



## Original article

# Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase

Muhammad Usman Mirza, Matheus Froeyen\*

Department of Pharmaceutical and Pharmacological Sciences, Rega Institute for Medical Research, Medicinal Chemistry, University of Leuven, B-3000, Leuven, Belgium

## ARTICLE INFO

## Article history:

Received 29 March 2020

Received in revised form

23 April 2020

Accepted 26 April 2020

Available online 28 April 2020

## Keywords:

SARS-CoV-2

COVID-19 outbreak

CoV-Mpro

CoV-Nsp12 polymerase

CoV-Nsp13 helicase

## ABSTRACT

Recently emerged SARS-CoV-2 caused a major outbreak of coronavirus disease 2019 (COVID-19) and instigated a widespread fear, threatening global health safety. To date, no licensed antiviral drugs or vaccines are available against COVID-19 although several clinical trials are under way to test possible therapies. During this urgent situation, computational drug discovery methods provide an alternative to tiresome high-throughput screening, particularly in the hit-to-lead-optimization stage. Identification of small molecules that specifically target viral replication apparatus has indicated the highest potential towards antiviral drug discovery. In this work, we present potential compounds that specifically target SARS-CoV-2 vital proteins, including the main protease, Nsp12 RNA polymerase and Nsp13 helicase. An integrative virtual screening and molecular dynamics simulations approach has facilitated the identification of potential binding modes and favourable molecular interaction profile of corresponding compounds. Moreover, the identification of structurally important binding site residues in conserved motifs located inside the active site highlights relative importance of ligand binding based on residual energy decomposition analysis. Although the current study lacks experimental validation, the structural information obtained from this computational study has paved way for the design of targeted inhibitors to combat COVID-19 outbreak.

© 2020 Xi'an Jiaotong University. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The ongoing 2019–2020 outbreak of coronavirus disease 2019 (COVID-19) [1], as of April 20, 2020, has claimed 157, 847 lives globally and has been declared as a pandemic. SARS-CoV-2 belongs to the coronaviruses family and is the seventh known human coronavirus (HCoV) from the same family after 229E, NL63, OC43, HKU1, MERS-CoV, and SARS-CoV [1].

Currently, there is no licensed drug or vaccine available for SARS-CoV-2. Although several clinical trials are in progress to test possible therapies, the treatment is focused on the alleviation of symptoms which may include dry cough, fever and pneumonia [2]. Following the SARS outbreak, existing antivirals, including nine protease inhibitors [3], nelfinavir blocked SARS-CoV-2 replication at the lowest concentration ( $EC_{50} = 1.13 \mu\text{M}$ ,  $CC_{50} = 24.32 \mu\text{M}$ ,  $SI = 21.52$ ). Following this, antiviral efficiency of several FDA-

approved drugs has been reported, including remdesivir (GS-5734, Gilead) ( $EC_{50} = 0.77 \mu\text{M}$ ) and chloroquine ( $EC_{50} = 1.13 \mu\text{M}$ ) against a clinical isolate of SARS-CoV-2 in vitro showing potential inhibition at low-micromolar concentration [4]. The efficacy of remdesivir is evident from a recent recovery of US patient infected with SARS-CoV-2 after intravenous treatment [5], while chloroquine is being evaluated in an open-label trial (ChiCTR2000029609). Moreover, Ivermectin, a broad spectrum anti-parasitic agent, was reported in vitro activity (~5000-fold reduction) in a model of Vero/hSLAM cells infected with a SARS-CoV-2 isolate (Australia/VIC01/2020) [6]. Others include Nafamostat ( $EC_{50} = 22.50 \mu\text{M}$ ), Nitazoxanide ( $EC_{50} = 2.12 \mu\text{M}$ ) and Favipiravir ( $EC_{50} = 61.88 \mu\text{M}$ ) [4]. Recently, hydroxychloroquine has shown to be effective in COVID-19 patients [7] and its efficiency has been reportedly reinforced by azithromycin for virus elimination [8], although no well-controlled, randomized clinical evidence supports azithromycin therapy in COVID-19 [9].

CoVs are single-stranded positive-sense RNA (+ssRNA) viruses with 5'-cap and 3'-poly-A tail. The ~30 kb SARS-CoV-2 genome contains at least six open reading frames (ORFs). The first ORF

Peer review under responsibility of Xi'an Jiaotong University.

\* Corresponding author.

E-mail address: [mathy.froeyen@kuleuven.be](mailto:mathy.froeyen@kuleuven.be) (M. Froeyen).

(ORF1a/b) is about two-thirds of the whole genome length and encodes 16 non-structural proteins (nsp1-16). ORFs near 3'-end of the genome encode four main structural proteins including spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins. Among non-structural proteins, most are known to play a vital role in CoV replication. Structural proteins, however, are important for virion assembly as well as for causing CoV infection. Additionally, specific structural and accessory proteins such as HE protein are also encoded by the CoV genome [10]. SARS-CoV-2 maintains ~80% nucleotide identity to the original SARS epidemic viruses [11].

