, we consider both compounds as false positives. The identification of Caspofungin, an

antifungal drug specifically used in HIV-infected individuals [20] is also unexpected. Considering the macrocyclic structure of Capsfungin, together with the large number of rotatable bonds, we expect the docking of such a molecule to be rather challenging and so we have some doubts upon the reliability of this results. Several other compounds were selected by docking but failed during the molecular dynamic simulation. Such a failure might be due to several causes, among all poor parameterization (the BCC charge parameterization method [21] is chosen for time constraints). However, the reason of the failure might also be due to severe steric clashes or mispositioning during the docking stage. Thus, although some good hit might be found within the "F" series, these selections are considered the least reliable. Further calculations with more reliable (but much more time-consuming) parameterization schemes should be made to confirm whether or not these compounds are potential COVID-19 main protease or 3CL-PRO proteinase inhibitors.

Acknowledgement

I thank Dr. Andrea Bazzoli for useful discussions that lead to the idea of performing this study.

References

- [1] WHO Situation Report -23, February 12th 2020.
- [2] <https://www.pharmaceutical-technology.com/news/who-forum-coronavirus-vaccine/>
- [3] Corman et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, *Euro Surveill.* 2020 Jan 23; 25(3): 2000045. doi: 10.2807/1560-7917.ES.2020.25.3.2000045
- [4] Liu, X., Zhang, B., Jin, Z., Yang, H., Rao, Z., The crystal structure of 2019-nCoV main protease in complex with an inhibitor N3. To be published. PDB code 6LU7, <https://www.rcsb.org/structure/6lu7>
- [5] A Roy, A Kucukural, Y Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5: 725-738 (2010)
- [6] Chengxin Zhang, Wei Zheng, Xiaoqiang Huang, Eric W. Bell, Xiaogen Zhou, Yang Zhang (2020) "Protein structure and sequence re-analysis of 2019-nCoV genome does not indicate snakes as its intermediate host or the unique similarity between its spike protein insertions and HIV-1." doi: <http://dx.doi.org/10.1101/2020.02.04.933135>
- [7] <https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3>
- [8] <https://www.selleckchem.com/screening/fda-approved-drug-library.html>
- [9] <https://www.targetmol.com/compound-library/FDA-approved-Drug-Library>
- [10] Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2019.
- [11] Turlington, M., Chun, A., Tomar, S., Egger, A., Grum-Tokars, V., Jacobs, J., Daniels, J.S., Dawson, E., Saldanha, A., Chase, P., Baez-Santos, Y.M., Lindsley, C.W., Hodder, P., Mesecar, A.D., Stauffer, S.R. (2013) *Bioorg.Med.Chem.Lett.* 23: 6172-6177
- [12] I. Maffucci, X. Hu, V. Fumagalli and A. Contini An Efficient Implementation of the Nwat-MMGBSA Method to Rescore Docking Results in Medium-Throughput Virtual Screenings. *Front. Chem.*, 05 March 2018 | <https://doi.org/10.3389/fchem.2018.00043>
- [13] Korb, O., Möller, H. M., and Exner, T. E. (2010). NMR-guided molecular docking of a protein-peptide complex based on ant colony optimization. *ChemMedChem* 5, 1001–1006. doi: 10.1002/cmdc.201000090
- [14] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2018), AMBER 2018, University of California, San Francisco.

- [15] a) Maffucci, I., and Contini, A. (2013). Explicit ligand hydration shells improve the correlation between MM-PB/GBSA binding energies and experimental activities. *J. Chem. Theory Comput.* 9, 2706–2717. doi: 10.1021/ct400045d. b) Maffucci, I., and Contini, A. (2016). Improved computation of protein–protein relative binding energies with the Nwat-MMGBSA method. *J. Chem. Inf. Model.* 56, 1692–1704. doi: 10.1021/acs.jcim.6b00196
- [16] Massova, I., and Kollman, P. A. (2000). Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov. Des.* 18, 113–135. doi: 10.1023/A:1008763014207
- [16] <https://pubchem.ncbi.nlm.nih.gov/compound/indinavir>
- [17] <https://www.ncbi.nlm.nih.gov/pubmed/11367365?dopt=Abstract>
- [18] <https://www.drugs.com/monograph/atazanavir-sulfate.html>
- [19] Xu L, Liu H, Murray BP, et al. Cobicistat (GS-9350): A Potent and Selective Inhibitor of Human CYP3A as a Novel Pharmacoenhancer. *ACS Med Chem Lett.* 2010;1(5):209–213. Published 2010 May 17. doi:10.1021/ml1000257
- [20] Waters L, Nelson M The use of caspofungin in HIV-infected individuals. *Expert Opin Investig Drugs.* 2007 Jun;16(6):899-908.
- [21] Jakalian A., Jack D.B., Bayly C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem.* 2002 Dec;23(16):1623-41.