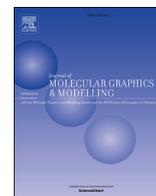




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



In silico approach for identifying natural lead molecules against SARS-COV-2



Shiv Shankar Gupta^{a, c, 1}, Ashwani Kumar^{b, 1}, Ravi Shankar^{b, c, **}, Upendra Sharma^{a, c, *}

^a Chemical Technology Division, CSIR-IHBT, Palampur, HP, 176 061, India

^b Biotechnology Division, CSIR-IHBT, Palampur, HP, 176 061, India

^c Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, 201002, India

ARTICLE INFO

Article history:

Received 1 February 2021

Received in revised form

24 March 2021

Accepted 29 March 2021

Available online 13 April 2021

Keywords:

COVID-19

Natural products

In silico

Main protease

Molecular dynamics

Pharmacokinetic parameters

ABSTRACT

The life challenging COVID-19 disease caused by the SARS-CoV-2 virus has greatly impacted smooth survival worldwide since its discovery in December 2019. Currently, it is one of the major threats to humanity. Moreover, any specific drug or vaccine unavailability against COVID-19 forces to discover a new drug on an urgent basis. Viral cycle inhibition could be one possible way to prevent the further genesis of this viral disease, which can be contributed by drug repurposing techniques or screening of small bioactive natural molecules against already validated targets of COVID-19. The main protease (M^{Pro}) responsible for producing functional proteins from polyprotein is an important key step for SARS-CoV-2 virion replication. Natural product or herbal based formulations are an important platform for potential therapeutics and lead compounds in the drug discovery process. Therefore, here we have screened >53,500 bioactive natural molecules from six different natural product databases against M^{Pro} (PDB ID: 6LU7) of COVID-19 through computational study. Further, the top three molecules were subjected to pharmacokinetics evaluation, which is an important factor that reduces the drug failure rate. Moreover, the top three screened molecules (C00014803, C00006660, ANLT0001) were further validated by a molecular dynamics study under a condition similar to the physiological one. Relative binding energy analysis of three lead molecules indicated that C00014803 possess highest binding affinity among all three hits. These extensive studies can be a significant foundation for developing a therapeutic agent against COVID-19 through vet lab studies.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

Severe acute respiratory syndrome-coronavirus-2, a positive-sense single-stranded RNA virus with 27–31 kb RNA genome size, largest one known to date, belongs to the β genus of coronaviridae family [1]. COVID-19 is a major global health emergency in the 21st century emanated from Wuhan city of China in December 2019, which is responsible for lead pandemic with high morbidity and 4–5% mortality rate [2,3]. It leads to over 2.6 million death & over 119 million cases until March 16, 2021 globally, as per the WHO

report and still counting [4].

The genesis of coronavirus is over 55 million years earlier, which was concurred with bats, and its most recent familiar progenitor [5]. The first infection by SARS-CoV in 2002 affected over 8000 human beings in China, at the rate of 9.2% fatality [6]. Primary infection site by a coronavirus (CoV) in humans and other animals include severe respiratory and gastrointestinal tract infections [7]. The scientific community worldwide is trying to develop an effective vaccine as well small drug molecules against COVID-19. Several druggable targets in order to inhibit threatening viral infection cycle can be targeted including viral entry inhibitor to cell, viral replication inhibitor, polymerase inhibitor, viral protein synthesis inhibitor, protease inhibitor, RNA-dependent RNA polymerase (RdRp) inhibitor, viral exit inhibitor etc. to develop a remedy [8]. Among them, viral main protease is one of the attractive target as it is responsible for producing polyprotein, which cleaved to different functional proteins of replicase and polymerase, responsible for viral RNA replication process [9,10]. M^{Pro} is comprised of three

* Corresponding author. Chemical Technology Division, CSIR-IHBT, Palampur, HP, 176 061, India.

** Corresponding author. Biotechnology Division, CSIR-IHBT, Palampur, HP, 176 061, India.

E-mail addresses: ravish@ihbt.res.in (R. Shankar), upendra@ihbt.res.in, upendraithbt@gmail.com (U. Sharma).

¹ Equal contribution.

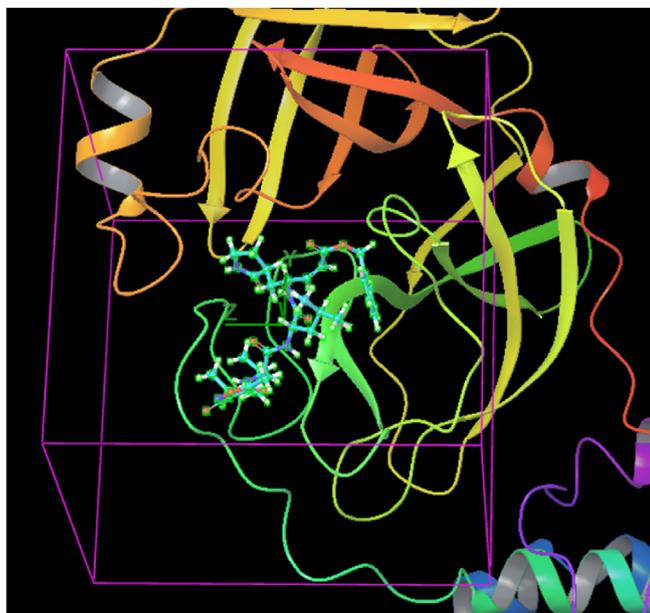


Fig. 1. Receptor grid generation at active site by specifying internal ligand.

domains (I-III) featured by active His-Cys catalytic dyad [11]. Main protease possesses several substrate binding pockets such as P1, P1', P2, P3, P4 and P5 along with S1, S1', S2 and S4 subsites [11].

Global research status is a vital step for catching an effective cure against COVID-19. Remdesivir has been conceded as a propitious antiviral agent against SARS/MERS-CoV virus in cultured cells and non-human primate models [12]. Remdesivir has shown its potential by treating first US case of COVID-19 successfully [13]. Apart from synthetic drug, there are natural products based active agents which can't be ignored, especially for the treatment of infectious diseases [14]. Many, nature derived anti-SARS-CoV molecule identified experimentally include glycyrrhizin, lycorin, luteolin, emodin, hesperetin, quercetin, sinigrin, savinin, myricetin and scutellarein, which could contribute significantly toward the discovery of a therapeutic agent against COVID-19 [15–20].

Wrapp et al. disclosed the cryo-electron microscopic structure of the COVID-19 spike protein target for antibody-mediated neutralization [21] and knowledge of spike protein structure at the atomic-level could help to discover drug molecules and vaccine design. Additionally, Liu et al. disclosed the structure of nCoV main protease with internal ligand (PDB ID: 6LU7), which greatly helped to identify new drug molecules against COVID-19 virus through *in-silico* study [22]. Based on *in silico* study, Wang J. repurposed five neutral drugs, including eravacycline, lopinavir, carfilzomib, elbasvir, and valrubicin as potential inhibitor of COVID-19 main protease [23]. Similarly, Contini A. also identified indinavir and cobicistat as potential binder of M^{pro} of COVID-19 [24]. In this series, Wu group contributed, few indispensable natural molecules e.g. series of andrographolide derivatives, isodecortinol, cervisterol & hesperidin against M^{pro} of nCoV through a structure

based virtual ligand screening [25]. Alqahtani et al. has also proposed few natural lead molecules such as chrysophanol 8-(6-galloylglucoside), 3,4,5-tri-*O*-galloylquinic acid, mulberrofuran G, withanolide A, isocodonocarpine and calonysterone as efficient binder of protease enzyme [26]. Mendanha et al. has screened ten anti-HIV drug molecules which have shown nelfinavir with better molecular dynamics parameters as compare to reference N3 [27]. Additionally, Prasad et al. has screened two hundred fifty-six FDA approved antimicrobial agents to reveal cefiderocol, plazomicin and vanganciclovir as active agent against main protease [28]. Further, Purohit et al. has identified oolonghomobisflavan-A lead molecule against COVID-19 protease via virtual screening of sixty-five molecules derived from tea plant [29]. Further virtual screening by AboulMagd et al. inferred chromene-2-one based active molecule against main protease [30].

Inspired by these studies along with the fact that most of the natural origin molecules possess high biocompatibility and low toxicity, high throughput virtual screening (Glide module of Schrodinger) of natural bioactive molecule databases (six different natural database) against the main protease of SARS-CoV-2 virus is performed. Subsequent filtration of a potent molecule through standard & extra precision algorithm afforded hit molecules. Moreover, top-ranked molecules were further validated through molecular dynamics study along with N3 (reference molecule) to propose a potent lead molecule against M^{pro} of COVID-19. *In-silico* pharmacokinetic studies were done for top candidates in order to reduce the drug failure rate. Through this study we have aimed to find lead molecules, which have potential to act as therapeutic candidate against SARS-CoV-2.