To address the current outbreak, the development of wide-spectrum inhibitors against CoV-associated diseases is an attractive strategy. However, this approach requires the identification of a conserved target region within entire coronavirus genus [12,13]. On the contrary, all structural proteins including S, E, M, HE, and N proteins among different CoVs have considerable variations as reported [14–16], thus adding more complexity towards identification of SARS-CoV-2 inhibitors. Subsequently, the Nsp12 RNA-dependent RNA polymerase (RdRp), Nsp13 helicase, and main protease (Mpro) or chymotrypsin-like protease (3CLpro) [17] constitute highly conserved regions in non-structural proteins among coronaviruses which can be targeted.

Although no structural data were available for SARS-CoV-2 proteins at the beginning of this study, considering high sequence similarity, SARS-CoV-2 proteins were modeled for rational drug design which can later lead to downstream modification for drug leads. The current study focusses on structure elucidation of a critically important nCoV-Nsp12 polymerase, Nsp13 helicase and Mpro together with virtual screening (VS). The study reports potential hits identified through an integrative VS and molecular dynamics simulation approach. Although the current study lacks experimental validation of proposed hits, structural information along with the identified potential hits may serve as the starting point for structure-guided drug discovery.

## 2. Methods

### 2.1. Protein modeling

Homology models of key SARS-CoV-2 proteins including Mpro, Nsp12 polymerase, and Nsp13 helicase were designed using SWISSMODEL [18] as there were no resolved crystal structures available at the time of this study. Later for molecular docking studies, a recently reported co-crystallized Mpro structure was utilized (PDB ID: 6LU7). Templates with the highest identity were selected, and respective models were generated. The residues inside the binding pocket were predicted using an analytical tool, COACH meta-server and compared after superimposing various closely related X-ray resolved co-crystallized structures with bound inhibitors. The generated homology models were refined using 20 ns molecular dynamics (MD) simulations to remove steric clashes and optimize side-chain geometry.

AMBER18 simulation package was utilized for unrestrained MD simulations using AMBER ff99SB force field [19]. The stepwise minimization and equilibration protocol were performed (details in [Supplementary information](#)), and solvated system for each model with explicit TIP3 water molecules was submitted to a production run of 20 ns at constant 300K temperature and 1 bar pressure. The protein backbone conformation was analysed using CPPTRAJ module and evaluated through Ramachandran plots. The most favourable conformation was selected for virtual screening.

### 2.2. Structure-based virtual screening and comparative docking

Small molecules dataset in SMILES format was downloaded

from ZINC database [20] and uploaded to Mcule which is a fully automated drug discovery platform [21]. Virtual screening pipeline was established using drug-like filters. Diversity selection was implemented to reduce the size of the library and to maximize the coverage of chemical space at the same time. This procedure excluded the closest analogues based on the Tanimoto similarity coefficient and maximized the diversity of the potential active scaffolds identified from the huge dataset. A thorough analysis of binding site residues was performed prior to docking, as depicted from COACH and structure comparisons. Docking grid was set up covering the key binding site residues to reduce the search space for ligand optimization.

An automated in-built Autodock (AD) Vina docking engine was utilized to screen the composed docking library, which uses gradient optimization method in its local optimization process to efficiently rank the best poses [22]. Individual compounds of the docking library together with Mcule database were docked iteratively into respective binding pockets of the representative SARS-CoV-2 Mpro, Nsp12 polymerase, and Nsp13 helicase structures. The compounds were ranked based on Vina empirical scoring function that approximates the ligand binding affinity in kcal/mol. Since various docking programs only estimate the real binding affinity, testing different docking programs add reliability. The top hits obtained from the first round of docking were then evaluated with Glide “extra precision” docking mode (Glide XP) as implemented in Schrödinger’s Maestro modeling package (a trial version was utilized for this study). SARS-CoV-2 proteins were prepared using protein preparation wizard by adding hydrogen atoms and missing side chains, while short minimization of only the hydrogen atoms was carried out using the OPLS 3 force field. Selected ligands ranked by AutoDock Vina were prepared using LigPrep as described previously [23]. The grid box was generated with dimensions covering the predicted active site residues. Top hits were selected after extensive post-docking analysis by comparing the binding poses predicted from AD Vina and Glide.

### 2.3. Molecular dynamics simulations and energy calculations

To gain a better understanding of the predicted molecular interactions, the best docked complexes were analysed over a period of 20 ns MD simulations using AMBER 18 simulation package. The antechamber module of AMBER18 was used to generate atomic partial charges for the selected hits. In this study, previously reported MD simulation protocol was followed [23,24] (details in [Supplementary data](#)). To gain rational insights into the different binding modes and residual contribution, the Molecular Mechanics/Generalized Born Solvent Area (MM/GBSA) method was employed and total binding free energies (kcal/mol) were calculated. The MM/GBSA approach is well detailed in binding free energy calculations [25] for antiviral inhibitors [26,27].

