ChemistrySelect

Supporting Information

Natural Products Database Screening for the Discovery of Naturally Occurring SARS-Cov-2 Spike Glycoprotein Blockers

Abdullah G. Al-Sehemi[+], Fisayo A. Olotu[+], Sanal Dev[+], Mehboobali Pannipara[+], Mahmoud E. Soliman[+], Simone Carradori,* and Bijo Mathew[+]*

1. Materials and methods

1.1. Natural compounds database construction

In brief, we employed high-throughput virtual screening followed by a post docking for binding free energy calculation. Molecular modeling studies were performed on a personal computer running on Windows 7. Schrödinger suite 2012 as well as and modules such as Ligprep, Glide, Prime/MMGBSA, and Virtual Screening Workflow were used in this study.[1-3] In the proposed study, virtual screening was performed to identify hit compounds from the NPASS (natural product activity and species source) database consisting of 30926 natural compounds (http://bidd.group/NPASS/downloadnpass.html). This is freely accessible with its contents searchable by keywords, physicochemical property range, structural similarity, species and target search facilities.

1.2. Natural product database

The structures downloaded from NPASS database were prepared using Ligprep module available in Schrödinger suite 2017.[4] All possible states at pH 7 ± 1 were generated using Epik. “Desalt” and “generate tautomers” tools were also selected on the Ligprep interface and the stereoisomer computation was kept to retain specific chiralities and to generate at most 32 isomers per ligand. For each molecule, a set of conformers were generated through Monte Carlo multiple minimum (MCMM)/low-mode conformational search method, followed by 100 minimization steps with the OPLS3 force field. The maximum number of conformers per compound was limited to 1000. After the generation of conformers, only one conformer was displayed and carried for the further analyses.[5]

1.3. Structure-Based Virtual Screening

Virtual Screening was performed using virtual screening workflow of Schrödinger suite 2017-2. Structure-based virtual screening method was employed for filtering out the compounds from the

database based on dock score. Structure-based virtual screening uses HTVS (high-throughput virtual screening) docking protocol followed by SP (standard precision) docking and XP (extra precision) docking.[6]

Since the protein was devoid of any co-crystallized ligand, we used site map tool to identify the binding pocket of the selected crystal structure and the site with the highest site score was selected and based on the identified site points grid box was created which in turn was used for grid generation and virtual screening studies were carried out as follows. In the first step, Glide was run in high-throughput virtual screen mode, top 50% of the compounds were retained hence a total of 15463 molecules were picked up after HTVS docking.

It was then allowed to go to the next, Glide Single Precision (SP), stage, and the criteria set was oriented to pick up the top 5%. The retrieved hits were narrowed down to 773 molecules and these were retained and docked using Glide Extra Precision (XP) mode keeping the final output to 5%. The 38 molecules obtained from XP method were taken further to calculate ΔG_{bind} values using MM/GBSA calculations. Prime was performed using the VSGB solvation model to estimate the binding affinity of the selected ligands to the active site. In prioritizing the potential compounds, docking score and ΔG_{bind} were taken into account and two lead candidates were identified. The sequential virtual screening performed in this study is schematically represented in a flowchart in Figure 2 of the main text.

1.4. Post Docking and Binding Free Energy Calculation

Post docking calculations for the docked poses were calculated by using Prime MM/GBSA (molecular mechanics based generalized born/surface area) post docking scoring protocol of Schrödinger software using VSGB solvation model. The Prime MM-GBSA approach in Schrödinger predicts the total free energy of binding for a given receptor and a set of ligands. The total free energy of binding is then expressed in the form below mentioned equation (1):

$$\Delta G_{bind} = E_{complex}\ (\text{minimized}) - (E_{protein}\ (\text{unbound, minimized}) + E_{lig}\ (\text{unbound, minimized})) \quad (1)$$

where, ΔG_{bind} is the calculated binding free energy, $E_{complex}$ (minimized) is the MM-GBSA energy of the minimized complex, $E_{protein}$ (unbound, minimized) is the MM-GBSA energy of the minimized protein after separating it from its bound ligand and E_{lig} (unbound, minimized) is the MM-GBSA energy of the ligand after separating it from the crystal complex and allowing it to relax.[7]