2. Material and methods

2.1. Protein selection and preparation

The solved structure of nCoV-19 main protease was retrieved from Protein Data Bank (PDB ID: 6LU7). Mentioned protein has resolution of 2.16 Å possess R-value free and work of 0.235 and 0.203, respectively. Internal co-crystallized ligand namely benzyl(3S,6S,9S,12R,Z)-9-isobutyl-6-isopropyl-3-methyl-1-(5-methylisoxazol-3-yl)-1,4,7,10-tetraoxo-12-(((R)-2-oxopyrrolidin-3-yl)methyl)-2,5,8,11-tetraazapentadec-13-en-15-oate or simply N3 was defined to generate receptor grid. The protein preparation wizard was utilized for optimization and minimization process for downloaded raw protein (PDB ID: 6LU7) with default option. Maestro operating environment was used to visualize the protein structure in Maestro software. Optimization, ionization and energy minimization action were performed by keeping default option in protein preparation wizard of Maestro [31].

2.2. Natural product databases and ligands preparation

An extensive literature survey has been carried out in search of safe naturally bioactive candidates. Bioactive natural molecules to be screened were extracted in sdf format from different natural product database such as NPASS (Natural product activity and species source database), TIPdb (Taiwan indigenous plants

Table 1
H-bonding of top ranked candidates and reference molecule with main protease (6LU7).

Compound	Docking score (kcal/mol)	H-bond donor (By ligand to amino acid residue)	H-bond acceptor (By ligand from amino acid residue)
C00014803 (TIPdb)	−12.85	THR24, THR26, SER46, CYS145, HID163, GLU166, GLN189	ASN142
C00006660 (TIPdb)	−11.68	THR26, LEU141, ASN142, THR190	CYS145
ANLT0001 (Analyticon)	−11.12	THR26, ASN142, HID163, GLU166	–
Internal Reference (N3)	−8.18	GLU166, GLN189, THR190	GLU166

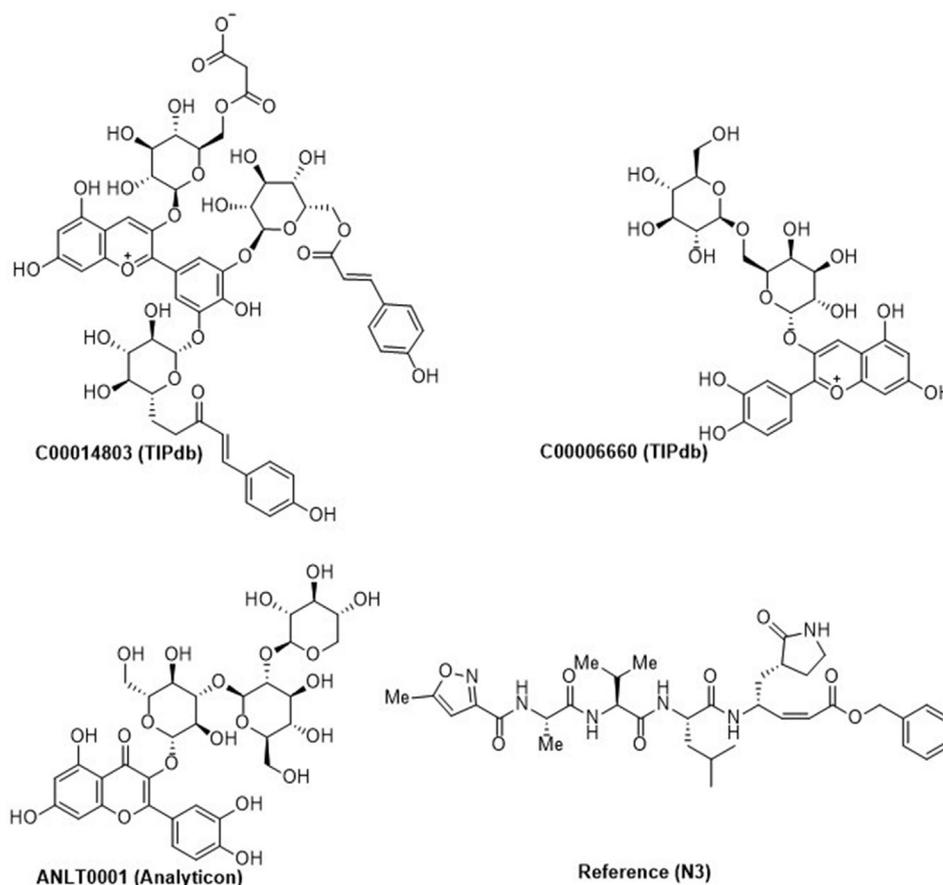


Fig. 2. Representative top three natural candidates and reference molecule.

database), Analyticon, ChemDiv, ChEBI (Chemical Entities of Biological Interest) and Life Chemicals. Reference molecule i.e. co-crystallized ligand molecule (N3) were drawn through ChemDraw 15.1 and saved in sdf format. All ligands were prepared into 3D structure by utilizing LigPrep module keeping 32 stereoisomer for each ligand as default. OPLS_2005 force field were utilized to generate all possible 3D-conformer of ligands [32].

2.3. Receptor grid generation

The active site of main protease was identified for binding of natural ligands and internal ligand of the receptor was chosen for grid box generation. The glide module of Maestro 2017-4 was utilized for receptor grid generation by selecting the atom of co-crystallized ligand of the protein (Fig. 1). The size of the grid box was centroid of internal ligand (N3) of the protein, and the docked ligands were similar in size to the N3 molecule. The atoms of protease were fixed within the default parameters of the radii of

Vander Waal's scaling factor of 1 Å along with partial charge cut-off of 0.25 Å using OPLS_2005 force field.

2.4. Virtual screening

Maestro program was utilized with default parameters in order to evaluate binding affinities of natural candidates within active site of main protease [31]. Receptor grid was produced by selecting atom of co-crystallized ligands. Docking score reflects binding affinity of ligands with the protein; therefore, docking score has been utilized to afford top hits [33]. Initially, HTVS (High Throughput Virtual Screening) was performed for retrieved molecules, eliminating most of the molecules against the main protease of COVID-19 based on their docking score. Molecules with considerable dock score were selected for SP (Standard Precision) & XP (Extra Precision) algorithm of docking, in order to get hit candidates. Top-ranked candidates docking pose were compared with reference molecule N3 (co-crystallized ligand).

Table 2

Pharmacokinetic parameters for top ranked candidates.

Top Ranked Candidates	MW	SASA	FOSA	Dipole	QPlogPw	QPlogS	QPlogHERG	QPlogKhsa	Rule of Five	ToxiM Score (<0.8, safe)
C00014803	1155.04	1650.33	375.99	13.46	56.44	-6.05	-8.03	-1.99	3	0.56
C00006660	612.54	865.81	181.94	1.94	39.99	-2.41	-6.31	-1.55	3	0.73
ANLT0001	758.64	1006.27	257.52	13.69	48.79	-1.83	-6.54	-2.23	3	0.68

ADME-T descriptors: MW, molecular weight; SASA, total solvent accessible surface area (SASA) in square angstroms using a probe with a 1.4 Å radius; FOSA, hydrophobic component of the SASA; Dipole, dipole moment; QPlogPw, predicted water/gas partition coefficient; QPlogS, predicted aqueous solubility; QPlogHERG, Predicted IC₅₀ value for blockage of HERG K⁺ channels; QPlogKhsa, prediction of binding to human serum albumin.