## 3. Results

### 3.1. Structural insights of SARS-CoV-2 proteins

In silico modeling work was initiated by analysing Wuhan sea-food market pneumonia virus genome (NCBI genome ID MN908947), which was recently deposited by Wu et al. [28]. For the current work, orf1ab polyprotein sequence (GenBank ID: QHD43415.1) was considered (a total of 7096 amino acids). In order to locate the potential Nsp12, Nsp13 and Mpro sequences, a multiple sequence alignment was performed with known SARS-CoV proteins (Uniprot ID: POC6X7). The corresponding regions in SARS-CoV-2 genome were identified based on the sequence identity and similarity score. Sequence alignment revealed that SARS-

CoV-2 Nsp12 polymerase, Nsp13 helicase and Mpro shared high similarity with that of SARS.

Homology modeling was performed using SWISSMODEL and generated models were validated accordingly. SARS-CoV proteins deposited in Protein Data Bank (PDB) indicated 96.35%, 99.83% and 96.08% identities with SARS-CoV Nsp12 (6NUR), Nsp13 (6JYT) and Mpro (2Z9J). As a result of high sequence similarity (>96%), the homology models revealed a strikingly conserved overall architecture. The superimposition of RdRp, helicase and Mpro with SARS-CoV proteins demonstrated high structural similarity, with root mean square deviation (RMSD) values of 1.35 Å, 0.25 Å and 0.46 Å, respectively. Model validation using PROCHECK indicated 96.51% (Mpro) and 97.5% (RdRp) of all residues in Ramachandran favoured regions with 3.27% and 0.28% being rotamer outliers, while helicase indicated 84.34% with 4.43% outliers. The models were further refined by 20 ns MD simulation and compared with the templates. With the availability of the crystal structure of SARS-CoV-2 Mpro (6LU7) complexed with N3 peptide, homology model was also compared and less than 0.5 Å RMSD was found, which indicated the reliability of the generated model with only subtle difference (Fig. S1–S3). Overall, SARS-CoV-2 Nsp12 polymerase and Nsp13 helicase models were reliable enough to perform virtual screening. For Mpro, co-crystallized structure (6LU7) was utilized. All models revealed the conserved features especially in functional regions due to strikingly similar protein conformation. However, some of the structural features have been elucidated together with the identification of inhibitor binding site in SARS-CoV-2 proteins as shown in Fig. 1.

### 3.1.1. SARS-CoV-2 Mpro

SARS-CoV-2 Mpro was predicted to contain 306 amino acids (located in the polyprotein, from 3264 to 3569 a.a). Structural analysis of SARS-CoV-2 Mpro revealed the same location of the active site as reported in SARS-Mpro, which is arranged in the cleft between domain I (8–99 a.a) and domain II (100–183 a.a). Both domains contribute one residue to the catalytic dyad (His41 and Cys145) (Fig. 1A), connected by a long loop (184–199 a.a) to the helical domain III (200–306 a.a). Superimposition of several co-crystallized SARS-Mpro structures bound with inhibitors [13], N1 (1WOF), I2 (2D2D), N3 (2AMQ), and N9 (2AMD), also revealed the similar position of S1, S2, and S4 subsites, especially in the active site close to His41 and Cys145, which is crucial for substrate recognition [13], along with Tyr161 and His163 in the substrate-binding pocket [29]. Docking grid was formed around these subsites for virtual screening by taking co-crystallized SARS-CoV-2 Mpro (6LU7), and top hits were identified based on the docking scores.

### 3.1.2. SARS-CoV-2 Nsp12 RdRp

The SARS-CoV-2 Nsp12 polymerase was predicted to contain 932 amino acids (located in the polyprotein, from 4393 to 5324 a.a). The nsp12 comprised N-terminal (1–397 a.a) and a polymerase domain (a.a. 398–919) when compared with SARS-CoV Nsp12 (6NUR). The polymerase domain adopted a structure resembling a cupped “right hand” like other polymerases [24]. Nsp 12 polymerase is composed of a finger (398–581 and 628–687 a.a), a palm (582–627 and 688–815 a.a), and a thumb subdomain (816–919 a.a) (Fig. 1B). Similar to MERS and SARS-Nsp12, the finger and the thumb subdomains of nCoV-RdRp contacted each other, which configured the RdRp active site in the centre for the substrate access through template entry, template-primer exit, and NTP tunnels [30]. Alongside, the SARS-CoV-2 Nsp12 also revealed seven conserved motifs (A – G) arranged in the polymerase active site chamber, which are involved in a template and nucleotide binding and catalysis [31]. The binding site of SARS-CoV-2 Nsp12 was further analysed through the superimposition of elongation

complex of poliovirus bound with CTP-Mn (3LO8) [32] and crystal structure of Japanese encephalitis RdRp in complex with ATP (4HDH) [33]. Docking grid was formed around the polymerase active site chamber covering the conserved motifs (A – G), and potential hits were identified.

