DOI: 10.14529/jsfi200305

In Search of Non-covalent Inhibitors of SARS–CoV–2 Main Protease: Computer Aided Drug Design Using Docking and Quantum Chemistry

Alexey V. Sulimov[1,2]***, Danil C. Kutov***[1,2]***, Anna S. Taschilova***[1,2]***,***
Ivan S. Ilin[1,2]***, Nadezhda V. Stolpovskaya***[3]***, Khidmet S. Shikhaliev***[3]***,***
Vladimir B. Sulimov[1,2,4]

Two stages virtual screening of a database containing several thousand low molecular weight organic compounds is performed with the goal to find inhibitors of SARS–CoV–2 main protease. Overall near 41000 different 3D molecular structures have been generated from the initial molecules taking into account several conformers of most molecules. At the first stage the classical SOL docking program is used to determine most promising candidates to become inhibitors. SOL employs the MMFF94 force field, the genetic algorithm (GA) of the global energy optimization, takes into account the desolvation effect arising upon protein-ligand binding and the internal stress energy of the ligand. Parameters of GA are selected to perform the meticulous global optimization, and for docking of one ligand several hours on one computing core are needed on the average. The main protease model is constructed on the base of the protein structure from the Protein Data Bank complex 6W63. More than 1000 ligands structures have been selected for further postprocessing. The SOL score values of these ligands are more negative than the threshold of −6.3 kcal/mol obtained for the native X77 ligand docking. Subsequent calculation of the protein-ligand binding enthalpy by the PM7 quantum-chemical semiempirical method with COSMO solvent model have narrowed down the number of best candidates. Finally, the diverse set of 20 most perspective candidates for the *in vitro* validation are selected.

Keywords: docking, global optimization, quantum docking, inhibitors, CADD, SARS–CoV–2, COVID–19, M^{pro}.

Introduction

The recent outbreak of the COVID–19 pandemic caused by the SARS–CoV–2 coronavirus (CoV) makes the development of appropriate drugs and vaccines extremely urgent. Currently, there are no direct-acting drugs for the SARS–CoV–2 virus that are specifically designed to inhibit the proteins of this particular coronavirus. At the same time, there are already three-dimensional structures of the main protease (M^{pro} or $3CL^{pro}$ – 3-chymotrypsin-like protease) of SARS–CoV–2 with high resolution. The molecular mechanisms of the SARS–CoV–2 replication life cycle are largely understood, and M^{pro} is considered an important therapeutic target for new drugs. Different test systems for *in vitro* experiments to determine the activity of compounds to inhibit M^{pro} of this coronavirus have been developed. Thus, there are all the necessary prerequisites for the development of new anti-SARS–CoV–2 drugs using computer modeling based on the known structure of the main protease of SARS–CoV–2.

Developing direct anti-CoV drugs was conducted long before the COVID–19 outbreak. Different CoVs cause a variety of respiratory diseases from the common cold [36] to the Severe Acute Respiratory Syndrome (SARS)–CoV [10, 29]. The coronavirus similar to SARS–CoV, the Mid-

[1]Research Computer Center of Lomonosov Moscow State University, Moscow, Russia
[2]Moscow Center of Fundamental and Applied Mathematics, Moscow, Russia
[3]Department of Organic Chemistry, Faculty of Chemistry, Voronezh State University, Voronezh, Russia
[4]Dimonta Ltd., Moscow, Russia

dle East respiratory syndrome coronavirus (MERS–CoV), was identified in 2012 [67]. Soon after discovery SARS–CoV the spatial atomistic models of its main protease have been determined by homology [3] and from experimental crystal structure [65] and the key role of M^{pro} [2] in the SARS–CoV and MERS–CoV replication has been revealed. During 2003–2015 many reversible and irreversible inhibitors of the main protease of SARS–CoV have been found [41], but most of them were weak inhibitors, and none of them reached clinical trials. However, these attempts prepared the basis for anti-SARS–CoV–2 computer aided structural based drug design. The main protease and several other proteins of SARS–CoV–2 have been rapidly identified as the therapeutic target for anti-COVID–19 drugs. Many high quality 3D structures of SARS–CoV–2 M^{pro} in the apo form and with different inhibitors have been already deposited in Protein Data Bank [5]. SARS–CoV–2 M^{pro} has a cysteine-histidine catalytic dyad: $Cys145$ and $His41$.

Docking is the most widely used molecular modeling method in the search of SARS–CoV and SARS–CoV–2 main protease inhibitors. Results of a broad search for SARS–CoV–2 non-covalent M^{pro} inhibitors are presented in [59] where the Deep Docking (DD) platform [16] is used. The effective combination of quantitative structure-activity relationship (QSAR) methods with docking by the Glide program [13, 57] allows to explore the ZINC15 library containing 1.36 billion compounds. The M^{pro} model is prepared on the base of the PDB 6LU7 structure containing the M^{pro} co-crystallized with a covalent inhibitor. DD uses a deep neural network (DNN) which is trained with docking scores calculated for a set of randomly selected ligands. The trained DNN predicts scores for other ligands and these estimates allow to avoid wasting time for docking of ligands with bad scores. The procedure is organized in four iterative steps with improving the training set at each iteration and docking of three million of best ligands at the final iteration. As a result 1000 ligand have been revealed with the best scores for experimental validation of their inhibitory activity but the experimental confirmation has not been published yet.

In the frame of the drug repurposing strategy the search of M^{pro} inhibitors is performed among existing and approved drugs. Such approach is possessed of two advantages. First, such compounds have acceptable solubility in water and this facilitates experimental measurements of their inhibitory activity *in vitro*. Second, it is easier and faster to perform all necessary toxicity studies, preclinical and clinical trials for approving the new drug on the base of such inhibitors of a new target. In the study [32], clinically approved drugs from the DrugBank database were virtually screened against M^{pro} by using the Libdock docking program (Discovery Studio 3.5, Accelrys Software Inc). The M^{pro} model is constructed by homology using the SARS–CoV M^{pro} structure (PDB ID: 1UJ1) and ten best candidates for the experimental verification are identified. Among them are *colistin* (a polymyxin antibiotic), *valrubicin* (an anthracycline antibiotic), *icatibant* (an antagonist of bradykinin receptors) and *bepotastine* (belongs to the family of antihistamines). Those are still waiting an experimental validation. The MOE docking software is used in for the search of SARS–CoV–2 M^{pro} inhibitors among 16 antiviral drugs targeting viral proteases and in in-house database, overall there are about 8000 molecules [26]. As a result two *flavone* and *coumarin* derivatives from the in-house database and three approved protease inhibitors (*Remdesivir*, *Daraunvir*, and *Saquinavir*) have been revealed for subsequent experimental validation. Motonory Tsuji performs virtual screening [61] of 1485144 known bioactive compounds by two docking programs, rDock [44] and AutoDock Vina [60]. The screened database is extracted from the ChEMBL26 [9] database and it includes 13308 approved drugs. The M^{pro} model is based on the structure with PDB ID 6Y2G and optimization of the protein is

