

Computational guided identification of novel potent inhibitors of NTD-N-Protein of SARS-CoV-2

Poonam Dhankhar[#], Vikram Dalal[#], Vishakha Singh[#], Shailly Tomar, and Pravindra Kumar*

Department of Biotechnology, Indian Institute of Technology Roorkee, India – 247667

*:To whom correspondence should be addressed: Prof. Pravindra Kumar, Department of Biotechnology, Indian Institute of Technology Roorkee, India – 247667, Telephone: +91-1332-286286, E-mail: pravinmcu@gmail.com

[#]: Authors contributed equally.

Abstract

The Coronavirus Disease 2019 (COVID-19), caused by the SARS-CoV-2 virus has raised severe health problems in china and across the world as well. CoVs encode the nucleocapsid protein (N-protein), an essential RNA-binding protein that performs different roles throughout the virus replication cycle and forms the ribonucleoprotein complex with viral RNA using the N-terminal domain (NTD) of N-protein. Recent studies have shown that NTD-N-protein is a legitimate target for the development of antiviral drugs against human CoVs. Owing to the importance of NTD, the present study focuses on targeting the NTD-N-protein from SARS-CoV-2 to identify the potential compounds. The pharmacophore model has been developed based on the guanosine monophosphate (GMP), a RNA substrate and further pharmacophore-based virtual screening was performed against ZINC database. The screened compounds were filtered by analysing the *in silico* ADMET properties and drug-like properties. The pharmacokinetically screened compounds (ZINC000257324845, ZINC000005169973, and ZINC000009913056) were further scrutinized through computational approaches including molecular docking and molecular dynamics simulations and revealed that these compounds exhibited good binding affinity as compared to GMP and provide stability to their respective complex with the NTD. Our findings could disrupt the binding of viral RNA to NTD, which may inhibit the essential functions of NTD. These findings may further provide an impetus to develop the novel and potential inhibitor against SARS-CoV-2.

Keywords: Coronavirus Disease 2019 (COVID-19); N-terminal domain (NTD); guanosine monophosphate (GMP); Pharmacophore modeling; Molecular Dynamics Simulation.

1. Introduction

Coronaviruses (CoVs) belong to the family of *Coronaviridae* and the order of *Nidovirales*, which include non-segmented, enveloped, single-stranded, positive-sense RNA viruses and cause various infectious diseases in vertebrates, and humans (1-3). According to the latest International Committee of Taxonomy of Viruses (ICTV), CoVs are classified into four genera, namely *Alphacoronaviruses*, *Betacoronaviruses*, *Gammacoronaviruses*, and *Deltacoronaviruses* (4). *Alphacoronaviruses* and *Betacoronaviruses* infect the humans and mammals mainly with the gastrointestinal, respiratory, and central nervous system, while the *gammacoronaviruses* and *deltacoronaviruses* mainly infect birds (2,3,5,6). Some members of coronaviruses such as human *alphacoronaviruses*, 229E (HCoV-229E) and NL63 (HCoV-NL63), and human *betacoronaviruses*, OC43 (HCoV-OC43) and HKU1 (HCoV-HKU1), are commonly found worldwide and cause mild upper respiratory diseases in most individuals, although some of them can suffer more serious illness (7-10). Nevertheless, the other two *betacoronaviruses*, including severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle-East respiratory syndrome coronavirus (MERS-CoV) have emerged in 2002-2003 and 2012, respectively, and both are highly pathogenic to humans (11-13). Both are of zoonotic origins, and due to the lack of effective treatment, they triggered a global epidemics with a high mortality rate (9.6% for SARS-CoV and 34.4% for MERS-CoV) by causing a severe respiratory diseases (12-14).

Recently in December 2019, a pneumonia outbreak of unknown cause emerged in Wuhan, Hubei province in China with clinical symptoms similar to viral pneumonia (15). On the deep analysis of sequencing results from the lower respiratory tract samples, a novel coronavirus named as 2019 novel coronavirus (2019-nCoV/ SARS-CoV-2) has been identified as a causing pathogen for the coronavirus disease 2019 (abbreviated as COVID-19) (15-17). 2019-nCoV belongs to the *betacoronavirus* and closely related more to SARS-CoV than MERS-CoV (16,18). As of May 7, 2020, more than 3,595,662 cases with 247,652 deaths of SARS-CoV2 were confirmed across the world. Currently, COVID-19 is spreading like wildfire, and there are no effective medications for this new virus. Thus, the development of effective treatment and vaccine are urgently needed to combat the disease.

HCoVs genome consists of open reading frames for five major proteins, which plays a significant role in the virus structure assembly and viral replication, including replicase complex (ORF1ab), spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins

(19). Among these, N-protein is a major structural protein as it produces the ribonucleoprotein complex by the binding of viral RNA genome (20). It plays an essential role in the regulation of viral RNA synthesis and also has functional importance in fundamental aspects of the CoV life cycle, such as encapsidation and replication of virus genomes (21). It also has the ability to regulate the cellular processes during viral pathogenesis, including actin reorganization, host cell cycle progression, and apoptosis (22,23). Moreover, N-protein is an early diagnostic marker and a key antigen for the diagnosis and detection of CoV (24,25). N-protein of CoV has been widely used as an antiviral drug target for CoV. Besides these roles, there is less variability in the viral N- gene sequence, and thus it is a genetically stable protein, an essential prerequisite of an effective drug target candidate (26). Studies reported that N-protein from CoVs and other viruses were used as a drug target to identify the potent compound (27-30).

Coronavirus N-proteins are organized into three domains, namely an N-terminal RNA-binding domain (NTD), a poorly structured central Ser/Arg (SR)-rich linker, and a C-terminal dimerization domain (CTD) (31-33). Previous studies have determined the crystal structures of N-NTDs from various viruses such as SARS, HCoV-OC43, infectious bronchitis virus (IBV), and mouse hepatitis viruses (MHV) (33-37). Furthermore, complex crystal structures of HCoV-OC43 N-NTD provide structural clues about the binding of ribonucleoside 5'-monophosphates, including guanosine monophosphate (GMP), uridine monophosphate (UMP), cytidine monophosphate (CMP), and adenosine triphosphate (AMP) to NTD of N-protein (28). Specific molecules (PJ34, H3) were designed by using this knowledge against NTD of HCoV-OC43 N-protein, which were verified by the *in-vitro* experiments (27,28). Some herbal phenolic products, namely catechin gallate and gallic acid, have been shown to inhibit the SARS-CoV (38). Due to its importance to the virus cycle and pathogenicity to the host, N-protein of SARS-CoV-2 has become an enticing and essential target for drug development.

In this paper, we utilized the crystal structure of NTD-N-protein of SARS-CoV-2 (PDB ID 6VYO) as a target for the identification of novel potent compounds. Active site residues were identified using the complex structure of NTD-N-protein of HCoV-OC43 with GMP (PDB ID 4LM9). To generate the pharmacophore model, GMP was docked into the NTD-N-protein of SARS-CoV-2 and validated by the molecular dynamics simulation. Further, screening was performed using the pharmacophore features of NTD-GMP complex against ZINC database. Screened compounds were further filtered through the ADMET profile and drug-like properties. To infer the binding affinity of these 3 pharmacokinetically hits

(ZINC000257324845, ZINC000005169973, and ZINC000009913056), molecular docking was performed using the AutoDock Tools. These molecules are stabilized by hydrogen bonding, and hydrophobic interaction. Molecular dynamics and Molecular mechanic/Poisson-Boltzmann Surface Area (MMPBSA) was employed to assess the stability of NTD-GMP and NTD-ligand(s) complexes. The sequential work done for the screening of potent molecules against NTD-N-protein of SARS-CoV-2 is represented in Figure 1.

2. Material methods

2.1. Retrieving crystallographic structures

The crystal structure of N-terminal domain (NTD) of nucleocapsid protein (N-protein) of SARS-CoV-2 (PDB ID 6VYO) was downloaded from the RCSB-PDB database to find its potential inhibitor. Due to the existence of this structure in native form, the protein sequence similarity search was performed by using NCBI-BLAST server against the PDB in order to screen the complex homologous structure. The structure of N-protein-NTD of HCoV-OC43 complexed with its substrate GMP (PDB ID 4LM9) was considered as a point of reference to analyse the active site (28). Both the protein structures were superimposed with PyMol, and the active site residues were identified for further *in-silico* analysis. The water and the other molecules were omitted, and the structure was prepared by adding the essential hydrogen atoms and Kollman charges (8.0) using the AutoDock Tools (39). The ligand of NTD-N-protein, GMP, was extracted from the PDB and processed by adding polar hydrogens and gasteiger charges (0.5003). The protein and ligand were saved in .pdbqt and the molecular docking was performed using AutoDock Vina and AutoDock Tools (39,40). Molecular grid was created around the active site residues (Ser51, Phe53, Ala55, Ala90, Arg107, Tyr109, Tyr111, Arg149) of NTD-N-protein of SARS-CoV-2. The grid box dimensions were set to be 60 x 60 x 60 Å and centre coordinates were 63.52 X 58.33 X 3.77 with 0.375 default grid space value. The 50 conformations of GMP were predicted using Lamarckian genetic algorithm. The best conformation of GMP with NTD was selected on the basis of binding energy and binding interactions with NTD. The multiple sequence alignment (MSA) of NTD-N-protein of SARS-CoV-2 with other NTD from other viruses was generated using Clustal Omega and illustrated using ESPript (41,42).

2.2. Pharmacophore modeling

The structure-based pharmacophore modeling was performed by using the Pharmit server, which directly works on the ligand-protein complex. Pharmit provides an interactive webserver

to screen the multiple compound databases by utilizing the pharmacophore template, molecular shapes, and energy minimization (43). The 3D structure of NTD-GMP complex was loaded into the Pharmit system as a query to identify the pharmacophore feature and to search the pre-built compound library. 123,399,574 conformers of 13,190,317 molecules from ZINC library were chosen for the pharmacophore model screening in order to find the most likely compound of GMP (44). Shape filters were applied for the protein and ligand to filter out the compounds in the database before conducting the pharmacophore search. Further, the total 4,576 compounds with RMSD less than 1.5 Å from the pharmacophore features of query molecules were saved and considered for virtual screening.

2.3. Virtual Screening

Structure-based virtual screening (SBVS) was used to screen the substantial databases in order to identify the novel and potent compounds. SBVS was performed by utilizing the AutoDock Vina, in PyRx 0.8 platform (40,46). 4,576 hit compounds, which has the pharmacophore properties of GMP, were preferred as ligands for the virtual screening against NTD. All the ligands were energy minimized by applying the Universal Force Field (UFF) and converted them into .pdbqt by OpenBabel (47). The grid map was centred with the dimension of 58 X 60 X 57 Å that covers the active site residues of the protein. Each ligand was docked in the binding site of NTD and scored with binding affinity. Top-ranked molecules were analyzed using PyMOL, and best conformations with higher affinity energy were saved (48).

