# Digging for the Discovery of SARS-CoV-2 nsp12 Inhibitors: A Pharmacophore-Based and Molecular Dynamics Simulation Study

**Fatemeh Sana Askari**
Vista Aria Rena Gene Inc

**Mohsen Ebrahimi**
Golestan University of Medical Sciences

**Jabbar Parhiz**
Golestan University of Medical Sciences

**Mina Hassanpour**
Vista Aria Rena Gene Inc

**Alireza Mohebbi** ( ✉ alirezaa2s@gmail.com )
Vista Aria Rena Gene Inc

**Abbas Mirshafiey**
Tehran University of Medical Sciences

---

**Research Article**

---

# Abstract

**Background:** The severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) is a grave global threat causing Coronavirus Disease 2019 (COVID-19). The therapeutics are urgently needed to cure patients severely infected with COVID-19. The aim of the study was to investigate for potential candidates of nsp12 inhibitors by searching for druggable cavity pockets within the viral protein and drug discovery.

**Methods:** The crystallographic structure of SARS-CoV-2 nsp12 was searched for strong druggable cavity pockets and pharmacophore features by the CavityPlus server. The features were selected for high-throughput screening (HTS) of a chemical library of ZINC natural products and hit identification database by ZINCPharmer. Autodock Vina was furthered utilized for estimation of hits' affinities to nsp12. A lead compound with the highest affinity to nsp12 was simulated dynamically by GROMACS for 10 nanoseconds (ns) to measure the hit stability in complex with nsp12 and conformational changes.

**Results:** 1 of 6 cavities with the highest score was selected for extraction of pharmacophore features and hit-identification. 9 pharmacophores were screened, and a total of unique 1274 hits were identified. One compound, ZINV03977803, with an -11.0 Kcal.mol$^{-1}$ affinity was selected as the lead compound for molecular dynamic simulation (MDS). The results showed stable interaction between ZINV03977803 and nsp12 during 10 ns of simulation. The room-mean-square of deviation (RMSD) measure showed dramatically high conformational changes in the complex of ZINV03977803 and nsp12 compare two the viral proteins alone.

**Conclusions:** The lead compound ZINV03977803 showed stable interaction with higher potential and hydrogen bonding with the catalytic subunit of SARS-CoV-2, nsp12. It could also inhibit the SARS-CoV-2 life cycle by direct interaction with nsp12 and inhibits RdRp complex formation.

# Background

The severe acute respiratory syndrome *Coronavirus* 2 (SARS-CoV-2), the etiological agent of *Coronavirus* disease 2019 (COVID-19), is an ongoing pandemic that has led to over 200 million infections and more than 4 million dead around the world till August 15, 2021 (https://www.worldometers.info/coronavirus/). A single-stranded, positive-sense RNA genome of ~ 30 kb codes for 14 open reading frames (ORF) [1]. At least 26 proteins are responsible for the structural, non-structural proteins (nsp) and accessory proteins required for the viral assembly, replication, and poly-protein modifications. Two ORFs, ORF1a and ORF1ab, are the largest transcriptional regions on the SARS-Cov-2 genome that are enclosed at both ends by two untranslated regions (UTR). The whole genome of the virus is comprised of two overlapping polyproteins denoted pp1a and pp1ab. These two polyproteins are post-processed by two virally coded accessory enzymes, PLpro (nsp3) and Mpro (nsp5), into 16 nsp (nsp1-16). The rest of the genome encodes 4 structural Spike (S), Envelope (E), Matrix (M), and Nucleocapsid (N) proteins, as well as six accessory proteins. All three categories of SARS-CoV-2 proteins have been targeted for antiviral measures. Structural protein S protein, which is responsible for the virus binding to the host cell receptor angiotensin-converting enzyme 2 (ACE2), has been one of the essential viral proteins for therapeutic studies [2–4]. Nsp3 and nsp5 proteases, as two mail accessory proteins, are also targeted for anti-COVID-19 studies [5–7]. Among all non-structural proteins, catalytic nsp12, nsp7, and nsp8 subunits form the RNA-dependent RNA polymerase (RdRp) [1]. RdRp is responsible for replication and transcription of the viral genome and is a very interesting target for in silico drug discovery methods. Therefore, targeting the nsp12 active site might suppress the formation of the RdRp complex, incorporate free nucleotides, translocation the resides nucleotides, and suppress the proofreading process.

As a suitable drug discovery target, the catalytic nsp12 subunit of the RdRp complex is comprised of 932 amino acid residues and is consisted of an N-terminal nidovirus-unique RdRp-associated nucleotidyltransferase (NiRAN) domain (amino acids 60–249) and RdRp domain (366–920). The RdRp domain is contained three fingers, palm, and thumb subunits [8]. The only FDA-approved therapeutic for COVID-19 is the active remdesivir triphosphate (RTP), which stalls RdRp from the growing RNA. The mechanism by which remdesivir inhibits viral polymerization is induced in both RNA-dependent manner and an RNA-independent mechanism [9]. Molnupiravir is another nucleoside analogue in the clinical trial phase II and recognized by cell-free CoV-RdRp-Gluc assay to inhibit the RdRp more efficiently than remdesivir [10]. Furthermore, the emergence of escape mutations at the catalytic nsp12 subunit is rare and very conserved among beta-coronaviruses [11]. Therefore, nucleot(s)ide therapy to inhibit the SARS-CoV-19 RdRp seems reasonable. However, recent findings demonstrated that exoribonuclease proofread activity of nsp14 co-expressed with nsp10 confer high resistance to nucleot(s)ide therapy in vitro [10].

The rise of highly contagious SARS-CoV-2 strains has challenged the vaccination efficiency due to the large gap of vaccination interval in developed countries and developing/under developing countries with higher population density. Accordingly, therapeutic approaches are necessitated to improve the mortality rates of acute SARS-CoV-2 infections. Different strategies are implemented to introduce anti-SARS-CoV-2 therapeutics. These approaches are repurposing FDA-approved Nucleot(s)ide analogues like Lamivudine to inhibit SARS-CoV-2 nsp12 [12], herbal and traditional herbal medicine, and medical plant extracts [13–15], and drug discovery of virtually available libraries of small

molecules [16, 17]. SARS-CoV-2 undergoes high rates of nucleotide mutations, and it may influence the therapeutic potential of Nucleot(s)ide analogues.

Furthermore, the binding of the small molecules to the targeted protein is more stable and less affected by nucleotide substitutions. We recently published an in silico work introducing a potential small molecule, Compound 38, interfering with the virus attachment and fusion, and experimental experiments are undergoing on Compound 38 [18]. Due to the probability of the emergence of drug-resistance strains of SARS-CoV-2, we have decided to introduce dual therapy of COVID-19 to reduce the chance of drug irresponsiveness. This approach could be implemented in the in-silico process of drug discovery and experimental works.

Therefore, following the previous study, we have decided to introduce a second potent inhibitor of SARS-CoV-2 to block nsp12 protein. Dual therapeutics are much more promising to avoid the emergence of evolved resistance mutants, and it also improves the chance of inhibiting the virus at two different levels of binding/attachment and replication. The results showed a promising lead compound, ZINC03977803, with stable binding to the nsp12 and less affected to the already known nsp12 mutants.

## Materials And Methods

## Preparation of SARS-CoV-2 Spike glycoprotein

The crystallographic structure of human SARS-CoV-2 nsp12 was obtained from the Protein databank (PDB) [19]. PDB databank was searched for the crystallographic structure of the full-length nsp12 protein. Two structures with PDB IDs 7BTF and 6M71 with resolutions 2.90 and 2.95 were provided for the study [20]. PDB files of the structure were curated by using UCSF Chimera v1.10.2. Accordingly, only chain(s) containing nsp12 amino acid residues were cleaned from other non-standard or other protein chains. To achieve better resolution, the structures were energy minimized and checked for amino acid atoms using Swiss PDB viewer (SPDBV) v4.1.0, as reported before [21, 22].

## Finding druggable cavity pockets of SARS-CoV-2 RdRp

CavityPlus server was used for precise and robust protein cavity detection and functional analysis of the energy minimized SARS-CoV-2 nsp12 3D structure [23]. The server's module Cavity was used to detect potential binding sites on the surface of the nsp12 protein. The binding sites were ranked with Druggability and DrugScore. The strongest cavity with the highest DrugScore was further analyzed by the server's modules ChemPharmer, CorrSite, and CovCys.

