Computational Modeling of the SARS-CoV-2 Main Protease Inhibition by the Covalent Binding of Prospective Drug Molecules

Alexander V. Nemukhin[1,2], *Bella L. Grigorenko*[1,2], *Igor V. Polyakov*[1,2], *Sofya V. Lushchekina*[2]

We illustrate modern modeling tools applied in the computational design of drugs acting as covalent inhibitors of enzymes. We take the Main protease (M^{pro}) from the SARS-CoV-2 virus as an important present-day representative. In this work, we construct a compound capable to block M^{pro}, which is composed of fragments of antimalarial drugs and covalent inhibitors of cysteine proteases. To characterize the mechanism of its interaction with the enzyme, the algorithms based on force fields, including molecular mechanics (MM), molecular dynamics (MD) and molecular docking, as well as quantum-based approaches, including quantum chemistry and quantum mechanics/molecular mechanics (QM/MM) methods, should be applied. The use of supercomputers is indispensably important at least in the latter approach. Its application to enzymes assumes that energies and forces in the active sites are computed using methods of quantum chemistry, whereas the rest of protein matrix is described using conventional force fields. For the proposed compound, containing the benzoisothiazolone fragment and the substitute at the uracil ring, we show that it can form a stable covalently bound adduct with the target enzyme, and thus can be recommended for experimental trials.

Keywords: SARS-CoV2 main protease, QM/MM, molecular docking, molecular dynamics, covalent inhibitor.

Introduction

Under the current public health emergency caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), it is imperative to propose prospective drugs for medical treatment of COVID-19. The cysteine protease, called the Main protease (M^{pro}), from the SARS-CoV-2 is considered as one of the targets in current research [11]. The goal is to inhibit the enzyme and to prevent its function required for viral replication and transcription. In this paper, we describe practical steps in computational modeling of the mechanism of M^{pro} inhibition by a chemical compound capable to form a covalent bond with the catalytic amino acid residue cysteine in the active site of the enzyme, which causes the inhibition. This approach differs from a common strategy to propose inhibiting molecules, which form non-covalently bound complexes with an enzyme, thus blocking the entrance to the active site for natural substrates. From the computational side, the latter strategy requires an application of algorithms of molecular docking and classical molecular dynamics, which are successfully used in high performance computer simulations [12]. If our aim is to predict a covalent inhibitor, a broader range of computational tools is necessary, including quantum chemistry based methods [1], which require supercomputer resources. These approaches are discussed in the Methodology section. A specific drug candidate capable to inhibit SARS-CoV-2 M^{pro} by the covalent binding is proposed in this work by the following reasons. The search of prospective candidates is often based on screening the data bases of drugs already approved for treatment of other diseases in attempts to find a new application for a corresponding chemical compound. In particular, the known antimalarial drugs

[1]Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia
[2]Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, Russia

are among those actively investigated to fight COVID-19. At preliminary steps, we attempted to attach the molecule of the famous hydroxychloroquine compound to M^{pro}. This substance is widely known, for example, from mass media communications. Our studies, including docking and molecular dynamics simulations, did not show positive trends in respect of its interaction with M^{pro}. More promising species were found among other proposed antimalarial drugs with the benzoisothiazolone moiety [10]. Some of these compounds could be docked to M^{pro} with reasonable binding energies. Second, a recent study of covalent inhibitors of another cysteine protease [5] suggests a set of molecules with the building blocks, which can be efficient covalent inhibitors of M^{pro}. Correspondingly, we designed computationally a compound by joining a benzoisothiazolone fragment, which is capable to dock at M^{pro}, with a moiety capable to interact with the active site of M^{pro} forming a stable adduct. A molecular model of this compound is shown in Fig. 1.
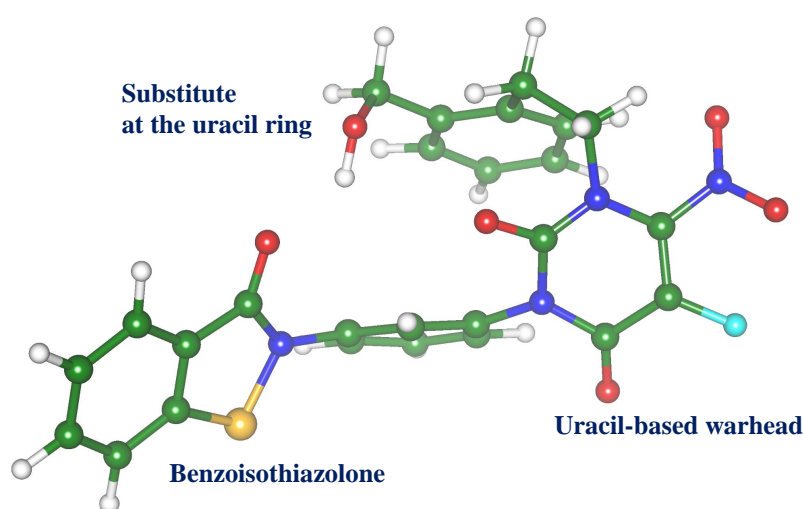

