CHALCONE DERIVATIVES AS POTENTIAL SARS-COV-2 INHIBITORS: A VIRTUAL SCREENING STUDY ON THE PAPAIN-LIKE PROTEASE ENZYME (PL^{PRO})

Nguyen Thi Thanh Thao[1], Nguyen Thi Phuong Truc[1]
(1) Thu Dau Mot University
Corresponding author: thanhthaont@tdmu.edu.vn

Abstract

The papain-like protease (PLPro) is a highly conserved, non-structural protein that plays a crucial role in the formation of the replication-transcription complex and the processing of polyproteins in SARS-CoV-2, as well as improving the host's antiviral immune responses against said virus. Chalcone is a common ingredient, which can be found in a multitude of natural substances, such as food and herbs. It has been proven to have various biological activities, including antiviral effects. Previous studies have identified several natural chalcone-based compounds with the ability to inhibit SARS-CoV-2 by targeting the PLPro enzyme. Based on these findings, this study investigated potential chalcone-derived PLPro inhibitors, as retrieved from Pubchem and in-house libraries. Virtual screening protocols, specifically molecular docking and molecular dynamics simulating filter, were applied to reach the desired goal. As a result, 1448 out of 1454 chalcone derivatives can effectively bind to SARS-CoV-2 via PLPro. The 5 substances with the most suitable docking score and binding mode were selected for the next step. Through MD, CID1021201513 and CID101585417 showed the greatest potential in targeting PLPro. However, further in vitro and in vivo studies must be conducted before the bio-activities of these chalcones against SARS-CoV-2 can be confirmed. Furthermore, the ligand-protein interaction mode analysed in this research can help design effective chalcone derivatives.

1. Introduction

Considered one of the most terrible pandemics of this century, the SARS-CoV-2 had infected more than 770 million and killed 6.9 million, as reported on February 11th, 2024 (World Health Organization, 2024). Though most people have been vaccinated, there is no guaranteed protection against SARS-CoV. Effects of the vaccine need to be regularly enforced with booster shots, which is not entirely feasible, given the short recommended interval between shots and the rapid rate at which the virus is mutating. Therefore, it is essential to procure effective, safe and cost-efficient inhibitors capable of binding to highly conserved proteins long-term. In this research, we focus on PL^{Pro}, a highly

conserved protein involved in virus replication, and active ingredients, preferably natural, capable of inhibiting this protein.

The papain-like protease (PL^{Pro} or nsp3) is a subunit which plays a major role in the replication-transcription complex (RTC). PL^{Pro} is responsible for hydrolysing the carboxyl group of glycine belonging to the non-structural proteins nsp1, nsp2, and nsp3. In addition, PL^{Pro} is also responsible for the catalytic activities of the deISGlating (deISGyl) and deUbiquitination (DUB) enzymes, which contribute to the suppression of the host's innate immune responses by acting on the IFNβ and NFκB signalling pathways (Mielech *et al.*, 2014).

Chalcone derivatives have been proven to have multiple advantageous biological effects, including cytotoxicity (Fogaça *et al.*, 2017), immunomodulatory (Lee *et al.*, 2015), anti-inflammatory (Herencia *et al.*, 1998), and antibacterial (Okolo *et al.*, 2021). The antiviral effects of synthetic and natural chalcones have been the subject of many studies, with several reports suggest that chalcone is capable of inhibiting HIV (Lee *et al.*, 2015), influenza A (Herencia *et al.*, 1998), rubella virus (Okolo *et al.*, 2021), and SARS-CoV (Duran *et al.*, 2021). During the COVID19 pandemic, drugs derived from chalcone have been extensively studied, revealing that chalcone can inhibit the activities of the RNAdependent RNA polymerase enzyme of the SARS-CoV-2 virus (Duran *et al.*, 2021). In the host's cells, chalcones inhibit the production cytokines such as TNFα, IL1β, IL6, and IκB kinase (IKKs) an essential regulator in the nuclear factor kappa B pathway (NFκB) and a central mediator of inflammatory immune responses (Pahl, 1999).

Therefore, inhibiting PL^{Pro} can potentially prevent viral replication while strengthening the host 's immune system. This, in turn, can help patients recover quickly and minimise the risk of a cytokine storm (Osipiuk *et al.*, 2021).

In 2022, Valipour (2022) gave evidence suggesting that the amide group in chalcone structure is a suitable backbone for the design and development of selective SARS-CoV-2 PLpro inhibitors. Through SAR analysis and biological activity testing, the study also showed that when this backbone loses its chalcone-amide configuration (either by being expanded, compressed, having the B-ring removed or converted to the "N, 2-diphenyl acetamide" configuration), its inhibitory activity against PL^{Pro} could be significantly reduced or completely eliminated. Therefore, in order to map out effective chalcone-derived PL^{Pro} inhibitors, the chalcone-amide backbone needs to be properly preserved.
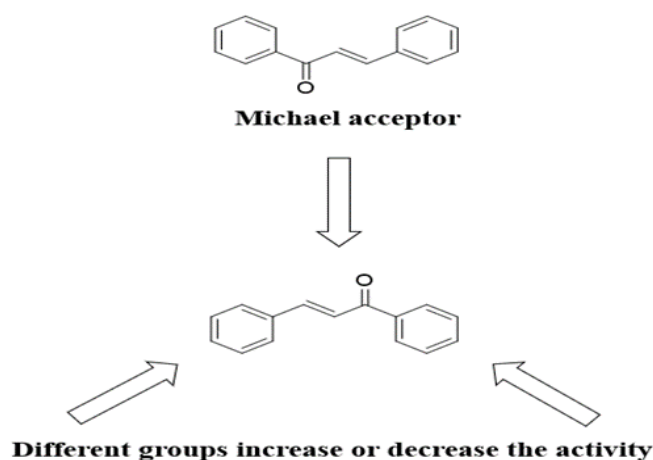

Figure 1. General structure of a chalcone

So far, the concept of finding chalcone derivatives capable of inhibiting 3CLpro is still new, prompting this study which aims to explore chalcone derivatives as potential PL^{Pro} inhibitors using virtual screening approaches, including molecular docking, molecular dynamics simulation, and binding free energy calculation. The substances that showed the most potential in this study will undergo in vitro testing in later research with the end goal of establishing effective chalcone derivatives in the fight against SARS-CoV-2.