2.4. Pharmacokinetic and ADMET analysis

The pharmacophore screened compounds were refined by means of Lipinski's rule of five for the drug-like criteria including molecular weight, log P, number of rotatable bonds, hydrogen bond acceptors, and donors (45). Further, to evaluate the possible impact of the selected compounds on humans, the ADMET studies were performed by using the pkCSM server (49). The molecule's structure was converted to SMILES notations and submitted to the online server for calculation. Various parameters including aqueous solubility, blood-brain barrier penetration (BBB), CYP2D6 inhibition, human intestinal absorption (HIA), plasma protein binding (PPB), and hepatotoxicity were assessed for the selected compounds, which defines the ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity). The compounds, which satisfied the Lipinski's rule of five and ADMET properties were further chosen for the molecular docking to determine the appropriate orientation and binding affinities with NTD.

2.5. Molecular docking

Molecular docking simulation studies for three lead compounds (ZINC000257324845, ZINC000005169973, and ZINC000009913056) were performed using the AutoDock tools. Hydrogen atoms and Kollman charges were added on the receptor. Polar hydrogen atoms and gasteiger charges for ZINC000257324845 (0.0007), ZINC000005169973 (0.0002), and ZINC000009913056 (-0.0002) were added and saved in .pdbqt format. The grid box for molecular docking was generated around the active site residues of the protein. All the generated 50 conformations for each lead compound during molecular docking were analyzed, and figures were prepared in PyMOL 2.0.5 (48).

2.6. Molecular dynamics

Molecular dynamics simulation of NTD with GMP, ZINC000257324845, ZINC000005169973, and ZINC000009913056 was performed to assess the structural and dynamics variation at an atomistic level. Protein topologies and coordinates files were generated using GRONingen Machine for Chemical Simulations (GROMACS) 2019.2 along with GROMOS96 43a1 force field on an Ubuntu-based workstation (50,51). Total four systems (NTD-GMP, NTD-ZINC000257324845, NTD-ZINC000005169973 and NTD-ZINC000009913056) were generated and kept for molecular simulation of 100 ns. Topology of ligands (GMP, ZINC000257324845, ZINC000005169973, and ZINC000009913056) were predicted by using PRODRG webserver (52). The partial atomic charges for ligands were computed by density functional theory (DFT) analysis using Lee-Yang-Parr correlation functional (B3LYP) method with a 6-311G (d,p) basis set in Gaussian 16 (53-56). The protein-ligand complexes were solvated using a simple point charge (SPC) water model in a triclinic box of volume (234.61 nm³) with a minimum distance of 1.0 nm between atoms of protein and edge of the box. Three chlorides (Cl⁻) ions were added using genion tool to neutralize the system. The steepest descent algorithm was utilized for 50,000 steps to minimize the protein up to an energy cut of 10.0 kJmol⁻¹. Two-phase of equilibration: constant number of particles, volume, and temperature (NVT) and a constant number of particles, pressure, and temperature (NPT) were performed for 5000000 steps. Systems were kept at NVT using Parrinello-Rahman barostat pressure coupling method at 300 K for 1 ns (57). NPT was done for 1 ns using Berendsen thermostat (58). The covalent bonds were constrained by using Linear Constraints Solver (LINCS) (59). Long-range interactions were determined using Particle Mesh Ewald (PME) (60). 12 Å radius cut-off was used to compute the short-range (Lennard-Jones and

Coulomb) interactions. The final production run was done for 100 ns, and coordinates were updated every 10 ps. The trajectories were analyzed by visual molecular dynamics (VMD) and XMGRACE (61). Root mean square deviations (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), hydrogen bond numbers, and principal component analysis (PCA) were determined to assess the stability of NTD-GMP and NTD-inhibitor(s) complexes.

2.7. MMPBSA binding free energy calculation

In the present study, the binding energy calculations of protein-ligand complexes were done using Molecular Mechanics/Position-Boltzmann Surface Area (MMPBSA) method (62). The trajectory from the equilibrium state was selected to calculate the corresponding binding free energy of NTD-GMP and NTD-inhibitor(s) complexes. In GROMACS, the `g_mmpbsa` tool was used to compute the binding free energy of protein-ligand complexes as:-

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - (\Delta G_{\text{protein}} + \Delta G_{\text{ligand}})$$

Where, $\Delta G_{\text{complex}}$, $\Delta G_{\text{protein}}$, and ΔG_{ligand} is the total free energy of protein-ligand complex, protein, and ligand in a solvent, respectively.

In the current study, last 20 ns (80-100 ns) of molecular dynamics snapshots generated at every 10 ps were used to calculate the binding affinity of NTD-GMP and NTD-inhibitor(s) complexes.

3. Results

3.1. Retrieving crystallographic structures

As mentioned earlier, NTD-N-protein CoV is a potential target for the treatment of CoV diseases. The crystal structure of NTD-N-protein of SARS-CoV-2 (PDB ID 6VYO) was obtained from the PDB and used as the target sequence. From the NCBI-BLAST search against PDB, NTD structure from HCoV-OC43 (PDB-ID 4LM9) complexed with RNA substrate (GMP) was found and indicated that residues Ser64, Phe66, Gly68, His104, Arg122, Tyr124, Tyr126, and Arg164 were noteworthy interacting residues for the formation of NTD-GMP complex.

For identifying the active site pocket of NTD-N-SARS-CoV-2, both structures (PDB ID 6VYO) were superimposed in PyMol, and have shown that residues (Ser51, Phe53, Ala55, Ala90, Arg107, Tyr109, Tyr111, Arg149) in the NTD-N-SARS-CoV-2 are the corresponding

to GMP interacting residues in NTD- HCoV-OC43. Figure 2 illustrates that these residues are conserved among all the NTD of CoV. Moreover, the molecular docking study of GMP with NTD-N-SARS-CoV-2 was performed using AutoDock Vina and AutoDock tools in PyRx. The analysis demonstrated that GMP was interacting with the expected residues of NTD-N-SARS-CoV-2 with -5.5 kcal/mol binding energy, as shown in Table 1. The best-docked conformation of GMP in the respective pocket of NTD-N-SARS-CoV-2 was shown in Figure 3A and 4A using PyMol. The complex of NTD-N-SARS-CoV-2 with GMP was validated and further used for *in-silico* studies.

3.2. Pharmacophore modeling

To perform the structure-based pharmacophore modeling, the validated complex of NTD-N-SARS-CoV-2 with GMP was uploaded to the Pharmit server. In this study, NTD-N-SARS-CoV-2 and GMP were used as receptor and ligand, respectively, to construct the pharmacophore model. Total 5 pharmacophore properties, including 3 hydrogen acceptors (orange), aromatic group (purple), and hydrophobic feature (green), with radius of 1.0 Å were considered to screen the compounds, as shown in Figure 5. Both protein and ligand shape filters refined the compounds in the ZINC data-set before performing a pharmacophore search. The ligand shape is indeed an inclusive feature that covers the entire ligand and should be superposed with the screened compounds in the database (63). While the receptor shape is an exclusive feature centered on the receptor structure around the GMP binding site that will be forbidden to be used by any compound in the database during the search. ZINC database comprising 13,190,317 molecules with 123,399,574 conformations was screened using the constructed pharmacophore model, and 8,192 compounds were identified, which shares the pharmacophore features. Further, these identified compounds were refined with the Lipinski's rule of five, RMSD, and energy minimization. With the RMSD of less than 1.5 Å and binding affinity of more than -5.0 kcal/mol, a total of 4576 hits were eventually saved.

3.3. Virtual screening

In order to identify novel compounds against the NTD-N-protein, virtual screening was performed using the AutoDock Vina in PyRx 0.8. The downloaded selected compounds from the ZINC database having a high negative score and less RMSD values were converted from .sdf format to .pdbqt format using Open Babel. Further, these compounds were docked into the GMP binding site and ranked on the basis of binding energy with the NTD-N-protein. Ligands showing binding affinity in range of -7.5 to -8.3 kcal/mol were considered for further study.

3.4. Pharmacokinetic and ADMET analysis

In-silico ADMET prediction was performed for the top-ranked results using the pkCSM server to predict the overall risks of absorption, distribution, metabolism, excretion, and toxicity. Three hits (ZINC000257324845, ZINC000005169973, and ZINC000009913056) were successfully fulfilled the drug-like properties as per the Lipinski's rule of five including molecular weight, LogP, hydrogen bond donor, and acceptor (Table 2). A promising ADMET profile is required for the compounds in drug discovery (64). For this purpose, pharmacokinetic parameters such as aqueous solubility (log S), skin permeability coefficient (logKp), logBB, CNS permeability, number of metabolic reactions, *etc.* were calculated and enlisted in Table 3. The results indicated that all three compounds were soluble in water, and the aqueous solubility (S) of a compound substantially affects its absorption and distribution properties. All compounds showed the skin permeability, and also absorbed by the human intestine. In addition, they were able to penetrate the BBB and CNS after oral administration. CYP enzymes, various CYP450 substrates and inhibitors, played a fundamental role in the metabolism of the drug. Further, metabolism analyses revealed that none of the compounds were substrates of CYP2D6 substrate. Toxicity assessment has shown that all compounds are non-toxic in nature to humans.

3.5. Molecular docking

The molecular docking study was performed to examine the binding pose, and binding affinity of all three pharmacokinetically screened compounds with NTD. All compounds showed higher binding energy than GMP with NTD and presented in Table 1. ZINC000257324845 was found stable at the active site of NTD by forming 3 H-bonds (Ser51, Ala55, and Tyr111), π - π (Tyr109), and π -alkyl (Tyr109, and Ala156), as shown in Figure 3B and 4B. ZINC000005169973 was able to make 2 H-bond (Arg107, and Arg149), π -donor (Tyr109, Arg107, and Arg149), π -sigma (Ala156) and π -alkyl (Ala50, and Ala90) (Figure 3C and 4C). ZINC000009913056 was able to form 4 H-bond (Thr57, Arg107, Tyr111, and Arg149), π -stacked (Tyr109), and π -alkyl (Ala50, and Ala90) (Figure 3D and 4D).

3.6. Molecular dynamics

Molecular dynamics simulation was carried out to assess the flexibility and stability of NTD with identified inhibitor(s). Therefore in this study, molecular dynamics were employed to evaluate the stable and static interactions of protein-ligand complexes by examination of

various molecular simulation results like RMSD, RMSF, Rg, SASA, hydrogen bond formation, and PCA.

3.6.1. Root Mean Square Deviation (RMSD)

The atomistic dynamics movements and conformational variations of C α backbone atoms of NTD-GMP and NTD-inhibitor(s) complexes were calculated by RMSD. In Figure 6A, it is clearly seen that all the protein-ligand complexes showed an initial sharp in RMSD values during 5-10 ns and attained equilibrium at 29 ns and systems remained stable throughout the molecular simulation of 100 ns. NTD-inhibitor(s) complexes exhibited an RMSD in range of 0.25 to 0.36 nm, which is lesser than NTD-GMP (0.39 nm) complex. The average RMSD values of NTD-GMP and NTD-inhibitor(s) complexes are shown in Table 4. Ligand RMSD of GMP and inhibitors are represented in Figure 6B. The average ligand RMSD for GMP, ZINC000257324845, ZINC000005169973, and ZINC000009913056 are 0.09 ± 0.008 , 0.03 ± 0.007 , 0.06 ± 0.009 and 0.06 ± 0.005 nm, as shown in Table 4. Overall RMSD results illustrated that binding of identified inhibitor(s) at the active site of NTD is stable and formed more stable NTD-inhibitor(s) complexes as compared to NTD-GMP complex.