### 3.1.3. SARS-CoV-2 Nsp13 helicase

The SARS-CoV-2 Nsp13 was predicted to contain 596 amino acids (located in orflab polyprotein from 5325 to 5925 a.a). Like that of SARS and MERS-Nsp13, the overall structure of SARS-CoV-2 Nsp13 adopted a triangular pyramid shape comprising five domains. Among these, two “RecA-like” domains, 1A (261–441 a.a) and 2A (442–596 a.a), and 1B domain (150–260 a.a) form the triangular base, while N-terminal Zinc binding domain (ZBD) (1–99 a.a) and stalk domain (100–149 a.a), which connects ZBD and 1B domain, are arranged at the apex of the pyramid (Fig. 1C). Small molecules able to inhibit the NTPase activity by interferences with ATP binding may serve as an ideal strategy to develop inhibitors. The SARS-CoV-2 Nsp13 revealed the similar conserved NTPase active site residues including Lys288, Ser289, Asp374, Glu375, Gln404 and Arg567 as present in SARS-Nsp13. All these residues were clustered together in the cleft located at the base between domains 1A and 2A, while docking grid was formed by locating bound ADP of crystalized yeast Upf1 and top hits were identified.

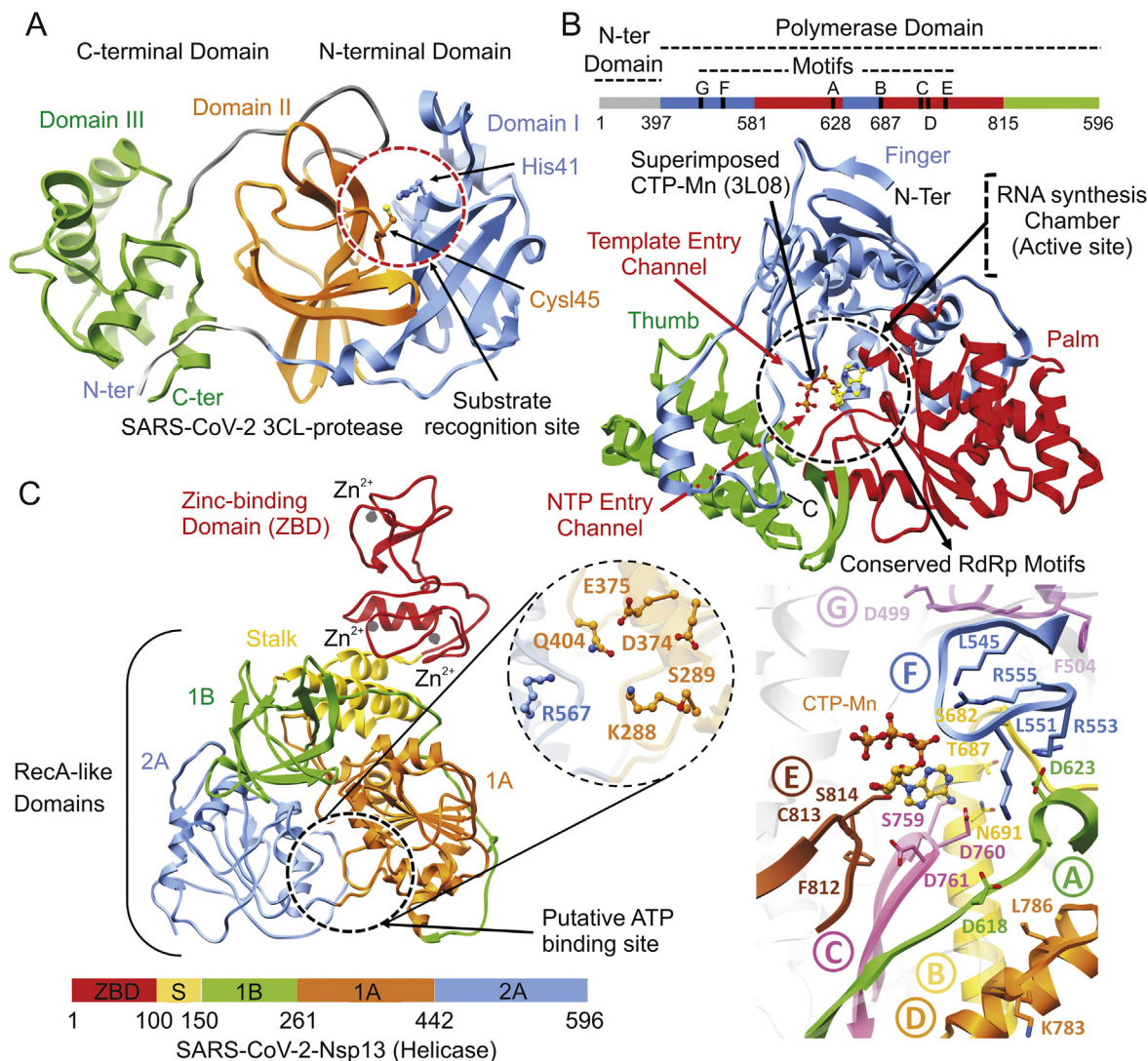
## 3.2. Identification of potential compounds against SARS-CoV-2 proteins