ChemPharmer was used to extract pharmacophore features within the selected cavity through the receptor-based pharmacophore modeling program, Pocket. Potential allosteric ligand binding sites were identified by the CorrSite module based on motion correlation analysis between allosteric and orthosteric cavities (cut-off Z-score of 0.5). CovCys was used to automatically detect druggable cysteine residues for covalent ligand design, which is especially useful for identifying novel binding sites for covalent allosteric ligand design. The server setting was kept as default, as we have reported before.

## Optimization of Pharmacophore features and hit identification

ChemPharmer may result in several pharmacophore features. We have manually selected features with relative coordinates to make different pharmacophores covering the surface cavity of the druggable binding site(s). Hit identification was performed by importing the .pdb files containing the pharmacophore features into the ZINCPharmer server [24]. The files were double-checked by the server, indicating pharmacophore feature classes are as same as the .pdb files.

ZINCPharmer contains eight databases of chemical compounds and small molecules, including ZINC Purchasable (206,433,075 conformers of 21,777,093 compounds), ZINK Purchasable Thiols (321,022 conformers of 39,257 compounds), Alex Doemling UDC (946,136 conformers of 95,075 compounds), ZINC Drug Database (20,432 conformers of 2,770 compounds), ZINC In Man (77,269 conformers of 10,813 compounds), ZINC Drug Database of Metabolites (32,339 conformers of 5,030 compounds), ZINC Natural Derivatives (185,393 conformers of 25,164 compounds) and ZINC Natural Products (1,374,813 conformers of 172,324 compounds).

Each database was applied to every pharmacophore. For pharmacophore with the least features that might lead to high numbers of hits, the server settings were changed to Molecular weight ≤ 500 da, Rotatable bonds ≤ 10, Max RMSD = 500, and Max Total Hit = 500. OpenBable command line was used to check the duplication of identified hits for each pharmacophore [25].

## Protein sequence retrieval and finding amino acid substitutions

Literature was searched for the amino acid substitution at the SARS-CoV-2 nsp12 protein till April 20, 2021. UCSF Chimera v1.10.2 was used for making computationally mutant constructs. Accordingly, the protein was structurally edited using Dunbrack's rotamer (8) to choose the

most probable rotamer(s) for the substituted amino acid residue(s). The computationally constructed mutants were checked for clashes with adjustment atom residues. The clashes were curated by energy minimization by Swiss-PBD view v4.1.0 [21].

# Molecular docking and hit optimization

The affinity of identified hits was further investigated by using molecular docking. For this purpose, the open-source tool Autodock Vina [26] was used in the setting of PaDEL-ADV (http://www.yapcwsoft.com/dd/padeladv/), as we reported before [18]. Vina is a fast and accurate tool for ligand-receptor docking, and with PaDEL-ADV, it allows high-throughput screenings of several ligands in one run. Molecular docking was performed on the druggable binding site with the highest DrugScore. In this regard, the viral nsp12 was treated as the receptor in the MGLTools 1.5.6 software (Molecular Graphics Laboratory, The Scripps Research Institute). Accordingly, a grid-box was defined in 3D dimension to encompass the entire cavity.

The hit(s) with the highest affinity to the nsp12 protein were selected for further analysis. Accordingly, the binding site(s) of the chosen hits were highlighted and compared to the viral nsp7 and nsp8 proteins in the crystallographic structure of the viral proteins (PDB ID: 7BTF) by using UCSF Chimera v1.10.2.

# ADMET profiling

Toxicity estimation software tool (TEST) v4.2.1 was used to predict the toxicity of the lead compounds. The highest prediction accuracy and improved applicable domain were obtained by the tool's consensus method. The validation of each predicted toxicity was evaluated using external statistical validation of training- and test sets within the software package. Six different predictive tests were evaluated for each compound, including 96-hour fathead minnow LC50, 48-hour Daphnia magna LC50, 40-hour Tetrahymena pyriformis IGC50, Oral rat LD50, Developmental toxicity, and Ames mutagenicity.

# Molecular dynamic simulation

The simulation of nsp12 in the presence and absence of the lead compound was performed using the GROMACS version 2021.2 [27]. The CHARMM36 force field was applied to the lead compound and complex structure of nsp12/lead [28]. The compound topology was generated by using the CGenFF server [29–32]. The explicit compound coordination was matched in the topology file by sorting the list of bonds by using a Perl script provided by the GROMACS team (sort_mol2_bonds.pl). All starting structures were solvated in a simple cubic water box under periodic boundary conditions using a 1.0 nm distance from the protein to the box faces. The system was then neutralized by $Cl^-$ or $Na^+$ ions. Following steepest descent energy minimization, the systems were equilibrated under a constant number of particles, volume, and temperature (NVT) conditions for 60 ps at 300 °K, followed by 60 ps under constant particle pressure and temperature (NPT) conditions. Finally, 10 ns MD was produced to analyze the stability of each system. Lenard Jones and Coulomb Potentials (Kj/mol) were estimated to analyze the lead compound interaction with the nsp21. MD simulation was run on a system utilizing a brand of Intel(R) Core(TM) i3-4160 CPU @ 3.60GHz accompanied by a GPU brand of AMD Radeon RX 5500 XT 8G OC GDDR6.

# Results

The crystallographic structure of nsp12 with PDB ID: 6M71 was discarded for the study due to missing amino acid sidechain atoms. These amino acids were including, Phe70, Lys73, Arg74, Glu83, Lys98, Phe101, Phe102, Ile114, Arg365, and Asp824. Therefore, screening for the druggable cavities was proceeded by the other crystallographic structure of SARS-CoV-2 nsp12 structure with PDB ID: 7BTF. Total energy was minimized to -31860.766 KJ.mol$^{-1}$. The energy minimized nsp12 structure was used for druggable cavity detection.

The detection of potential cavities within the nsp12 protein showed four strong druggable binding sites (Fig. 1). Furthermore, six cavities binding sites with minimum possibility of being druggable with minimum and maximum DrugScore 161.0 and 22.0 were highlighted. Twenty-two other binding site cavities had a weak possibility of being druggable. The only cavity No. 2 with the highest DrugScore (23739.0) was selected for pharmacophore modeling. As shown in Fig. 1, the druggable cavity pocket resides in the catalytic site of the nsp12 protein, where fingers, palm, and thumb meet each other.

Pharmacophore class of cavity No. 2 had several pharmacophore features including, twenty-one H-bond acceptor (HBA) centers, twenty H-bond donor (HBD) centers, forty-two H-bond roots, eighteen Hydrophobic centers (HC), twelve negative electrostatic centers (NEC), and five Positive electrostatic centers (PEC). As shown in Fig. 2, we manually selected adjacent pharmacophore features surrounding the cavity No. 2 surface. For simplicity of selection of neighboring pharmacophore features, H-bond roots, PEC, and NEC were ignored. The pharmacophore features and their coordinates are provided in the Supplemented material (Table S1). Ten pharmacophores were adopted based on the neighboring pharmacophore features, each containing 5 to 9 features (Fig. 2). The largest pharmacophore was comprised of three HBD centers, three HBA centers, and three hydrophobic centers. CorrSite results demonstrated no allosteric site within cavity No. 2. However, cavities 3, 8, 7, 6, 12, and 9 contained allosteric sites with Z-score of 2.68, 2.5, 1.6, 1.23, 1.05, and 0.77. CovCys results also demonstrated

Cys622 within the SARS-CoV-2 nsp12 has a 92% possibility of being a covalent targetable cysteine residue, which had a significant ligandablility potential (pKdAvg = 6.74).

The virtual high-throughput screening (HTS) of the ZINCPharmer databases resulted in 3,932 hits from the ten different pharmacophores. Table 1 shows the unique number of identified hits from HTS. After duplication removal, 1,274 unique hits were remained to be docked with the SARS-CoV-2 nsp12.