Figure 1. Molecular model of the compound considered as a covalent inhibitor of SARS-CoV-2 M^{pro}. The uracil-based warhead is responsible for the chemical reaction with the enzyme, the benzoisothiazolone fragment and the substitute at the uracil ring provide the enhanced binding to the protein. Green – carbon atoms, red – oxygen, blue – nitrogen, yellow – sulfur, cyan – fluorine

It should be pointed out that actually a set of molecules can be derived on the base of the used template, namely, a construct from the benzoisothiazolone moiety and cysteine protease inhibitors, which differ by substitutes at the uracil warhead. Below, we describe the computational approaches, which are utilized in modeling mechanisms of the covalent inhibition of M^{pro} taking this particular compound, but these approaches are similar for the entire series of prospective drugs.

1. Methodology

The following steps should be performed to predict computationally a covalent inhibitor of an enzyme.

Step 1. One should construct molecular models of the compound and the protein. To design a molecule of the type shown in Fig. 1, computer programs of molecular mechanics (MM) and quantum chemistry (QC) are the major tools. Typically, organic molecules of prospective drugs contain up to a hundred of atoms, and MM-based constructors do not require expensive

calculations. There are a plenty of algorithms and computer programs employing conventional force field parameters, which allow one to create a preliminary model. If necessary, equilibrium geometry parameters of the molecule and the atomic charges can be cleaned in QC calculations. Although a high accuracy of these parameters is not expected in this application, and expensive QC algorithms can be avoided, nevertheless, supercomputer resources can be requested if a problem of multiple conformations of polyatomic flexible molecule should be solved. Initial coordinates of an enzyme macromolecule usually are taken from a suitable crystal structure deposited to the Protein Data Bank [2]. To convert these raw materials to a relevant molecular model one should add hydrogen atoms, which are not recognized in crystallography experiments, and often restore missing amino acid residues or correct artificial molecular groups needed to perform measurements. Also, the protein molecule should be surrounded by a proper amount of solvent water molecules. These actions are carried out using either MM-based computer programs, or classical molecular dynamics (MD) simulations. The latter can require supercomputer resources because of an enormous amount of atoms in a model molecular system.

Step 2. Molecular docking is one of the most popular molecular modeling techniques, particularly wide spread in drug design. While in certain cases docking can be performed using inexpensive equipment, the use of the corresponding programs for virtual screening of large databases containing many thousands of ligands requires supercomputer resources. Moreover, for flexible ligands with many torsion degrees of freedom (like, for instance for the compound shown in Fig. 1), the multi-processor performance of the docking program at several dozen or hundred computing cores of a supercomputer is important [8, 12]. Therefore, molecular docking of the selected compound to the enzyme surface is a major step in characterization of non-covalent inhibitors and it constitutes a preliminary step in the design of covalent inhibitors. These algorithms are very helpful in finding location of the compound near the enzyme active site to obtain a first approach to construct an enzyme-substrate (ES) complex. Usually, the ES complex, which is required in subsequent calculations of reaction profiles, differs considerably from a hint in docking experiments; however, analysis of such obtained protein-ligand configurations is practically important.

Step 3. Making and breaking chemical bonds, which accompany formation of covalently bound adducts in modeling of covalent inhibition of enzymes, cannot be described at the level of force fields and require application of quantum-based algorithms. The most suitable approach, the hybrid quantum mechanics / molecular mechanics (QM/MM) method, was proposed almost half-century ago. Its application to enzymes assumes that energies and forces in the active sites are computed using methods of quantum chemistry, whereas the rest of protein matrix is described using conventional force fields. Since first applications, this methodology developed rapidly and presently several computer packages can be used for QM/MM calculations. For instance, we make a reference to the work [3], in which the entire reaction energy profile in enzyme catalysis from the ES complex to the enzyme-product (EP) complex was evaluated using QM/MM. In our case, the ES complex corresponds to the structure optimized in QM/MM calculations initiated by the atomic coordinates preliminary obtained in molecular docking and classical molecular dynamics simulations. The optimization means that all geometry parameters of the complex corresponding to the minimum energy point on the potential energy surface are evaluated using the QM/MM approach, which requires multiple calculations of the energy and energy gradients with respect to coordinates of nuclei. If reliable QM/MM approaches are applied, including a reasonable size of the QM-subsystem, trustworthy functionals in the density