A stepwise structure-based virtual screening pipeline was adopted considering the putative binding sites of SARS-CoV-2 Mpro, Nsp12 polymerase and Nsp13 helicase to identify potential hits as described in previous studies [24,34,35]. The identified potential hits against each target were subjected to in silico ADMET (absorption, distribution, metabolism, elimination, toxicity) predictions, which includes drug-likeness and toxicity potential. This filtering removed substantial hits that exhibited poor ADMET due to inhibitory effects on the renal organic cation transporter and CYP450 isozymes. While a large subset of hits was found to contain high-risk chemical groups like epoxides and quinones, the potential hits against each target with the lowest docking energy were analysed for their binding pose inside the binding site. After careful inspection, best hits based on AD Vina that predicted binding energies were re-docked with Glide and best docked complexes were further processed through 20 ns MD simulations. MMGBSA method was employed to analyse electrostatic and vdW energy contribution to total free energy of binding. The MD-simulated complexes indicating protein backbone stability (during the last 5 ns in a 20 ns production run) were considered for further molecular interaction analysis. Based on molecular interactions and binding free energy calculations, the subsequent screening resulted in 5, 4 and 3 hits for SARS-CoV-2 Mpro, Nsp12 and Nsp13, respectively (Detailed ADMET profile of these compounds is tabulated in Table S1, estimated from Swiss-ADME server [36]). The top hits with AD Vina, Glide XP-score together with contribution of electrostatic ( $\Delta E_{ele}$ ), van der Waals ( $\Delta E_{vdw}$ ), and solute-solvent energies ( $\Delta G_{np}$  and  $\Delta G_p$ ) towards total binding free energy ( $\Delta G_{tot}$ ) were estimated through Amber-MMGBSA module as tabulated in Table 1. All energy contributions were calculated in kcal/mol.

### 3.3. SARS-CoV-2 Mpro hits

As a validation docking protocol, the cocrystallized N3 peptide was removed and redocked with the substrate-binding site of SARS-CoV-2 Mpro (6LU7) using the same docking parameters. The generated redocked pose was quite similar to the co-crystallized





**Fig. 1.** Structural representation of SARS-CoV-2 proteins. **(A)** SARS-CoV-2 main protease monomer (ribbon representation) composed of: (i) N-terminal domain I (cornflower blue), (ii) domain II (orange), and (iii) C-terminal domain III (green). Substrate recognition site in circle (red) and catalytic dyad residues, His41 and Cys145 are highlighted and labelled. **(B)** Linear schematic description of domain architecture of SARS-CoV-2 Nsp12 polymerase followed by its structure composed of thumb (green), palm (red), and finger (cornflower blue) subdomains. The active site of Nsp12 polymerase is highlighted and arrangement of structurally conserved RdRp motifs in Nsp12 polymerase model coloured green (A), yellow (B), pink (C), orange (D), brown (E), cornflower blue (F) and magenta (G) for motifs A-G respectively is displayed in bottom right. Superposition of the polio virus elongation complex structure (PDB: 3OL8) CTP (orange) inside the predicted binding site also displayed. **(C)** Overall structure of SARS-CoV-2 Nsp13 helicase composed of ZBD (red), stalk (golden), 1B (green), 1A (orange) and 2A (cornflower blue) domains. Three zinc atoms in ZBD are shown as dark grey spheres. The binding pocket residues are zoomed in and labelled. Linear schematic diagram of the domain organization of SARS-CoV-2 Nsp13 helicase is displayed at the bottom.