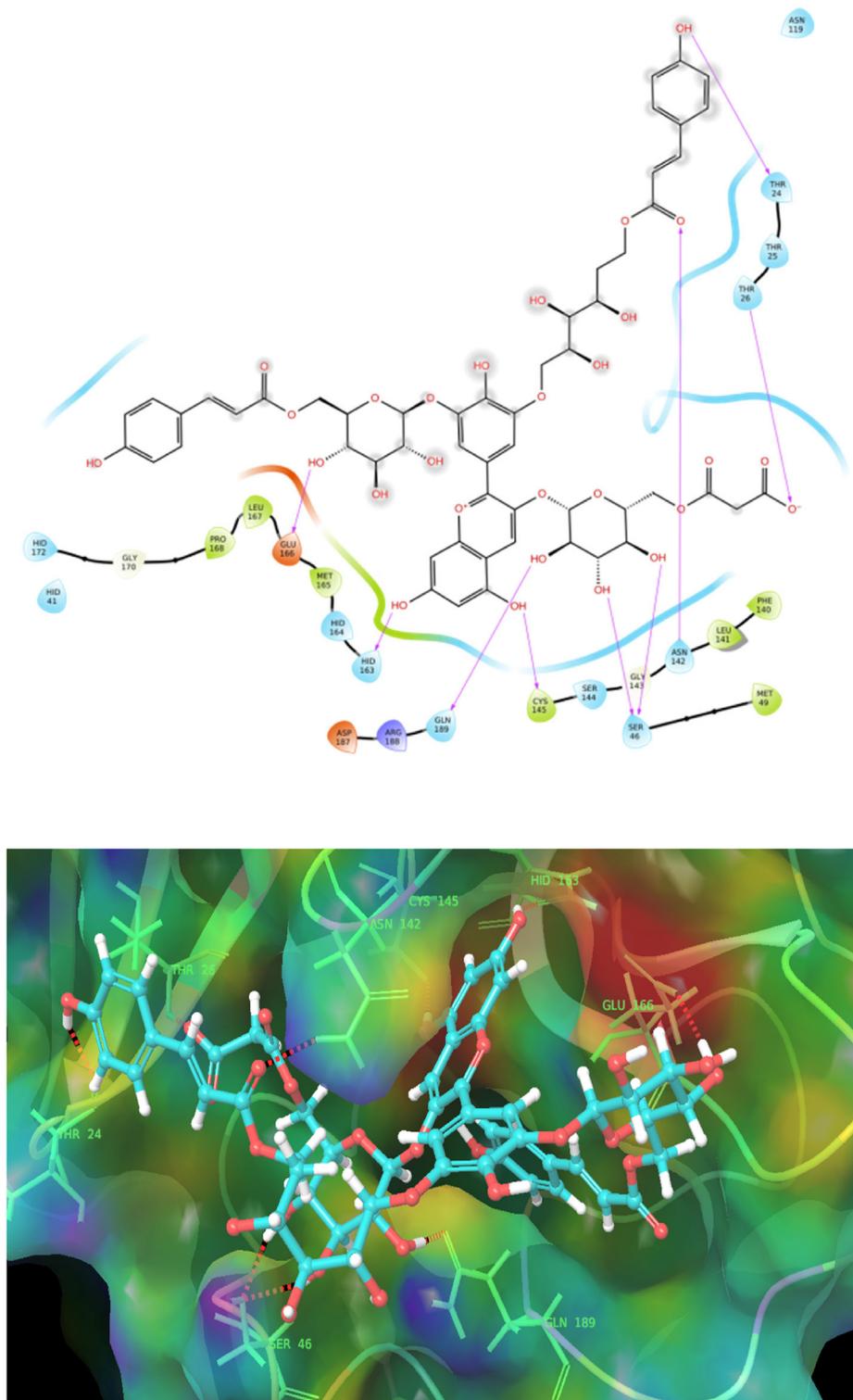


Fig. 3. 2D & 3D interaction diagram of docked C00014803 with main protease (6LU7).

2.5. Ligand-receptor binding mode visualization

In order to know dock pose and ligand interaction for apex molecule, Ligplot option of Maestro software was used [33]. Ligand interactions such as H-bonding, pi-pi interaction & hydrophobic interactions can be seen through this tool. These interactions are important criteria for any ligand candidates to be a lead molecule.

3-D interaction diagram for apex hits that were executed directly from Maestro suite workspace visualizer.

2.6. Pharmacokinetic study

Drug likeness parameters were estimated as stipulated in Qik-Prop tool [34]. ToxiM web server was utilized in order to predict

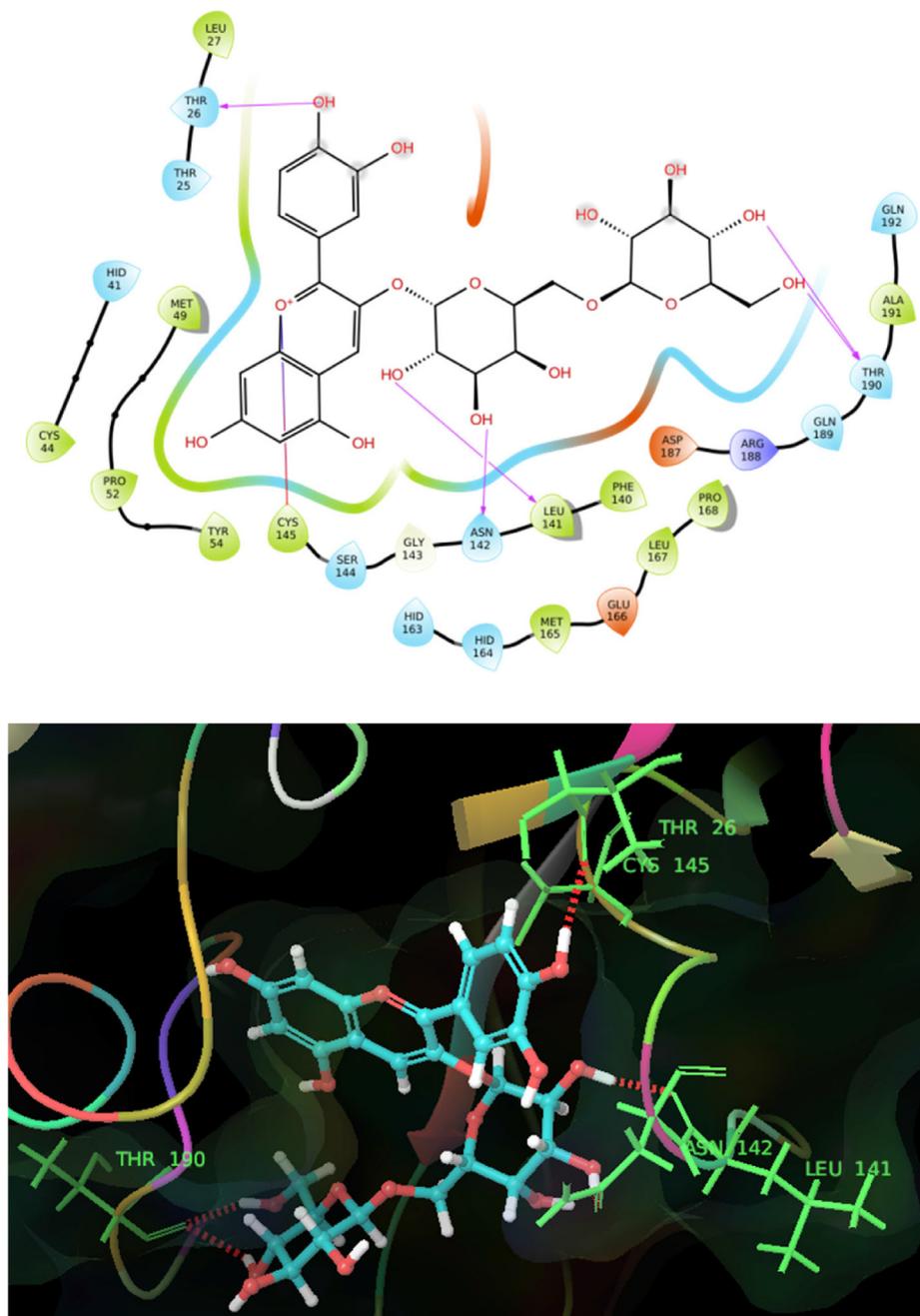


Fig. 4. 2D & 3D interaction diagram of docked C00006660 with main protease (6LU7).

toxicity of hit molecules [34,35] The ADME-T (Absorption, distribution, metabolism, excretion-Toxicity) profile is important aspect for a candidate to behave like drug molecule. Bioavailability and bioactivity greatly depends upon pharmacokinetic properties. Toxicity is another huge factor, therefore safety of hit candidates in human physiology do matter for further experimental validation.