performed with the AMBER99 force field and subsequent molecular dynamics low-temperature annealing. First, rDock reveals 64 ligands candidates to become M^{pro} inhibitors. Second, these best ligands are docked with AutoDock Vina and 29 hits are determined. Among these hits is only one approved drug (*eszopiclone*). The author considers that AutoDock Vina is more accurate than rDock but we notice the lack of Coulomb interactions in AutoDock Vina. No results on experimental confirmation of these findings are published. Authors of [17] go further in the screening strategy and use three docking programs: Glide, FRED [33] and AutoDock Vina. Only the equivalent high affinity binding modes predicted simultaneously by the three docking programs were considered to correspond to bioactive poses. A complex of M^{pro} with the ligand Glide pose is minimized by applying the MM-GBSA minimization in the Prime module (Schrödinger, LLC). Two libraries of approved drugs were screened: eDrug3D [40] and Reaxys-marketed [12]. On the base of docking results and the visual inspection of best ligand poses in the active site of M^{pro} seven candidates to become inhibitors of M^{pro} are selected: *perampanel*, *carprofen*, *celecoxib*, *alprazolam*, *trovafloxacin*, *sarafloxacin* and *ethyl biscoumacetate*. Two of them, *carprofen* and *celecoxib*, demonstrate 4% and 12% inhibition of M^{pro} at 50 μM respectively – certainly, it is too weak activity. AutoDock Vina is used also in [6] where four SARS–CoV–2 proteins, including M^{pro}, are targeted by 16 antiviral compounds. The atomistic models of targets are constructed by homology modelling with the I-Tasser server [66]. Docking into M^{pro} is performed with nine moveable residues in the active site. *Simeprevir*, a $HCVNS3/4A$ protease inhibitor, is found the best for M^{pro} inhibition. AutoDock Vina is used for docking of 62 alkaloids and 100 terpenoids from African plants into the M^{pro} model constructed using the complex with PDB ID 6LU7 [19]. The two best alkaloids are *10-hydroxyusambarensine* and *cryptoquindoline*, and two best terpenoids are *6-oxoisoiguesterin* and *22-hydroxyhopan-3-one*. Authors of [15] in search of M^{pro} inhibitors screen marine natural products (14064 compounds) by a combination of the pharmacophore filter, molecular docking and molecular dynamics (MD) simulations. The model of M^{pro} is constructed by the Pharmit server using the complex with PDB ID 6LU7. After the pharmacophore filter 180 compounds are docked by AutoDock Vina. 17 best docked compounds are subjected to MD simulation and the most promising candidates to become M^{pro} inhibitors are phlorotannins which are oligomers of *phloroglucinol* (1,3,5-trihydroxybenzene). Among these phlorotannins is *Dieckol* which has been already proven to be the M^{pro} inhibitor with IC50 = 2.7 μM. SARS–CoV–2 M^{pro} inhibitors are found in [11] but without an experimental validation. Thirty four approved and on-trial inhibitors of different proteases are docked by AutoDock 4.2 into the SARS–CoV–2 M^{pro} models prepared from the PDB complex 6LU7 and from the PDB complex 6M2N. Several drugs are identified as candidates to become SARS–CoV–2 M^{pro} inhibitors, within the classes of the HCV protease, DPP-4, α-thrombin and coagulation Factor Xa inhibitors. Several conclusions can be made from this short review. First, the bottleneck of new inhibitors discovery is the experimental *in vitro* validation of predicted inhibitors of SARS–CoV–2. Actually such validation will possible be made soon and new SARS–CoV–2 M^{pro} inhibitors will be published. Second, after selection of best compounds using results of fast docking more accurate and slow docking programs are used for the identification of best ligand-candidates for experimental validation. Third, for the acceleration of screening of large databases some a priori considerations are widely used. They are different pharmacophore filtering, narrowing sets of molecules for screening by other methods including "scientific intuition" as well as discarding non-active compounds using predictions on the base of neural networks as in Deep Docking platform (see above).

All existing inhibitors targeting SARS–CoV–2 M^{pro} can be classified into two unequal groups: larger set of covalent inhibitors and smaller group of non-covalent inhibitors. The chemical description of compounds in the first group is mainly based on the nature of the reactive group (a warhead) forming a covalent bond with $Cys145$. Development of covalent inhibitors has its own specific features but we focus here only on non-covalent inhibitors. At the time of this writing only several weak non-covalent inhibitors of SARS–CoV–2 M^{pro} are published (Fig. 1): *tideglusib* [23], *emedastine* [14], *X77* [35], *carprofen* and *celecoxib* [17], *quercetin* [1], and *baicalein* [42].

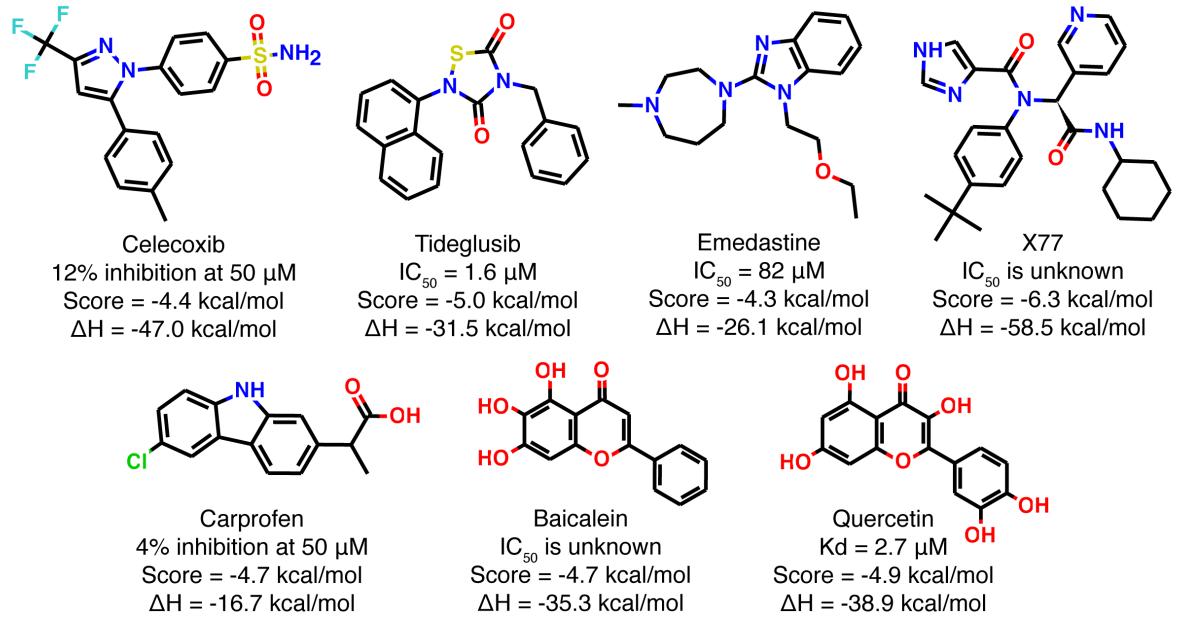

Figure 1. Non-covalent inhibitors of SARS–CoV–2 M^{pro}

Tideglusib is naphthalene-containing compound with thiadiazolidine-3,5-dione as a scaffold. *Emedastine* is an antihistamine drug consisting of a diazepane ring and benzimidazole. Compound *X77* – is an only non-covalent drug-like inhibitor crystallized with M^{pro}. It possesses X-like shape with imidazole, iPr-benzene, pyridine and cyclohexene as structural moieties. *Carprofen* is a non-steroidal anti-inflammatory drug now used in veterinary which is simply di-substituted carbazole. *Celecoxib* is also a non-steroidal anti-inflammatory agent based on pyrazole scaffold substituted with methylbenzene and benzenesulfonamide. *Quercetin* and *baicalein* both belong to a group of flavonoids and contain a flavone fragment substituted with a few hydroxyl groups.

