DOI 10.18699/SBB-2022-513

Finding possible inhibitors for SARS-CoV-2 proteins using virtual ligand screening

Gorelov S.V.[1, 3]*, Shvetsov A.V.[1, 3], Egorov V.V.[1, 2], Isaev-Ivanov V.V.[1]

[1] *Saint Petersburg Nuclear Physics Institute named by B.P. Konstantinov, NRC "Kurchatov Institute", Gatchina, Russia*

[2] *Smorodintsev Research Institute of Influenza, St. Petersburg, Russia*

[3] *St. Petersburg Polytechnic University of Peter the Great, St. Petersburg, Russia*

** Gorelov_sv@pnpi.nrcki.ru*

Key words: nsp, molecular docking, molecular dynamics, virtual screening, structure, energy minimization, active center, box, AutoDock Vina, GROMACS

Motivation and Aim: SARS-CoV-2 is an enveloped, single positive-stranded RNA belonging to the genus Betacoronavirus, a virus with large viral RNA genomes that causes COVID-19 disease. It was first registered in 2019. Although SARS-CoV-2 belongs to the group of betacoronaviruses, it differs from MERS-CoV and SARS-CoV. So far, it has already affected more than 210 million people worldwide. Existing virtual screening and computer simulation methods allow for research that can detect new potential inhibitors of the main protease of the SARS virus, thereby identifying substances with possible therapeutic properties against COVID-19. Coronavirus replication cycle begins with translation of ORF1a and ORF1b after penetration and release of the genome. First, two large replicase polyproteins (pp1a and pp1ab) are synthesized. They are then cleaved into 16 non-structural proteins (nsps), including nsp12, viral RNA-dependent RNA polymerase (RdRp), papain-like protease (nsp3), and 3C-like protease (nsp5). Viral unstructured proteins engage the host cell's membrane structure to assemble into replication and transcription complexes that are involved in the synthesis of the RNA minus strand. Then subgenomic RNAs are synthesized and structural and accessory proteins are expressed. The pathogenesis of severe acute respiratory syndrome coronavirus SARS-CoV-2 is an important issue for the treatment and prevention of SARS, and the current state of the problem under study allows finding new potential inhibitors [1, 2]. Methods and Algorithms: This work is devoted to the virtual screening of ligands for Sars-CoV-2 proteins, in particular, for unstructured proteins nsp5 and nsp12. For this, the methods of molecular docking and molecular dynamics were used. Molecular docking was calculated using the AutoDock Vina program, while the GROMACS universal software package was used for molecular dynamics and energy minimization. Substances from the ZINK15 database in the amount of 5901 units served as ligands for molecular docking. Results: Using the methods of molecular docking and molecular modeling, a virtual screening of replicase ligands of SARS-CoV-2 polyprotein 1ab was performed. Carrying out test runs showed that the AutoDock Vina software for molecular docking and virtual screening really has high accuracy and speed of calculations. Software algorithms were created for visualizing search areas and carrying out stream docking with subsequent sorting of the results obtained, which can be used in further work. Structured lists of substances have been obtained that can be used for future calculations of binding constants in solution with recombinant proteins. Since the ligands from the ZINC15 database, with small

differences in structure, had an insignificant spread in affinity values, it can be assumed that docking was applied correctly. Using the method of molecular dynamics, the possible binding of the nsp7 protein to the NILYRSLAETR peptide sequence was analyzed. The natural nsp7 protein was found to have at least 6 potential binding sites for the putative ligand. The nearest amino acid residues that are supposedly involved in intermolecular interactions have been characterized. The process of attaching a peptide sequence to nsp7 was also visualized. It was found that NILYRSLAETRs are indeed able to form fibril-like structures among themselves. Conclusion: A workflow algorithm was developed for calculating molecular docking for any number of ligands with almost any receptor protein, followed by structuring the output data. A similar algorithm was used to calculate molecular docking for structured proteins nsp5 and nsp12 in different regions. For each ligand, 20 stacking conformations with their corresponding affinity values were obtained at the output. Of these, the 3 highest affinity values were selected and, having calculated their arithmetic mean, sorted in descending order with their corresponding ligands. This made it possible to create a list of the 25 most optimal ligands for each receptor. Using the data obtained from the molecular dynamics procedure using the GROMACS universal modeling software package, it was possible to visualize the trajectory of 10 molecules of fragments of the sequence of feline, canine and porcine coronaviruses and the unstructured nsp7 protein. The trajectory of protein movement was analyzed, due to which some potential binding sites of nsp7 with molecules of peptide sequences NILYRSLAETR in the amount of 6 pieces were identified.

References

1. Wang W., Zhou Z., Xiao X. et al. *Cell Mol Immunol.* 2021 Apr;18(4):945-953.
2. Mariano G., Farthing R.J., Lale-Farjat S.L.M., Bergeron J.R.C. *Front Mol Biosci*. 2020 Dec;17(7):605236.