2.7. Molecular dynamics simulation

Molecular dynamics (MD) study was performed for four protein-ligand complexes, including three-hit ligand and one reference molecule, i.e. N3, using GROMACS (GRONingen Machine for Chemical Simulation) 2018.1. Molecular dynamics was utilized to predict the stability of top hits with target protein under similar

to physiological condition. The CHARMM General Force Field (CgenFF) server was utilized to produce desirable ligand topology, while 'pdb2gmx' script was implied for protein topology preparation. Further, a complex of both topologies i.e. ligands and protein, were subjected under Charmm36-July-2017 force field to produce the energy minimized conformation of all the processed complexes [36]. After that, ligand-protein complexes were solvated in a cubic simulation box through one-point water model (SPC216) [37]. The particle mesh Ewald method was taken into account for specifying long-range electrostatic interactions. Steepest decent function was incorporated for the energy minimization of the complexes (50,000 steps) [38]. The system was equilibrated under NVT ensemble with a constant number of particles, volume and temperature for 2 ns, followed by NPT ensemble with a constant number of particles,

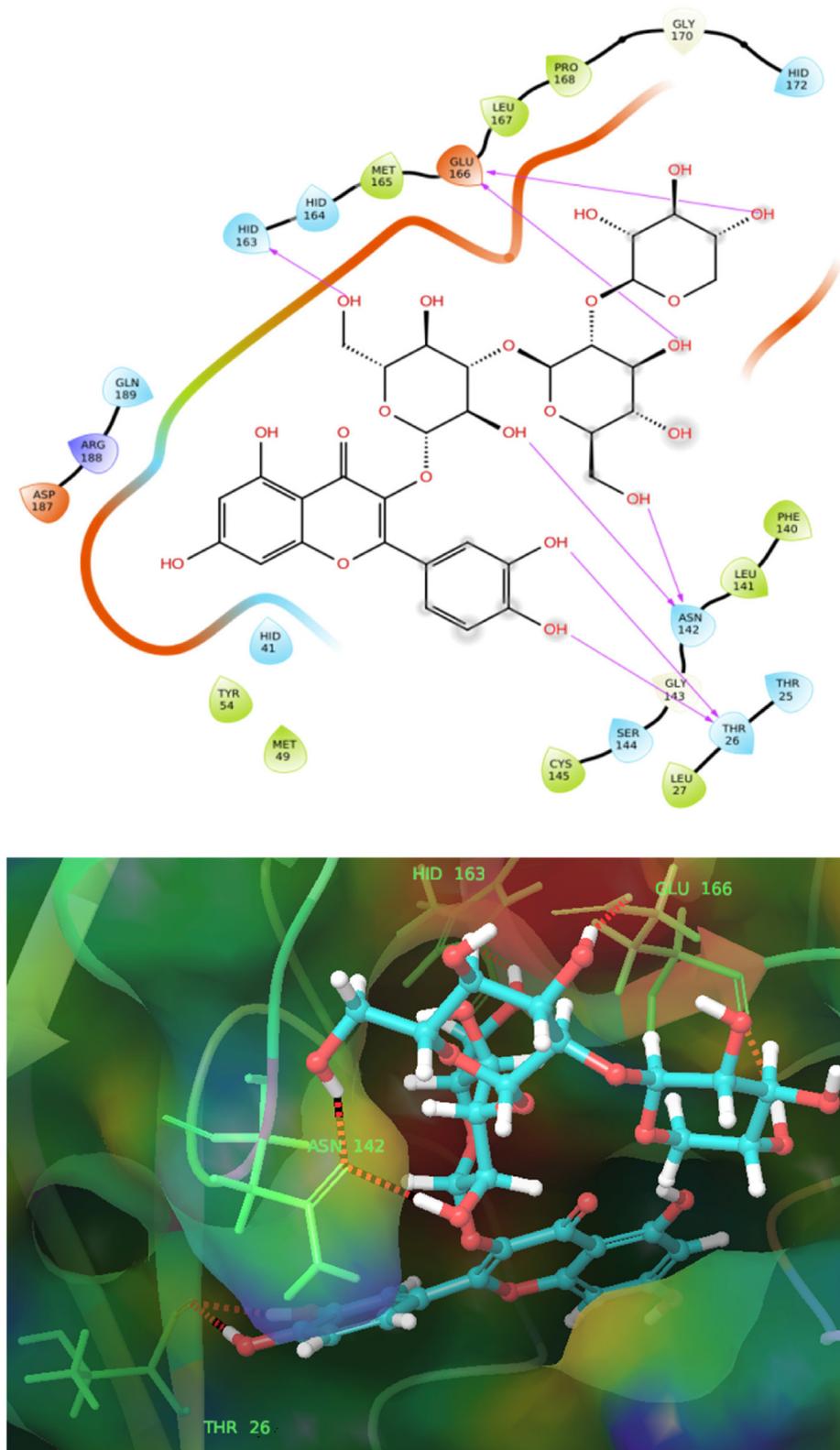


Fig. 5. 2D & 3D interaction diagram of docked ANLT0001 with main protease (6LU7).

pressure and temperature for 10 ns. In terms of geometry and solvent orientation, the well-equipped and solvated structures were subjected to further simulation. Sequentially, production phase was executed for all equilibrated complexes without any

restriction for 200 ns and system coordinates were saved after every 2 ps [39]. The trajectories generated by the simulation were inspected using in-built GROMACS scripts in the form of quantitative parameters such as the root mean square deviation (RMSD),

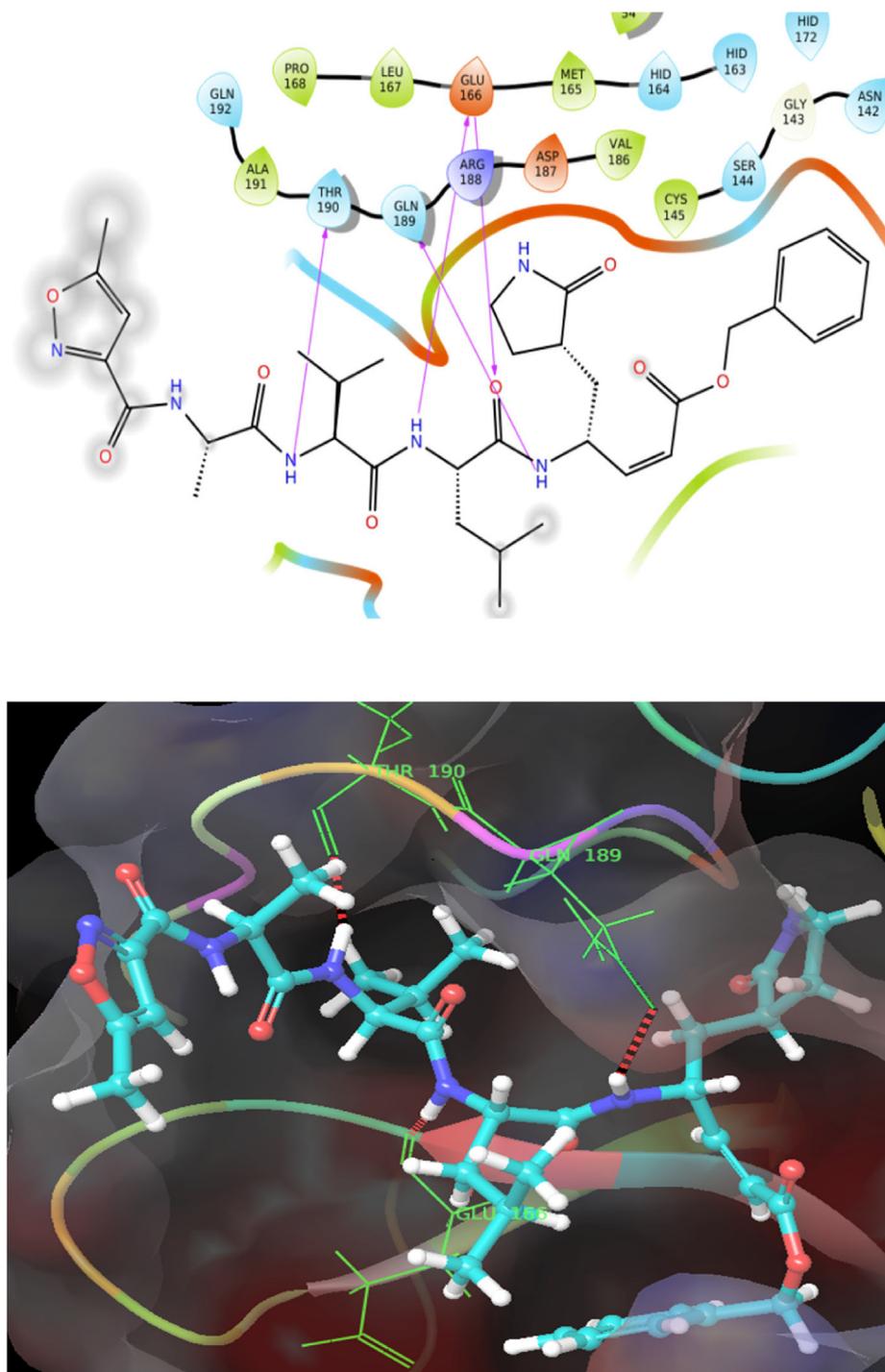


Fig. 6. 2D & 3D interaction diagram of docked internal reference (N3) with main protease (6LU7).

root mean square fluctuation (RMSF), gyration radius (Rg) and hydrogen bond contact profile [40].