2. Method

2.1. Library preparation

In this study, the screening database included a library of 1184 chalcone derivatives from the Pubchem library using the search term "Chalcone", and 270 inhouse chalcones. The compounds were saved under *.sdf format and later converted to *.mdb format using MOE 2015.10 software.

2.2. Molecular docking –based virtual screening

The virtual screening process is shown in Figure 2. The binding ability of 1454 chalcone substances towards PL^{Pro} was evaluated via Molecular docking. The 5 compounds with the most suitable docking score and binding mode were selected to undergo Molecular dynamics simulations (MDs) and Free binding energy calculations via the Molecular mechanics generalised born surface area (MM/GBSA) method (Genheden and Ryde, 2015).
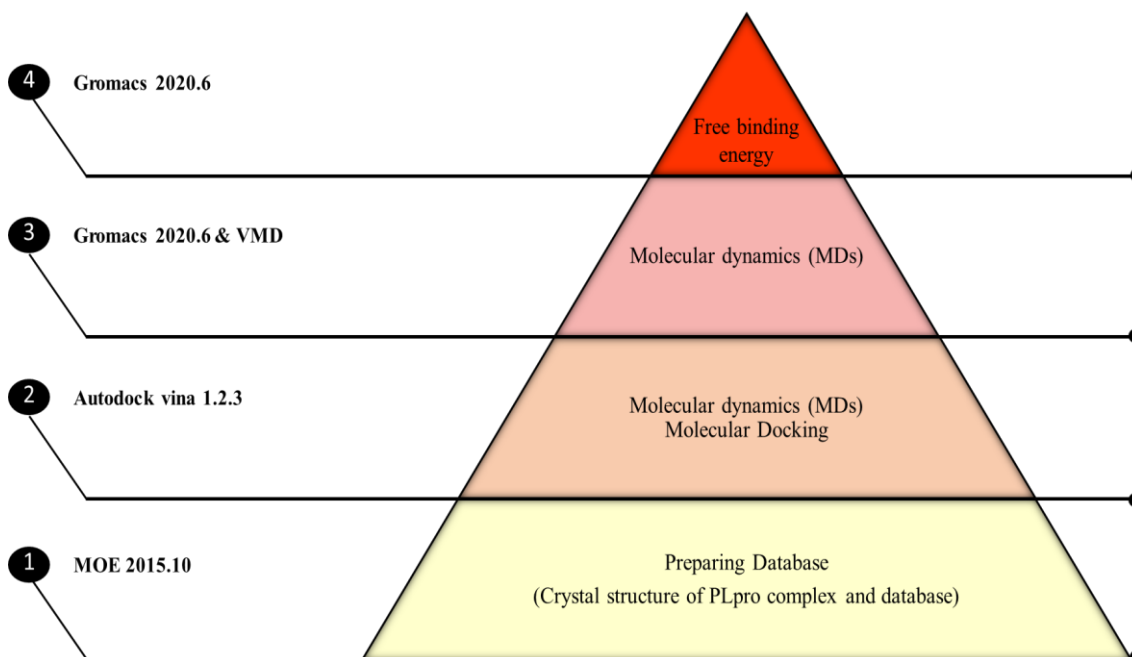

Figure 2. Screening procedure for reversible inhibitors of PL^{Pro}

2.3. Protein structure and docking database preparation

The crystal structure of PL^{Pro} (Protein Data Bank ID: 7JN2) was retrieved from the Protein Data Bank (Osipiuk *et al.*, 2020). The co-crystallization ligand PLP_Snyder441 was separated from the complex using AutoDockTools (Morris *et al.*, 2009). Water and irrelevant molecules were then removed and hydrogen atoms were added to the protein

receptor. Finally, the 3D structure of the protein was minimised energetically and optimised interactively using the MOE 2015.10 software.

The chalcone structures from two databases were prepared via salt removal and hydrogen bond optimization. The chalcone derivatives and reference ligand PLP_Snyder441 were prepared for molecular docking analysis using MOE 2015.10 software, which minimises the energy of the ligands to the stable gradient parameter of 0.0001 kcal.mol^{-1}.

2.4. Identifying binding site and molecular docking

AutoDock4 software was used to simulate the binding site by extending a suitable radius from the interactive region of re-docked ligand PLP_Snyder441 into the PL^{Pro} (Morris *et al.*, 2009). The redocking process was conducted via Autodock vina 1.2.3 to assess the "reproducibility" of the binding site with suitable parameters. The model is considered reliable when the re-docked ligand has a good docking score, can successfully bind to the protein in the correct binding pocket, and has the RMSD value of under 2 Å (Ramírez and Caballero, 2018). In the docking stage, the binding affinity of chalcone substances towards PL^{Pro} was evaluated using a docking score (kcal.mol^{-1}). In addition, the ligand-protein interaction was analysed using Discovery Studio 2021 software, which focuses on the hydrogen, ionic, hydrophobic, and van der Waals interactions.

2.5. Molecular dynamics simulation

The stability of ligand-protein complex was assessed throughout molecular dynamics simulation using GROMACS software (Abraham *et al.*, 2015). This study also utilises the CHARMM27 force field, TIP3P water model, and dodecahedron box. The ligand topology was created using SwissParam tool (http://www.swissparam.ch) (Zoete *et al.*, 2011) whereas the protein topology was created using GROMACS software. After solvation, the protein-ligand complex was neutralised by adding Na$^+$ or Cl$^-$. The energy was then minimised at 100 ps intervals. After that, solvent and ions around the protein were equilibrated in the NVT and NPT phases at 100 ps intervals. The temperature during the NVT step was 300K in the Berendsen thermostat. The pressure in the NPT step was 1 bar (0.987 atm) in the ParinelloRahman pressurizer. The MD process was recorded for a period of 100 ns and the output was recorded frame-by-frame at 0.01 ns intervals. The output trajectories were analysed throughout RMSD, RMSF, Rg and SASA values, hydrogen bond frequency, and binding free energy calculation.