Table 1
The results of unique identified hits obtained by HTS of the pharmacophores through ZINCPharmer database

| Pharmacophores | ZINCPharmer databases | | | | | | | | Duplications§ | Unique Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | ZINC Purchasable | ZINK Purchasable Thiols | Alex Doemling UDC | ZINC Drug Database | ZINC In Man | ZINC Drug Database of Metabolites | ZINC Natural Derivatives | ZINC Natural Products | | |
| Pharma1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Pharma2 | 13 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 13 |
| Pharma3 | 2 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 9 | 4 |
| Pharma4 | 34 | 0 | 9 | 1 | 19 | 3 | 19 | 5 | 33 | 57 |
| Pharma5 | 210 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 130 | 83 |
| Pharma6¥ | 500 | 51 | 2 | 271 | 500 | 74 | 500 | 500 | 1,906 | 492 |
| Pharma7 | 212 | 0 | 0 | 4 | 8 | 1 | 3 | 50 | 146 | 132 |
| Pharma8 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 27 | 10 |
| Pharma9¥ | 500 | 121 | 9 | 6 | 12 | 0 | 43 | 182 | 396 | 477 |
| Pharma10 | 6 | 0 | 0 | 0 | 2 | 0 | 0 | 5 | 8 | 5 |
| Total | 1,502 | 172 | 11 | 286 | 541 | 77 | 567 | 760 | 2,659 | 1,274 |

¥ For pharmacophores no. 6 and 9 ZINCPharmer search limits, including Molecular weight ≤ 500 da, Rotatable bonds ≤ 10, Max RMSD = 500 and Max Total Hit = 500 were applied.

§ Duplication was automatically selected by OpenBabel command line based on the same strings and conformers.

For molecular docking of the identified hits, a gridbox was defined by having the center coordinates of 125.14, 126.874, and 147.48 along the x-, y-, and z-axis, respectively. The gridpoints' size was 76, 88, and 58 along the x-, y- and z-axis, respectively. The spacing was kept to default values. The docking results of the hits within the predicted druggable cavity pocket were interesting. The hits obtained from virtual screening resulted in a lead (ZINC03977803) from the pharmacophore number four with the affinity of -11 Kcal.mol$^{-1}$ (Fig. 3). Analyzing the interaction of ZINC03977803 to the nsp12 showed the involvement of amino acid residues proximate to the finger motifs of the viral protein. The contact between amino acid residues of the viral protein with the hits was investigated with VDM overlap ≥ -0.4 angstroms. The amino acids were including, Ser501, Ala502. Gly503, Phe504, Pro505, Asn507, Ile539, Thr540, Gln541, Met542, Asn543, Leu544, Met668, and Cys669. The results of interactive residues within nsp12 and other lead compounds of other pharmacophores are provided in the Supplemented materials (Table S2). Additionally, only one compound was found due to screening of the chemical library through the pharmacophore number 1. The compound had been discarded because of the lowest interaction energy (-4.7 Kcal.mol$^{-1}$) with the nsp12. The predicted toxicity (Table 1) of the lead compounds showed the lowest toxicity of ZINC03977803. Ames mutagenicity was positive for ZINC04096710 and ZINC54120801 (Table 2).

Table 2
The predicted toxicity of the lead compounds by using T.E.S.T

| Leads | Pharmacophore No.¥ | Formula | Fathead minnow $LC_{50}$ (96 h) (mg/L) (mg/L) | Daphnia magna $LC_{50}$ (48 hr) (mg/L) | T. pyriformis $IGC_{50}$ (40 hr) (mg/L) | Oral rat $LD_{50}$ (mg/kg) | Developmental toxicity | Ames mutagenicity |
|---|---|---|---|---|---|---|---|---|
| ZINC63303082 | 2 | $C_{19}H_{21}N_5O_2$ | 0.14 | 12.25 | N/A | 1732.74 | 0.90 | Negative |
| ZINC04096710 | 3 | $C_{27}H_{48}O_3$ | 4.31 | 3.89 | 4.43 | 194.50 | 0.75 | Positive |
| ZINC03977803 | 4 | $C_{28}H_{32}O_{15}$ | 1.12E-03 | 340.43 | N/A | 2972.04 | 0.52 | Negative |
| ZINC13450627 | 5 | $C_{25}H_{41}NO_6$ | N/A | N/A | N/A | 207.69 | 0.79 | Negative |
| ZINC19796022 | 6 | $C_{22}H_{21}N_2O_8Cl$ | N/A | 26.14 | N/A | 2402.53 | 1.00 | Negative |
| ZINC11592622 | 7 | $C_{22}H_{24}N_2O_9$ | 1.12 | 4.29 | N/A | 1165.80 | 0.88 | Negative |
| ZINC35570685 | 8 | $C_{30}H_{48}O_5$ | 1.06 | 1.92 | N/A | 491.96 | 0.91 | Negative |
| ZINC54120801 | 9 | $C_{20}H_{15}N_3O_2S$ | 0.15 | 6.17 | N/A | 2028.77 | 0.84 | Positive |
| ZINC67912716 | 10 | $C_{19}H_{28}O_7$ | 1.51 | 21.36 | N/A | 155.24 | 0.71 | Negative |
| ¥ Pharmacophore No.1 was discarded due to its low affinity to the nsp12 | | | | | | | | |

ZINC03977803 was chosen for continuing the study. The literature was searched for reported amino acid substitutions of SARS-CoV-2's nsp12 worldwide. As a result, forty-five nsp12 mutants were created (Table 3). The lead compound had the lowest affinity to mutant N489K (-9.6 Kcal.mol$^{-1}$). Additionally, the affinity of the compound to the mutants S6L, T141I, P328S, A526V and T806I was – 11.8 Kcal.mol$^{-1}$, -11.5 Kcal.mol$^{-1}$, -11.3 Kcal.mol$^{-1}$, -11.8 Kcal.mol$^{-1}$ and – 11.5 Kcal.mol$^{-1}$.

Table 3
The nsp12 amino acid substitutions and their impact on
the binding affinity of ZINC03977803 to the mutant nsp12

| Mutant | Position | Reference | Affinity (Kcal.mol$^{-1}$) |
|---|---|---|---|
| A/V | 97 | (1,2) | -11.0 |
| P/L | 323 | (1−3) | -10.8 |
| P/L$^{¥}$ | 314 | (3,4) | -10.5 |
| S/L | 6 | (2) | -11.8 |
| G/Y | 25 | (2) | -11.0 |
| T/I | 26 | (2) | -10.7 |
| G/V | 44 | (2) | -10.0 |
| Y/C | 80 | (2) | -11.0 |
| T/I | 85 | (2) | -10.6 |
| K/R | 91 | (2) | -10.8 |
| K/R | 103 | (2) | -10.6 |
| M/V | 110 | (2) | -10.4 |
| T/I | 141 | (2) | -11.5 |
| D/G | 154 | (2) | -10.3 |
| D/V | 161 | (2) | -10.8 |
| G/S | 179 | (2) | -11.0 |
| G/C | 228 | (2) | -10.2 |
| S/N | 229 | (2) | -10.8 |
| K/N | 263 | (2) | -10.9 |
| T/M | 276 | (2) | -10.7 |
| P/S | 328 | (2) | -11.3 |
| A/V | 382 | (2) | -10.5 |
| A/S | 400 | (2) | -10.6 |
| V/I | 405 | (2) | -10.6 |
| V/F | 405 | (2) | -10.5 |
| A/V | 406 | (2) | -10.8 |
| A/V | 443 | (2) | -11.0 |
| A/V | 449 | (2) | -10.5 |
| I/V | 466 | (2) | -10.5 |
| N/K | 489 | (2) | -9.6 |
| A/V | 526 | (2) | -11.8 |
| R/H | 640 | (2) | -10.6 |
| A/V | 656 | (2) | -10.7 |
| M/I | 668 | (2) | -10.5 |
| A/S | 699 | (2) | -10.3 |
| N/T | 734 | (2) | -10.6 |

| Mutant | Position | Reference | Affinity (Kcal.mol$^{-1}$) |
|--------|----------|-----------|----------------------------|
| G/S | 774 | (2) | -10.5 |
| D/Y | 804 | (2) | -10.9 |
| T/I | 806 | (2) | -11.5 |
| K/R | 807 | (2) | -10.7 |
| H/L | 810 | (2) | -10.2 |
| G/S | 823 | (2) | -10.8 |
| H/Y | 872 | (2) | -11.0 |
| D/Y | 879 | (2) | -10.9 |
| W/C | 916 | (2) | -10.7 |

The residues of nsp12 with close contacts to the hit compound were Ser501, Phe504, Pro505, Asn507, Ile539, Gln541, Met542, Asn543, Leu544, Met668, and Cys669. The mean distance between the O and C atoms of ZINC03977803 in contact with nsp12 residues was 2.46 ± 0.73 Å (Fig. 4). As shown in Fig. 4B, Pro505 in the nsp12 protein was in close contact with ZINC03977803 with is also resided in the neighboring site of nsp12 and nsp7/8 proteins.