3.6.2. Root Mean Square Fluctuation (RMSF)

Residue wise fluctuation was calculated to determine the motion in C α atoms from its average position during the molecular simulation. It is very well known that secondary structures like helix and sheet possess less RMSF as compared to the non-secondary structures such as loops and turns. The RMSF profile of NTD-inhibitor(s) complexes is almost comparable to NTD-GMP, as shown in Figure 7. The average RMSF values for NTD-GMP, NTD-ZINC000257324845, NTD-ZINC000005169973, and NTD-ZINC000009913056 are 0.16 ± 0.07 , 0.15 ± 0.07 , 0.16 ± 0.07 and 0.16 ± 0.07 nm, respectively as shown in Table 4. RMSF analysis suggested that inhibitor(s) were well fitted at the active site of NTD form a stable NTD-inhibitor(s) complex.

3.6.3. Radius of gyration (Rg)

Radius of gyration was calculated to assess the compactness and stability of protein-ligand complex during the molecular simulation. Smaller Rg value suggests a stably folded protein. NTD-inhibitor(s) complex showed a lesser Rg value than NTD-GMP complex, as shown in Figure 8. NTD-inhibitor(s) complexes showed average Rg values in the range of 1.47 ± 0.014 to 1.48 ± 0.011 nm, which was lesser than NTD-GMP (1.50 ± 0.013 nm) complex, as shown

in Table 4. Rg results indicated that the binding of inhibitor(s) to NTD form a higher stable NTD-inhibitor(s) complex as compared to NTD-GMP complex.

3.6.4. Solvent Accessible Surface Area (SASA)

Solvent Accessible Surface Area (SASA) of a protein is defined as the surface area of a protein covered by the interaction of solvent molecules. SASA value of a protein decline with an increment in compactness of a protein. SASA plot reveals that values of NTD-inhibitor(s) complexes are smaller than NTD-GMP complex, as shown in Figure 9. Average SASA of NTD-GMP, NTD-ZINC000257324845, NTD-ZINC000005169973, and NTD-ZINC000009913056 complexes were 72.08 ± 1.66 , 70.20 ± 2.41 , 69.42 ± 2.09 and 69.02 ± 2.43 nm², respectively as shown in Table 4. SASA analysis implied that NTD-inhibitor(s) complexes were more stable and compact than NTD-GMP complex.

3.6.5. Hydrogen Bond Analysis

Hydrogen bond numbers and distributions were calculated to determine the stability of the protein-ligand complex(s) during the molecular simulation of 100 ns. The g_hbond tool of GROMACS was used to compute the intra-protein and inter-molecular hydrogen bonds within a distance and an angle of 3.5 Å and 120°, respectively. Intra-protein and inter-molecular hydrogen bond plots of NTD-GMP and NTD-inhibitor(s) complexes are shown in Figure 10. NTD-inhibitor(s) complexes have higher intra-hydrogen bonds as compared to NTD-GMP as shown in Figure 10A. The average number of intra-protein hydrogen bonds between protein-ligand(s) complexes were in a range of 64.86 ± 5.39 to 71.93 ± 6.08 , as shown in Table 4. As shown in Figure 10B, inter-molecular hydrogen bonds between NTD and inhibitor(s) were sustained during the course of molecular simulation. Hydrogen bonding affinity of NTD-inhibitor(s) complexes were similar to NTD-GMP complex, as shown in Figure 10C and 10D. Overall hydrogen bond results affirmed that binding of inhibitor(s) to NTD form stable NTD-inhibitor(s) complexes.

3.6.6. Principal Component Analysis (PCA)

Essential dynamics were used to differentiate the motion between NTD-GMP and NTD-inhibitor(s) complexes. Principal component analysis (PCA) reveals the overall expansion of protein during the molecular simulation. The dynamical differences of 371 eigenvectors were generated and framed in a covariance matrix. Directional movements contributed by first eigenvector (PC1) and second eigenvector (PC2) were represented in a two-dimensional

projection, as shown in Figure 11. The two-dimensional projection PCA results showed that binding of inhibitor(s) to NTD results in the formation of stable NTD-inhibitor(s) complexes.

3.7. MMPBSA binding free energy calculation

The binding free energy of the protein-ligand complex was evaluated using MMPBSA within the GROMACS. The last 20 ns molecular dynamics trajectory consisting of 2000 snapshots were used to generate the binding affinity of NTD-GMP and NTD-inhibitors(s) complexes. The binding affinity of NTD-GMP, NTD-ZINC000257324845, NTD-ZINC000005169973, and NTD-ZINC000009913056 complexes were -99.48 ± 2.48 , -157.02 ± 1.99 , -147.04 ± 1.78 and -129.88 ± 1.68 kJmol⁻¹, respectively as shown in Table 5. In addition, Van der Waals interactions, electrostatic interactions, and non-polar solvations energy overall decline the binding energy, while polar solvation energy complement the binding energy positively. The MMPBSA results confirmed that the binding of identified compounds with NTD results in the formation of a higher stable NTD-inhibitor(s) complexes as compared to NTD-GMP complex.

4. Discussion

Currently, the global pandemic disease called Coronavirus Disease 2019 (COVID-19) is posing serious survival threats to the world population (18). COVID-19 is a zoonotic disease and caused by a third highly pathogenic human *betacoronavirus* named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (16,18). Human coronaviruses (HCoV) are the most important human virus, which affects global health. In the past two decades, SARS-CoV-2 is the third HCoV after SARS-CoV and MERS-CoV that caused severe problems in humans, and it is spreading across the world (12,13). There is no effective treatment or medicine for this disease. Therefore, it is imperative to deploy new antiviral drugs for SARS-CoV-2.

This study focuses on the finding of potent molecules against the N-terminal domain of nucleocapsid (NTD-N-protein) from SARS-CoV-2. As it was reported that viral nucleocapsid is often multifunctional and plays a vital role in the viral cycle. Previous studies have shown that N-protein can be used as a target for the development of an antiviral drug against viral infections (28,32,65,66). To inhibit the function of N-protein, two strategies were reported in which one strategy target the RNA binding site of NTD, which having conserved residues while other inhibit the oligomerization of C-terminal domain (67).

Here, NTD-N-protein from SARS-CoV-2 was used as a target to inhibit the viral cycle of SARS-CoV-2 by finding potential compounds and crystal structure of RNA binding domain

of N-protein from SARS coronavirus 2 (PDB ID 6VYO) was obtained from the Protein Data Bank. As this structure was reported in the native form, we choose the crystal structure of NTD-N-protein from HCoV-OC43 (PDB ID 4LM9), which was reported with GMP, an RNA substrate to find the active site for the substrate or inhibitor in our target structure. On the sequence comparison, residues (Ser51, Phe53, Ala55, Ala90, Arg107, Tyr109, Tyr111, Arg149) were found, which involved in the interaction of substrate and inhibitor as well. Further docking of GMP was performed using the same residues as an active site of NTD-N-protein from SARS-CoV-2, and results confirmed that identified residues were involved in holding the GMP at this site and form significant interactions. In addition, molecular dynamics simulation confirmed the stability of complexed NTD-N-protein with GMP. Following the validation of GMP interactions with NTD-N-protein, GMP was used by the ligand structure-based strategy of pharmacophore modeling to build a pharmacophore model.

Previous studies have reported various inhibitors based on the aromatic rings of GMP (27,28). Similarly, in our study, we considered one aromatic ring of GMP, 3 hydrogen acceptors, and hydrophobic features, which were involved in the stabilization of interaction between the GMP and the binding residues of the protein. Before conducting the pharmacophore search, shape filters for protein and ligand were applied to refine the compounds in the database. As a result of the pharmacophore search, 8,192 compounds were found, which has the pharmacophore features of our model. A total of 4,576 molecules were downloaded after applying the RMSD and energy minimization filters. The final hits from pharmacophore screening were used for virtual screening using the crystal structure of NTD-N-protein. Virtual screening is a powerful technique for identifying the potential lead compounds in drug discovery (68-71). Total 50 molecules bound at the active site of NTD and having higher binding energy than GMP were used for further study.

Further, ADMET properties were analyzed using the pkCSM server. Many FDA approved drugs in the later stages were failed due to unacceptable pharmacokinetic and toxicity properties. Thus, inhibitor must be investigated for long term effectiveness on the basis of ADMET properties (54,55). Three molecules (ZINC000257324845, ZINC000005169973, and ZINC000009913056) were successfully passed the ADMET filters and also having the drug-like properties, which were analysed by Lipinski's rule of five. All three pharmacokinetically screened (fulfilling the ADMET and Lipinski's rule of five criteria) compounds were evaluated using the molecular docking approach for their efficacy of binding to NTD. Various molecular docking studies have been used to assess the efficiency of the ligand with protein (72-74). Our

docking analysis indicated that all compounds were bound to the active site with high binding energy in the range of -5 to -7.42 kcal/mol, which was higher than the binding energy (-2.6 kcal/mol) of GMP. All three compounds and GMP form a stable protein-ligand complex by establishing a network of molecular interactions including hydrogen bonds, and the hydrophobic interactions with the key-residues (Ser51, Phe53, Ala55, Ala90, Arg107, Tyr109, Tyr111, Arg149) of NTD as shown in Figure 3 and 4. ZINC000257324845, ZINC000005169973, and ZINC000009913056 showed the binding interactions at the active site of NTD through 3, 2, and 4 hydrogen bonds, respectively, as shown in Figure 3. Henceforth, we speculate that these compounds may bind specifically to the NTD protein to inhibit its viral cycle activity.

Molecular dynamics simulation has been widely used to predict the stability of protein-ligand or protein-protein complexes (71,75-77). Molecular dynamics were employed to determine the structural and conformational changes in the protein-ligand complex. Trajectories retrieved from the molecular simulation were used to analyze the RMSD, RMF, Rg, SASA, hydrogen bond numbers, and PCA of NTD-GMP and NTD-inhibitor(s) complexes. RMSD results indicated that binding of identified molecules (ZINC000257324845, ZINC000005169973, and ZINC000009913056) to NTD tends to form the higher stable NTD-inhibitor(s) complexes than NTD-GMP complex. RMSF results suggested that predicted inhibitor(s) were well bound at the active site of NTD. The smaller Rg and SASA values of NTD-inhibitor(s) complex as compared to NTD-GMP complex revealed that identified compounds result in the formation of a higher stable protein-ligand complex than NTD-GMP complex. Further hydrogen bonding analysis affirmed that binding of identified molecules with NTD forms the stable and compact protein-ligand complex. Dynamics movements along PC1 and PC2 for protein-ligand complexes illustrated that identified molecules tend to form stable NTD-inhibitor(s) complexes. MMPBSA results confirmed that binding of identified molecules to NTD results in the formation of higher stable and lower energy NTD-inhibitor(s) complexes as compared to NTD-GMP complex. Overall molecular dynamics results concluded that all the identified compounds are potent molecules and might be used to inhibit the binding of viral RNA genome to the RNA binding region of NTD, and it may ultimately stop the essential functions of NTD including ribonucleoprotein formation, viral cycle, which are required for the survival of virus into the host.