2.6. Binding free energy calculation

The binding free energy of the interaction between chalcone substances and PL^{Pro} was analysed using the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method at the temperature of 298K and the salt concentration of 0.15 M. This study employ the gmx_MMPBSA program, in which the binding free energy (ΔGbind) was calculated using formula (1):

$$\Delta G_{bind} = \Delta G_{complex} - \left(\Delta G_{protein} + \Delta G_{ligand} \right) \ (1)$$

The binding free energy of each component in formula (1) is determined according to formula (2), (3) and (4):

$$\Delta G_x = \Delta H - T\Delta S = \ \Delta E_{MM} + \ \Delta G_{solv} - T\Delta S \ (2)$$

$$\Delta E_{MM} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{dihedral} + \Delta E_{vdW} + \Delta E_{ele} \ (3)$$

$$\Delta G_{solv} = \Delta G_{GB} + \Delta G_{SA} \ (4)$$

x: protein, ligand, or protein complex; ΔE-MM: molecular mechanics energy; ΔG-solv: solvation free energy; -TΔS: conformational entropy change; ΔE-bond, ΔE-angle, ΔE-dihedral: internal energies for fixed geometries before and after binding; ΔE-vdW: van der Waals bond energy; ΔE-ele: electrostatic energy; ΔG-solv: total electrostatic solvation energy; ΔG-GB: polar solvent contribution; ΔG-SA: nonpolar solvent contribution (Genheden and Ryde, 2015).

3. Results and discussion

3.1. Redocking results and molecular docking model

Redocking results suggest the model in this study was qualified with the RMSD value of 0.57 Å (< 2 Å). The best binding score between ligand and protein was 9.907 kcal.mol^{-1}. The ligand also interacted with ligands similar to the co-crystallized ligand, forming hydrophobic bonds with Asp164, Pro248, Tyr264, Gln269 and hydrogen bonds with Asp164, Gln269, Tyr273 (Figure 3). From the configuration of the referenced ligand PLP_Snyder441, the binding site of PLpro receptor was expanded to cover the whole pocket, the specific parameters being: center_X= 51.4, center_Y= 31.92, center_Z= 2.31; box size: 14.91 x 17.49 x 14.04; spacing = 1 Å. Nummode and exhaustiveness parameters were set to 9 and 8, respectively.
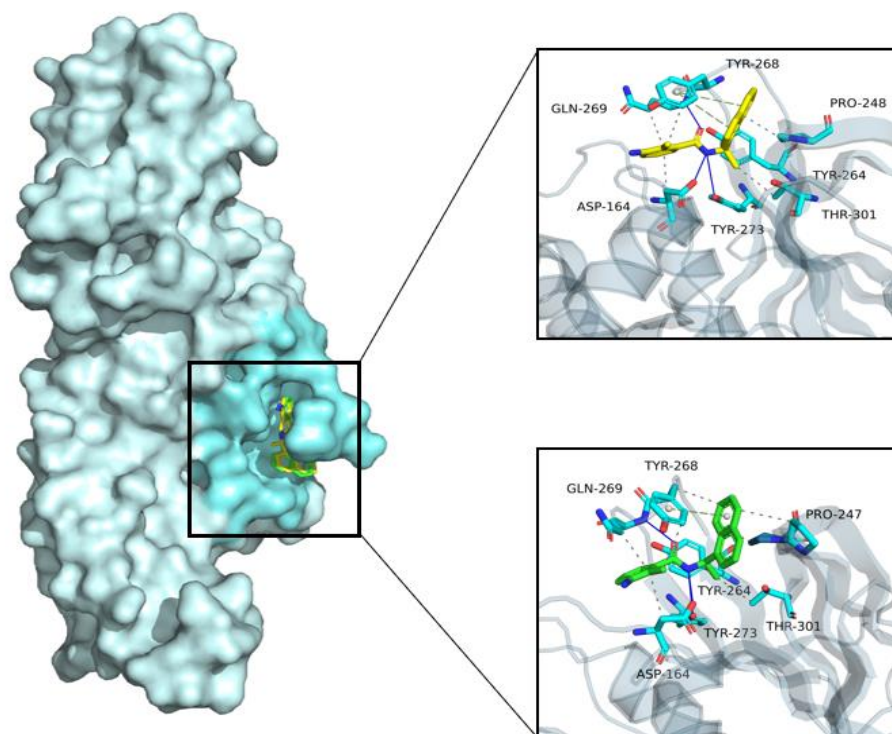

Figure 3. Configuration and interaction between referenced ligand and amino acids of redocking model in co-crystallization complex

3.2. Screening using molecular docking

From 1454 chalcone substances in the input database, 1448 compounds (99.59%) successfully docked into the binding site of PL^{Pro}. The docking score distribution is as presented in Figure 4.
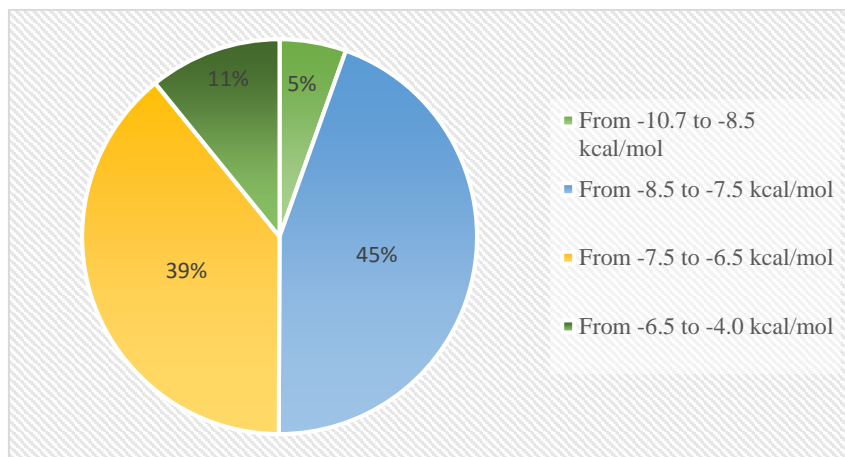
Figure 4. Docking score distribution chart