functional theory (DFT), and well-defined basis sets to represent molecular orbitals, then the use of supercomputer resources is inevitable. Next, a reaction coordinate should be specified to describe the pathway from ES to EP. Multiple series of constrained QM/MM minimizations, i.e. optimizations of all structural parameters for each value of the reaction coordinate should be carried out to find the points of reaction intermediates and transition states. Needless to say, that those are extremely time- and resource-consuming calculations. In our case, an important quantity is the energy difference between the ES and EP points, which indicates, how stable is the covalently bound adduct, i.e. the product of the interaction of a possible drug molecule with an enzyme.

2. Implementation for the M^{pro} Covalent Inhibition

This paper describes all steps listed in the Methodology section for the case of the covalent inhibition of the SARS-CoV-2 M^{pro}.

Step 1. The compound shown in Fig. 1 was constructed using the Discovery Studio Visualizer MM-based program. QC calculations at the DFT PBE0/6-31G* level using the NWChem program [13] allowed us to correct few geometry parameters of the molecule. Initial set of atomic coordinates was taken from the structure of the ligand-free SARS-Cov-2 M^{pro} deposited to the Protein Data Bank [2] as the 6yb7 entry. Restoration of the three-dimensional all-atom molecular model was preformed using the MM and MD tools. Solvent water molecules were added using the Visual Molecular Dynamics (VMD) package [4] to prepare a solvent box border at least 1000 pm from the protein. Relaxation of the structure was performed in MD simulations using the NAMD program [9] with the CHARMM36 force field [14] for the protein and TIP3P for water molecules with the NPT ensemble at 298 K temperature and 1 atm pressure. Positions of all atoms found in the X-ray structure, including oxygen atoms of crystallographic water molecules, were fixed. First, added solvent shells were equilibrated during 1 ns MD simulation, until cell volume stabilized. Next, the system was minimized during 1000 steps. In QM/MM calculations the solvent shell was reduced to the distance of 600 pm from the protein. Figure 2 illustrates a ligand-free model protein system showing the enzyme active site in the inset. The mechanism of this enzyme includes the nucleophilic attack of S_{Cys} on the carbon atom of a substrate coupled with the proton transfer from the S-H group of Cys145 to the nitrogen atom N_{His} of His41. The inset in the right part in Fig. 2 shows the catalytic dyad, Cys145 and His41, and the groups from the 143-145 chain forming the oxyanion hole, N_{Cys}-H and N_{Gly}-H.

Step 2. The active site of M^{pro} spans far beyond the catalytic site illustrated in Fig. 2. Four subsites (pockets) S1–S4 are distinguished on the enzyme surface [16]. Hence the molecular docking grid box included the whole area with the all these pockets. This requires 80 grid points in each direction with grid spacing 375 pm (30000 pm × 30000 pm × 30000 pm). Affinity maps for this grid box were prepared with Autogrid 4.2.6 tool of AutoDock suite [8]. Considering an amount of torsional degrees of freedom of the considered compound and a size of the grid box, the following parameters of the main Lamarckian Genetic Algorithm [7] were used: 256 runs, 25×10^6 evaluations, 27×10^4 generations, and a population size of 3000. Docking was performed with AutoDock 4.2.6 program [8]. The protein was fully rigid during molecular docking. Analyzing the obtained docked poses, a particular attention was payed to the distance between the carbon atom of the inhibitor and the sulfur atom of Cys145 of the enzyme, i.e. the nucleophilic attack distance. Position with the shortest distance of 450 pm was in TOP-5 poses, ranged by estimated binding free energies, difference with them was within 0.2 kcal/mol and the binding