MDS was performed in 10 ns to evaluate the conformational changes of nsp12 protein in complex with the lead compound, ZINC03977803. RMSD and Rg measures were evaluated for speculating the stability and proper folding of the protein (Fig. 5). The only significant fluctuation of RMSD and Rg was observed at the beginning of the simulation due to the system's energy minimization and imperfect position restraints. GROMACS rms, energy, hbond, and mindist utility modules were used to evaluate the stability of the lead compound within the cavity, the total interaction energy between the lead compound and nsp12, hydrogen bonding, and distance between groups of the lead and protein backbone during 10 ns of MD simulation.

As it is shown in Fig. 6A, RMSD fluctuation of the lead compound was according to the nsp12's, suggesting stable binding of ZINC03977803 lead compound within the binding cavity pocket of nsp12. The interaction energy of ZINC03977803 to the nsp12 was measured by short-range (SR) Lennard-Jones (LJ-SR) and Coulomb (Coul-SR) potentials. The Sum of LJ-SR and Coul-SR was reported as total interaction energy (Fig. 6B). Accordingly, the mean of LJ-SR and Coul-SR were − 123.53 ± 17.94 KJ/mol and − 90.38 ± 37.03 KJ/mol. Total interaction energy was estimated as -213.92 ± 27.49 KJ/mol.

Hbond showed H19 of ZINC03977803 as the donor site was involved in Ala406N and Gly671N residues on the nsp12. The distance between H19 of the lead compound with amine groups of Ala406 and Gly671 was 0.31625 ± 0.02 nm. The hydrogen bond pairing during 10 ns of MDS is shown in Fig. 6C. Distance and number were also measured between the lead compound ZINC03977803 and the nsp12 backbone (Fig. 6D, E). The results showed a mean distance of 2.86E-01 ± 0.02 nm and 29.12 ± 10.57 number of contacts < 0.4 nm between ZINC03977803 and the nsp12 backbone. These results suggest stable close contacts between ZINC03977803 and nsp12 during 10 ns of MDS.

## Discussion

There is an excellent opportunity to use natural products and plant-derived small antiviral molecules to deal with challenging viral diseases [33–35]. In this regard, in silico approaches have the advantage of using different chemical libraries for drug discovery against SARS-CoV-2 with fewer resources. Accordingly, some studies demonstrated the potential of plant-derived natural products in suppressing the SARS-CoV-2 RdRp [36–39]. In the present study, a rational drug discovery process was implemented to find the possible druggable cavities within SARS-CoV-2 nsp12, the catalytic subunit of the RdRp complex. As a result, ten druggable sites were screened to fining hits of an extensive chemical library of natural compounds and small molecules of the ZINC database. Only one lead compound, ZINC03977803, was highlighted to inhibit the formation of RdRp potentially. As it has been shown that amino acid substitution might interfere with drug binding, we measured the impact of known nsp12 substitution on the binding affinity of ZINC03977803 to nsp12.

The substitutions at the amino acid residues of the nsp12 protein may affect the affinity of drugs or even confer drug resistance. Therefore, the role of the reported substitutions at the amino acid sites of the nsp12 was measured in the affinity of the lead compound ZINC03977803 to the nsp12. The results demonstrated that the residue N489 is very important at the druggable cavity of nsp12 since its substitution to N489K resulted in decreased affinity to the protein. The substitution is only reported in a genome-wide associated study in a sample from Iceland [40]. The rest of the mutants did not substantially affect the binding affinity of the lead compound ZINC03977803 to the nsp12,

suggesting a conserved unique druggable cavity for drug discovery of SARS-CoV-2 nsp12. As a finding, amine groups of Ala406 and Gly671 at the close contact in the binding site of the lead compound were the hydrogen bond donors. The hydrogen bindings were also stable during dynamic simulation. This suggests a stable, strong contact between the lead compound and nsp12, which makes it a potent inhibitor of catalytic activity of SARS-CoV-2 replication and transcription (-213.92 ± 27.49 KJ/mol).

Further important finding was the presence of the residue Pro505 at the binding site of the ZINC03977803, where the nsp12 make the multimeric RdRp complex with heterodimer nsp7/nsp8 subunits. This would highlight the possible role of the lead compound, ZINC03977803, in suppressing the formation of the SARS-CoV-2 RdRp complex. This is also supported by the stable close RMSD fluctuation of the lead compound alongside the nsp12 and significant RMSD changes while they are in the complex.

The introduced lead compound, ZINC03977803, had no known activity in the ChEMBL or ZINC database. So, in another study, we aim to investigate the anti-SARS-CoV-2 activity of the compound. Recently, we have reported Compound 38 as a possible inhibitor of SARS-CoV-2 attachment and fusion. Due to the high rate of amino acid substitution of SARS-CoV-2, dual therapy by targeting two viral proteins would be promising in the future. Therefore, we highlighted ZINC03977803 as the second potent anti-SARS-CoV-2 compound.

Furthermore, the process implemented for the presented study can be utilized for further drug discovery of other SARS-CoV-2 main exoribonuclease proteins nsp14 and nsp12. This might enhance the efficiency of nucleot(s)ide analogues in the treatment of COVID-19 infection. In the present study, eight potent nsp12 inhibitors were highlighted to bind to the different druggable pharmacophore cavities within the nsp12 protein. The role of each of those compounds could also be investigated in other studies.

## Conclusion

The lead compound ZINV03977803 showed stable interaction with higher potential and hydrogen bonding with the catalytic subunit of SARS-CoV-2, nsp12. The implemented mechanisms by which the lead compound would inhibit nsp12 activity could directly interfere with the interaction of nsp12 with nsp7/nsp8, inhibiting the whole enzyme completion. It may also induce suppression of the proofreading activity of nsp12.

## Abbreviations

Severe acute respiratory syndrome *Coronavirus* 2, SARS-CoV-2; *Coronavirus* disease 2019, COVID-19; non-structural proteins, nsp; open reading frames, ORF; untranslated regions, UTR; angiotensin-converting enzyme 2, ACE2; N-terminal nidovirus-unique RdRp-associated nucleotidyltransferase, NiRAN; RNA-dependent RNA polymerase, RdRp; remdesivir triphosphate, RTP; Protein databank, PDB; Swiss PDB viewer, SPDBV; Toxicity estimation software tool, TEST; H-bond donor, HBD; H-bond acceptor, HBA; Hydrophobic centers, HC; negative electrostatic centers, NEC; Positive electrostatic centers, PEC; high-throughput screening, HTS; root-mean square of deviation, RMSD; molecular dynamic simulation, MDS; short-range Lennard-Jones potential, LJ-SR ; short-range Coulomb potential, Coul-SR.

## Declarations

## References

1. Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R. A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. Cells. 2020;9. doi:10.3390/CELLS9051267.

2. AB D, E D, E D, A B, N N, L L, et al. Identification of SARS-CoV-2 Receptor Binding Inhibitors by In Vitro Screening of Drug Libraries. Molecules. 2021;26. doi:10.3390/MOLECULES26113213.

3. Bojadzic D, Alcazar O, Buchwald P. Methylene Blue Inhibits the SARS-CoV-2 Spike–ACE2 Protein-Protein Interaction–a Mechanism that can Contribute to its Antiviral Activity Against COVID-19. Front Pharmacol. 2021;0:2255.

4. Mohebbi A, Askari FS, Ebrahimi M, Zakeri M, Yasaghi M, Bagheri H, et al. Susceptibility of the Iranian population to severe acute respiratory syndrome coronavirus 2 infection based on variants of angiotensin i converting enzyme 2. Future Virol. 2020;15:507–14.