The 5 ligands with the highest docking scores, ranging from -9.705 to -10.72 kcal.mol^{-1}, are shown in Figure 5.
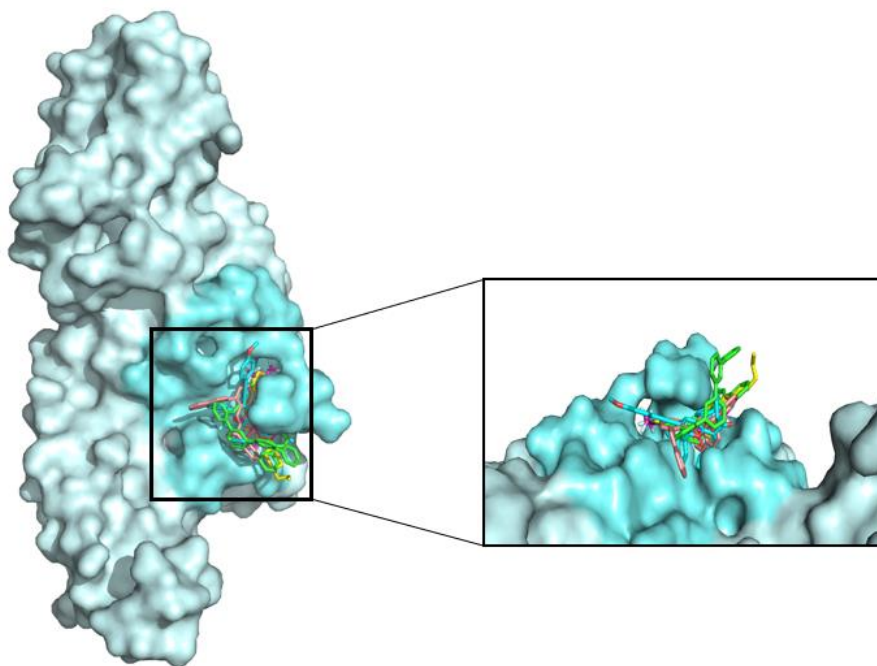

Figure 5. Configurations of the 5 ligands with the highest docking scores with PL^{Pro}, in ligand-protein model

Analysis of the 5 ligands with the highest docking scores with PL^{Pro}.

CID102120513

The chalcone CID102120513 showed good binding power towards PL^{Pro}, with a docking score of 10.72 kcal.mol^{-1}. This ligand had a similar binding mode with the co-crystalized ligand, forming hydrophobic bonds with residues Asp164, Pro247, Pro248, Tyr264, Tyr268 and Gln269 (Figure 6A). CID102120513 also formed an additional πStacking bond with Tyr264, which plays a vital role in maintaining the desired configuration of the active site of PL^{Pro}.

CID102261900

CID102261900 formed hydrophobic bonds with the PL^Pro via the Asp164, Arg166, Tyr264, Tyr268, Tyr273, Thr301 residues and two H-bonds with Lys157 and Arg166 residues. CID102261900 had a docking score of 9.667 kcal.mol^{-1} (Figure 6B).

CID101585417

The docking score of CID101585417 was 9.765 kcal.mol^{-1}. This chalcone formed hydrophobic bonds with Leu162, Asp164, Pro248, Tyr264, Tyr268 residues, H-bonds with Thr301 and πStacking bonds with Tyr268 (Figure 6C).

CID101544859

CID101544859 had a binding affinity of 9.76 kcal.mol^{-1} with a binding site located deep within the PL^Pro receptor. This compound formed H-bonds with Thr301 residue and hydrophobic bonds with Leu162, Asp164, Pro248, Tyr264, Tyr268, Gln269 residues (Figure 6D).