5. Conclusion

This study utilized various computational approaches, including pharmacophore-based virtual screening, molecular docking, and molecular dynamics to identify the potent molecules, which target the RNA binding region named NTD of N-protein of SARS-CoV-2 to inhibit its function. The crystal structure of NTD-N-protein of SARS-CoV-2 was retrieved from the PDB, and further GMP was docked around the RNA binding region. Based on the NTD-N-GMP complexed structure, the pharmacophore model of GMP was generated by considering the features of GMP, including one aromatic ring of GMP, 3 hydrogen acceptors, and a hydrophobic region. After the virtual screening against ZINC database, three molecules (ZINC000257324845, ZINC000005169973, and ZINC000009913056) were screened, which follow the Lipinski's rule of five and fulfill the ADMET properties. Further, molecular docking results revealed that these molecules interact with the NTD-N-protein via hydrogen bonding and hydrophobic interactions. The molecular dynamics and MMPBSA study verified that the selected compounds efficiently bind to NTD and form stable NTD-ligand complexes. These lead molecules can be further investigated their importance through *in vitro* studies and can be used for the development of antiviral compounds against SARS-CoV-2 in the future.

6. Disclosure statement

No conflict of interest is reported by the authors.

7. Acknowledgment

PD, VD, and VS thanks Ministry of Human Resource Development, Department of Biotechnology [DBT/2015/IIT-R/349], and University Grant Commission for financial support. We are thankful to the Department of chemistry, IIT Roorkee to provide the Gaussian software. Authors thank Macromolecular Crystallographic Unit (MCU), a Central Facility at Institute Instrumentation Centre (IIC), IIT Roorkee for computational work.

8. Conflict of Interest

The authors declare that they have no conflicts of interest associated with the manuscript.

9. References

1. Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J., and Snijder, E. J. (2006) Nidovirales: evolving the largest RNA virus genome. *Virus research* **117**, 17-37
2. Fehr, A. R., and Perlman, S. (2015) Coronaviruses: an overview of their replication and pathogenesis. in *Coronaviruses*, Springer. pp 1-23

3. Cavanagh, D. (2007) Coronavirus avian infectious bronchitis virus. *Veterinary research* **38**, 281-297
4. King, A. M., Lefkowitz, E. J., Mushegian, A. R., Adams, M. J., Dutilh, B. E., Gorbalenya, A. E., Harrach, B., Harrison, R. L., Junglen, S., and Knowles, N. J. (2018) Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). *Archives of virology* **163**, 2601-2631
5. Perlman, S., and Netland, J. (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nature reviews microbiology* **7**, 439-450
6. Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., Tsang, A. K., Lau, J. H., Bai, R., Teng, J. L., Tsang, C. C., and Wang, M. (2012) Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *Journal of virology* **86**, 3995-4008
7. Jevšnik, M., Uršič, T., Žigon, N., Lusa, L., Krivec, U., and Petrovec, M. (2012) Coronavirus infections in hospitalized pediatric patients with acute respiratory tract disease. *BMC infectious diseases* **12**, 365
8. Lau, S. K., Woo, P. C., Yip, C. C., Tse, H., Tsoi, H.-w., Cheng, V. C., Lee, P., Tang, B. S., Cheung, C. H., and Lee, R. A. (2006) Coronavirus HKU1 and other coronavirus infections in Hong Kong. *Journal of clinical microbiology* **44**, 2063-2071
9. Pyrc, K., Berkhout, B., and van der Hoek, L. (2007) The novel human coronaviruses NL63 and HKU1. *Journal of virology* **81**, 3051-3057
10. van der Hoek, L., Sure, K., Ihorst, G., Stang, A., Pyrc, K., Jebbink, M. F., Petersen, G., Forster, J., Berkhout, B., and Überla, K. (2005) Croup is associated with the novel coronavirus NL63. *PLoS medicine* **2**
11. Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., Khattra, J., Asano, J. K., Barber, S. A., and Chan, S. Y. (2003) The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399-1404
12. Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G. M., Ahuja, A., Yung, M. Y., Leung, C., and To, K. (2003) A major outbreak of severe acute respiratory syndrome in Hong Kong. *New England Journal of Medicine* **348**, 1986-1994
13. Lancet, T. (2013) MERS-CoV: a global challenge. *Lancet (London, England)* **381**, 1960

14. Denison, M. R. (2004) Severe acute respiratory syndrome coronavirus pathogenesis, disease and vaccines: an update. *The Pediatric infectious disease journal* **23**, S207-S214
15. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., and Lu, R. (2020) A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*
16. Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., and Huang, C.-L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273
17. Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., and Wei, Y. (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* **395**, 507-513
18. Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., and Pei, Y.-Y. (2020) A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269
19. Forni, D., Cagliani, R., Clerici, M., and Sironi, M. (2017) Molecular evolution of human coronavirus genomes. *Trends in microbiology* **25**, 35-48
20. Hsieh, P.-K., Chang, S. C., Huang, C.-C., Lee, T.-T., Hsiao, C.-W., Kou, Y.-H., Chen, I.-Y., Chang, C.-K., Huang, T.-H., and Chang, M.-F. (2005) Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. *Journal of virology* **79**, 13848-13855
21. Schelle, B., Karl, N., Ludewig, B., Siddell, S. G., and Thiel, V. (2005) Selective replication of coronavirus genomes that express nucleocapsid protein. *Journal of virology* **79**, 6620-6630
22. Surjit, M., Liu, B., Chow, V. T., and Lal, S. K. (2006) The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex and blocks S phase progression in mammalian cells. *Journal of Biological Chemistry* **281**, 10669-10681
23. Du, L., Zhao, G., Lin, Y., Chan, C., He, Y., Jiang, S., Wu, C., Jin, D.-Y., Yuen, K.-Y., and Zhou, Y. (2008) Priming with rAAV encoding RBD of SARS-CoV S protein and boosting with RBD-specific peptides for T cell epitopes elevated humoral and cellular immune responses against SARS-CoV infection. *Vaccine* **26**, 1644-1651

24. Che, X.-Y., Hao, W., Wang, Y., Di, B., Yin, K., Xu, Y.-C., Feng, C.-S., Wan, Z.-Y., Cheng, V. C., and Yuen, K.-Y. (2004) Nucleocapsid protein as early diagnostic marker for SARS. *Emerging infectious diseases* **10**, 1947
25. Liang, F.-Y., Lin, L.-C., Ying, T.-H., Yao, C.-W., Tang, T.-K., Chen, Y.-W., and Hou, M.-H. (2013) Immunoreactivity characterisation of the three structural regions of the human coronavirus OC43 nucleocapsid protein by Western blot: Implications for the diagnosis of coronavirus infection. *Journal of virological methods* **187**, 413-420
26. Chang, C.-k., Hou, M.-H., Chang, C.-F., Hsiao, C.-D., and Huang, T.-h. (2014) The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral research* **103**, 39-50
27. Chang, C.-k., Jeyachandran, S., Hu, N.-J., Liu, C.-L., Lin, S.-Y., Wang, Y.-S., Chang, Y.-M., and Hou, M.-H. (2016) Structure-based virtual screening and experimental validation of the discovery of inhibitors targeted towards the human coronavirus nucleocapsid protein. *Molecular BioSystems* **12**, 59-66
28. Lin, S.-Y., Liu, C.-L., Chang, Y.-M., Zhao, J., Perlman, S., and Hou, M.-H. (2014) Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target. *Journal of medicinal chemistry* **57**, 2247-2257
29. Kao, R. Y., Yang, D., Lau, L.-S., Tsui, W. H., Hu, L., Dai, J., Chan, M.-P., Chan, C.-M., Wang, P., and Zheng, B.-J. (2010) Identification of influenza A nucleoprotein as an antiviral target. *Nature biotechnology* **28**, 600
30. Musah, R. A. (2004) The HIV-1 nucleocapsid zinc finger protein as a target of antiretroviral therapy. *Current topics in medicinal chemistry* **4**, 1605-1622
31. Chang, C.-k., Chen, C.-M. M., Chiang, M.-h., Hsu, Y.-l., and Huang, T.-h. (2013) Transient oligomerization of the SARS-CoV N protein—implication for virus ribonucleoprotein packaging. *PloS one* **8**
32. Lo, Y.-S., Lin, S.-Y., Wang, S.-M., Wang, C.-T., Chiu, Y.-L., Huang, T.-H., and Hou, M.-H. (2013) Oligomerization of the carboxyl terminal domain of the human coronavirus 229E nucleocapsid protein. *FEBS letters* **587**, 120-127
33. Chen, I.-J., Yuann, J.-M. P., Chang, Y.-M., Lin, S.-Y., Zhao, J., Perlman, S., Shen, Y.-Y., Huang, T.-H., and Hou, M.-H. (2013) Crystal structure-based exploration of the important role of Arg106 in the RNA-binding domain of human coronavirus OC43 nucleocapsid protein. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1834**, 1054-1062

34. Saikatendu, K. S., Joseph, J. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C., and Kuhn, P. (2007) Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *Journal of virology* **81**, 3913-3921
35. Fan, H., Ooi, A., Tan, Y. W., Wang, S., Fang, S., Liu, D. X., and Lescar, J. (2005) The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties. *Structure* **13**, 1859-1868
36. Jayaram, H., Fan, H., Bowman, B. R., Ooi, A., Jayaram, J., Collisson, E. W., Lescar, J., and Prasad, B. V. (2006) X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *Journal of virology* **80**, 6612-6620
37. Grosseohme, N. E., Li, L., Keane, S. C., Liu, P., Dann III, C. E., Leibowitz, J. L., and Giedroc, D. P. (2009) Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. *Journal of molecular biology* **394**, 544-557
38. Roh, C. (2012) A facile inhibitor screening of SARS coronavirus N protein using nanoparticle-based RNA oligonucleotide. *International journal of nanomedicine* **7**, 2173
39. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* **30**, 2785-2791
40. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **31**, 455-461
41. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., and Söding, J. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**
42. Gouet, P., Courcelle, E., Stuart, D. I., and M²toz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics (Oxford, England)* **15**, 305-308
43. Sunseri, J., and Koes, D. R. (2016) Pharmit: interactive exploration of chemical space. *Nucleic acids research* **44**, W442-W448