We present here the results of virtual screening of a database containing more than 40 000 structures of low molecular weight ligands using our own SOL [53, 56] docking program. For the ligands with best docking scores the binding enthalpy is calculated using the PM7 quantum-chemical semiempirical method and the implicit COSMO solvent model. The choice of thresholds separating active compounds from inactive ones is made with respect to the docking score and the binding enthalpy obtained by the same methods for the native ligand crystallized with M^{pro} in the respective Protein Data Bank. 20 compounds have been selected for the further experimental *in vitro* validation on the base of best docking scores as well as best binding enthalpy values.

1. Materials and Methods

Virtual screening requires construction of the atomistic 3D model of the target protein, preparation of 3D structures of all ligands from the database to be screened. The protein model as well as ligand models define accuracy of docking calculations and they should be prepared carefully. The target protein models are constructed on the base of protein-ligand complexes which 3D structures are stored in Protein Data Bank (PDB). These structures contain Cartesian coordinates of all heavy, i.e. non-hydrogen, atoms. Several thousand hydrogen atoms should be added to the protein taking into account the pH condition of solvent where the protein is working. Protonation states of amino acid residues are well known for any pH (usually it is neutral condition pH = 7.4) if the influence of neighboring atoms in the protein globule is not taken into account. Even for equal protonation states different programs add hydrogen atoms in different ways resulting in unequal hydrogen atoms spatial positions. The latter leads to varied results of docking performance, different best ligand poses and diverse scores predicting the protein-ligand binding affinity [30]. Ligands in most of databases are stored as 2D structures and some programs should be used to construct low energy conformers, tautomers and protonation states for each given ligand. Below we explain how all these preliminary works are done in the present case of virtual screening as well as we describe shortly the docking program and quantum-chemical method used.

1.1. Protein Spatial Model

Many structures of SARS–CoV–2 main protease have been already deposited in the Protein Data Bank. The structures with best quality have following PDB ID: 5R7Z, 5R83 and 6W63. These are structures of M^{pro} crystallized with non-covalent inhibitors, they do not contain missing residues or atoms and have a good resolution < 2.2 $\AA$. Protein models are prepared from these PDB structures by removing all atoms, ions and molecules which do not belong to the protein, and then hydrogen atoms are added by our APLITE program [30]. Hydrogen atoms are added to the native ligands extracted from the PDB structures by Avogadro [21]. After local optimization of the energy of these complexes in the frame of the MMFF94 force field [20] with the variation of Cartesian coordinates of all native ligand atoms the position of the ligand does not change significantly: RMSD values calculated over all ligand atoms between native crystallized ligand pose and the optimized ligand pose are less than 1.6 $\AA$ in all these three models. This is the simplest check of applicability of the MMFF94 force field used in SOL to modeling of protein-ligand interactions in these complexes. Next, we check the ability of SOL to reproduce the native ligand pose crystallized with the protein. For the Mpro model prepared from the 5R7Z complex the native docking fails: RMSD between best docked ligand pose and the crystallized one is more than 7 $\AA$. For models constructed on the base of 5R83 and 6W63 complexes the native docking is successful: RMSD = 1.19 $\AA$ and RMSD = 1.31 $\AA$. The visual inspection of the active sites of these protein models reveals that one residue (MET49) of the active site is mobile and adapts to the bound inhibitor. Finally, the model based on the 6W63 PDB structure (the X77 native ligand has 7 torsions) has been selected as the target protein model for virtual screening because the protein active site is more open than one of the model used 5R83 PDB complex (the native ligand has 4 torsions). The SOL score of the native ligand docked into the 6W63 model is equal to -6.3 kcal/mol. This value gives a threshold

for the selection of best ligands after the docking step of virtual screening. The Lomonosov supercomputer [62] of Lomonosov Moscow State University is employed.

1.2. Database and Ligand 3D Structure Preparation

The database of compounds of the Department of Organic Chemistry of Voronezh State University [58] is used for the present virtual screening. A wide range of nitrogen-, oxygen- and sulfur-containing heterocyclic compounds are presented in the database. The compounds are small drug-like molecules. Among them there are hydroquinoline derivatives with antibacterial, antifungal, anticoagulant activity [22, 25, 34, 37, 38], aminopyrimidines and pyrrolo[3,2,1-ij]quinolin-2-ones, which are factor Xa and protein kinases inhibitors [50, 54], various plant growth stimulants of the getarylcarboxylic acid class [63, 64].

The database contains approximately 19 000 molecules, and after 2D→3D transformation we obtain 41 000 molecular 3D-structures. The main source of using different 3D-structures for one molecule is different low energy conformations of non-aromatic rings including macrocycles. The SOL docking program treats the ligand flexibility, as many other docking programs do, only by variations of all torsions, i.e. degrees of freedom describing internal rotations of ligand molecular groups around ordinary covalent bonds, keeping fixed covalent bond lengths and valence angles. Different conformations of non-aromatic rings and macro-cycles of one and the same molecule should be docked as different molecules. The ligands from the database are protonated with ChemAxon Protonation module [7] at pH = 7.4. OpenBabel was used for generating 3D coordinates.

1.3. SOL Docking Program

SOL [53, 56] is a classic docking program with many features used in popular docking programs [57]. However, developing this program we tried to make as few model simplifications as possible and to take into account the most important effects determining the accuracy of docking. Performance of SOL is based on the docking paradigm connecting docking with the global optimization problem: the best ligand pose should be near the global energy minimum of the protein-ligand complex. The energy is calculated in the frame of the MMFF94 force field [20] almost without any simplifications and the genetic algorithm (GA) is used for the global optimization. The preliminary calculated grid of potentials describing interactions of a probe ligand atom with the rigid protein is used. Coulomb and van der Waals interactions as well as desolvation energy potentials are calculated in nodes of the grid covering the docking cube (its edge is equal to 22 $\AA$) with the distance between neighboring nodes of 0.22 $\AA$. The desolvation potentials are calculated in the frame of a simplified Generalized Born implicit solvent model. In the process of the global optimization the ligand internal strain energy is calculated also using MMFF94. In the frame of the genetic algorithm niching is used. Niching provides diversity of probe ligand poses preventing premature concentration of best ligand poses near a local energy minimum. The default values of main parameters of GA are sufficiently large: the population size 30 000 and the number of generation 1000. 50 independent runs (by default) of GA are performed, the clustering of 50 solutions is made and the relatively large occupation number of the cluster containing ligand poses with lowest energies is the indication of the successful finding of the ligand pose with the lowest energy. SOL was used for CSAR 2011–2012 docking competition with other docking programs Gold, AutoDock, AutoDock Vina, ICMVLS, Glide