5. Amin SA, Banerjee S, Ghosh K, Gayen S, Jha T. Protease targeted COVID-19 drug discovery and its challenges: Insight into viral main protease (Mpro) and papain-like protease (PLpro) inhibitors. Bioorg Med Chem. 2021;29:115860. doi:10.1016/J.BMC.2020.115860.

6. Rajpoot S, Alagumuthu M, Baig MS. Dual targeting of 3CLpro and PLpro of SARS-CoV-2: A novel structure-based design approach to treat COVID-19. Curr Res Struct Biol. 2021;3:9–18.

7. Ibrahim TM, Ismail MI, Bauer MR, Bekhit AA, Boeckler FM. Supporting SARS-CoV-2 Papain-Like Protease Drug Discovery: In silico Methods and Benchmarking. Front Chem. 2020;0:996.

8. Arya R, Kumari S, Pandey B, Mistry H, Bihani SC, Das A, et al. Structural insights into SARS-CoV-2 proteins. J Mol Biol. 2021;433:166725. doi:10.1016/J.JMB.2020.11.024.

9. Kokic G, Hillen HS, Tegunov D, Dienemann C, Seitz F, Schmitzova J, et al. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. Nat Commun 2021 121. 2021;12:1–7. doi:10.1038/s41467-020-20542-0.

10. Zhao J, Guo SS, Yi D, Li Q, Ma L, Zhang Y, et al. A cell-based assay to discover inhibitors of SARS-CoV-2 RNA dependent RNA polymerase. Antiviral Res. 2021;190:105078.

11. Mari A, Roloff T, Stange M, Søgaard KK, Asllanaj E, Tauriello G, et al. Global surveillance of potential antiviral drug resistance in SARS-CoV-2: proof of concept focussing on the RNA-dependent RNA polymerase. medRxiv. 2021;:2020.12.28.20248663. doi:10.1101/2020.12.28.20248663.

12. García-Trejo JJ, Ortega R, Zarco-Zavala M. Putative Repurposing of Lamivudine, a Nucleoside/Nucleotide Analogue and Antiretroviral to Improve the Outcome of Cancer and COVID-19 Patients. Front Oncol. 2021;0:2618.

13. SA K, K A-B. Combating COVID-19: The role of drug repurposing and medicinal plants. J Infect Public Health. 2021;14:495–503. doi:10.1016/J.JIPH.2020.10.012.

14. Alam S, Sarker MMR, Afrin S, Richi FT, Zhao C, Zhou J-R, et al. Traditional Herbal Medicines, Bioactive Metabolites, and Plant Products Against COVID-19: Update on Clinical Trials and Mechanism of Actions. Front Pharmacol. 2021;0:1248.

15. Caputo L, Lentini G, Habtemariam S. Repurposing therapeutic agents and herbal medicines to defeat viral nemesis. Drug Dev Res. 2020;81:641–2. doi:10.1002/DDR.21668.

16. Tian D, Liu Y, Liang C, Xin L, Xie X, Zhang D, et al. An update review of emerging small-molecule therapeutic options for COVID-19. Biomed Pharmacother. 2021;137:111313.

17. Savi C De, Hughes DL, Kvaerno L. Quest for a COVID-19 Cure by Repurposing Small-Molecule Drugs: Mechanism of Action, Clinical Development, Synthesis at Scale, and Outlook for Supply. Org Process Res Dev. 2020;24:940–76. doi:10.1021/ACS.OPRD.0C00233.

18. Mohebbi A, Askari FS, Sammak AS, Ebrahimi M, Najafimemar Z. Druggability of cavity pockets within SARS-CoV-2 spike glycoprotein and pharmacophore-based drug discovery. Future Virol. 2021;:fvl-2020-0394. doi:10.2217/fvl-2020-0394.

19. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The protein data bank. Acta Crystallogr Sect D Biol Crystallogr. 2002;58 6 I:899–907.

20. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, … LC-, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. science.sciencemag.org. https://science.sciencemag.org/content/368/6492/779.abstract. Accessed 6 May 2021.

21. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis. 1997;18:2714–23.

22. Mohebbi A, Mohammadi S, Memarian A. Prediction of HBF-0259 interactions with hepatitis B Virus receptors and surface antigen secretory factors. VirusDisease. 2016;27:234–41.

23. Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, et al. CavityPlus: A web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. Nucleic Acids Res. 2018;46:W374–9.

24. Koes DR, Camacho CJ. ZINCPharmer: Pharmacophore search of the ZINC database. Nucleic Acids Res. 2012;40:W409–14.

25. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An Open chemical toolbox. J Cheminform. 2011;3.

26. Allouche A. Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares. J Comput Chem. 2012;32:174–82.

27. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1–2:19–25.

28. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, et al. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. Nat Methods. 2016;14:71–3. doi:10.1038/nmeth.4067.

29. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010;31:671–90. doi:10.1002/jcc.21367.

30. Yu W, He X, Vanommeslaeghe K, MacKerell AD. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. J Comput Chem. 2012;33:2451–68. doi:10.1002/jcc.23067.

31. Vanommeslaeghe K, MacKerell AD. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. J Chem Inf Model. 2012;52:3144–54. doi:10.1021/ci300363c.

32. Vanommeslaeghe K, Raman EP, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. J Chem Inf Model. 2012;52:3155–68. doi:10.1021/ci3003649.

33. Mohebbi A, Azadi F, Hashemi MM, Askari FS, Razzaghi N. Havachoobe (Onosma dichroanthum Boiss) Root Extract Decreases the Hepatitis B Virus Surface Antigen Secretion in the PLC/PRF/5 Cell Line. Intervirology. 2021;64:22–6. doi:10.1159/000512140.

34. Mushtaq S, Abbasi BH, Uzair B, Abbasi R. Natural products as reservoirs of novel therapeutic agents. EXCLI Journal. 2018;17:420–51. doi:10.17179/excli2018-1174.

35. Khani H, Ghorbani M, Nojoomi F, Mohebbi A. Honey Bee Dry Venom Reduces Hepatitis B Virus Surface Antigen Secretion in PLC/PRF/5 Cell Line. Int J Med Lab. 2019.

36. Singh S, Sk MF, Sonawane A, Kar P, Sadhukhan S. Plant-derived natural polyphenols as potential antiviral drugs against SARS-CoV-2 via RNA-dependent RNA polymerase (RdRp) inhibition: an in-silico analysis. J Biomol Struct Dyn. 2020;:1. doi:10.1080/07391102.2020.1796810.

37. Islam SS, Midya S, Sinha S, Saadi SMAI. Natural medicinal plant products as an immune-boosters: A possible role to lessen the impact of Covid-19. Case Stud Chem Environ Eng. 2021;4:100105.

38. Chakravarti R, Singh R, Ghosh A, Dey D, Sharma P, Velayutham R, et al. A review on potential of natural products in the management of COVID-19. RSC Adv. 2021;11:16711–35. doi:10.1039/D1RA00644D.

39. Koulgi S, Jani V, N. MU V., Sonavane U, Joshi R. Natural plant products as potential inhibitors of RNA dependent RNA polymerase of Severe Acute Respiratory Syndrome Coronavirus-2. PLoS One. 2021;16:e0251801. doi:10.1371/JOURNAL.PONE.0251801.