energy was equal to –9.4 kcal/mol. Thus, we conclude that there are several binding poses competing with each other, and one of them is productive. With the warhead (see Fig. 2) in the active site, an inhibitor occupies S2, S3 and S4 pockets of the active site (Fig. 3).
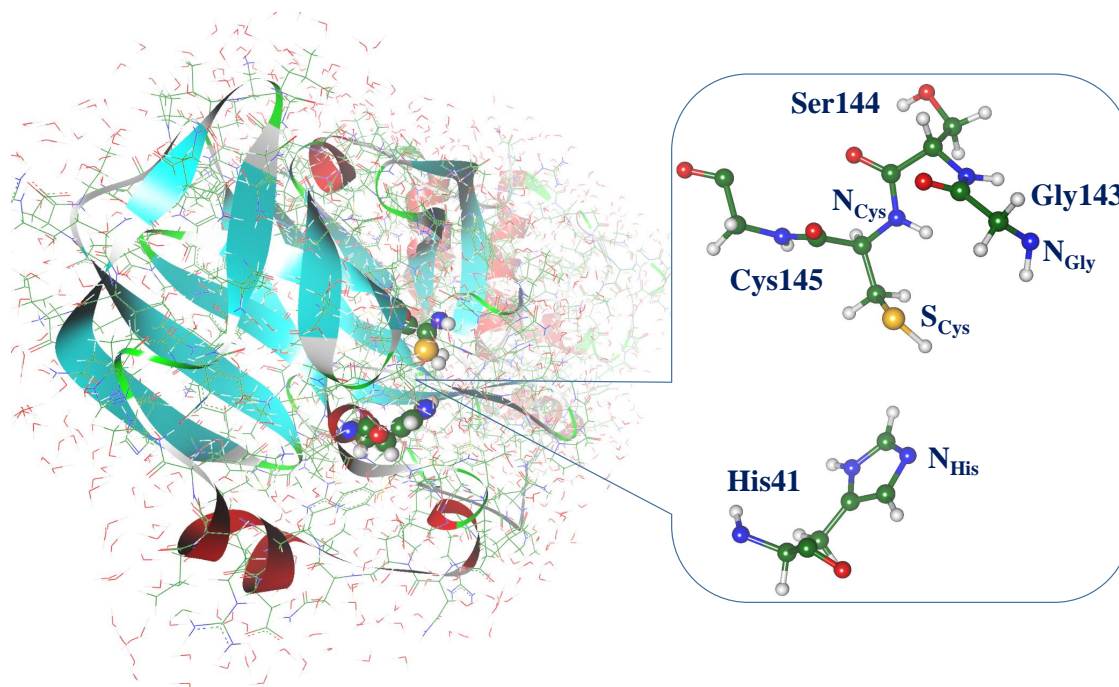

Figure 2. A model system for the SARS-CoV-2 M$^{\text{pro}}$. Left: a view on the ligand-free protein. Right: the substrate-free active site. Green – carbon atoms, red – oxygen, blue – nitrogen, yellow – sulfur
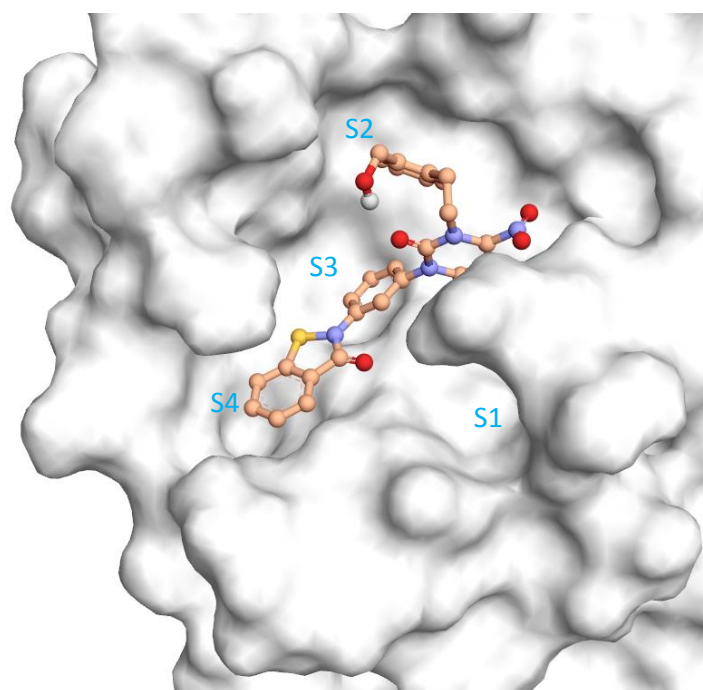