2.8. MM-PBSA study

The free energy calculation provides a quantitative estimation of interactions between protein and ligand that help to understand the stability of that protein–ligand complex. The binding free energy including the free solvation energy (polar and nonpolar solvation energies) and potential energy (electrostatic and Vander

Waal's interactions) of each protein–ligand complexes was calculated by the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) method. The MD trajectories were processed before doing MM-PBSA calculations for last 10 ns. The MM-PBSA binding free energy calculation was done with 'g_mmpbsa' script [41].

Maestro software 2017-4 [31] were employed for HTVS, SP and XP mode of docking algorithm. Molecular dynamics simulations and mm-pbsa study were done on the ubuntu 18.04 LTS supported by 256 GB RAM (64 processor) & hadoop cluster machine assisted by 1 TB RAM (224 processor) by employing GROMACS 2018.1

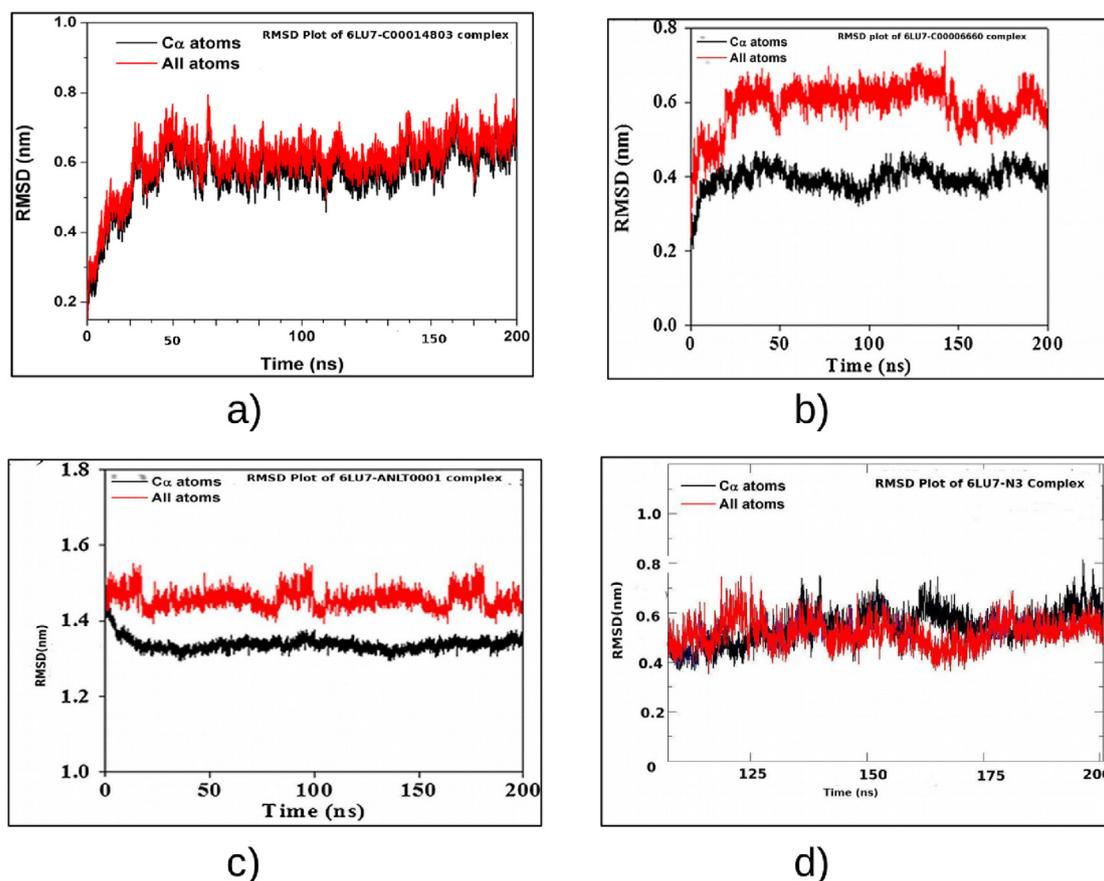


Fig. 7. Molecular dynamics (RMSD) plots for top ranked ligands a) C00014803; b) C0006660; c) ANLT0001 d) Internal reference (N3) with main protease (6LU7).

version. Pharmacokinetic parameters i.e. ADME-T were estimated with the aid of QikProp module of Schrodinger suit 2017-3 and ToxiM web server. Fujitsu brand monitor was utilized for docking & ADME (QikProp) studies with a Intel® Xeon(R) CPU E5-2620 v4(2.1 GHz; 48 GB RAM) 64-bit equipped Ubuntu16.04 LTS Linux operator.

3. Results

3.1. Molecular docking and ADME-T studies

Availability of nCoV main protease crystal structure was the main driving force for screening of natural molecule databases toward finding lead molecule against COVID-19. Molecular docking represent comparable affinity parameters of tested chemical entities with validated inhibitor molecule or drugs toward target receptor [32]. Three mode of docking i.e. HTVS, SP & XP were performed to identified lead molecules from diverse natural molecule taken from six different databases, NPASS (30,926), ChemDiv (3937), Analyticon (6833), ChEBI (2042), TIPdb (8,856) & Life Chemicals (936) against M^{Pro} of COVID-19. The LigPrep program of Maestro software produced 2,88,708 conformer/isomer out of 53,530 natural molecule. Ranking of top molecules was relied on their binding affinity to main protease & docking score. Firstly, top 50 molecules were selected from HTVS mode of docking which were then subjected to SP algorithm of docking. After that, top 10 molecules were selected (supporting information) for XP algorithm of docking. Finally, top three candidates i.e. C00014803 (TIPdb), C00006660 (TIPdb) & ANLT0001 (Analyticon) were found to have lowest docking score and strong binding character as compare to reference inhibitor (Table 1 and Fig. 2).

3.2. Molecular dynamics and ADME-T

Molecular docking allows us to determine the partial static bonding stability of the complex, but to determine full descriptors responsible for dynamic stability of complex in its biological environments; molecular dynamic simulation is required. The static picture of the complexes does not provide full descriptions of all the factors responsible for providing stability to the protein-ligand complexes [42]. RMSD is an important parameter for measuring the dynamic stability and equilibration of protein-ligand complex. To validate the docking study, we further analyzed basic MD parameters like RMSD and RMSF of C- α atom of complexed protein and fluctuation range of protein residues, respectively. Three top scored natural candidates based on virtual screening were subjected for simulation by using GROMACS-2018.1 for 200 ns. Different simulation parameters such as RMSD, RMSF, etc. resulting from GROMACS for protein-ligands complexes have been shown in Figs. 7–9.

Additionally, test candidates were also investigated for H-bonding interaction (% occupancy) against amino acid of protease enzyme for the last 50ns time-period trajectory (Table 3). Another important simulation parameter viz. Radius of gyration of protease were calculated to measure its consistency [43]. If a protein is efficiently folded, a relatively steady Rg value will likely be retained and vice-versa. Rg had been calculated for the whole protein structure. We observed that Rg value of C00014803 and N3 was almost identical and also greater than Rg value of C00006660 and ANLT0001. It means interaction of C00014803 and reference molecule i.e. N3 with protease made complex structure more compact and stable as shown in Fig. 9.