and others and for two of three target-proteins SOL results are among the best ones [8, 53]. SOL is successfully used for development of new inhibitors of several targets with the experimental *in vitro* confirmation: thrombin [45], urokinase (uPA) [4, 55] and the blood coagulation factors Xa and XIa [22, 38, 54]. SOL is adapted to screen large databases of ligands on Lomonosov supercomputer [62] of M.V. Lomonosov Moscow State. Although SOL is a parallel program it is more effective to perform virtual screening of large ligand databases distributing jobs of ligand docking over hundreds and thousands of computing cores, one ligand per one core. Some auxiliary scripts and programs are utilized to submit and queue up docking jobs and to analyze results. Docking of one ligand per one core needs from 1 to several hours depending of the size and flexibility of the ligand. Usually, the docking SOL score values for inhibitors of different target proteins are in the range from -5 kcal/mol to -7 kcal/mol.

1.4. Protein-ligand Binding Enthalpy

Quantum-chemical semiempirical methods are developed since 70–90th of 20th century but their main feedback was until recently the description of non-polar intermolecular interactions including H-bonds formations. But recently the new PM6 [43, 46] and PM7 [47] methods have been developed in the frame of the NDDO (Neglect of Diatomic Differential Overlap) approximation utilizing ideas which have been used in DFT methods [18, 24] for the proper description of dispersion interactions and the hydrogen and halogen bonds formation. Moreover, the PM7 method includes these novelties at the parameterization stage which is based on an extremely large set of molecular data. PM7 as well as PM6 with respective corrections are realized in the MOPAC package [48] and the validation on a broad set of test molecular systems demonstrates that these methods work no worse than DFT ones. In addition, MOPAC includes the MOZYME module [49], where the localized molecular orbitals method is used instead of the LCAO (Linear Combination of Atomic Orbitals) approximation. This allows calculations of whole protein-ligand complexes and to do this quickly. The solute-solvent interaction is responsible for the desolvation energy which is a large term in the protein-ligand binding energy. Due to a large electrostatic permittivity of water (78.5) desolvation term screens strongly Coulomb interactions between protein and ligand atoms. The use of solvent model is extremely important for docking and the binding-energy estimations [51, 52]. So, the binding enthalpy calculations we use here PM7 with the COSMO [27, 28] implicit solvent model implemented also in MOPAC.

After virtual screening by docking and determination of ligands with best docking scores for these best ligands the binding enthalpy ΔH_{bind} is calculated as follows:

$$\Delta H_{bind} = H(PL) - H(P) - H(L), \tag{1}$$

where $H(X)$ is the enthalpy of formation of the molecular system X (where X=PL, P or L) calculated by MOPAC: PL, P and L are the protein-ligand complex, the unbound protein and the unbound ligand, respectively. The protein-ligand binding free energy consists of enthalpy and entropy terms. The latter is positive in most cases due to the loss of degrees of freedom when the ligand is bound to the protein. So, we consider that the more negative is the binding enthalpy the more negative is the binding free energy and respective protein-ligand affinity.

The $H(PL)$ is calculated as follows. The best docked ligand pose in the protein-ligand complex is used as the initial ligand position at the local optimization of the energy of the protein-ligand complex. The optimization is made by the gradient L-BFGS (limited-memory Broyden-Fletcher-Goldfarb-Shanno) method implemented in MOPAC, positions of all ligand

atoms are varied while all protein atoms are kept fixed. During this optimization the energy of the complex is calculated by PM7 without solvent. The final enthalpy of formation of the complex in the local minimum is recalculated by PM7 with the COSMO solvent (PM7+COSMO) without optimization using 1SCF keyword of MOPAC. The initial conformation of the unbound ligand is generated by the Open Babel program [39] and the local energy optimization is performed by PM7 with variations of positions of all ligand atoms, and the enthalpy of formation $\Delta H(L)$ is recalculated by PM7+COSMO for the optimized configuration. The enthalpy of formation of the unbound protein is calculated by PM7+COSMO without energy optimization.

Overall about 178000 CPU*hours have been spent for this virtual screening.

2. Results

Screening of the prepared database results in 1045 primary docking hits with SOL score values more negative than −6.3 kcal/mol which is the value of the SOL score for the native ligand of the complex 6W63 (see above). All these virtual hits are subjected to quantum-chemical postprocessing to perform local optimization of best docked poses and to estimate their binding enthalpies with a more rigorous computational method. According to calculated enthalpies, the list of potential hits is narrowed down – only those compounds are kept which have binding enthalpies better (more negative) or slightly worse than binding enthalpy calculated for the X77 native ligand (−58.5 kcal/mol). Selection of molecules with worse enthalpies is justified by values obtained for other known non-covalent M^{pro} inhibitors which turn to be dramatically less negative than enthalpy of X77 (see Fig. 1). For the final set of 87 candidates to become M^{pro} inhibitors, predicted bound conformations are visually checked for the presence of specific contacts with SARS–CoV–2 M^{pro} as well as the ability to block its catalytic dyad. Moreover, similarity between these compounds is also visually assessed to form a list of best chemically diverse candidates. The inspection of similarity reveals the high number of hexahydro-1H-furo[3,4-c]pyrrole-4,6-diones: 37 compounds of 87 virtual hits possess this fragment. The second most common chemical class among top compounds is derivatives of pyrazolo[3,4-b]quinolin-5-one: seven compounds contain this scaffold. Alike compounds with similar binding modes are removed keeping the best one in terms of the presence of specific contacts with M^{pro} and calculated values of the SOL score and the binding enthalpy.

Relying upon results of modeling, observations of docking poses and visual estimation of similarity, we select 20 best candidate molecules as potential inhibitors of M^{pro}. All compounds in its predicted conformations with SARS–CoV–2 M^{pro} block the catalytic dyad of the enzyme. Their structures are listed in Fig. 2 and Fig. 3. Some molecules can be grouped into clusters according to the nature of their scaffold. Four compounds (VGY-0002900, VGY-0006802, VGY-0005103, and VGY-0015946) constitute class of 4,4-dimethyl-dithioloquinoline derivatives. Two compounds (VGY-0002425, VGY-0013355) are related to derivatives of hexahydro-1H-furo[3,4-c]pyrrole-4,6-dione. The similar scaffold, pyrrolidine-2,5-dione, can be found in other two candidates: VGY-0016826, VGY-0214225. It is noteworthy that one of known non-covalent M^{pro} inhibitors, *tideglusib* [23] (see Fig. 1), contains the similar aliphatic ring with a nitrogen atom placed between two carbonyl groups. Two candidates (VGY-0015689, VGY-0015693) contain 1,3,5-triazine-2,4,6-triamine as a central fragment. Two other compounds (VGY-019038 and VGY-0031572) belong to derivatives of pyrazolo[3,4-b]quinolin-5-one. Other selected molecules are unique and constitute singletons.