44. Sterling, T., and Irwin, J. J. (2015) ZINC 15–ligand discovery for everyone. *Journal of chemical information and modeling* **55**, 2324-2337
45. Lipinski, C. A. (2004) Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technologies* **1**, 337-341
46. Dallakyan, S., and Olson, A. J. (2015) Small-molecule library screening by docking with PyRx. in *Chemical biology*, Springer. pp 243-250
47. O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011) Open Babel: An open chemical toolbox. *Journal of cheminformatics* **3**, 33
48. DeLano, W. L. (2002) Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography* **40**, 82-92
49. Pires, D. E., Blundell, T. L., and Ascher, D. B. (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of medicinal chemistry* **58**, 4066-4072
50. Tran, F., Billeter, S., Eising, A., Hünenberger, P., Kruger, P., Mark, E., Scott, R., and Tironi, I. (1996) Biomolecular Simulation: The GROMOS96 manual and user guide.
51. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free. *Journal of computational chemistry* **26**, 1701-1718
52. Schüttelkopf, A. W., and Van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallographica Section D: Biological Crystallography* **60**, 1355-1363
53. Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Petersson, G., and Nakatsuji, H. (2016) Gaussian 16 revision a. 03. 2016; gaussian inc. Wallingford CT **2**, 4
54. Lee, C., Yang, W., and Parr, R. G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* **37**, 785
55. Schlegel, H. B. (1982) Optimization of equilibrium geometries and transition structures. *Journal of Computational Chemistry* **3**, 214-218
56. Jain, R., Ahuja, B., and Sharma, B. (2004) Density-Functional Thermochemistry. III. The Role of Exact Exchange. *Indian Journal of Pure & Applied Physics* **42**, 43-48
57. Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **52**, 7182-7190

58. Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **81**, 3684-3690
59. Hess, B., Bekker, H., Berendsen, H. J., and Fraaije, J. G. (1997) LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **18**, 1463-1472
60. Abraham, M. J., and Gready, J. E. (2011) Optimization of parameters for molecular dynamics simulation using smooth particle-mesh Ewald in GROMACS 4.5. *Journal of computational chemistry* **32**, 2031-2040
61. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33-38
62. Kumari, R., Kumar, R., Consortium, O. S. D. D., and Lynn, A. (2014) g_mmpbsa A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of chemical information and modeling* **54**, 1951-1962
63. Shiri, F., Pirhadi, S., and Rahmani, A. (2018) Identification of new potential HIV-1 reverse transcriptase inhibitors by QSAR modeling and structure-based virtual screening. *Journal of Receptors and Signal Transduction* **38**, 37-47
64. Hop, C. E., Cole, M. J., Davidson, R. E., Duignan, D. B., Federico, J., Janiszewski, J. S., Jenkins, K., Krueger, S., Lebowitz, R., and Liston, T. E. (2008) High throughput ADME screening: practical considerations, impact on the portfolio and enabler of in silico ADME models. *Current drug metabolism* **9**, 847-853
65. Chenavas, S., Crépin, T., Delmas, B., Ruigrok, R. W., and Slama-Schwok, A. (2013) Influenza virus nucleoprotein: structure, RNA binding, oligomerization and antiviral drug target. *Future microbiology* **8**, 1537-1545
66. Monod, A., Swale, C., Tarus, B., Tissot, A., Delmas, B., Ruigrok, R. W., Crépin, T., and Slama-Schwok, A. (2015) Learning from structure-based drug design and new antivirals targeting the ribonucleoprotein complex for the treatment of influenza. *Expert opinion on drug discovery* **10**, 345-371
67. Ciani, C., Gerritz, S. W., Deminie, C., and Krystal, M. (2013) Influenza nucleoprotein: promising target for antiviral chemotherapy. *Antiviral Chemistry and Chemotherapy* **23**, 77-91
68. Saini, G., Dalal, V., Savita, B. K., Sharma, N., Kumar, P., and Sharma, A. K. (2019) Molecular docking and dynamic approach to virtual screen inhibitors against Esbp of

- Candidatus Liberibacter asiaticus. *Journal of Molecular Graphics and Modelling* **92**, 329-340
69. Dhankhar, P., Dalal, V., Kotra, D., and Kumar, P. (2020) In-silico approach to identify novel potent inhibitors against GraR of *S. aureus*. *Frontiers in Bioscience (Landmark Edition)* **25**, 1337-1360
 70. Kumar, P., Dalal, V., Kokane, A., Singh, S., Lonare, S., Kaur, H., Ghosh, D. K., Kumar, P., and Sharma, A. K. (2020) Mutation studies and structure-based identification of potential inhibitor molecules against periplasmic amino acid binding protein of Candidatus Liberibacter asiaticus (CLasTcyA). *International Journal of Biological Macromolecules* **147**, 1228-1238
 71. Malik, A., Dalal, V., Ankri, S., and Tomar, S. (2019) Structural insights into Entamoeba histolytica arginase and structure-based identification of novel non-amino acid based inhibitors as potential antiamebic molecules. *The FEBS journal* **286**, 4135-4155
 72. Singh, N., Dalal, V., Mahto, J. K., and Kumar, P. (2017) Biodegradation of phthalic acid esters (PAEs) and in silico structural characterization of mono-2-ethylhexyl phthalate (MEHP) hydrolase on the basis of close structural homolog. *Journal of hazardous materials* **338**, 11-22
 73. Singh, N., Dalal, V., and Kumar, P. (2018) Structure based mimicking of Phthalic acid esters (PAEs) and inhibition of hACMSD, an important enzyme of the tryptophan kynurenine metabolism pathway. *International journal of biological macromolecules* **108**, 214-224
 74. Pandit, S., Dalal, V., and Mishra, G. (2018) Identification of novel phosphatidic acid binding domain on sphingosine kinase 1 of Arabidopsis thaliana. *Plant Physiology and Biochemistry* **128**, 178-184
 75. Dalal, V., Kumar, P., Rakhaminov, G., Qamar, A., Fan, X., Hunter, H., Tomar, S., Golemi-Kotra, D., and Kumar, P. (2019) Repurposing an ancient protein core structure: structural studies on FmtA, a novel esterase of staphylococcus aureus. *Journal of molecular biology* **431**, 3107-3123
 76. Singh, N., Dalal, V., and Kumar, P. (2020) Molecular docking and simulation analysis for elucidation of toxic effects of dicyclohexyl phthalate (DCHP) in glucocorticoid receptor-mediated adipogenesis. *Molecular Simulation* **46**, 9-21
 77. Kesari, P., Pratap, S., Dhankhar, P., Dalal, V., Mishra, M., Singh, P. K., Chauhan, H., and Kumar, P. (2020) Structural characterization and in-silico analysis of Momordica charantia 7S globulin for stability and ACE inhibition. *Scientific reports* **10**, 1-13

Figures legends:

Figure 1. Schematic representation implemented to identify the lead molecules against N-terminal domain (NTD) from SARS-CoV 2 (Severe acute respiratory syndrome coronavirus 2).

Figure 2. Multiple sequence alignment (MSA) of N-terminal domain (NTD) from SARS-CoV 2 (Severe acute respiratory syndrome coronavirus 2) with NTD from different viruses. MSA was done using Clustal Ω , and ESPript 3 was used to generate the figure. The active site residues (Ser51, Phe53, Ala55, Ala90, Arg107, Tyr109, Tyr111, Arg149) are marked in a black triangle. The PDB ID 5N4K belongs to the NTD of human Coronavirus (HCoV) NL63 nucleocapsid protein; 2BXX belongs to the NTD of Infectious bronchitis virus (IBV) coronavirus nucleocapsid protein; 4LM9 belongs to the NTD of HCoV-OC43 nucleocapsid protein; 4UD1 belongs to NTD of Middle East respiratory syndrome (MERS) CoV nucleocapsid protein; 1SSK belongs to the NTD of SARS CoV nucleocapsid protein; and 2OFZ belongs to the RNA Binding Domain of SARS nucleocapsid protein.

Figure 3. Binding interaction of best-docked pose of compounds with NTD protein. (A) GMP (cyan), (B) ZINC000257324845 (orange), (C) ZINC000005169973 (yellow), and (D) ZINC000009913056 (grey). NTD is displayed as cartoon and key residues are shown in green. Hydrogen bonds are represented with gray dashed lines.

Figure 4. The 2D representation of interactions between protein and compound within 4 Å. (A) GMP, (B) ZINC000257324845, (C) ZINC000005169973, and (D) ZINC000009913056. The color-coding features: green for hydrophobic, red for acidic, blue for basic, and cyan for polar residues of NTD SARS-CoV 2.

Figure 5. The structure-based pharmacophore model of GMP inside the active site of NTD, generated with Pharmit server. The pharmacophore features are three hydrogen-bond acceptors, HA (orange); one hydrophobic, HY (green) and one aromatic group, AR (purple) with radius 1.0 Å.

Figure 6. Root Mean square deviation (RMSD) of protein-ligand (s) complexes and ligands only. RMSD plots: A) NTD-GMP (black), NTD-ZINC000257324845 (red), NTD-ZINC000005169973 (green) and NTD-ZINC000009913056 (blue); B) Ligands: GMP (black), ZINC000257324845 (red), ZINC000005169973 (green) and ZINC000009913056 (blue) for the molecular dynamics of 100 ns at 300K.

Figure 7. Root Mean square fluctuation (RMSF) for 123 amino acid residues of NTD-GMP (black), NTD-ZINC000257324845 (red), NTD-ZINC000005169973 (green) and NTD-ZINC000009913056 (blue) complex.

Figure 8. Radius of gyration (Rg) profile of NTD with GMP, ZINC000257324845, ZINC000005169973 and ZINC000009913056 from the molecular simulation of 100 ns at 300K.

Figure 9. Solvent accessible surface area (SASA) graph for protein-ligand (NTD-GMP, NTD-ZINC000257324845, NTD-ZINC000005169973 and NTD-ZINC000009913056) complexes.

Figure 10. Hydrogen bond numbers and distribution patterns of NTD-GMP (black), NTD-ZINC000257324845 (red), NTD-ZINC000005169973 (green) and NTD-ZINC000009913056 (blue) for molecular dynamics of 100 ns at 300K. Number of hydrogen bonds analysis: A) Intra-protein hydrogen bonds and B) inter-molecular hydrogen bonds. Hydrogen bonds distribution pattern for: C) Intra-protein hydrogen bonds and D) inter-molecular hydrogen bonds.

Figure 11. Principal component analysis (PCA) of protein-ligand complexes along PC1 and PC2 for molecular dynamics of 100 ns. The color code for NTD-GMP, NTD-ZINC000257324845, NTD-ZINC000005169973 and NTD-ZINC000009913056 is black, red, green and blue color, respectively.