conformation (Fig. S4). Top-ranked SARS-CoV-2 Mpro hits were filtered based on the interaction with the catalytic dyad (at least one H-bond with either His41 or Cys145 or strong vdW interactions). A total of 13 compounds were analysed after 20 ns MD simulation based on the stability for at least last 5 ns. MMGBSA calculations were performed for the most stable complexes. Among top hits, total binding free energies ( $\Delta G_{\text{tot}}$ ) were very promising for cmp12, cmp14, cmp17 and cmp18 which exhibited  $\Delta G_{\text{tot}} = -45.22, -44.91, -41.72,$  and  $-44.55$  kcal/mol. In all complexes, ligand stayed inside the binding pocket and backbone RMSD of Mpro remained stable throughout simulation period (Fig. S5). All these compounds contributed well through H-bonds and exhibited favourable electrostatic energy for cmp3 ( $\Delta G_{\text{ele}} = -21.38$  kcal/mol), cmp12 ( $\Delta G_{\text{ele}} = -26.16$  kcal/mol), cmp14 ( $\Delta G_{\text{ele}} = -33.56$  kcal/mol) and cmp18 ( $\Delta G_{\text{ele}} = -36.65$  kcal/mol) except cmp17, which displayed weak electrostatic energy ( $\Delta G_{\text{ele}} = -5.39$  kcal/mol) due to unstable H-bonds over a simulation period (Table 1). Further in-

depth molecular interaction analysis after 20 ns MD simulations unveiled significant findings with respect to Mpro-subsites (Fig. 2A) and per-residue decomposition is displayed in Fig. 2B. Among all compounds, cmp12 and 14 showed some consensus features, where pentacyclic moiety of cmp12 anchored right inside the subsites and formed H-bonds with Ser144 of S1, Cys145 of catalytic dyad, Met165 of S4, and terminal 4-methylaniline established two H-bonds with two neighbouring glutamine residues (Gln189 and Gln192) of S4. Similarly, biphenyl interacted hydrophobically with S2, triazole moiety established one H-bond with Cys145 of catalytic dyad, and pyrazole-pyrimidine moiety formed three H-bonds with Gln189, Thr190, and Gln192 located at S4. The remaining compounds (cmp3, cmp17 and cmp18) demonstrated distinct H-bond patterns, even though the central moieties superimposed well near the catalytic dyad. For instance, the terminal oxadiazol-2-amine moiety of cmp3 mainly interacted through H-bonds with Cys145 of catalytic dyad, Asn142, Ser144, Glu166 and

**Table 1**  
Molecular docking predicted binding energies (kcal/mol) and molecular mechanics generalized born surface area (MM-GBSA) binding free energy calculation of identified hits in complex with SARS-CoV-2 Mpro, Nsp12 polymerase and Nsp13 helicase.

Contributions		AD Vina	XP-Score	$\Delta E_{ele}$	$\Delta E_{vdw}$	$\Delta E_{MM}$	$\Delta G_p$	$\Delta G_{np}$	$\Delta G_{sol}$	$\Delta G_{tot}$
Mpro	cmp3	-9.2	-7.9	-21.38	-36.26	-57.64	33.04	-4.53	28.51	-29.13
	cmp12	-8.8	-9.87	-26.16	-55.63	-81.79	43.44	-6.87	36.57	-45.22
	cmp14	-8.7	-8.51	-33.56	-44.15	-77.71	38.93	-6.13	32.8	-44.91
	cmp17	-8.7	-8.27	-5.39	-53.87	-59.26	23.26	-5.72	17.54	-41.72
	cmp18	-8.6	-9.2	-36.65	-41.9	-78.55	38.44	-4.44	34	-44.55
Helicase	cmp1	-10.9	-9.91	-15.73	-55.49	-71.22	40.73	-6.8	33.93	-37.29
	cmp3a	-10.6	-11.1	-26.89	-51.08	-77.97	44.25	-6.36	37.89	-40.08
	cmp11	-10.2	-10.5	-31.17	-67.39	-98.56	56.76	-7.93	48.83	-49.73
	cmp15	-10.2	-8.4	-27.13	-63.39	-90.52	66.81	-6.85	59.96	-30.56
RdRp	cmp2	-8.8	-8.6	-58.59	-48.12	-106.71	71.28	-6.31	64.97	-41.74
	cmp17a	-8.4	-8.9	-49.44	-43.18	-92.62	65.14	-6.89	58.25	-34.37
	cmp21	-8.2	-7.1	-38.06	-47.73	-85.79	66.01	-5.12	60.89	-24.9

Note:  $\Delta G_{tot}$  is the sum of molecular mechanics energy ( $\Delta E_{MM}$ ) and solvation free energy ( $\Delta G_{sol}$ ). Both  $\Delta E_{MM}$  and  $\Delta G_{sol}$  are further divided into internal energy ( $\Delta E_{int}$ ), electrostatic energy ( $\Delta E_{ele}$ ), and van der Waals ( $\Delta E_{vdw}$ ) energy in the gas phase, and polar ( $\Delta G_p$ ) and non-polar ( $\Delta G_{np}$ ) contributions to the solvation free energy. The units of these parameters are calculated in kcal/mol.