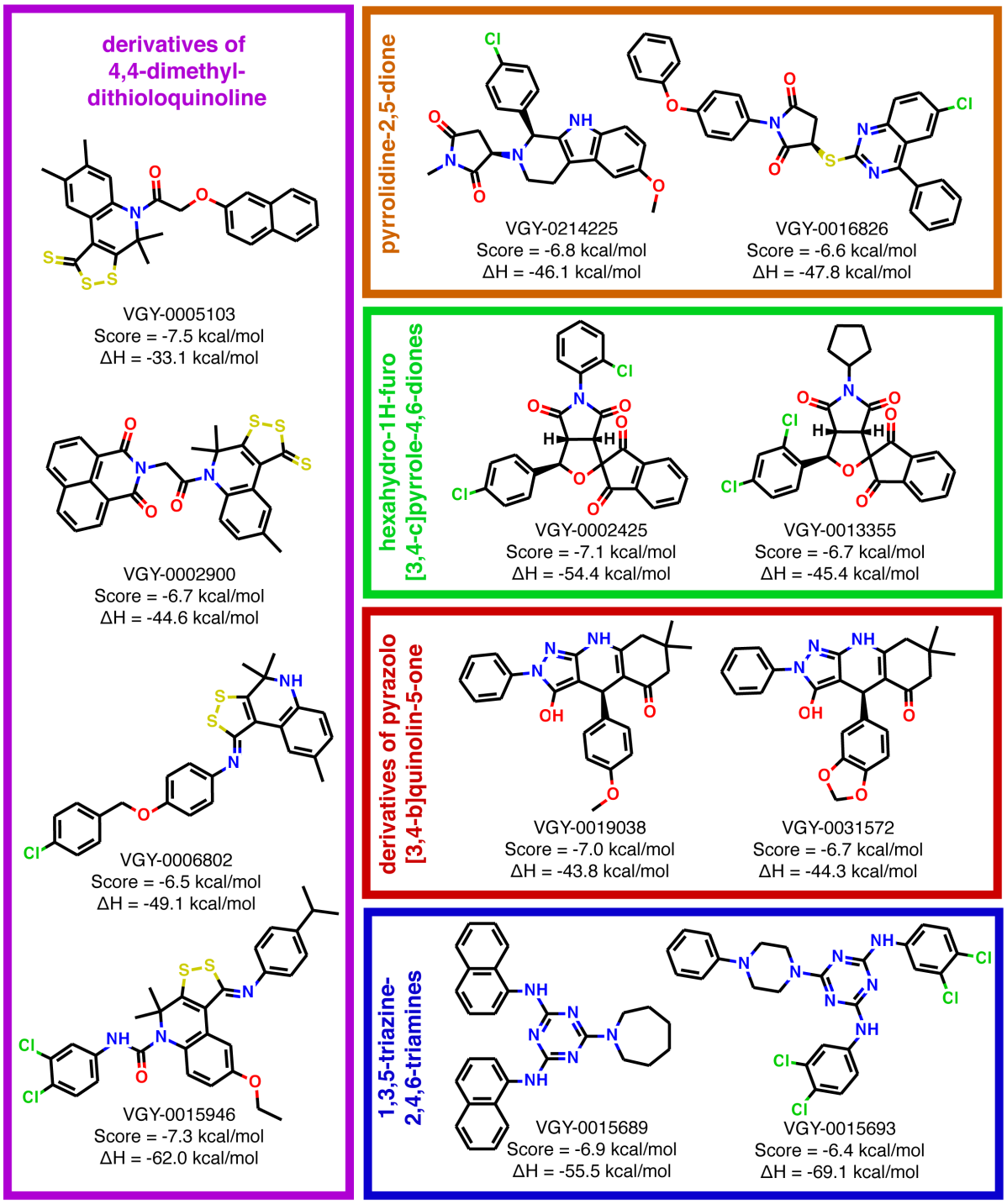

Figure 2. Structures of best potential M^{pro} inhibitors grouped according to the structure of the scaffold

VGY-0224133 represents the best scored compound and belong to derivatives of imidazo[2,1-b]quinazolin-6-one. In its bound conformation predicted by docking and subsequent local optimization by PM7, it forms one H-bond with *Gly*143 and pi-stacking interactions with three residues: *His*41, *Tyr*54, and *His*163. Its docking pose is shown in Fig. 4. The compound with the best predicted binding enthalpy, VGY-0015693, contains triazine-2,4,6-riamine scaffold linked to two benzene rings disubstituted with chlorine atoms which makes the molecule symmet-