40. Islam MR, Hoque MN, Rahman MS, Alam ASMRU, Akther M, Puspo JA, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. Sci Reports 2020 101. 2020;10:1–9. doi:10.1038/s41598-020-70812-6.
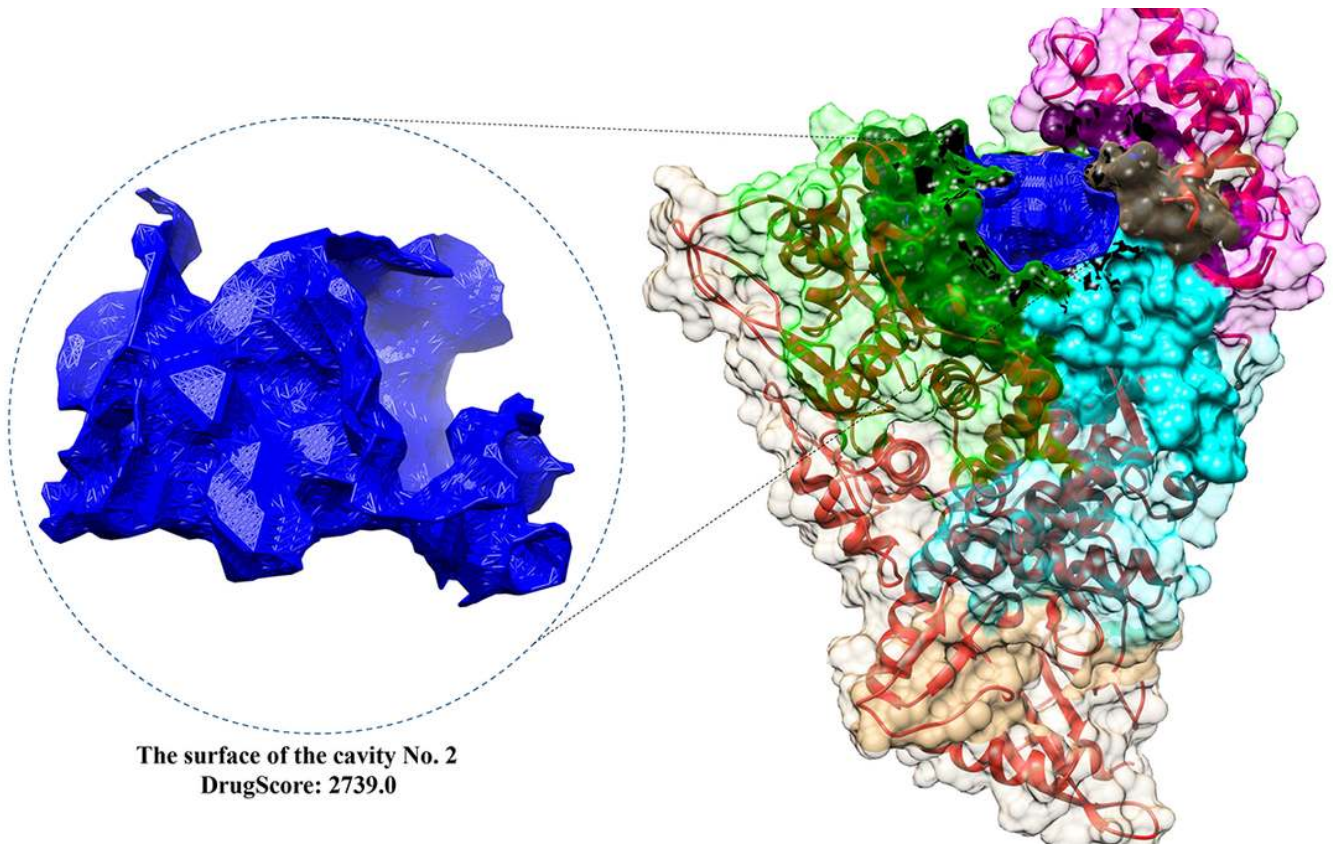
# Figures

**Figure 1**

Cavity No. 2 and its surface within the SARS-CoV-2 protein. The nsp12 fingers are presented in Green color. Palm and thumb domains are also illustrated in Cyan and Magenta colors, respectively. As it is illustrated, cavity No. 2 (colored in Black) is located adjustment to the left-handed finger (residues 366-581), palm (residues 582-620 and 680-815), and thumb (residues 816-920) domains.
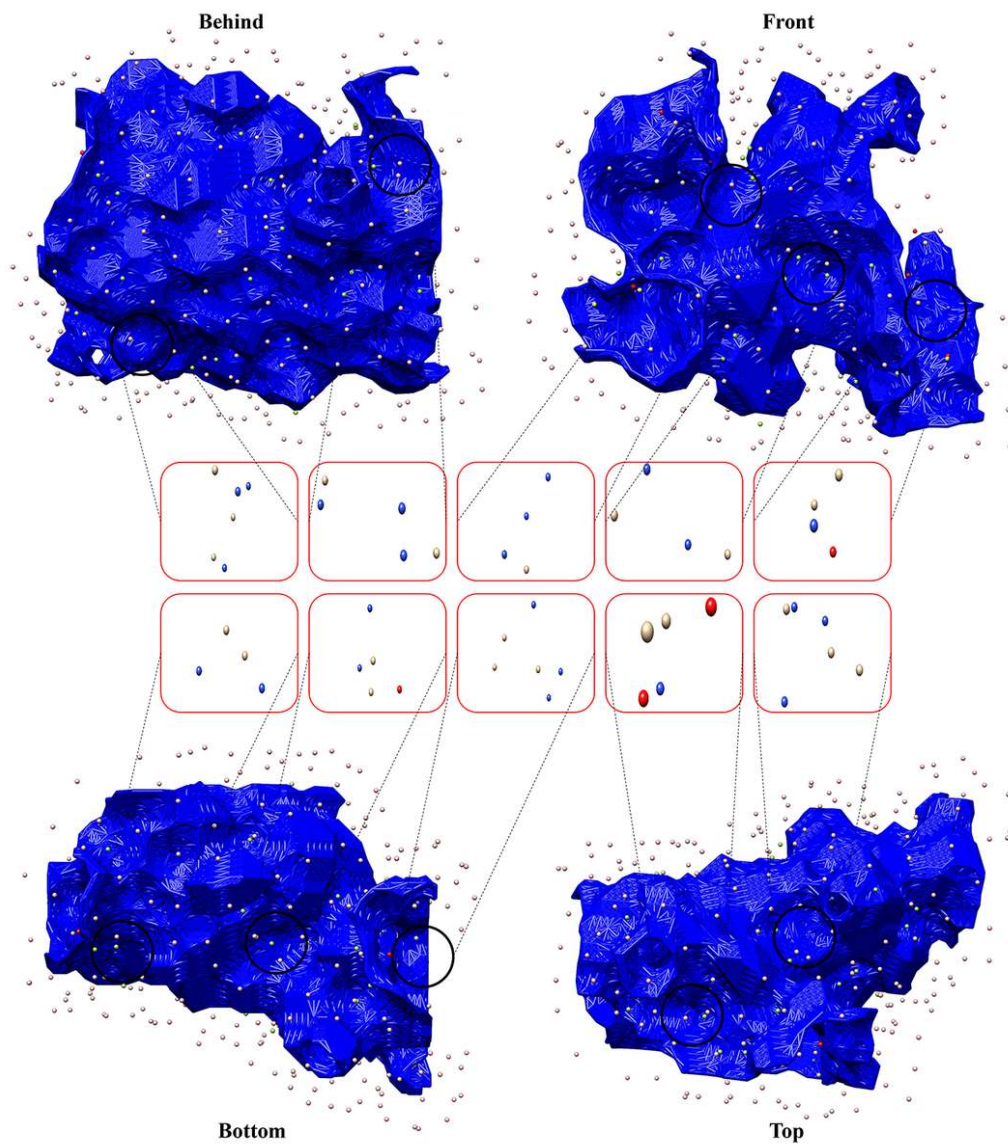
## Figure 2