CID102120512

CID102120512 is located deep within the binding site of the PL^Pro receptor. CID102120512 had a docking score of 9,705 kcal.mol^{-1}. The ligand formed hydrophobic bonds with Asp164, Glu167, Pro247, Tyr268, Gln269, Thr301 residues, one Hbond with Arg166 residues, and two Stacking bonds with Tyr264 and Tyr268 residues (Figure 6E).
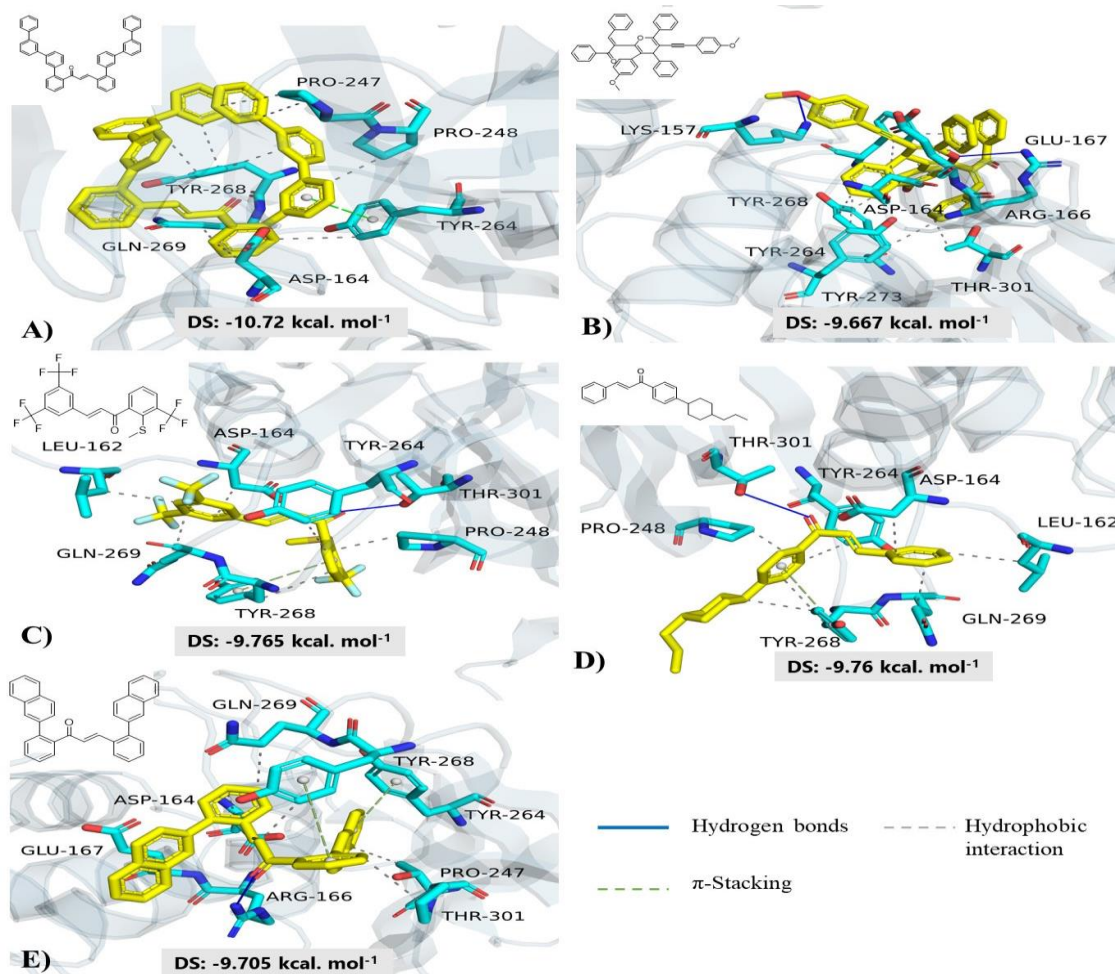

Figure 6. Ligand-protein complexes of 5 ligands and PL^Pro in interaction mode

3.3. Ligand-protein stability in 100 ns

Table 1 shows the average RMSD, Rg and SASA values of the protein backbone in apoprotein form and in complexes with ligands. In general, RMSD, Rg and SASA values of apoprotein and ligand-protein complex were similar for most of these chalcones, indicating stable bonds between PL^{Pro} and ligands. Only CID102120513 had a high RMSD value of 2.38 ± 0.81 Å in the ligand-protein complex, which can be explained by the bulky structure of this chalcone. However, it was still considered acceptable since structures with RMSD values below 3Å can still be deemed stable (Mohamad *et al.*, 2019). In addition, the RMSD value was lowest in the reference ligand. However, the chalcone scaffold was considered stable with RMSD values ranging from 1.01 - 2.15 Å.

TABLE 1. Stability of complexes during 100 ns MD simulation

Complex	Protein			Ligand
	RMSD (Å)	**Rg (Å)**	**SASA (nm²)**	**RMSD (Å)**
Apoprotein	1.76 ± 0.34	23.38 ± 0.14	164.83 ± 1.24	-
PLP snyder441 (Ref.)	1.75 ± 0.44	23.47 ± 0.16	165.83 ± 1.17	0.43 ± 0.22
CID101585417	1.67 ± 0.24	23.54 ± 0.07	268.77 ± 1.64	2.15 ± 0.23
CID102120513	2.38 ± 0.81	23.61 ± 0.11	269.61 ± 1.67	1.94 ± 0.94
CID102120512	1.73 ± 0.23	23.61 ± 0.08	268.24 ± 1.54	1.01 ± 0.29
CID102261900	1.54 ± 0.18	23.51 ± 0.08	268.53 ± 1.55	1.63 ± 0.35
CID101544859	1.42 ± 0.13	23.52 ± 0.07	267.67 ± 1.47	1.52 ± 0.29

The interaction between amino acids of PL^{Pro} and ligands in the complex was illustrated in Figure 7. This study focused on 5 important residues found in the binding site, Arg164, Pro248, Tyr264, Gln269 and Tyr273. The RMSF values of five chalcones were low and similar to that of the referenced.
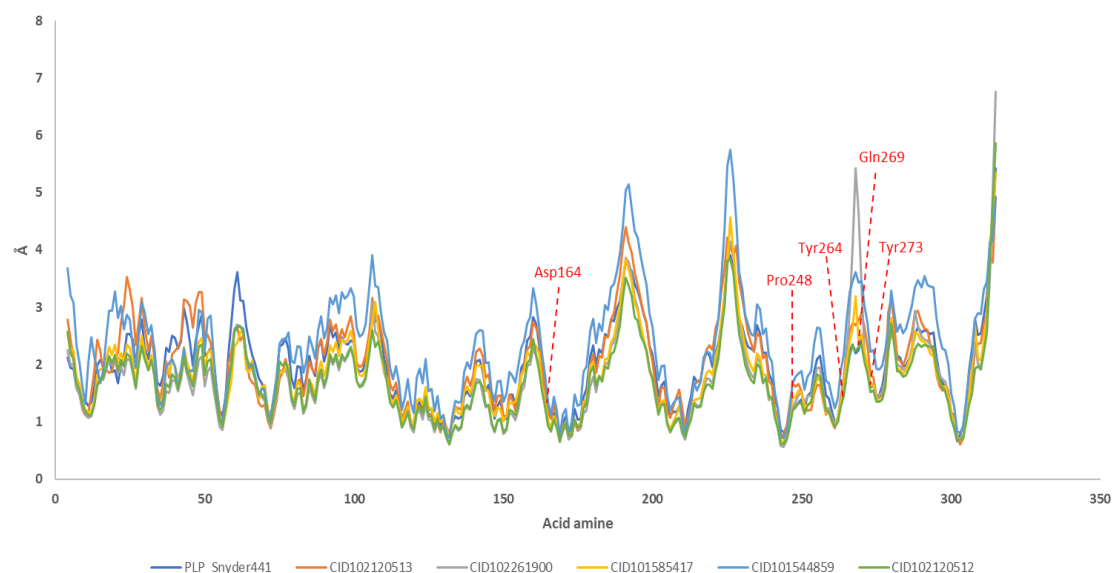