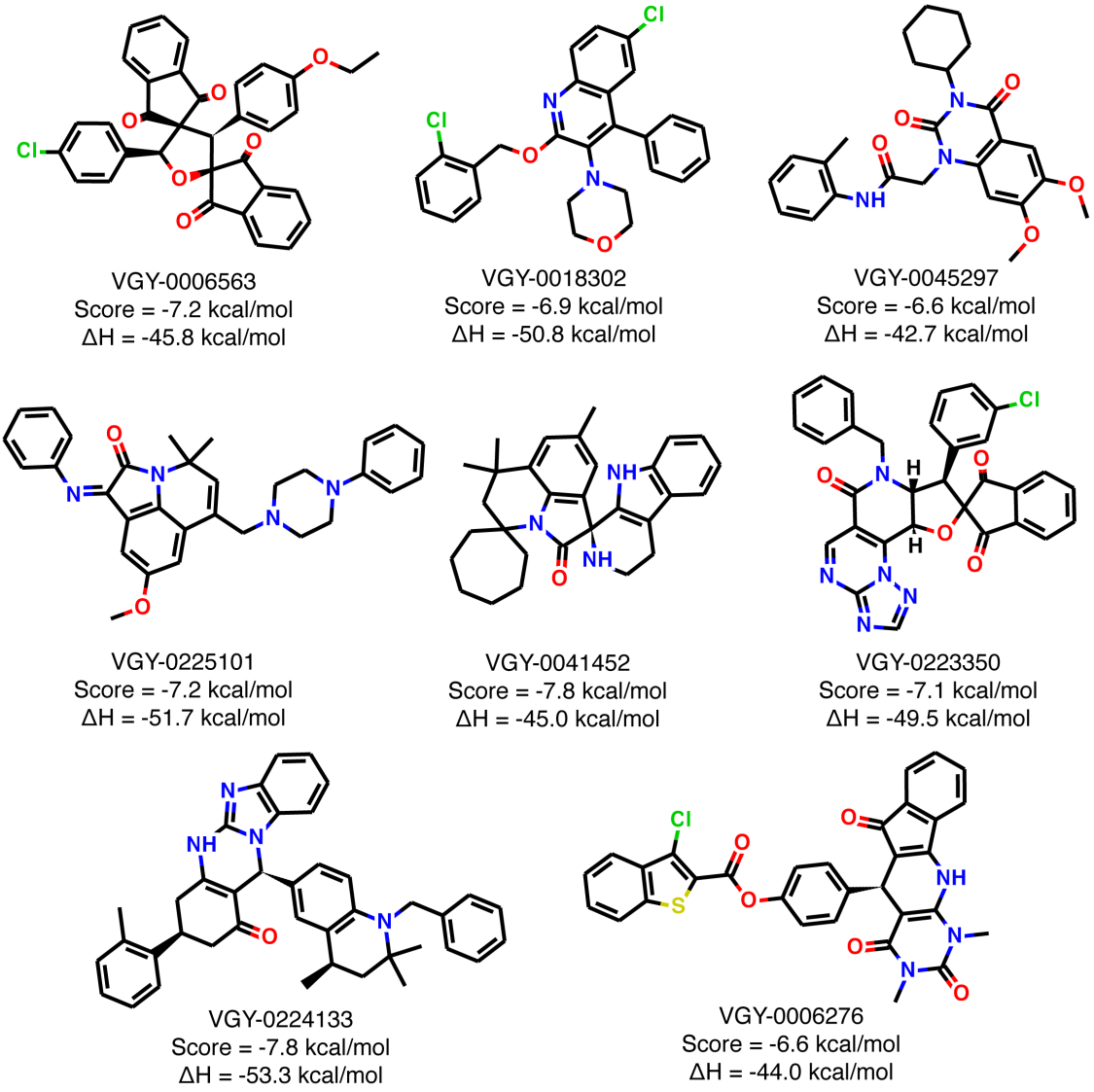

Figure 3. Structures of best potential M^{pro} inhibitors with no grouping to any cluster

rical. According to geometry complex predicted by docking, VGY-0015693 forms three specific contacts: anion-pi interaction with *Glu*166, a halogen bond with *Phe*140-O, two pi-stacking with *His*41 and *His*163 (see Fig. 5).

Another peculiarity related to selected candidates is that above mentioned derivatives of 4,4-dimethyl-dithioloquinoline: VGY-0002900, VGY-0006802, VGY-0005103, VGY-0015946, share a disulfide moiety which similar to one found in disulfiram – a known covalent inhibitor of SARS–CoV–2 M^{pro} with high thiol-reactiveness against cysteine residues [23, 31]. Because of this fact these compounds can be possible covalent modulators of M^{pro} activity. Two other compounds (VGY-0015693, VGY-0225101) contain the piperazine fragment, the common moiety among potential virtual inhibitors of SARS–CoV–2 M^{pro} we found by drug repurposing strategy with using the same computational protocol as we applied here. Activity of selected candidates is supposed to be checked *in vitro* against SARS–CoV–2 M^{pro}.

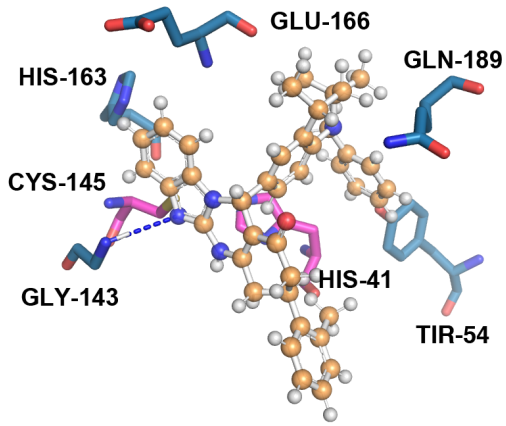

Figure 4. Docking pose of VGY-0224133 after local optimization by PM7

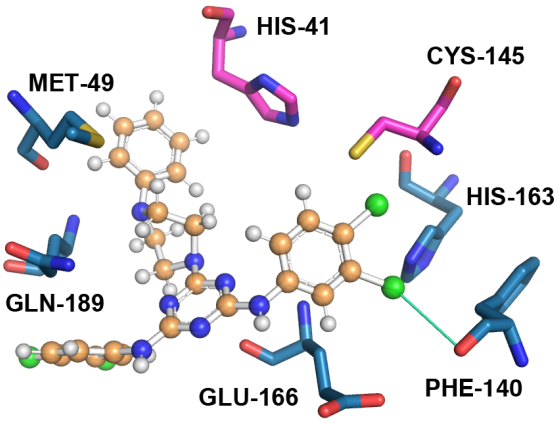

Figure 5. Docking pose of VGY-0015693 after local optimization by PM7

Conclusion

In search of non-covalent inhibitors of SARS–CoV–2 main protease virtual screening of the database of drug-like molecules is performed. There are two criteria of the selection of most promising molecules candidates to become inhibitors. First criterion is the sufficiently negative value of the SOL docking score and the second one is the sufficiently negative value of the binding enthalpy calculated by the PM7 quantum-chemical semiempirical method with the COSMO implicit solvent model. Thresholds of the criteria are defined by the respective values of the native inhibitor crystallized with the SARS–CoV–2 M^{pro} in the PDB complex 6W63. 20 compounds are selected for further experimental validation which molecules are satisfied the two criteria both.

Acknowledgments

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University, including the Lomonosov supercomputer [62].

References

1. Abian, O., Ortega-Alarcon, D., Jimenez-Alesanco, A., et al.: Structural stability of SARS–CoV–2 $3CL^{pro}$ and identification of quercetin as an inhibitor by experimental screening. International journal of biological macromolecules 164, 1693–1703 (2020), DOI: 10.1016/j.ijbiomac.2020.07.235

2. Anand, K., Yang, H., Bartlam, M., et al.: Coronavirus main proteinase: target for antiviral drug therapy BT – Coronaviruses with Special Emphasis on First Insights Concerning SARS, pp. 173–199. Birkhäuser Basel (2005), DOI: 10.1007/3-7643-7339-3_9

3. Anand, K., Ziebuhr, J., Wadhwani, P., et al.: Coronavirus main proteinase ($3CL^{pro}$) structure: basis for design of anti–SARS drugs. Science (New York, N.Y.) 300(5626), 1763–1767 (2003), DOI: 10.1126/science.1085658

4. Beloglazova, I.B., Plekhanova, O.S., Katkova, E.V., et al.: Molecular modeling as a new approach to the development of urokinase inhibitors. Bulletin of Experimental Biology and Medicine 158(5), 700–704 (2015), DOI: 10.1007/s10517-015-2839-3

5. Berman, H.M., Westbrook, J., Feng, Z., et al.: The Protein Data Bank. Nucleic Acids Research 28(1), 235–242 (2000), DOI: 10.1093/nar/28.1.235

6. Calligari, P., Bobone, S., Ricci, G., Bocedi, A.: Molecular Investigation of SARS–CoV–2 Proteins and Their Interactions with Antiviral Drugs. Viruses 12(4), 445 (2020), DOI: 10.3390/v12040445

7. ChemAxon software. `http://www.chemaxon.com`, accessed: 2020-10-01

8. Damm-Ganamet, K.L., Smith, R.D., Dunbar Jr., J.B., et al.: CSAR Benchmark Exercise 2011–2012: Evaluation of Results from Docking and Relative Ranking of Blinded Congeneric Series. J. Chem. Inf. and Model. 53, 1853–1870 (2013), DOI: 10.1021/ci400025f