Further, these molecules were subjected for ADME-T and molecular dynamics study in order to validate its drug likeness

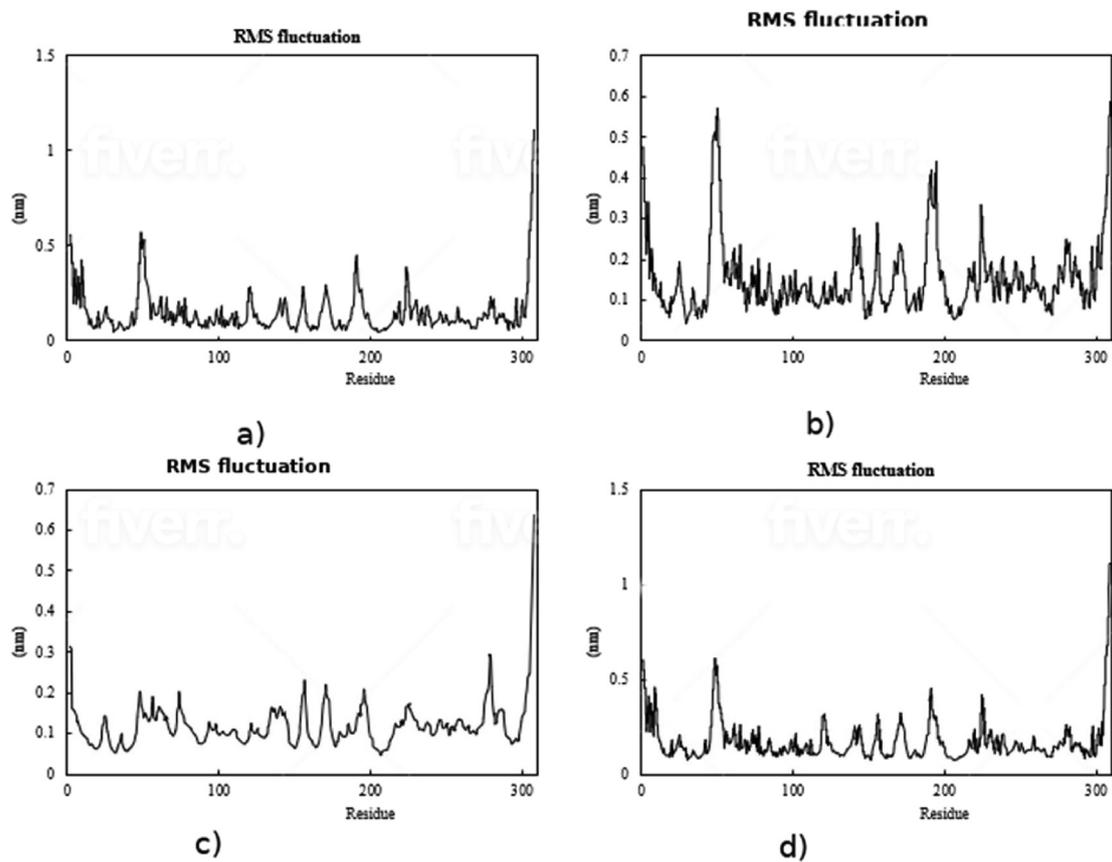


Fig. 8. Molecular dynamics (RMSF) plots for top ranked ligands a) C00014803; b) C00006660; c) ANLT0001 d) Internal reference (N3) with main protease (6LU7).

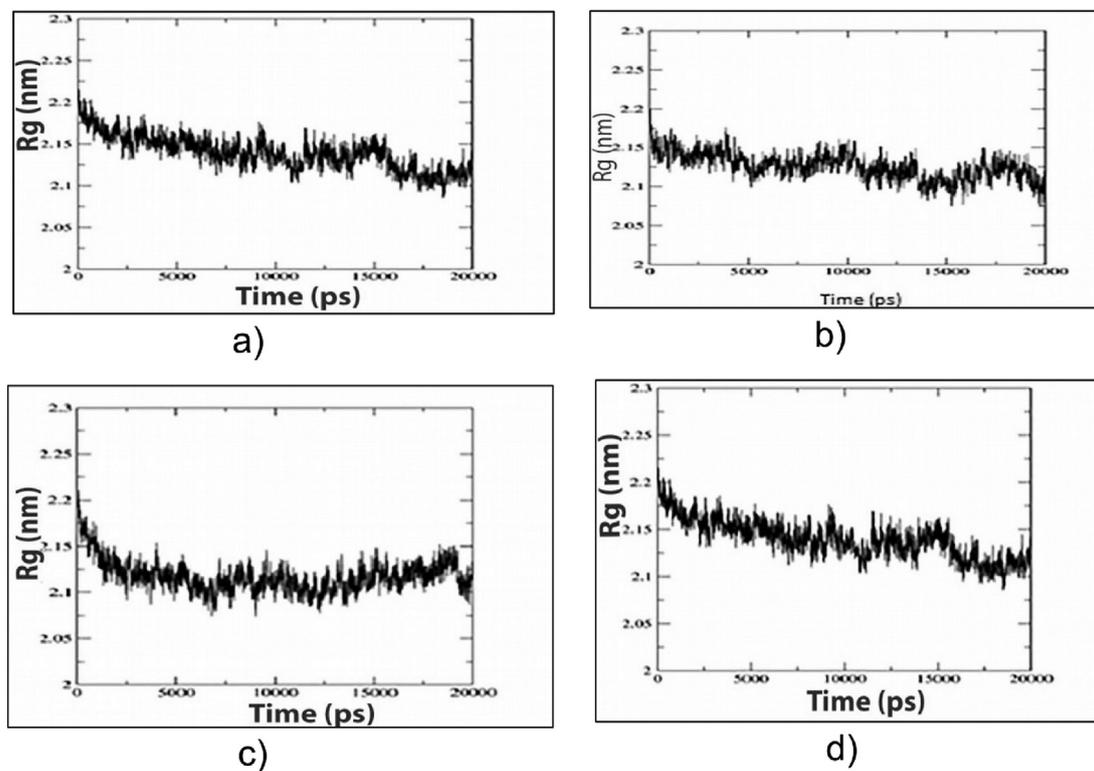


Fig. 9. Molecular dynamics (radius of gyration i.e. Rg) plots for top ranked ligands (A) C00014803; (B) C00006660; (C) ANLT0001; and (D) reference ligand N3 with main protease.

Table 3

H-bonding contact of molecular dynamics with significant percentage occupancy of test and reference compounds against residues of nCOV protease during last 50 ns simulation.

Compounds	Participating amino acid residues (% occupancies)
C00014803	PRO108 (38), GLY109 (31), THR111(26), VAL125 (23) ARG131(53), GLN127(16.6), ALA129 (6.7)
C00006660	PRO132(38), ASN133(15), ILE136(15), LYS137(15) SER139(6), ASN151(6), THR169(3)
ANLT0001	THR111(47), VAL125(23), GLN127(16.6), CYS128(9) ALA129(7), PRO108(38), LYS137(15)
N3 (Internal Reference)	ARG131(53), ASN133(45), GLN127(16.6), THR169(28) VAL202(25), TRP207(23), MET235(13), GLU240(11) PRO241 (36)

Table 4

Relative binding energy profile of top hits along with reference molecule.

Sr. no	Compounds	Vander Waal's energy	Electrostatic energy	Polar solvation energy	SASA energy	Binding energy
1	C00014803	-285.12	-9.83	175.33	-24.01	-143.63
2	C00006660	-298.23	-12.91	193.85	-24.04	-141.33
3	ANLT0001	-190.60	-7.22	107.89	-12.63	-102.56
4	N3	-362.22	-17.84	165.88	-21.66	-235.84

properties. ADME-T study is an important criteria for knowing whether a molecule can act as drug or not. Here, ADME-T parameter for top three hits found to be approximately in safe range. Toxicity behaviour of these hits were evaluated through ToxIM web server, and all parameters found in safe range (Table 2) [35].

3.3. Relative free binding energy

The binding affinity of protein-ligand complexes is also defined by the binding free energy measured using g_MM-PBSA software. The results of the MM-PBSA studies are shown in the Table 4. For the measurement of binding free energy, only 10ns MD trajectories were used. The binding energy calculations in this method were calculated by adding all type of energy i.e. Vander Waal's, electrostatic, polar solvation and SASA (Solvent accessible surface area) energy.

The binding affinities of all three expected hits were outstanding, but they were lower than the reference compound N3 (-235.84 kJ/mol). C00014803 has the highest binding energy (-143.63 kJ/mol), while C00006660 and ANLT0001 possess low binding affinities i.e. -141.33 kJ/mol and -102.56 kJ/mol respectively). Furthermore, the electrostatic interactions, non-polar solvation energy, and van Der Waals interactions all contributed negatively to the overall interaction energy, while the polar solvation energy enriched the binding energy positively. Based on these observations, the best compound-complexes among the four possible inhibitors are N3, C00014803 and C00006660. These findings corroborate our previous MD simulation study.