His172 of S1, and central carbaldehyde also formed H-bond with Glu166 of S1. In contrast, cmp17 only formed H-bonds with residues of S4 through terminal carbazole-1-carboxamide moiety and cmp18 established H-bonds with residues of S2 and S4 through urea and terminal imidazole moieties. Although the compounds mainly interacted electrostatically, nonpolar solvation energies also slightly contributed to ligand binding. The residual decomposition analysis revealed the favourable contribution of catalytic dyad to total binding free energy, where Cys145 exhibited -1.058, -1.19, -2.12, -1.015, and -0.892 kcal/mol and His41 exhibited -0.423, -1.528, -0.656, -1.032, and -0.532 kcal/mol with cmp3, 12, 14, 17 and 18, respectively. Other residues located in close vicinity to catalytic dyad which also contributed significantly included Glu189 (ranging from -3.505 to -1.219 kcal/mol), Glu166 (-2.711 to -0.845 kcal/mol), and Met165 (-3.859 to -0.988 kcal/mol) (Fig. 2B). Such favourable residual energy contribution was evident from the H-bonds formed by Glu166, Glu189 and Met165. The 2D chemical structure representation of each compound is displayed in connection with interactions of structural moieties with binding pocket subsites (Fig. 2C and D).

### 3.4. SARS-CoV-2 Nsp12 RdRp

For SARS-CoV-2 Nsp12 RdRp, hits were selected based on establishing H-bond with conserved motifs, especially with signature motif C (SDD), which has been experimentally evaluated in altering polymerase activity supported by various mutational studies [37–39]. A total of 17 hits were analysed through 20ns MD simulations, and complexes with stable RMSD were further processed through MMGBSA binding free energy calculations. Among all hits, cmp2, cmp17a and cmp21 showed promising  $\Delta G_{tot}$  (-41.74, -34.37, and -24.9 kcal/mol) which mainly interacted through H-bonds, thus contributing electrostatically (Table 1). All these complexes remained stable throughout 20 ns MD simulations (Fig. S6). A careful molecular inspection after MD revealed the similar conformation of these compounds, which were found to interact extensively with the conserved motifs (Fig. 3).

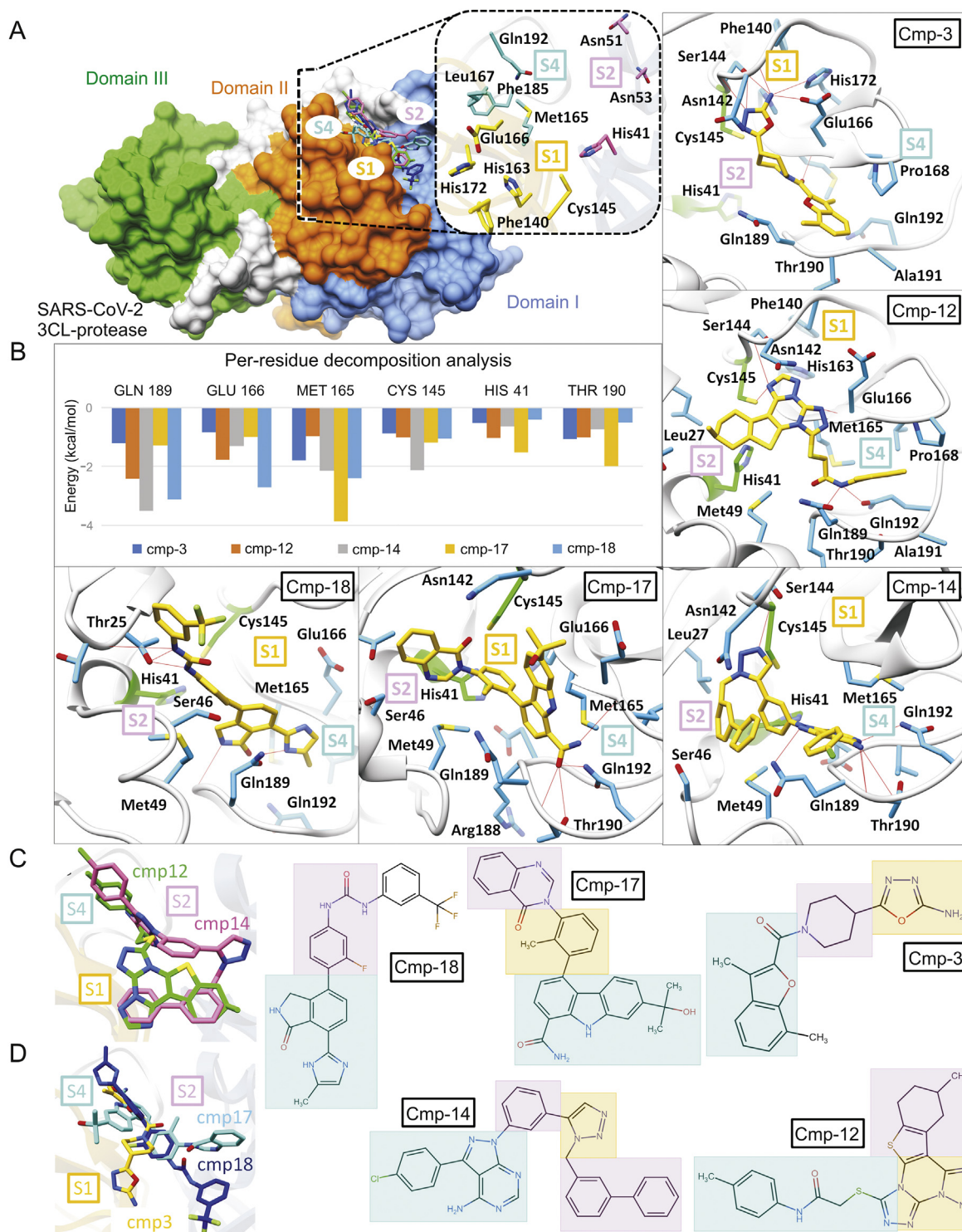
Cmp2 established three H-bonds with the residues belonging to signature motif C (Ser759), motif D (Lys798) and motif E (Ser814) (Fig. 3B), Cmp17a formed five H-bonds with Arg624 (motif A), Thr680 and Ser682 (motif B), Ser756 (motif C) and Lys798 (motif D) (Fig. 3C), whereas cmp21 interacted mainly with the residues of motif A (Asp623), motif F (Arg553 and Arg555), and motif C (Ser759) through four H-bonds (Fig. 3D). Together various residues also contributed through strong hydrophobic interactions. For instance, Tyr455 established strong stacking interaction with terminal difluorobenzene of cmp21. Per-residue binding energy