9. Davies, M., Nowotka, M., Papadatos, G., et al.: ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic acids research 43(W1), W612–W620 (2015), DOI: 10.1093/nar/gkv352

10. Drosten, C., Günther, S., Preiser, W., et al.: Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. New England Journal of Medicine 348(20), 1967–1976 (2003), DOI: 10.1056/NEJMoa030747

11. Eleftheriou, P., Amanatidou, D., Petrou, A., Geronikaki, A.: In Silico Evaluation of the Effectivity of Approved Protease Inhibitors against the Main Protease of the Novel SARS–CoV–2 Virus. Molecules 25(11) (2020), DOI: 10.3390/molecules25112529

12. Elsevier. Reaxys Database. 2020. `https://www.reaxys.com`, accessed: 2020-10-01

13. Friesner, R.A., Banks, J.L., Murphy, R.B., et al.: Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of Medicinal Chemistry 47(7), 1739–1749 (2004), DOI: 10.1021/jm0306430

14. Gao, J., Zhang, L., Liu, X., et al.: Repurposing Low-Molecular-Weight Drugs against the Main Protease of Severe Acute Respiratory Syndrome Coronavirus 2. The Journal of Physical Chemistry Letters 11(17), 7267–7272 (2020), DOI: 10.1021/acs.jpclett.0c01894

15. Gentile, D., Patamia, V., Scala, A., et al.: Putative Inhibitors of SARS–CoV–2 Main Protease from A Library of Marine Natural Products: A Virtual Screening and Molecular Modeling Study. Marine drugs 18(4) (2020), DOI: 10.3390/md18040225

16. Gentile, F., Agrawal, V., Hsing, M., et al.: Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. ACS Central Science 6(6), 939–949 (2020), DOI: 10.1021/acscentsci.0c00229

17. Gimeno, A., Mestres-Truyol, J., Ojeda-Montes, M.J., et al.: Prediction of Novel Inhibitors of the Main Protease (M-pro) of SARS–CoV–2 through Consensus Docking and Drug Reposition. International Journal of Molecular Sciences 21(11), 3793–3821 (2020), DOI: 10.3390/ijms21113793

18. Grimme, S.: Accurate description of van der Waals complexes by density functional theory including empirical corrections. Journal of Computational Chemistry 25(12), 1463–1473 (2004), DOI: 10.1002/jcc.20078

19. Gyebi, G.A., Ogunro, O.B., Adegunloye, A.P., et al.: Potential inhibitors of coronavirus 3-chymotrypsin-like protease ($3CL^{pro}$): an *in silico* screening of alkaloids and terpenoids from African medicinal plants. Journal of Biomolecular Structure and Dynamics pp. 1–13 (2020), DOI: 10.1080/07391102.2020.1764868

20. Halgren, T.A.: Merck molecular force field. Journal of Computational Chemistry 17(5-6), 490–641 (1996), DOI: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P

21. Hanwell, M.D., Curtis, D.E., Lonie, D.C., et al.: Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of Cheminformatics 4(1), 17 (2012), DOI: 10.1186/1758-2946-4-17

22. Ilin, I.S., Lipets, E.N., Sulimov, A.V., et al.: New factor Xa inhibitors based on 1,2,3,4-tetrahydroquinoline developed by molecular modelling. Journal of Molecular Graphics and Modelling 89, 215–224 (2019), DOI: 10.1016/j.jmgm.2019.03.017

23. Jin, Z., Du, X., Xu, Y., et al.: Structure of M^{pro} from SARS–CoV–2 and discovery of its inhibitors. Nature 582(7811), 289–293 (2020), DOI: 10.1038/s41586-020-2223-y

24. Jurecka, P., Cerny, J., Hobza, P., Salahub, D.R.: Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with *ab initio* quantum mechanics calculations. Journal of Computational Chemistry 28(2), 555–569 (2007), DOI: 10.1002/jcc.20570

25. Kartsev, V., Shikhaliev, K.S., Geronikaki, A., et al.: Appendix A. dithioloquinolinethiones as new potential multitargeted antibacterial and antifungal agents: Synthesis, biological evaluation and molecular docking studies. European Journal of Medicinal Chemistry 175, 201–214 (2019), DOI: 10.1016/j.ejmech.2019.04.046

26. Khan, S.A., Zia, K., Ashraf, S., et al.: Identification of chymotrypsin-like protease inhibitors of SARS–CoV–2 via integrated computational approach. Journal of Biomolecular Structure and Dynamics pp. 1–10 (2020), DOI: 10.1080/07391102.2020.1751298

27. Klamt, A., Schuurmann, G.: COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. Journal of the Chemical Society, Perkin Transactions 2 (5), 799–805 (1993), DOI: 10.1039/P29930000799

28. Klamt, A.: Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. The Journal of Physical Chemistry 99(7), 2224–2235 (1995), DOI: 10.1021/j100007a062

29. Ksiazek, T.G., Erdman, D., Goldsmith, C.S., et al.: A novel coronavirus associated with severe acute respiratory syndrome. The New England journal of medicine 348(20), 1953–1966 (2003), DOI: 10.1056/NEJMoa030781

30. Kutov, D.C., Katkova, E.V., Kondakova, O.A., et al.: Influence of the method of hydrogen atoms incorporation into the target protein on the protein-ligand binding energy. Bulletin of the South Ural State University, Ser. Mathematical Modelling, Programming & Computer Software 10(3), 94–107 (2017), DOI: 10.14529/mmp170308

31. Lin, M.H., Moses, D.C., Hsieh, C.H., et al.: Disulfiram can inhibit MERS and SARS coronavirus papain-like proteases via different modes 150, 155–163 (2018), DOI: 10.1016/j.antiviral.2017.12.015

32. Liu, X., Wang, X.J.: Potential inhibitors against 2019–nCoV coronavirus M protease from clinically approved medicines. Journal of Genetics and Genomics 47(2), 119–121 (2020), DOI: 10.1016/j.jgg.2020.02.001

33. McGann, M.: FRED and HYBRID docking performance on standardized datasets. J Comput Aided Mol Des 26(8), 897–906 (2012), DOI: 10.1007/s10822-012-9584-8

34. Medvedeva, S.M., Potapov, A.Y., Gribkova, I.V., et al.: Synthesis, docking, and anticoagulant activity of new factor–Xa inhibitors in a series of pyrrolo[3,2,1-ij]quinoline-1,2-diones. Pharmaceutical Chemistry Journal 51(11), 975–979 (2018), DOI: 10.1007/s11094-018-1726-4

35. Mesecar, A.D.: Structure of COVID–19 main protease bound to potent broad-spectrum non-covalent inhibitor X77, DOI: 10.2210/pdb6w63/pdb, accessed: 2020-10-01

36. Myint, S.H.: Human Coronavirus Infections BT – The Coronaviridae, pp. 389–401. Springer US (1995), DOI: 10.1007/978-1-4899-1531-3_18