Figure 7. Interaction of amino acids of PLPro with ligands

3.4. Investigation into the top hits during 100 ns MD simulation

The top five chalcones with the most suitable docking scores and binding modes were selected to further examine their stability in the protein-ligand complex for a period of 100 ns using molecular dynamics simulation. Judging by VMD visualisation, compounds

with the most inhibitory potential were CID101585417 and CID1021201513. These compounds remained stable throughout the investigated time period and showed normal binding interaction during the MD process (Figure 8A and B).
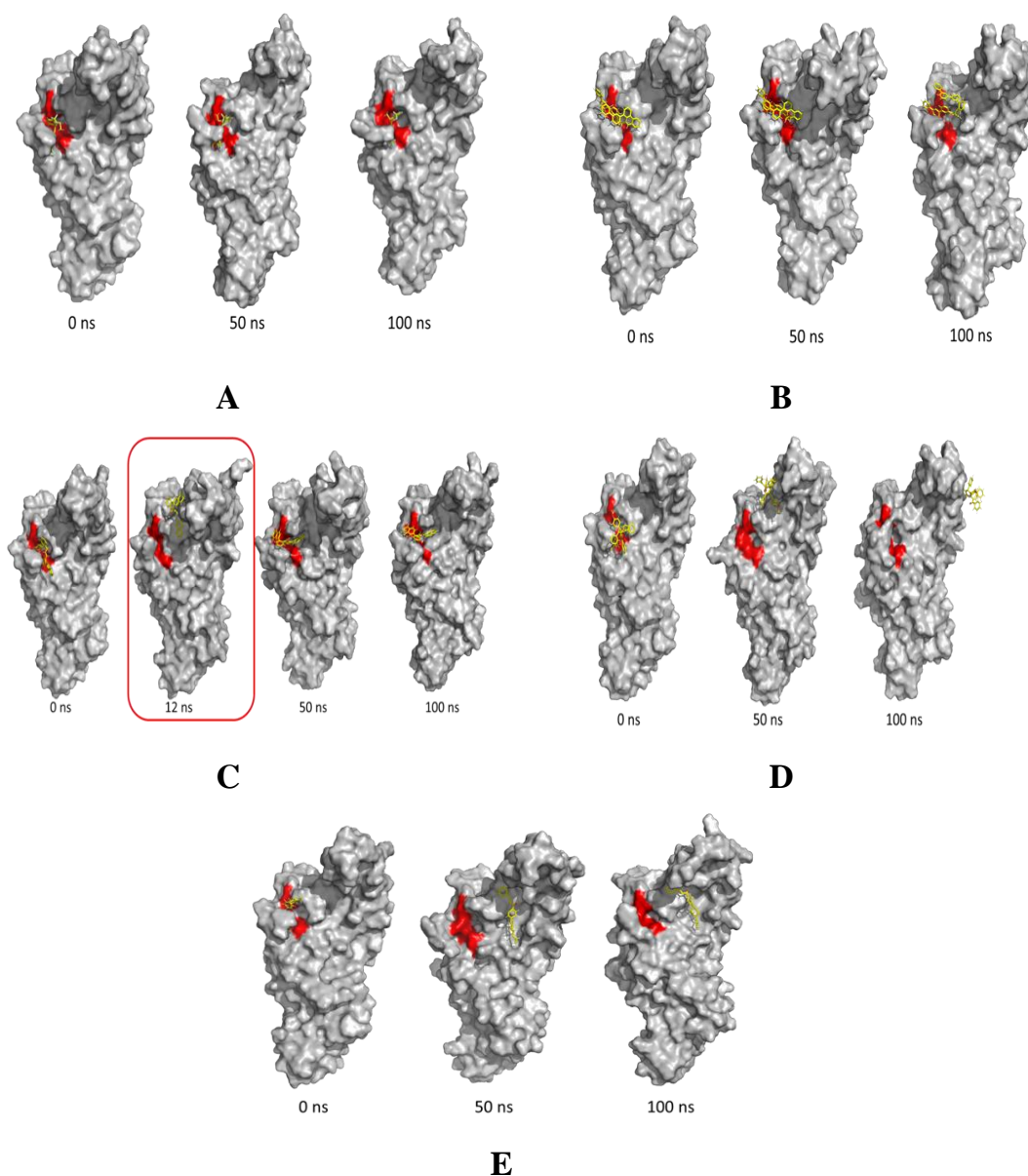

Figure 8. Position of ligand in reference to PLpro during 100 ns
A. CID101585417   *B.* CID1021201513   *C.* CID102120512   *D.* CID102261900
E. CID101544859

In contrast, CID102120512 stayed binded most of the simulation time, but escaped the binding pocket from 11 to 18 ns (Figure 8C). CID102261900 and CID101544859 moved away from the binding pocket during the MD entire process (Figure 8D and E).