Schematic representation of adjunct pharmacophore features distributed around the cavity surface no.2. The figure demonstrates the distribution of the pharmacophore features around the surface of cavity No. 2. H-bond acceptor (HBA) centers are shown in Red, H-bond donor (HBD) centers are in Blue, H-bond root is in Cyan, Hydrophobic centers (HC) are in Green, Negative electrostatic center is in Yellow, and Positive electrostatic center is in Gray. Pharmacophore features are shown in spears. The coordinates of HBA centers are overlapping with HBDs; therefore, they are not visible.
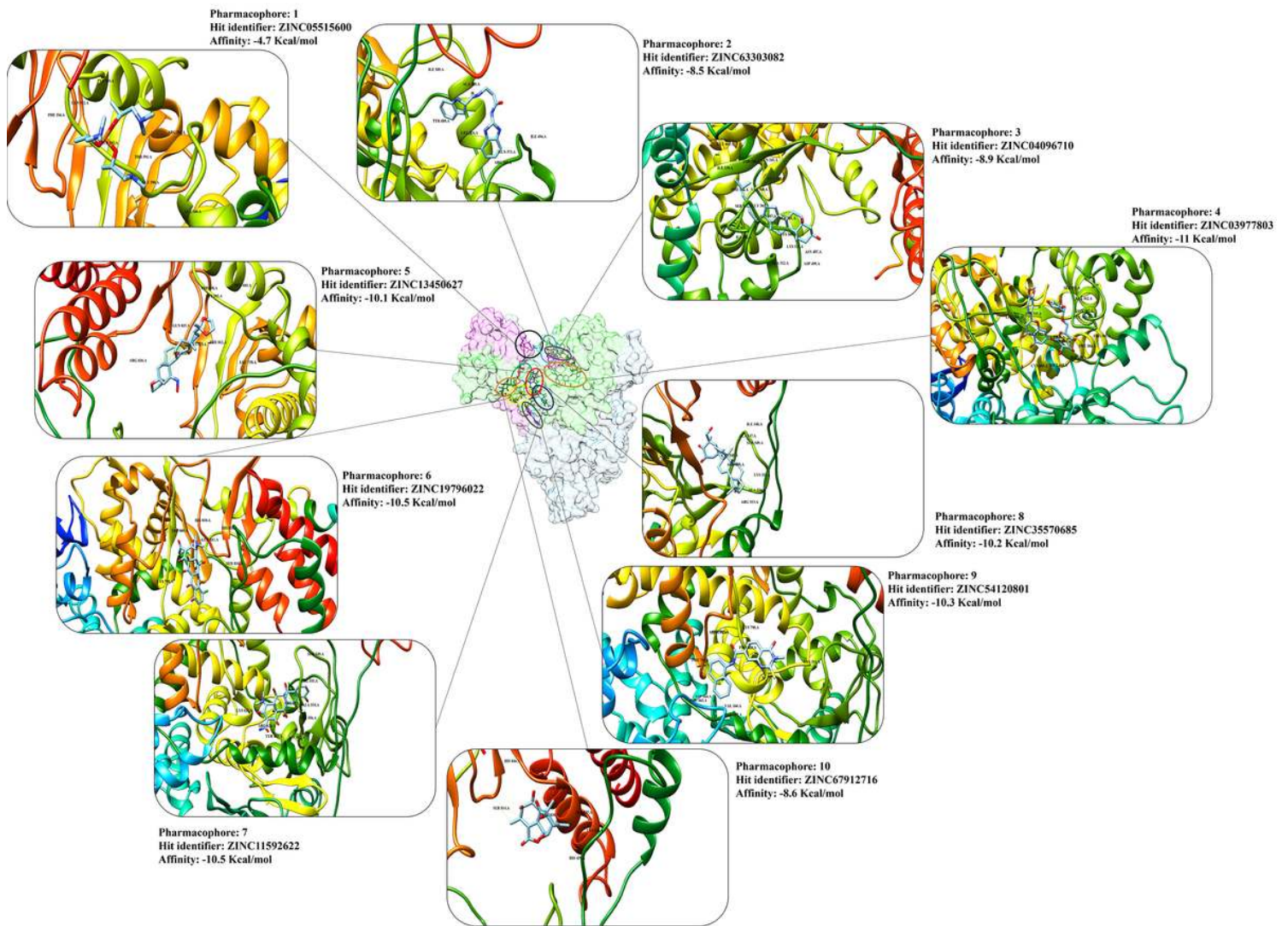
**Figure 3**

Schematic illustration of highlighted hits from different pharmacophores close or within the finger domain of the nsp12 protein. The exact site of binding of the selected hits is presented with different colors on the surface of the nsp12 at the center of the figure. Further information, including pharmacophore number and their relative hit's affinity, is also presented. In this regard, the hit identified for pharmacophore no. 1 had the lowest affinity to its nsp12, and the hit highlighted for pharmacophore no. 4 had the highest affinity to nsp12 protein.
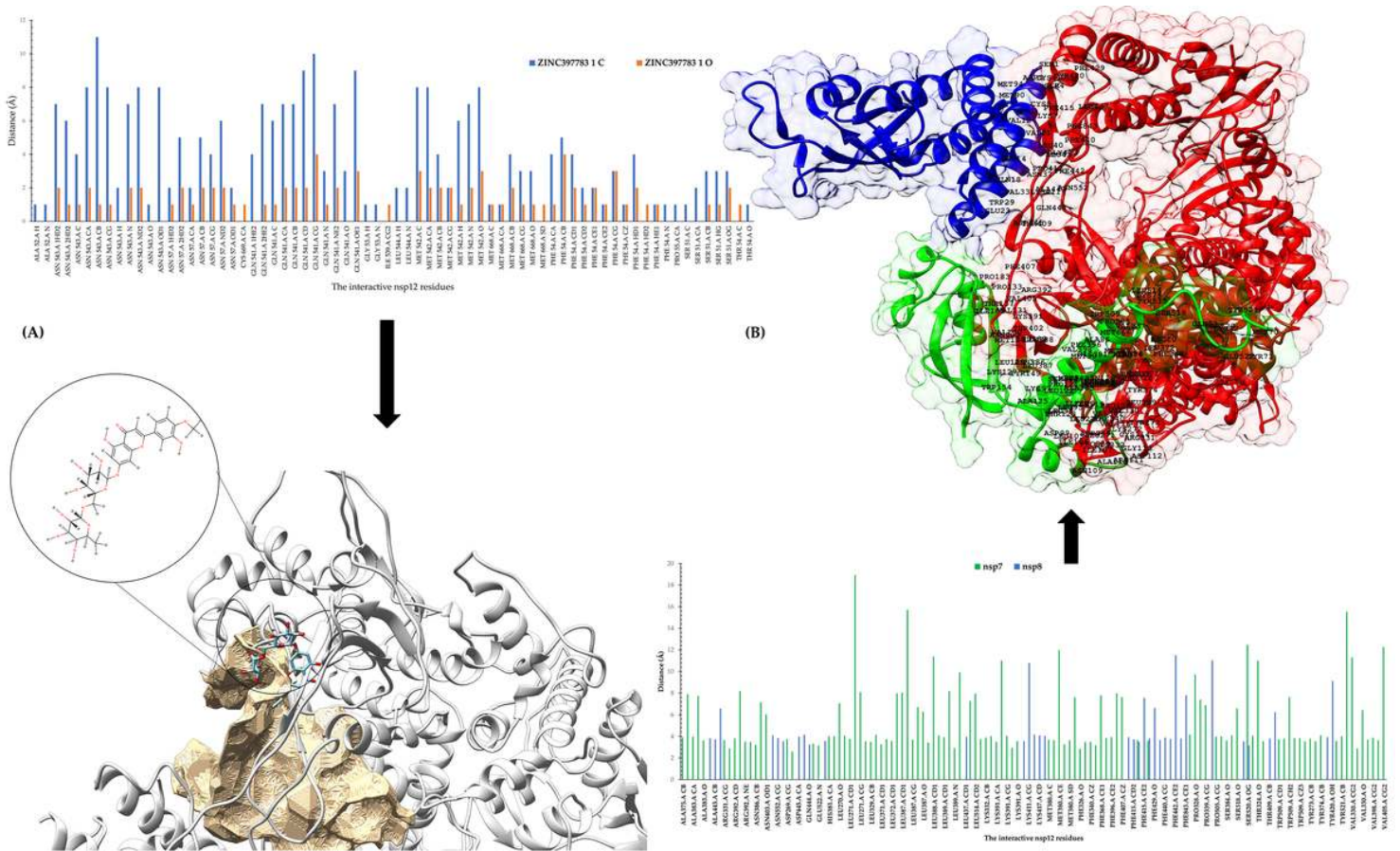
**Figure 4**