37. Novichikhina, N.P., Shestakov, A.S., Potapov, A.Y., et al.: Synthesis of 4H-pyrrolo[3,2,1-ij]quinoline-1,2-diones containing a piperazine fragment and study of their inhibitory properties against protein kinases. Russ Chem Bull 4, 787–792 (2020), DOI: 10.1007/s11172-020-2834-3

38. Novichikhina, N., Ilin, I., Tashchilova, A., et al.: Synthesis, Docking, and In Vitro Anticoagulant Activity Assay of Hybrid Derivatives of Pyrrolo[3,2,1-ij]Quinolin-2(1H)-one as New Inhibitors of Factor Xa and Factor XIa. Molecules 25(8), 1889 (2020), DOI: 10.3390/molecules25081889

39. O'Boyle, N.M., Banck, M., James, C.A., et al.: Open Babel: An open chemical toolbox. Journal of Cheminformatics 3(1), 33 (2011), DOI: 10.1186/1758-2946-3-33

40. Pihan, E., Colliandre, L., Guichou, J.F., Douguet, D.: e-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. Bioinformatics 28(11), 1540–1541 (2012), DOI: 10.1093/bioinformatics/bts186

41. Pillaiyar, T., Manickam, M., Namasivayam, V., et al.: An Overview of Severe Acute Respiratory Syndrome-Coronavirus (SARS–CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. Journal of Medicinal Chemistry 59(14), 6595–6628 (2016), DOI: 10.1021/acs.jmedchem.5b01461

42. Rathnayake, A.D., Zheng, J., Kim, Y., et al.: 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS–CoV-infected mice. Science Translational Medicine 12(557), eabc5332 (2020), DOI: 10.1126/scitranslmed.abc5332

43. Rezac, J., Fanfrlik, J., Salahub, D., Hobza, P.: Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. J Chem Theory Comput 5(7), 1749–1760 (2009), DOI: 10.1021/ct9000922

44. Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., et al.: rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLOS Computational Biology 10(4), 1–7 (2014), DOI: 10.1371/journal.pcbi.1003571

45. Sinauridze, E.I., Romanov, A.N., Gribkova, I.V., et al.: New Synthetic Thrombin Inhibitors: Molecular Design and Experimental Verification. PLOS ONE 6(5), 1–12 (2011), DOI: 10.1371/journal.pone.0019969

46. Stewart, J.J.: Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. J Mol Model 13(12), 1173–1213 (2007), DOI: 10.1007/s00894-007-0233-4

47. Stewart, J.J.: Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. Journal of Molecular Modeling 19(1), 1–32 (2013), DOI: 10.1007/s00894-012-1667-x

48. Stewart, J.J.P.: Stewart Computational Chemistry. MOPAC2016. `http://openmopac.net/MOPAC2016.html` (2016), accessed: 2020-10-01

49. Stewart, J.J.P.: Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. International Journal of Quantum Chemistry 58(2), 133–146 (1996), DOI: 10.1002/(SICI)1097-461X(1996)58:2<133::AID-QUA2>3.0.CO;2-Z

50. Stolpovskaya, N.V., Kruzhilin, A.A., Zorina, A.V., et al.: Synthesis of Substituted Aminopyrimidines as Novel Promising Tyrosine Kinase Inhibitors. Russian journal of organic chemistry 55(9), 1322–1328 (2019), DOI: 10.1134/S1070428019090094

51. Sulimov, A.V., Kutov, D.C., Katkova, E.V., Sulimov, V.B.: Combined docking with classical force field and quantum chemical semiempirical method PM7. Advances in Bioinformatics 2017 (2017), DOI: 10.1155/2017/7167691

52. Sulimov, A.V., Kutov, D.C., Katkova, E.V., et al.: New generation of docking programs: Supercomputer validation of force fields and quantum-chemical methods for docking. Journal of Molecular Graphics and Modelling 78, 139–147 (2017), DOI: 10.1016/j.jmgm.2017.10.007

53. Sulimov, A.V., Kutov, D.C., Oferkin, I.V., et al.: Application of the docking program SOL for CSAR benchmark. Journal of Chemical Information and Modeling 53(8), 1946–1956 (2013), DOI: 10.1021/ci400094h

54. Sulimov, V.B., Gribkova, I.V., Kochugaeva, M.P., et al.: Application of molecular modeling to development of new factor Xa inhibitors. BioMed Research International 2015 (2015), DOI: 10.1155/2015/120802

55. Sulimov, V.B., Katkova, E.V., Oferkin, I.V., et al.: Application of molecular modeling to urokinase inhibitors development. BioMed Research International 2014, 625176 (2014), DOI: 10.1155/2014/625176

56. Sulimov, V.B., Ilin, I.S., Kutov, D.C., Sulimov, A.V.: Development of docking programs for Lomonosov supercomputer. Journal of the Turkish Chemical Society Section A: Chemistry 7(1), 259–276 (2020), DOI: 10.18596/jotcsa.634130

57. Sulimov, V.B., Kutov, D.C., Sulimov, A.V.: Advances in docking. Current Medicinal Chemistry 26(42), 7555–7580 (2019), DOI: 10.2174/0929867325666180904115000

58. The Department of Organic Chemistry of Voronezh State University. `http://www.chem.vsu.ru/?req=/department/5/application/page.html`, accessed: 2020-10-01

59. Ton, A.T., Gentile, F., Hsing, M., et al.: Rapid Identification of Potential Inhibitors of SARS–CoV–2 Main Protease by Deep Docking of 1.3 Billion Compounds. Molecular Informatics 39, 2000028 (2020), DOI: 10.1002/minf.202000028

60. Trott, O., Olson, A.J.: AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 31(2), 455–461 (2010), DOI: 10.1002/jcc.21334

61. Tsuji, M.: Potential anti-SARS-CoV-2 drug candidates identified through virtual screening of the ChEMBL database for compounds that target the main coronavirus protease. FEBS Open Bio 10(6), 995–1004 (2020), DOI: 10.1002/2211-5463.12875

62. 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

63. Vostrikova, T.V., Kalaev, V.N., Medvedeva, S.M., et al.: Synthesized organic compounds as growth stimulators for woody plants. Periodico tche quimica 17(35), 327–337 (2020)

64. Vostrikova, T.V., Kalaev, V.N., Potapov, A.Y., et al.: Use of new compounds of the quinoline series as effective stimulants of growth processes. Periodico tche quimica 17(35), 781–790 (2020)

65. Yang, H., Yang, M., Ding, Y., et al.: The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proceedings of the National Academy of Sciences of the United States of America 100(23), 13190–13195 (2003), DOI: 10.1073/pnas.1835675100

66. Yang, J., Zhang, Y.: I-TASSER server: new development for protein structure and function predictions. Nucleic acids research 43(W1), W174–W181 (2015), DOI: 10.1093/nar/gkv342

67. Zaki, A., Boheemen, S., Bestebroer, T., et al.: Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. The New England journal of medicine 367 (2012), DOI: 10.1056/NEJMoa1211721