3.5. Hydrogen bonds frequency and binding free energy

Although suitable RMSD, Rg and RMSF values were recorded for CID102120512, CID102261900, and CID101544859, these ligands did not remain bound to PL^{Pro} for the duration of MDs. Through MD simulations, CID1021201513 and CID101585417 were

considered most suitable. Then, hydrogen bonds frequency and binding free energy of these 2 substances were analysed. The two compounds had strong affinity towards PL^{Pro}, evident by the high frequency of H-bonds being recorded (Figure 9A and B). Regarding binding free energy, the average values for CID1021201513 and CID101585417 were 21.02 ± 3.78 kcal.mol^{-1} and 19.82 ± 3.15 kcal.mol^{-1}, respectively (Figure 9C and D).

A   **B**

C   **D**
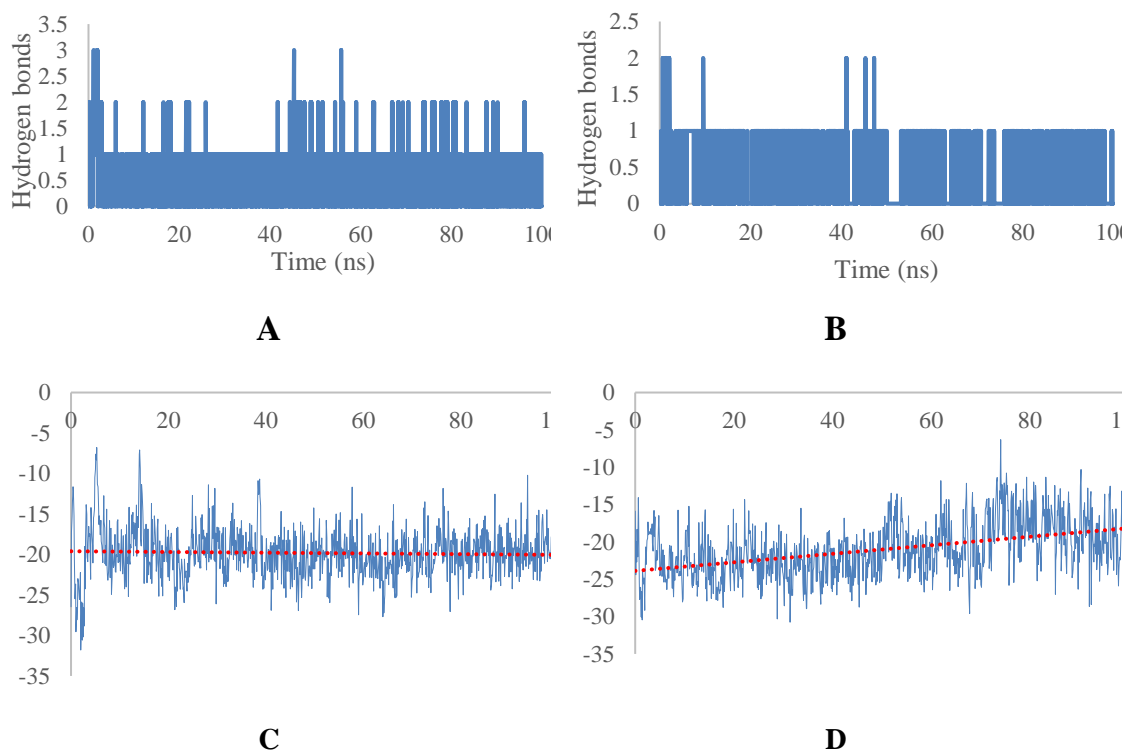
Figure 9. The hydrogen bonds frequency of *A*. CID101585417 *B*. CID102120513, Binding free energy of *C*. CID101585417 *D*. CID102120513

3.6. The role of functional groups in PL^{Pro} binding

Most of the chalcone derivatives in this study had substituents attached to the A and/or B ring of the chalcone scaffold. There were few substances with substituents at position C, which allows these substances to interact with more residues. For example, CID1022611900 had pyran heterocyclic, phenyl, and benzylic aromatic rings at position C (Figure 10). These structures helped form H and hydrophobic bonds with 5 amino acids, Asp164, Arg166, Tyr264, Tyr268, and Tyr273, resulting in a good docking score. Substituents attached to the A and B positions of the chalcone scaffold were more frequently found at o and p positions. Electron withdrawing groups, such as phenyl, naphthalene, and halogen-containing C chains could bond with various amino acids and help the aromatic rings form bonds with PL^{Pro} by reducing the charge density of the rings. This effect was observed in CID102120512, CID1021201513 and CID101585417, whose functional groups bound to Asp264, Pro248, Gln269, and Tyr268 residues, resulting in strong binding affinity towards PL^{Pro}. In contrast, electron donating groups like cyclohexyl found in CID101544859 was unable to bound to protein residues.
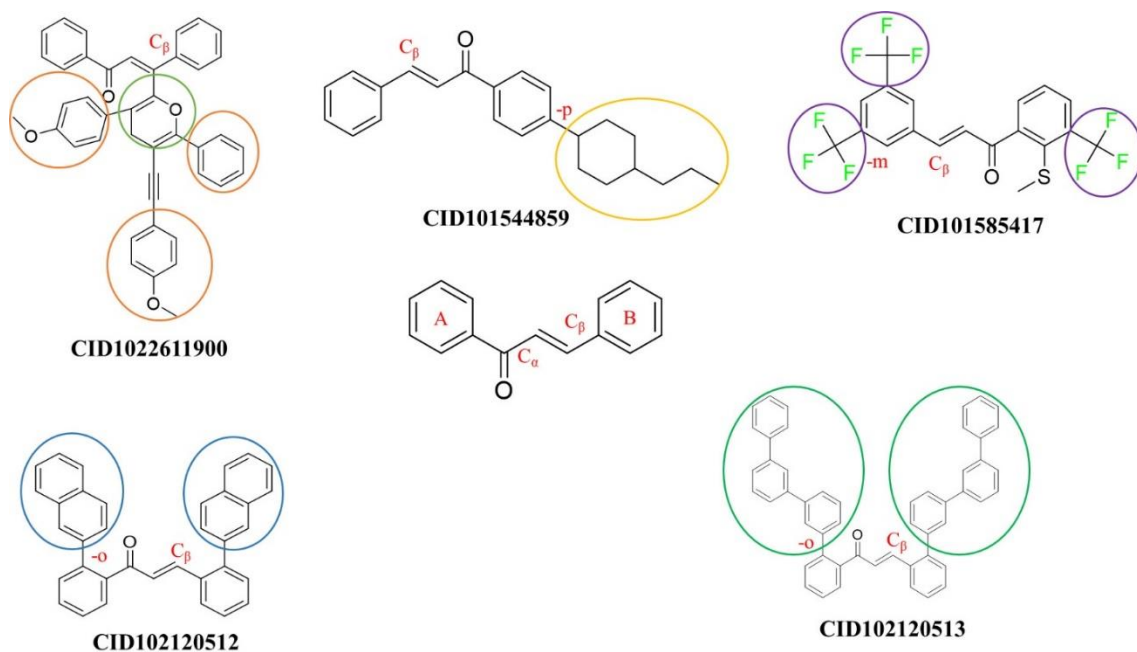
Figure 10. The role of functional groups in PLPro binding