1.5. Molecular dynamics simulation

The crystallographic structure of SARS-CoV-2 spike (S) protein was downloaded from the Protein Data Bank (PDB) with entry 6VSB (release date: 26-02-2020).[8] A S1/S2 S-protein monomer was prepared on UCSF Chimera GUI to minimize computational cost, while other co-crystalized molecules not applicable to this study were also removed. Missing residues were further filled using the integrated MODELLER module.[9] 2D structures of the natural compounds, Karuquinone B (KQB) and Castanospermine (CTN), were prepared on MarvinSketch and were then optimized on Avogadro-based UFF force field.[10] Ligand binding coordinates were defined using grid boxes on Autodock Vina prior to the docking of the compounds at the SiteMap predicted site.[11] The most favorable poses (highest negative) were selected for both compounds with values -6.4 and -5.0 kcal/mol for KQB and CTN, respectively. These complexes were then used for an all-atom molecular dynamics (MD) simulation run and they include KQB-S-protein and CTN-S-protein. Using in-house protocols previously implemented successfully based on their efficiencies,[12] MD simulations were performed with AMBER18 software package for 65 ns production runs. Resulting simulation trajectories were analyzed upon the completion of the MD run using the integrated CPPTRAJ module.[13]

Conformational variations across the protein structures were measured using parameters such as Cα-root mean square deviation (Cα-RMSD), root mean square fluctuation (Cα-RMSF), radius of gyration (Cα-RoG), surface accessible surface area (SASA), and principal component analysis (PCA). These metrics have been successfully used in previous studies to measure structural changes in protein systems across MD simulation trajectories and were also used herein to determine the inhibitory effects of the compounds on the target protein.[14] Visual analyses were also performed on the GUI of Discovery studio client 2016, Maestro 11.5 and UCSF Chimera.[15]

The affinities with which the compounds bind to the S-protein were also assessed using the Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) method, which estimated their relative binding free energies (ΔG_{bind}). To minimize entropic effects, the 3000 snapshots from the more stable terminal 10 ns time-frame were used for MM/PBSA calculation. Moreover, ligand interaction mechanisms were measured by decomposing the binding energies into those contributed by individual residues of the target site (per-residue decomposition).[16, 17]

1.6. ADME Prediction

Lastly, *in silico* ADME parameters were predicted to identify some physical-chemical properties, lipophilicity, water solubility, and pharmacokinetic data, through the login-free website http://www.swissadme.[18]

References

[1] Schrödinger Release 2018-2: Prime, Schrödinger, LLC, New York, NY, 2018.

[2] Schrödinger Release 2018-2: Glide, Schrödinger, LLC, New York, NY, 2018.

[3] S. Dev, S. R. Dhaneshwar, B. Mathew, *Comb. Chem. High Throughput Screen*. **2018**, *21*, 138-148.

[4] LigPrep, Schrödinger. "LLC." New York, NY, **2017**.

[5] D. G. T. Parambi, J. M. Oh, S. C. Baek, J. P. Lee, A. R. Tondo, O. Nicolotti, H. Kim, B. Mathew, *Bioorg. Chem*. **2019**, *93*, 103335.

[6] M. Oh, T. M. Rangarajan, R. Chaudhary, R. P. Singh, M. Singh, R. P. Singh, A. R. Tondo, N. Gambacorta, O. Nicolotti, B. Mathew, H. Kim, *Molecules* **2020**, *25*, 2356.

[7] B. Lakshminarayan, S. C. Baek, J. P. Lee, N. Kannappan, G. F. Mangiatordi, O. Nicolotti, T. Subburaju, H. Kim, B. Mathew, *ChemistrySelect* **2019**, *4*, 6614-6619.

[8] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, *Science* **2020**, *367*, 1260-1263.

[9] M. A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali, *Annu. Rev. Biophys. Biomol. Struct*. **2000**, *29*, 291-325.

[10] ChemAxon. Marvin Sketch. Https://WwwChemaxonCom/Products/Marvin/ **2013**.

[11] O. Trott, A. J. Olson. *J. Comp. Chem*. **2010**, *31*, 455-461.

[12] F. A. Olotu, M. Joy, M. A. Abdelgawad, S. E. Narayanan, M. E. Soliman, B. Mathew, *J. Biomol. Struct. Dyn*. **2020**, doi: 10.1080/07391102.2020.1796803.

[13] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.

[14] A. A. El Rashedy, F. A. Olotu, M. E. S. Soliman. *Chem Biodivers*. **2018**, e1700533.

[15] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.

[16] B. Mathew, A. A. Adeniyi, S. Dev, M. Joy, G. Ucar, G. E. Mathew, A. Singh-Pillay, M. E. Soliman. *J. Phys. Chem. B*. **2017**, *121*, 1186-1203.

[17] F. A. Olotu, M. E. Soliman, *Molecules* **2020**, *25*, E4247.

[18] A. Daina, O. Michielin, V. Zoete, *Sci. Rep*. **2017**, *7*, 42717.