Schematic representation and plots of interaction between ZINC03977803 and nsp7/8 with nsp12. A) shows the contacts of atom residues O and C with hydrogen donor and acceptor sites of the given amino acid residues within nsp12. Further 2D view and rotation of ZINC03977803 around the cavity No. 2 is presented at the bottom-left side of the figure. B) right-top of the figure represents the nsp7 (Green) and nsp8 (Blue) interactive residues with nsp12 (Red) in the crystallographic data of the SARS-CoV-2 proteins (PDB ID: 7BTF). Data showed higher numbers of nsp7 residues in close contacts with nsp12. Further information of the interactive nsp12 residues in plotted at the bottom-right of the figure.
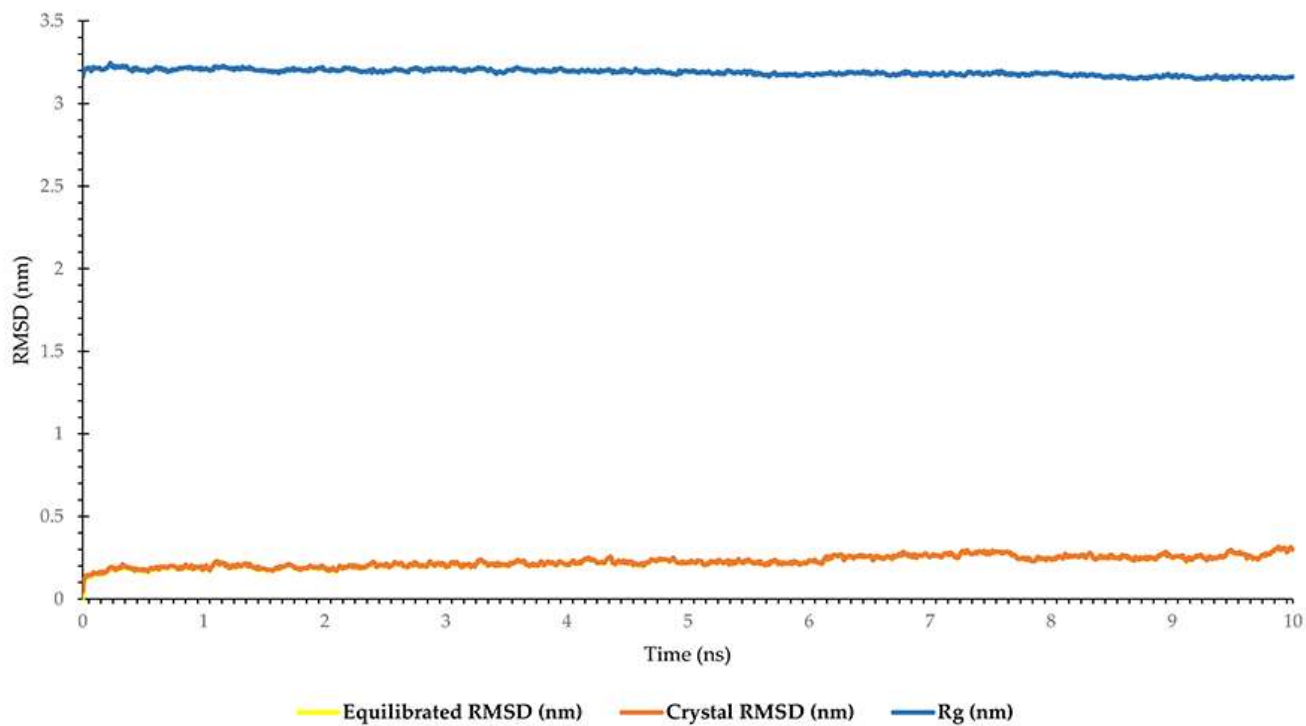
**Figure 5**

Root mean square of deviation (RMSD) and Radius of gyration (Rg) at nm scale within 10 ns of MD production. It has been shown that nsp12 RMSD relative to the crystal structure. RMSD fluctuation differences between the equilibrated protein and the crystal protein are slight indications that the structure is stable favorably. Rg of nsp12 protein showed slight fluctuation along with simulation time, suggesting proper folding of the protein.
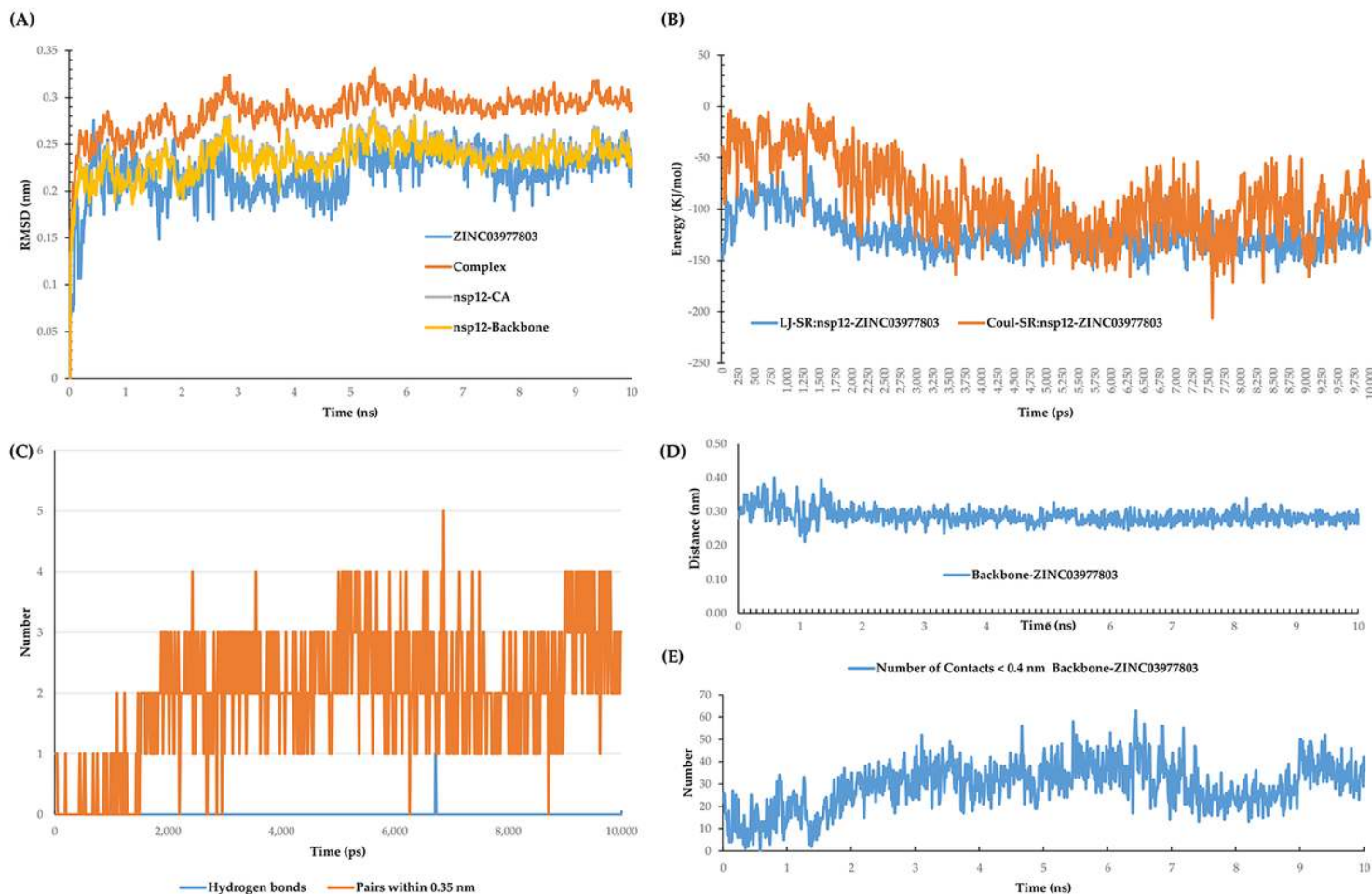
## Figure 6

The graphs of rms, energy, hbond, and mindist modules of GROMACS for ZINC03977803 and nsp12 backbone. A) RMSD fluctuation of the lead compound ZINC03977803 and its complex with nsp12 protein. The difference of the complex structure was substantially higher than that observed in the nsp12-CA and nsp12-Backbone. B) Plot of the short-range potentials of LJ and Coul shows lower electrostatic and higher inter intermolecular interactions at the first 2ns of MDS, which is converged after 2ns of MD simulation. C) H-bonding between nsp12 amino acid residues and the lead compounds showed the N-terminal residues, Ala52, was the hydrogen bond donor. D demonstrates the distance between nsp12-backbone interactive residues and the lead compound, and E shows the number of interactive residues of nsp12-backbone in close contact (< 0.4 nm) with the lead compound during 10 ns of MDS.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx
- TableS2.docx