4. Conclusion

This study employed the help of virtual screening tools in order to determine potential inhibitors against Papain-like Protease of SARS-CoV-2 from a chalcone database. Through docking, the study selected 5 ligands with the highest binding affinity and most suitable interaction mode. From there, 2 chalcones, CID1021201513 and CID101585417, were deemed the most promising based on their binding ability, recorded for a period of 100 ns using MD simulation. The study also analysed the role of functional groups in chalcone derivatives which contributed to the compounds' ability to bind to residues found in PL^{Pro}. Future research could benefit from deeper investigation into the biological activities of these inhibitory substances against SARS-CoV-2. Knowledge involving the effects functional groups had on the chalcone scaffold can be utilised to design optimal chalcones with desirable binding power against PL^{Pro} and similar proteins.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX, 1-2*, 19-25.

Duran, N., Polat, M. F., Aktas, D. A., Alagoz, M. A., Ay, E., Cimen, F., Tek, E., Anil, B., Burmaoglu, S., & Algul, O. (2021). New chalcone derivatives as effective against SARS-CoV-2 agent. *International Journal of Clinical Practice, 75(12)*, e14846.

Fogaça, T. B., Martins, R. M., Begnini, K. R., Carapina, C., Ritter, M., Pereira, C. M. Pd., Seixas, F. K., & Collares, T. (2017). Apoptotic effect of chalcone derivatives of 2-acetylthiophene in human breast cancer cells. *Pharmacological Reports*, 69(1), 156-161.

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov, 10*(5), 449-461.

Herencia, F., Ferrándiz, M. L., Ubeda, A., Domínguez, J. N., Charris, J. E., Lobo, G. M., Alcaraz, M. J. (1998). Synthesis and anti-inflammatory activity of chalcone derivatives. *Bioorg Med Chem Lett, 8*(10*)*, 1169-1174.

Lee J. S., Bukhari S. N., & Fauzi N. M. (2015) Effects of chalcone derivatives on players of the immune system. *Drug Des Devel Ther, 9*, 4761-4778.

Mielech, A. M., Chen, Y., Mesecar, A. D., & Baker, S. C. (2014). Nidovirus papain-like proteases: multifunctional enzymes with protease, deubiquitinating and deISGylating activities. *Virus Res*, *194*, 184-190.

Mohamad, Y., Tamer, A., & Khaled, E. (2019). Performance comparison of ab initio protein structure prediction methods. *Ain Shams Engineering Journal, 10*(4), 713-719.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem, 30*(16), 2785-2791.

Okolo, E. N., Ugwu, D. I., Ezema, B. E., Ndefo, J. C., Eze, F. U., Ezema, C. G., Ezugwu, J. A., & Ujam, O. T. (2021). New chalcone derivatives as potential antimicrobial and antioxidant agent. *Sci Rep, 11*(1), 21781.

Osipiuk, J., Azizi, S. A., Dvorkin, S., Endres, M., Jedrzejczak, R., Jones, K. A., Kang, S., Kathayat, R. S., Kim, Y., Lisnyak, V. G., Maki, S. L., Nicolaescu, V., Taylor, C. A., Tesar, C., Zhang, Y. A., Zhou, Z., Randall, G., Michalska, K., Snyder, S. A., Dickinson, B. C., & Joachimiak, A. (2021). Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat Commun*, *12*(1), 743.

Osipiuk, J., Tesar, C., Endres, M., Lisnyak, V., Maki, S., Taylor, C., Zhang, Y., Zhou, Z., Azizi, S. A., Jones, K., Kathayat, R., Snyder, S. A., Dickinson, B. C., Joachimiak, A., & Center for Structural Genomics of Infectious Diseases (2020). The crystal structure of Papain-Like Protease of SARS CoV-2 in complex with PLP_Snyder441 inhibitor. RCSB Protein Data Bank. Retrieved October 7, 2023 from https://www.rcsb.org/structure/7JN2.

Pahl, H. L. (1999). Activators and target genes of Rel/NF-κB transcription factor. *Oncogene, 18*(49), 6853-6866.

Ramírez, D., & Caballero, J. (2018). Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data?. *Molecules, 23*(5), 1038.

Valipour, M. (2022). Chalcone-amide, a privileged backbone for the design and development of selective SARS-CoV/SARS-CoV-2 papain-like protease inhibitors. *European Journal of Medicinal Chemistry, 240*, 114572.

World Health Organization (2024). *WHO Coronavirus (COVID-19) Dashboard*. Retrieved February 28, 2024 from: https://covid19.who.int/.

Zoete, V., Cuendet, M. A., Grosdidier, A., & Michielin, O. (2011). SwissParam: a fast force field generation tool for small organic molecules. *J Comput Chem, 32*(11), 2359-2368.