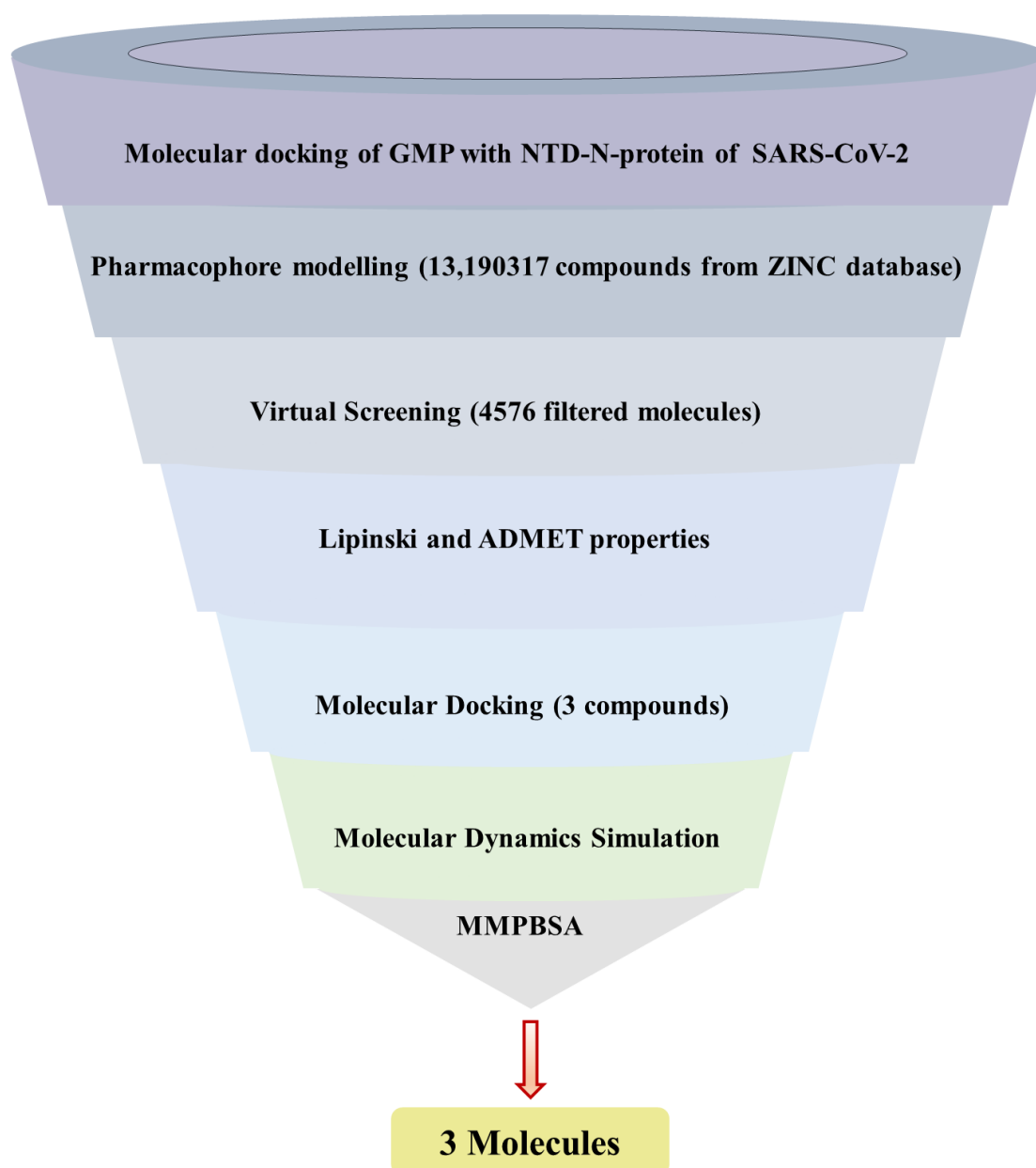


Figure 1

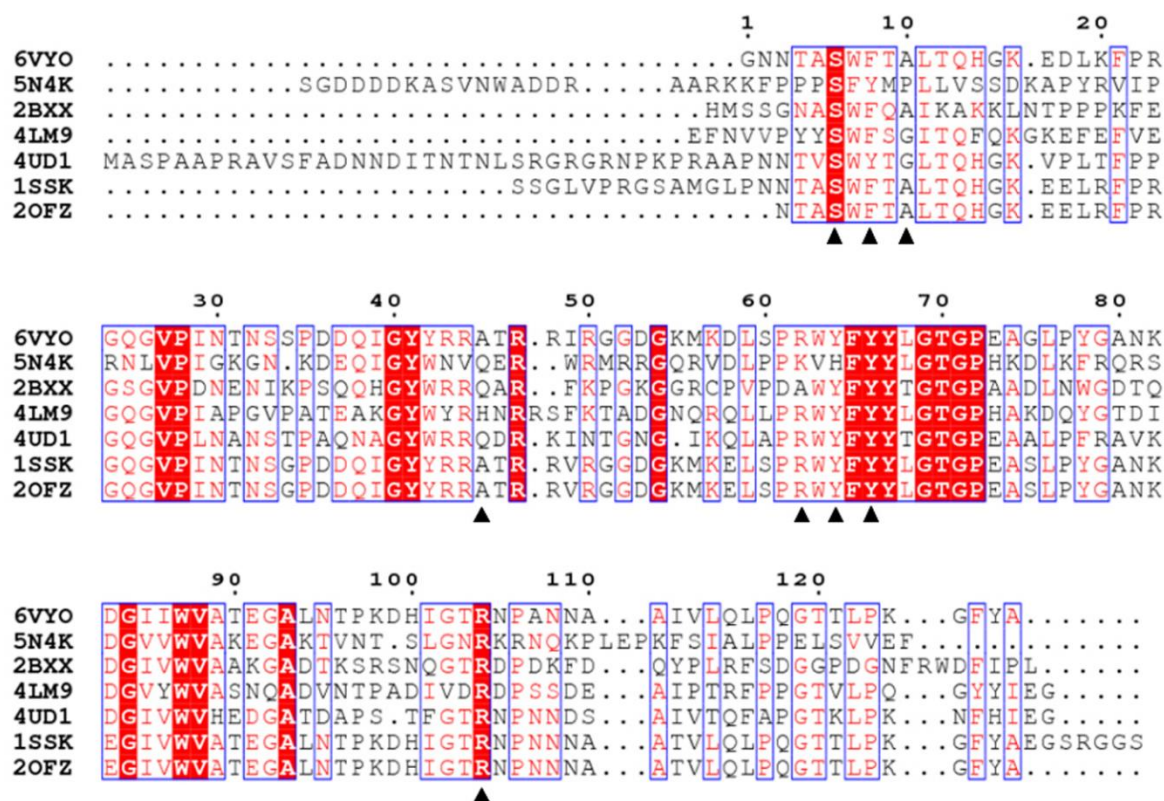


Figure 2

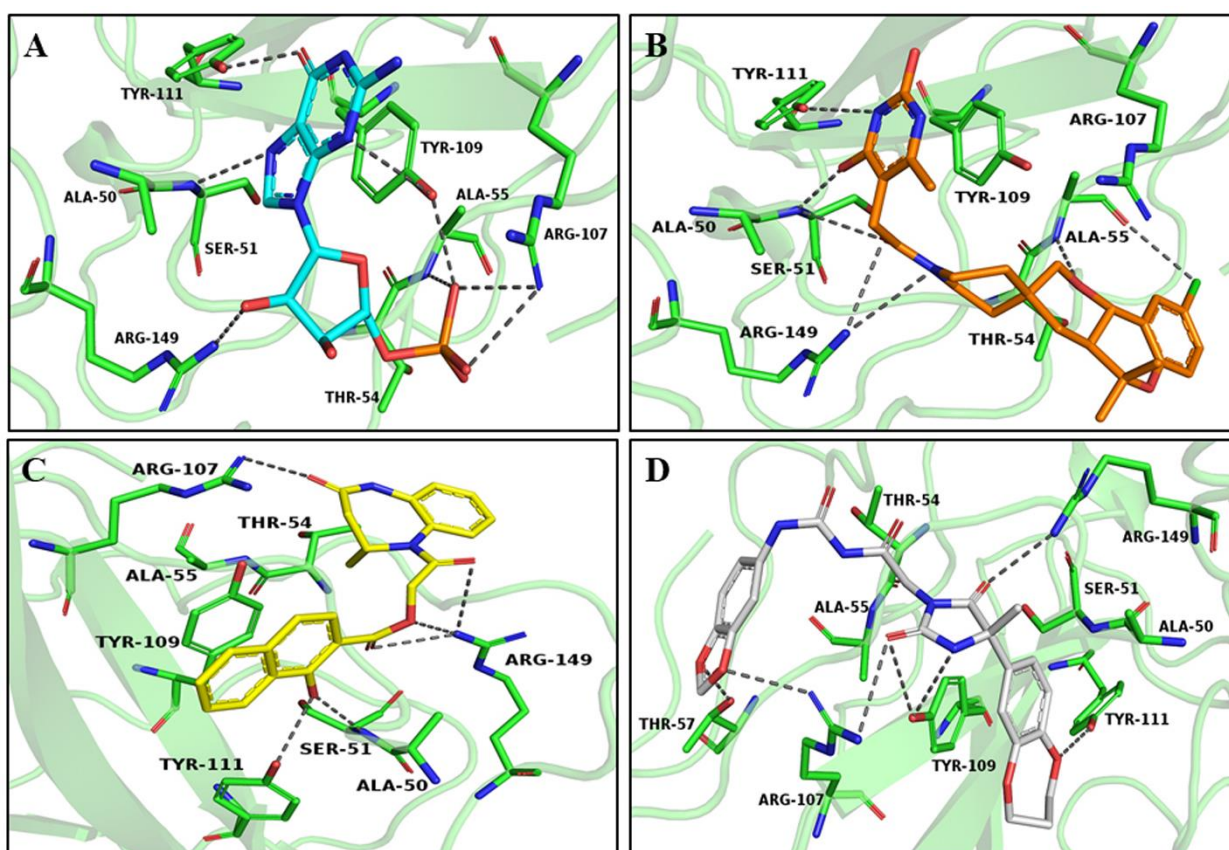


Figure 3

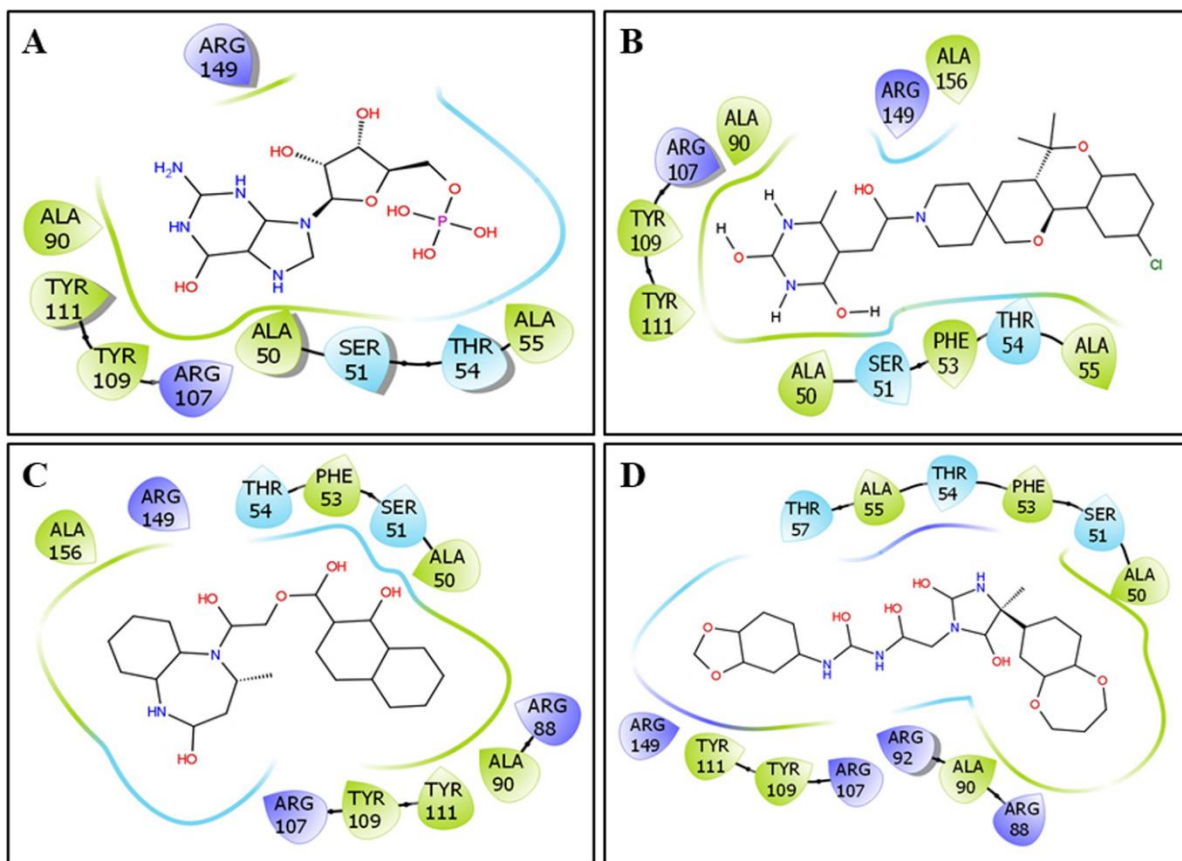


Figure 4

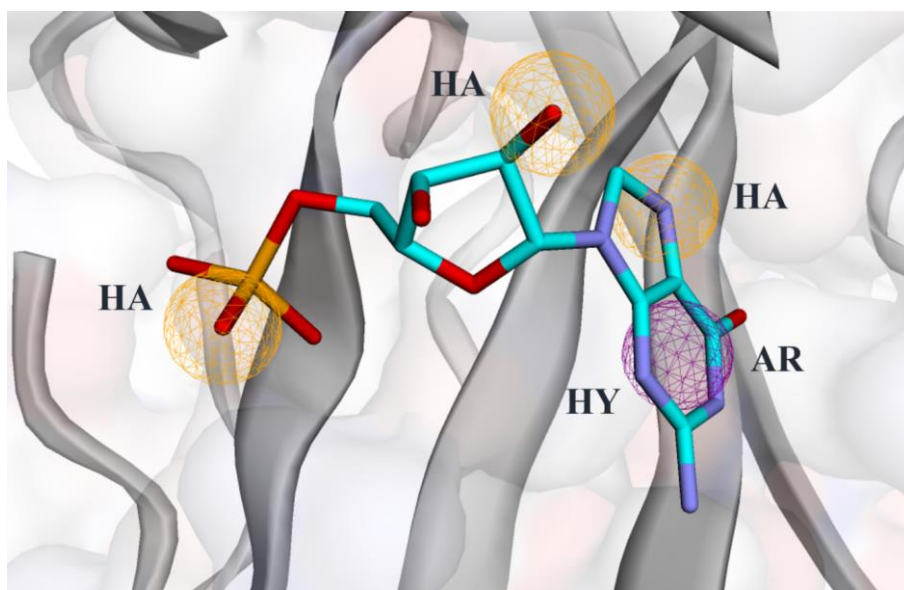


Figure 5

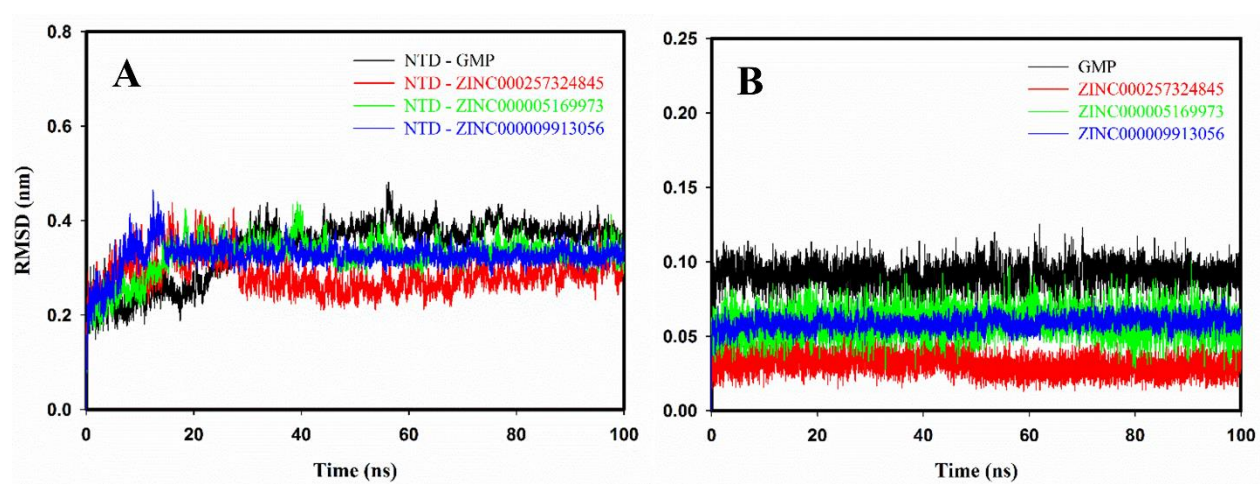


Figure 6

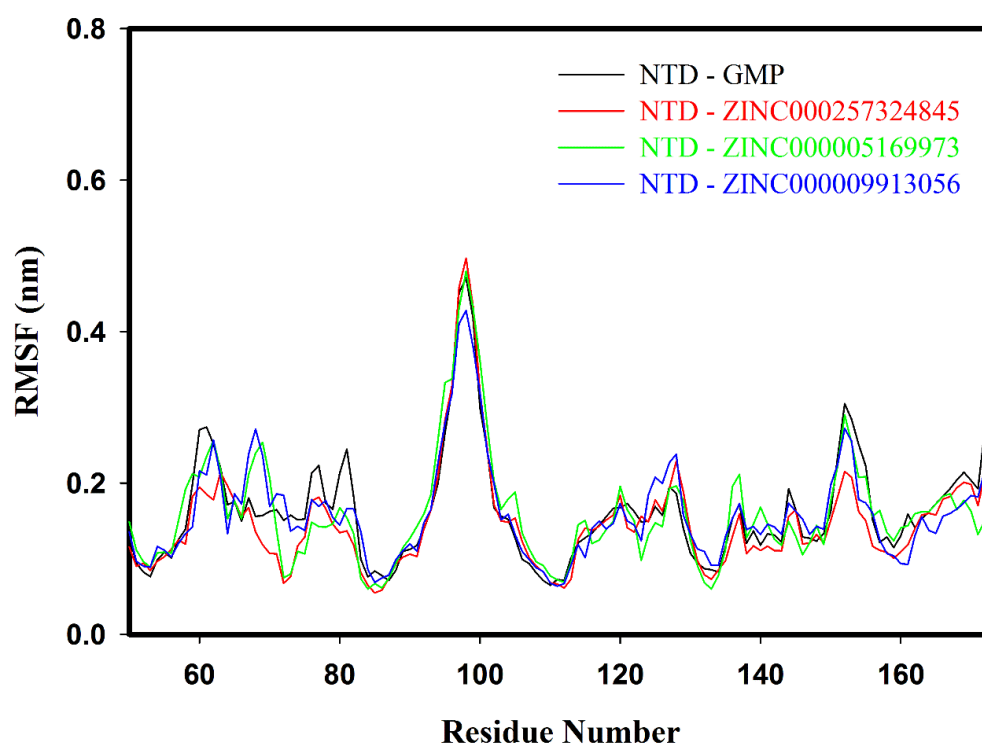


Figure 7

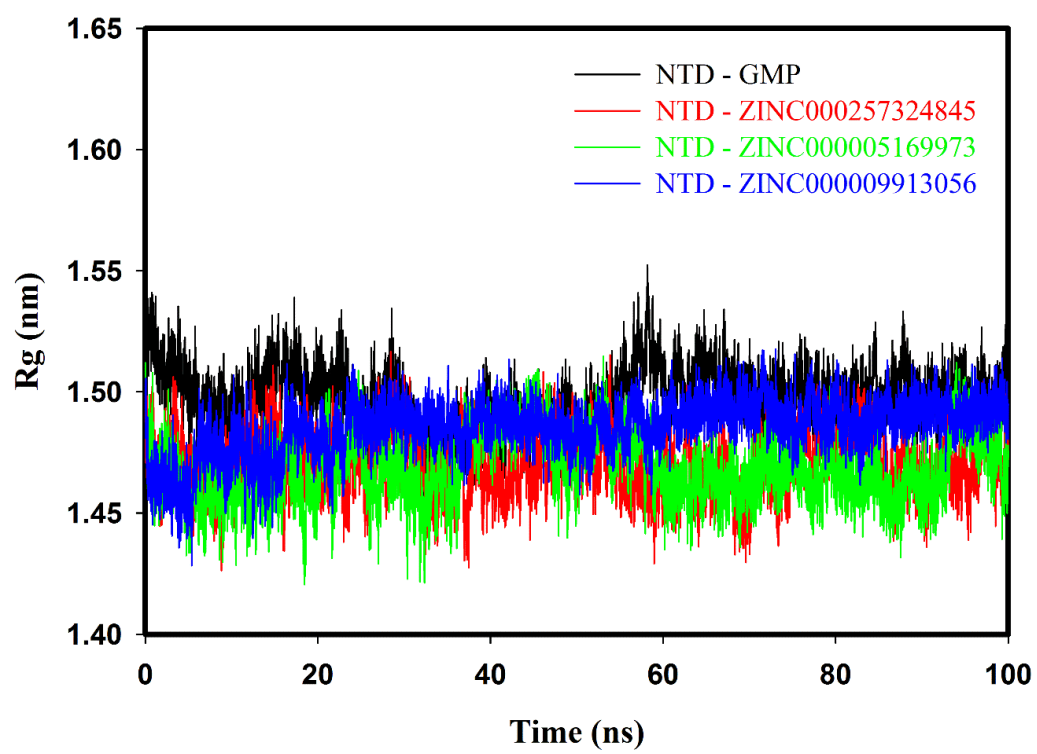


Figure 8

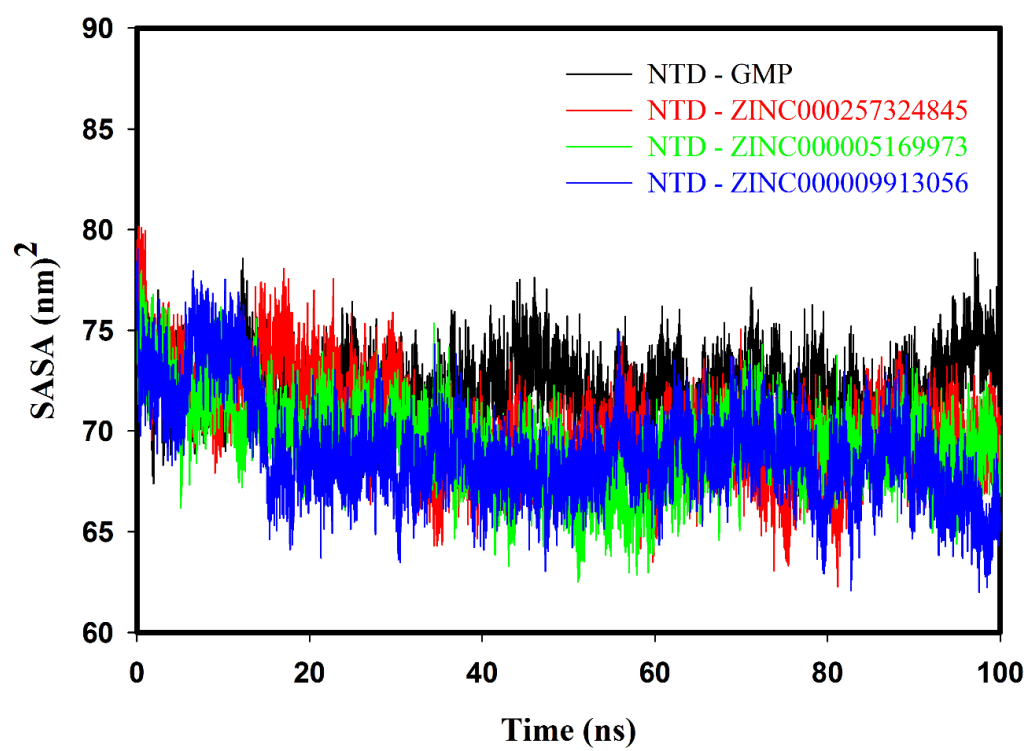


Figure 9

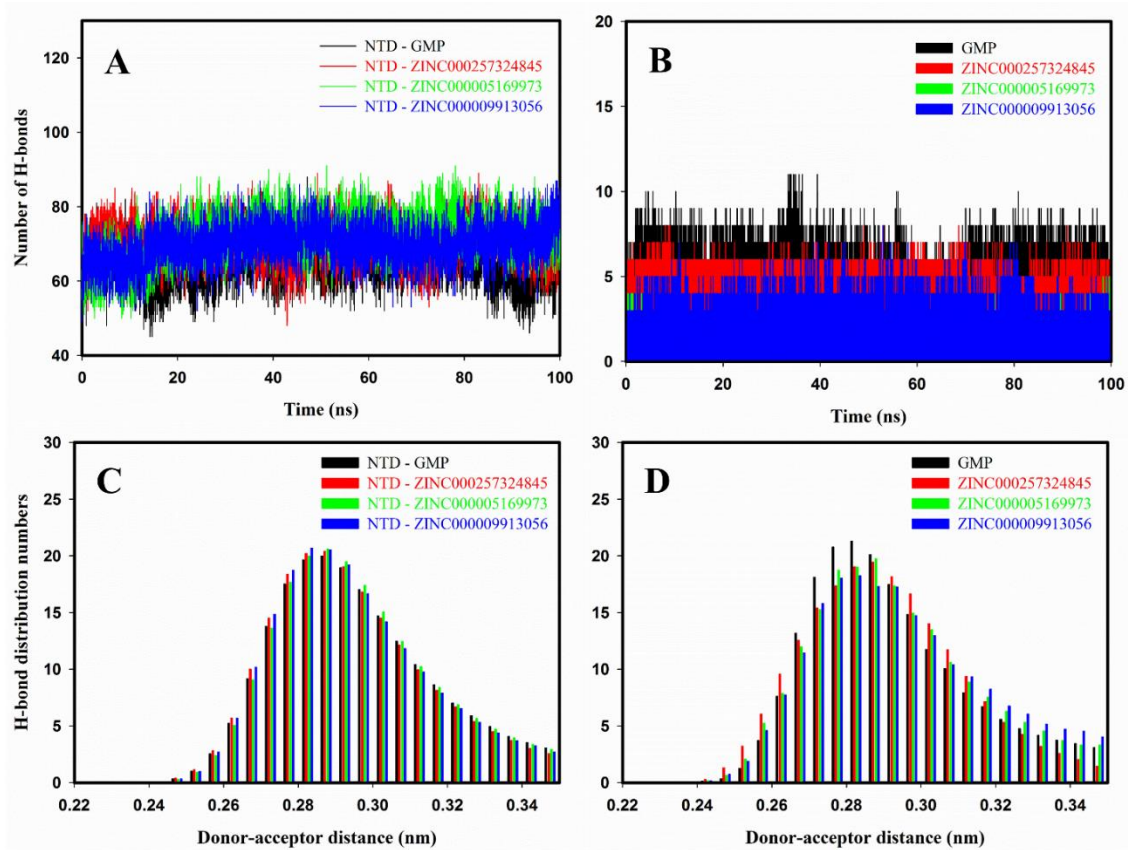


Figure 10

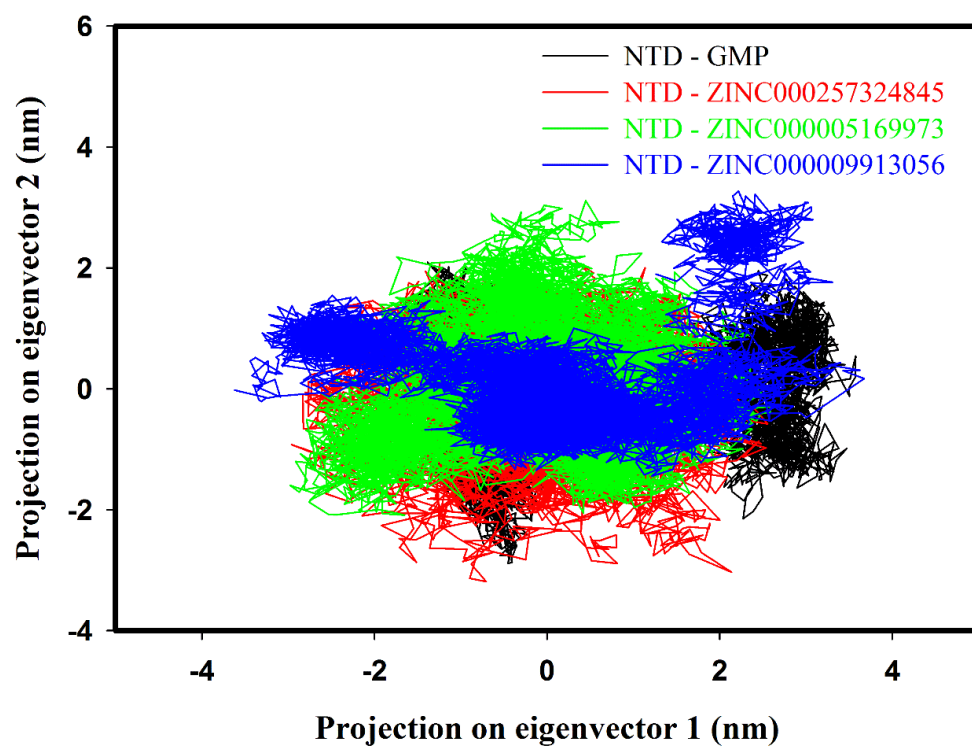


Figure 11

Tables

Table 1. The binding affinities (kcal/mol) and molecular interactions of the selected compounds with active site residues of NTD.

S. No.	Compound	Binding energy (kcal/mol)	
		AutoDock VINA	AutoDock Tool
1	GMP	-5.5	-2.6
2	ZINC000257324845	-8.3	-7.4
3	ZINC000005169973	-8.1	-6.4
4	ZINC000009913056	-8.0	-5.6

Table 2. Drug-likeness properties of the retrieved hit compounds by Lipinski's rule of five.

S. No.	ZINC ID	Molecular Weight (Da)	LogP	Rotatable Bonds	H-bond Acceptors	H-bond Donors
1.	ZINC000257324845	487.984	3.94972	2	7	2
2.	ZINC000005169973	404.422	3.4661	3	5	2