4. Discussion

Computational tool is an emerging technique in drug design and discovery process [44]. In order to search a new antiviral drug protease enzyme is most validated target, which is responsible for producing protein involved in viral replication [45]. In our study we have subjected thousands of natural molecules against main protease of COVID-19, in order to identify lead molecules, which can act as antiviral agent. Docking analysis afforded top three hits based on their binding affinity toward viral protease. First hit i.e. C00014803 has docking score of -12.76 kcal/mol possess nine hydrogen bonding with THR24, THR26, SER46 (2 H-bonding),

ASN142, CYS145, HID163, GLU166, GLN189 with no hydrophobic interaction (Table 1 and Fig. 3). Second hit molecule i.e. C00006660 has docking score of -11.59 kcal/mol possess six H-bonding with THR26, LEU141, ASN142, CYS145, THR190 (2 H-bonding) (Table 1 and Fig. 4). Third ranked molecule i.e. ANLT0001 has docking score of -11.09 kcal/mol possess seven H-bonding with THR26 (2 H-bonding), ASN142 (2 H-bonding), HID163, GLU166 (2 H-bonding) (Table 1 and Fig. 5).

All top three hit molecules share identical H-bonding with THR26 & ASN142 residues. Moreover, top two hit molecules i.e. C00014803 and C00006660 have shown identical H-bonding with CYS145, while HID163 and GLU166 residue have shown common H-bonding with top first (C00014803) and third hit candidate (C00006660). Reference molecule i.e. N3 molecule found to have less binding affinity with target protein reflected by low docking score (-8.178 kcal/mol) as compare to top hits. Reference molecule has shown four H-bonding contributed by GLU166 (2 H-bonding), GLN189 and THR190 residue of protein (Table 1 and Fig. 6). All top three hit molecules were found to have similar pose in active binding pocket of protein as reference molecule N3 (Supporting information, Figure S3). The three top-hits of docked protein-ligand complexes were studied further for understanding structural details, dynamic behaviour and stability by using MD simulation. When we simulated the protein with each ligands and compared against without ligand individually for 200 ns, we observed their rmsd graph and noticed the variation associated with them. The fluctuation of C α -backbone atom of simple protein shows less fluctuation (initial stage). RMSD initially swung slight for 4–12 ns for protein bound with ligand C00014803 followed by minute disturbance seen at 60 ns and eventually resulted in significant uniformity for the last 50 ns, suggesting stability of the complex.

In contrast, as compared to C00014803 and reference N3 molecules, C00006660 and ANLT0001 showed more RMSD fluctuations for the entire MD run. RMSF was also evaluated to measure the atomic mobility of structural fluctuating backbone atoms over a span of 200 ns. Overall, it was noted that, relative to other top docked hits, molecule specifically C00014803 docked protein residues showed less fluctuation as compared to remaining residues of protein. Initially, without any checked ligands, it shown alike reasonable spike generation at 50–90 and 150–200 residue index numbers as mentioned in the 6LU7 (backbone). The RMSF captured,

for each atom and alteration about its average position, which gives insight into the pliability in regions of the protease. Specifically, by looking at the RMSF plots for protease with 306 amino acids and three possible drug candidates and one reference molecule, RMSF was calculated; the values confirmed that the residues of the binding site showed less fluctuation. For first hit (C00014803), second hit (C00006660), third hit (ANLT0001) and N3 (reference) the average RMSF values were 0.13, 0.14, 0.11 and 0.03 nm respectively. The active site residue PRO108 and GLN127 reflects minute variation in almost all the structures. This residues are responsible for catalytic action of protein against all computationally simulated ligands. H-bonding contact were also evaluated for last 50 ns. We observed from Table 3 that there was high occurrence of some residues like PRO which play important role in molecular recognition and structurally provide more conformational rigidity. Along with PRO, VAL and GLN were other two most occurred amino acids which also provide structural stability. Radius of gyration were checked in order to see consistency of protein ligand complex, which have shown top hit i.e. C00014803 possess higher Rg value indicate more compact interaction with protein. Relative binding energy analysis has shown first hit (C00014803) possess higher binding affinity as compare to other two hits, but all hits have shown lower binding free energy as compare to reference molecule (Table 4). Pharmacokinetic analysis for lead candidates were found to be in acceptable range with negligible toxicity profile, which could increase its drug likeness properties.

5. Conclusion and future perspective

In current study, a library of natural compounds retrieved from natural product databases was screened against nCoV-protease in order to propose possible therapeutic candidates for current crisis i.e. COVID-19. Our studies provide detailed information about the inhibitory ability of top-ranked three natural candidates. Molecular docking algorithms against main protease afforded hit molecules, which were further subjected under different pharmacokinetic and toxicity parameters, in order to check their drug likeness behavior. ADME-T studies have shown acceptable pharmacokinetic descriptor with safe profile. Further, these top hits were subjected to MD simulation study in order to validate stability behavior of top natural candidates against COVID-19 protease. MD simulation studies confirmed the stability of screened hits with protease of nCoV, which were evaluated on the basis of the RMSD, RMSF and Rg trajectories of dynamic simulation for 200 ns. Hydrogen bonding contact were further evaluated to confirm their stability. Current studies afforded top 3 candidates in form of C00014803, C00006660, ANLT0001 (natural candidates) which possess potential against COVID-19 main protease. These top three hits also possess decent free binding energy with main protease. The current findings conclude that the design of novel class of COVID-19 protease inhibitors and the need for more experimental validation can be considered for these three top hits.

Author contribution

S.S.G. and A.K. performed molecular docking, molecular dynamics and pharmacokinetic analysis. U.S. and R.S. supervised and analyzed the data and wrote the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the Director, CSIR-IHBT, Palampur for providing necessary facility. This work is supported by CSIR (MLP0154/MLP0203). The authors are thankful to Dr. Prajwal Nandekar for interactive discussion and suggestions. Mr. Shiv Shankar Gupta acknowledges CSIR, New Delhi for Junior Research Fellowship and Dr. Ashwani Kumar acknowledges DBT, New Delhi for Research Associate Fellowship. CSIR-IHBT communication no. for this publication is 4770.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmngm.2021.107916>.

References

- [1] E.C. Smith, M.R. Denison, Coronaviruses as DNA Wannabes: a new model for the regulation of RNA virus replication fidelity, *PLoS Pathog.* 9 (2013) 1–4, <https://doi.org/10.1371/journal.ppat.1003760>.
- [2] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, F.C.-C. Discovery, U, Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2, *Cell Discov.* 6 (2020) 1–18, <https://doi.org/10.1038/s41421-020-0153-3>, 2020.
- [3] J. Chen, Pathogenicity and transmissibility of 2019-nCoV—a quick overview and comparison with other emerging viruses, *Microb. Infect.* 22 (2020) 69–71, <https://doi.org/10.1016/j.micinf.2020.01.004>.
- [4] WHO, Weekly epidemiological update on COVID-19 – 16 March 2021. <https://www.who.int/publications/m/item/weekly-epidemiological-update-16-march-2021>.
- [5] J.O. Wertheim, D.K.W. Chu, J.S.M. Peiris, S.L. Kosakovsky, L.L.M. Poon, A case for the ancient origin of coronaviruses, *Am. Soc. Microbiol.* (2013) 7039–7045, <https://doi.org/10.1128/JVI.03273-12>.
- [6] W.H. Organization, Summary of Probable SARS Cases with Onset of Illness from 1 November 2002 to 31 July 2003, 2003. *CiNii*.
- [7] L. Enjuanes, F. Almazan, I. Sola, S. Zuñiga, Biochemical aspects of coronavirus replication and virus-host interaction, *Annu. Rev. Microbiol.* 60 (2006) 211–230, <https://doi.org/10.1146/annurev.micro.60.080805.142157>.
- [8] Z. Said, K.A.-V. Replication, U, Antiviral replication agents, *Viral Repl.* (2013) 127–144, 2013.
- [9] A. Salonen, T. Ahola, L. Kaariainen, Viral RNA replication in association with cellular membranes, membrane trafficking in viral replication, *Curr. Top. Microbiol. Immunol.* 285 (2004) 139–173, https://doi.org/10.1007/3-540-26764-6_5.
- [10] T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S.H. Jung, An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy, *J. Med. Chem.* 59 (2016) 6595–6628, <https://doi.org/10.1021/acs.jmedchem.5b01461>.
- [11] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (2020) 289–293, <https://doi.org/10.1038/s41586-020-2223-y>.
- [12] T. Sheahan, A. Sims, R.G.-S. translational, U, Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses, *Sci. Transl. Med.* 9 (2017), <https://doi.org/10.1126/scitranslmed.aal3653>, 2017.
- [13] Y. Cao, Q. Deng, S. Dai, Remdesivir for severe acute respiratory syndrome coronavirus 2 causing COVID-19: an evaluation of the evidence, *Trav. Med. Infect. Dis.* 35 (2020), 101647, <https://doi.org/10.1016/j.tmaid.2020.101647> (2020).
- [14] J.A. Beutler, Natural products as a foundation for drug discovery, *Curr. Protoc. Pharmacol.* 86 (2019) e67, <https://doi.org/10.1002/cpph.67>.
- [15] J. Cinatl, B. Morgenstern, G. Bauer, P. Chandra, H. Rabenau, H.W. Doerr, Glycyrrhizin, an active component of liquorice roots, and replication of SARS-associated coronavirus, *Lancet* 361 (2003) 2045–2046, [https://doi.org/10.1016/S0140-6736\(03\)13615-X](https://doi.org/10.1016/S0140-6736(03)13615-X).
- [16] L. Yi, Z. Li, Kehu Yuan, Xiuxia Qu, J. Chen, G. Wang, H. Zhang, H. Luo, L. Zhu, P. Jiang, L. Chen, Y. Shen, M. Luo, G. Zuo, J. Hu, D. Duan, Y. Nie, X. Shi, W. Wang, Y. Han, T. Li, Y. Liu, M. Ding, H. Deng, X. Xu, Small molecules blocking the entry of severe acute respiratory syndrome coronavirus into host cells, *J. Virol.* 78 (2004) 11334–11339, <https://doi.org/10.1128/JVI.78.20.11334-11339.2004>.
- [17] S. Li, C. Chen, H. Zhang, H. Guo, H. Wang, L.W.-A. Research, U, Identification of natural compounds with antiviral activities against SARS-associated coronavirus, *Antivir. Res.* 67 (2005) 18–23, <https://doi.org/10.1016/j.antiviral.2005.02.007>, 2005.
- [18] C. Lin, F. Tsai, C. Tsai, C. Lai, L. Wan, T.H.-A. Research, U, Anti-SARS coronavirus 3C-like protease effects of Isatis indigotica root and plant-derived phenolic compounds, *Antivir. Res.* 68 (2005) 36–42, <https://doi.org/10.1016/j.antiviral.2005.07.002>, 2005.