Figure 3. Position of the inhibitor on the Mpro according to the docking studies

Step 3. QM/MM calculations of the reaction energy profile were carried out on the "Lomonosov-2" supercomputer [15]. The NWChem software package [13] was compiled man-

ually with the Intel® Parallel Studio XE 2019 Update 5 and linked with Intel® Math Kernel Library and OpenMPI 4.0.1 message-passing library. Our previous experience benchmarking the NWChem [6] proved that a storage-based approach was much faster than the "direct" self-consistent field (SCF) procedure. The only disadvantage is that one needs an adequate amount of nodes to hold the intermediate data (mostly, the two-electron integrals) in the fast memory (RAM in this case, the nodes are diskless). Hence, we used 8 nodes per task to be able to hold all the data in RAM and achieve a proper performance. For example, a typical calculation with 1250 basis set functions, 50 optimization cycles (each consists of 5 steps in the quantum subsystem and 100 classical steps) takes around 12 hours of wall time or 1344 CPU×hours. In total, it took over 150000 CPU×hours* to compute the entire reaction energy profile. The selected reaction coordinate is a distance between S_{Cys} of the catalytic Cys145 of the enzyme and the carbon atom covalently bound to the fluorine atom in the inhibitor molecule (Fig. 1). Figure 4 illustrates the QM/MM optimized structure of the stable reaction intermediate on the reaction route with the covalent bond between the inhibitor molecule and catalytic cysteine Cys145 of the enzyme. The adduct is tightly bound in the protein matrix. The C-S_{Cys} distance is 188 pm corresponding to the newly formed chemical bond. The F-N distances (387 pm and 397 pm) evidence that the adduct enters the oxyanion hole.
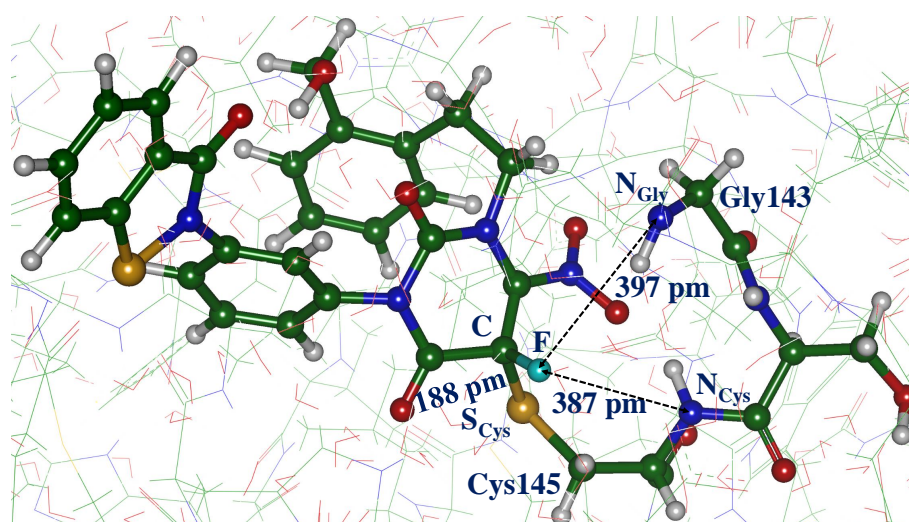

Figure 4. Stable reaction intermediate formed upon the interaction of the inhibitor with the M^pro. Green – carbon atoms, red – oxygen, blue – nitrogen, yellow – sulfur, cyan – fluorine

The reaction pathway includes the points of the enzyme-substrate (ES) complex, reaction intermediate (INT) and the enzyme-product (EP) complex. The ES complex is a non-covalent complex of a substrate in the active site of an enzyme; the reaction intermediate is a transient covalent complex of these two species and the EP is the final stable covalent complex. By using the DFT approach PBE0/6-31G* in the QM-subsystem and the AMBER force field parameters in the MM-subsystem we obtain that the energy of the reaction intermediate shown in Fig. 4 is lower than the level of ES by about 15 kcal/mol, whereas the enzyme-product (EP) complex, separated from ES by energy barriers of about 10 kcal/mol, is lower than the level of ES by about 19 kcal/mol. These energy gaps are large enough to conclude that the covalently bound adduct is a fairly stable complex, and the enzyme cannot function after capturing the inhibitor. This indicates that the goal of simulations is achieved, and the computationally designed compound is able to block M^pro.