3.	ZINC000009913056	482.449	1.6919	4	8	3

Table 3. The ADMET properties of the selected compounds predicted by using pkCSM.

Properties		ZINC000257324 845	ZINC000005169 973	ZINC000009913 056
Adsorption	Water solubility	-3.782	-4.714	-3.716
	Intestinal absorption (human)	82.038	87.004	73.787
	Skin Permeability	-2.805	-2.748	-2.769
	P-glycoprotein substrate	Yes	Yes	Yes
Distribution	VDss (human)	0.192	-0.483	-0.326
	Fraction unbound (human)	0.127	0.12	0.136
	BBB permeability	-1.205	-0.353	-1.085
	CNS permeability	-3.267	-2.153	-3.451
Metabolism	CYP2D6 substrate	No	No	No
Excretion	Total Clearance	0.082	0.026	-0.455
	Renal OCT2 substrate	No	No	No
Toxicity	AMES toxicity	No	No	No
	hERG I inhibitor	No	No	No
	Oral Rat Acute Toxicity (LD50)	2.538	1.854	1.598
	Oral Rat Chronic Toxicity (LOAEL)	2.147	2.196	1.875
	Skin Sensitisation	No	No	No
	Minnow toxicity	1.338	-0.955	3.862

Table 4. Average values of RMSD (Protein and Ligand), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), and intra-hydrogen bond numbers for NTD-GMP and NTD-inhibitor(s) complexes for the duration of 100 ns.

Averages	GMP	ZINC000257324845	ZINC000005169973	ZINC000009913056
Protein RMSD (nm)	0.34 ± 0.06	0.28 ± 0.03	0.31 ± 0.04	0.31 ± 0.03