- [19] C.-C. Wen, Y.-H. Kuo, J.-T. Jan, P.-H. Liang, S.-Y. Wang, H.-G. Liu, C.-K. Lee, S.-T. Chang, C.-J. Kuo, S.-S. Lee, C.-C. Hou, P.-W. Hsiao, S.-C. Chien, L.-F. Shyur, N.-S. Yang, Specific plant terpenoids and lignoids possess potent antiviral activities against severe acute respiratory syndrome coronavirus, *J. Med. Chem.* 50 (2007) 4087–4095, <https://doi.org/10.1021/jm070295s>.
- [20] M. Yu, J. Lee, J. Lee, Y. Kim, Y. Chin, J.J. B. U. medicinal, Identification of myricetin and scutellarein as novel chemical inhibitors of the SARS coronavirus helicase, *nsP13, Bioorg. Med. Chem. Lett.* 22 (2012) 4049–4054, <https://doi.org/10.1016/j.bmcl.2012.04.081>, 2012.
- [21] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science* 367 (2020) 1260–1263, <https://doi.org/10.1126/science.abb2507>.
- [22] Z. Liu, X. B. Zhang, Z. Jin, H. Yang, Rao, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (7811) (2020) 289–293, <https://doi.org/10.1038/s41586-020-2223-y>.
- [23] J. Wang, Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study, *J. Chem. Inf. Model.* 60 (2020) 3277–3286, <https://doi.org/10.1021/acs.jcim.0c00179>.
- [24] A. Contini, Virtual Screening of an FDA Approved Drugs Database on Two COVID-19 Coronavirus Proteins, 2020.
- [25] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharm. Sin. B.* 10 (2020) 766–788, <https://doi.org/10.1016/j.apsb.2020.02.008>, 2020.
- [26] M.A. Alamri, A. Altharawi, A.B. Alabbas, M.A. Alossaimi, S.M. Alqahtani, Structure-based virtual screening and molecular dynamics of phytochemicals derived from Saudi medicinal plants to identify potential COVID-19 therapeutics, *Arab. J. Chem.* 13 (2020) 7224–7234.
- [27] W.B. Cardoso, S.A. Mendanha, Molecular dynamics simulation of docking structures of SARS-CoV-2 main protease and HIV protease inhibitors, *J. Mol. Struct.* 1225 (2021), 129143.
- [28] R. Yadav, R.D. Parihar, U. Dhiman, P. Dhamija, S. Kumar, Docking of fda approved drugs targeting nsp-16, n-protein and main protease of sars-cov-2 as dual inhibitors, *Biointerf. Res. Appl. Chem.* 11 (2020), <https://doi.org/10.33263/BRIAC113.98489861>.
- [29] V.K. Bhardwaj, R. Singh, J. Sharma, V. Rajendran, R. Purohit, S. Kumar, Identification of bioactive molecules from tea plant as SARS-CoV-2 main protease inhibitors, *J. Biomol. Struct. Dyn.* (2020) 1–10, <https://doi.org/10.1080/07391102.2020.1766572>.
- [30] N.M. Mohamed, E.M. Ali, A.M. AboulMagd, Ligand-based design, molecular dynamics and ADMET studies of suggested SARS-CoV-2 M pro inhibitors, *RSC Adv.* 11 (2021) 4523–4538, <https://doi.org/10.1039/D0RA10141A>.
- [31] Schrodinger, Maestro, (2017–4).
- [32] M. Gupta, R. Sharma, M. Singh, A. Kumar, Docking techniques in pharmacology: how much promising? *Comput. Biol. Chem.* 76 (2018) 210–217, <https://doi.org/10.1016/j.compbiolchem.2018.06.005>.
- [33] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem.* 47 (2004) 1739–1749, <https://doi.org/10.1021/jm0306430>.
- [34] L. Ioakimidis, L. Thoukydidis, A. Mirza, S. Naeem, J. Reynisson, Benchmarking the reliability of QikProp. correlation between experimental and predicted values, *QSAR Comb. Sci.* 27 (2008) 445–456, <https://doi.org/10.1002/qsar.200730051>.
- [35] A.K. Sharma, G.N. Srivastava, A. Roy, V.K. Sharma, ToxiM: a toxicity prediction tool for small molecules developed using machine learning and chemoinformatics approaches, *Front. Pharmacol.* 8 (2017), <https://doi.org/10.3389/fphar.2017.00880>.
- [36] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. Mackerell, CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, *J. Comput. Chem.* 31 (2010) 671–690, <https://doi.org/10.1002/jcc.21367>.
- [37] P. Mark, L. Nilsson, Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K, *J. Phys. Chem.* 105 (2001) 9954–9960, <https://doi.org/10.1021/jp003020w>.
- [38] M. Ganahl, J. Rincon, G. Vidal, Continuous matrix product states for quantum fields: an energy minimization algorithm, *Phys. Rev. Lett.* 118 (2017), 220402, <https://doi.org/10.1103/PhysRevLett.118.220402>.
- [39] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics* 29 (2013) 845–854, <https://doi.org/10.1093/bioinformatics/btt055>, 2013.
- [40] T. Makarewicz, R. Kaźmierkiewicz, Molecular dynamics simulation by GROMACS using GUI plugin for PyMOL, *J. Chem. Inf. Model.* 53 (2013) 1229–1234, <https://doi.org/10.1021/ci400071x>.
- [41] R. Kumari, R. Kumar, g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations, *J. Chem. Inf. Model.* 54 (2014) 1951–1962, <https://doi.org/10.1021/ci500020m>.
- [42] K. Peele, P. Chandrasai, T.S. I. M. U, Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: a computational study, *Inf. Med.* 19 (2020) (2020), 100345, <https://doi.org/10.1016/j.imu.2020.100345>.
- [43] M. Lobanov, N. Bogatyreva, O.G.-M. Biology, U, Radius of gyration as an indicator of protein structure compactness, *Mol. Biol.* 42 (2008) 623–628, <https://doi.org/10.1134/S0026893308040195>, 2008.
- [44] S.P. Leelananda, S. Lindert, Computational methods in drug discovery, *Beilstein J. Org. Chem.* 12 (2016) 2694–2718, <https://doi.org/10.3762/bjoc.12.267>.
- [45] P.G. Par, A1, Three decades of the hepatitis C virus from the discovery to the potential global elimination: the success of translational researches, *Orv. Hetil.* 159 (2018) 455–465, <https://doi.org/10.1556/650.2018.30997>.