Conclusion

We show that the computationally predicted compound (Fig. 1) created by motifs of the known antimalarial drug molecules and cysteine protease inhibitors can be considered as a prospective drug capable to inhibit the critical enzyme M^{pro} from the SARS-CoV-2 virus. Molecular dynamics and molecular docking approaches evidence that this molecule can be attached to the surface of the enzyme and obstruct delivery of natural substrates. A more intriguing option is a covalent binding of the compound to the catalytic cysteine residue from the M^{pro} active site. Results of high performance QM/MM calculations show that the energy of the enzyme-product complex is about 19 kcal/mol lower that the level of the enzyme-substrate complex, which leads to complete blocking the enzyme.

Acknowledgements

References

1. Arodola, O.A., Soliman, M.E.: Quantum mechanics implementation in drug-design workflows: does it really help? Drug Design, Development and Therapy 11, 2551–2564 (2017), DOI: 10.2147/DDDT.S126344

2. Berman, H.M.: The Protein Data Bank. Nucleic Acids Research 28(1), 235–242 (2000), DOI: 10.1093/nar/28.1.235

3. Grigorenko, B.L., Khrenova, M.G., Nilov, D.K., Nemukhin, A.V., Švedas, V.K.: Catalytic cycle of penicillin acylase from *Escherichia coli*: QM/MM modeling of chemical transformations in the enzyme active site upon penicillin G hydrolysis. ACS Catalysis 4(8), 2521–2529 (2014), DOI: 10.1021/cs5002898

4. Humphrey, W., Dalke, A., Schulten, K.: VMD: Visual molecular dynamics. Journal of Molecular Graphics 14(1), 33–38 (1996), DOI: 10.1016/0263-7855(96)00018-5

5. Klein, P., Johe, P., Wagner, A., et al.: New cysteine protease inhibitors: Electrophilic (het)arenes and unexpected prodrug identification for the *Trypanosoma* protease rhodesain. Molecules 25(6), 1451 (2020), DOI: 10.3390/molecules25061451

6. Mironov, V.A., Grigorenko, B.L., Polyakov, I.V., Nemukhin, A.V.: Benchmarking quantum chemistry methods in calculations of electronic excitations. Supercomputing Frontiers and Innovations 5(4), 62–66 (2018), DOI: 10.14529/jsfi180405

7. Morris, G.M., Goodsell, D.S., Halliday, R.S., et al.: Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. Jour-

nal of Computational Chemistry 19(14), 1639–1662 (1998), DOI: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

8. Morris, G.M., Huey, R., Lindstrom, W., et al.: AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry 30(16), 2785–2791 (2009), DOI: 10.1002/jcc.21256

9. Phillips, J.C., Braun, R., Wang, W., et al.: Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 26(16), 1781–1802 (2005), DOI: 10.1002/jcc.20289

10. Price, K.E., Armstrong, C.M., Imlay, L.S., et al.: Molecular mechanism of action of antimalarial benzoisothiazolones: species-selective inhibitors of the Plasmodium spp. MEP pathway enzyme, IspD. Scientific Reports 6(1), 36777 (2016), DOI: 10.1038/srep36777

11. Sivasankarapillai, V.S., Pillai, A.M., Rahdar, A., et al.: On facing the SARS-CoV-2 (COVID-19) with combination of nanomaterials and medicine: Possible strategies and first challenges. Nanomaterials 10(5), 852 (2020), DOI: 10.3390/nano10050852

12. Sulimov, A., Kutov, D., Sulimov, V.: Supercomputer docking. Supercomputing Frontiers and Innovations 6(3), 26–50 (2019), DOI: 10.14529/jsfi190302

13. Valiev, M., Bylaska, E., Govind, N., et al.: NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. Computer Physics Communications 181(9), 1477–1489 (2010), DOI: 10.1016/j.cpc.2010.04.018

14. Vanommeslaeghe, K., MacKerell, A.D.: Charmm additive and polarizable force fields for biophysics and computer-aided drug design. Biochimica et Biophysica Acta 1850(5), 861–871 (2015), DOI: 10.1016/j.bbagen.2014.08.004

15. Voevodin, V.V., Antonov, A.S., Nikitenko, D.A., et al.: Supercomputer Lomonosov-2: Large scale, deep monitoring and fine analytics for the user community. Supercomputing Frontiers and Innovations 6(2), 4–11 (2019), DOI: 10.14529/jsfi190201

16. Zhang, L., Lin, D., Sun, X., et al.: Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. Science 368(6489), 409–412 (2020), DOI: 10.1126/science.abb3405