Ligand RMSD (nm)	0.09 ± 0.008	0.03 ± 0.007	0.06 ± 0.009	0.06 ± 0.005
RMSF (nm)	0.16 ± 0.07	0.15 ± 0.07	0.16 ± 0.07	0.16 ± 0.07
Rg (nm)	1.50 ± 0.013	1.47 ± 0.013	1.47 ± 0.014	1.48 ± 0.011
SASA (nm) ²	72.08 ± 1.66	70.20 ± 2.41	69.42 ± 2.09	69.02 ± 2.43
Intra protein H Bonds	64.86 ± 5.39	70.27 ± 5.08	71.93 ± 6.08	69.80 ± 5.19

Table 5. Binding free energy (kJ/mol) calculations of NTD-GMP and NTD-inhibitor(s) complexes computed from the last 20 ns (80-100 ns) from molecular dynamics. Van der Waal, electrostatic, polar solvation, SASA and binding energy in kJ/mol of NTD-GMP, NTD-ZINC000257324845, NTD-ZINC000005169973 and NTD-ZINC000009913056 complexes predicted by MMPBSA.

Energy (kJ/mol)	GMP	ZINC000257324845	ZINC000005169973	ZINC000009913056
Van der Waals energy	-166.09 +/- 2.59	-187.83 +/- 1.78	-151.55 +/- 1.48	-116.32 +/- 1.69
Electrostatic energy	-63.84 +/- 2.30	-54.18 +/- 1.36	-57.54 +/- 1.58	-48.81 +/- 1.01
Polar solvation energy	140.87 +/- 3.13	101.17 +/- 1.93	75.10 +/- 1.61	52.88 +/- 1.0
SASA energy	-10.54 +/- 0.19	-16.23 +/- 0.13	-13.01 +/- 0.11	-17.91 +/- 0.11
Binding energy	-99.48 +/- 2.48	-157.02 +/- 1.99	-147.04 +/- 1.78	-129.88 +/- 1.68