decomposition analysis revealed a key role of important residues located in conserved polymerase motifs (Fig. 3F). These include the residues: (i) Arg553 and Arg555 of motif F which exhibited most favourable  $\Delta G_{tot}$  with cmp2 (-2.05 and -2.14 kcal/mol), cmp17a (-1.84 and -2.57 kcal/mol) and cmp21 (-2.4 and -1.04 kcal/mol), (ii) conserved aspartate residues of motif A (Asp618 and Asp623) and C (Asp760), (iii) conserved serine residues of motif B (Ser682) and motif C (Ser759), and (iv) Lys798 of motif D. Among these, Ser759 and Asp760 of motif C exhibited significant binding free energy with all compounds (ranging from -0.988 to -1.76 kcal/mol and -1.03 to -1.47 kcal/mol), while most of the residues interacted electrostatically. By the end of long-run simulation process, an X-ray resolved SARS-CoV-2 Nsp12 polymerase structure was deposited (6M71) [40]. We compared crystal structure of SARS-CoV-2 and found only 0.41 Å deviation (Fig. S7). In further support, MD simulations of homology model and 6M71 also showed consensus RMSD over a period of 50ns. Moreover, the conformations of cmp2, cmp17a and cmp21 showed identical binding modes when docked inside the polymerase active site of crystal structure. Additionally, the binding mode and molecular interaction profile of cmp2, cmp17a and cmp21 agreed with the reported comparative binding model of remdesivir (7BV2) and sofosbuvir diphosphate (HCV-RdRp, 4WTG) [41] with polymerase (Fig. S7) [40].

### 3.5. SARS-CoV-2 Nsp13 helicase

For SARS-CoV-2 Nsp13 helicase, top hits were selected based on significant interactions with six key residues (Lys288, Ser289, Asp374, Glu375, Gln404 and Arg567) including neighbouring residues involved in NTP hydrolysis [42]. A 20 ns MD simulation on top ranked complexes led to the identification of 4 best compounds based on considerable interactions with key residues and demonstrated favourable MMGBSA binding free energy (Table 1). All these complexes remained stable throughout 20 ns MD simulations except cmp3a, which fluctuated in the start and adopted a more favourable conformation inside the binding pocket (Fig. S8). The superimposition of Upf1-ADP showed the overall same binding mode as identified with these 4 compounds (Fig. 4).

Two dimensional (2D) interaction plot of yeast Upf1-ADP revealed a network of H-bonds established by phosphate groups with Lys436, Thr434, Gly433, Arg639 Gly435 which were also conserved in nCoV-helicase (Fig. 4A). MMGBSA method estimated favourable  $\Delta G_{tot}$  in cmp1 (-37.29 kcal/mol), cmp3a (-40.08 kcal/mol) and cmp11 (-49.73 kcal/mol) except cmp15 (-30.56 kcal/mol) and contributed mainly through vdW energies, while nonpolar solvation energies also slightly contributed (Table 1). In-depth molecular analysis of cmp11 showed most significant

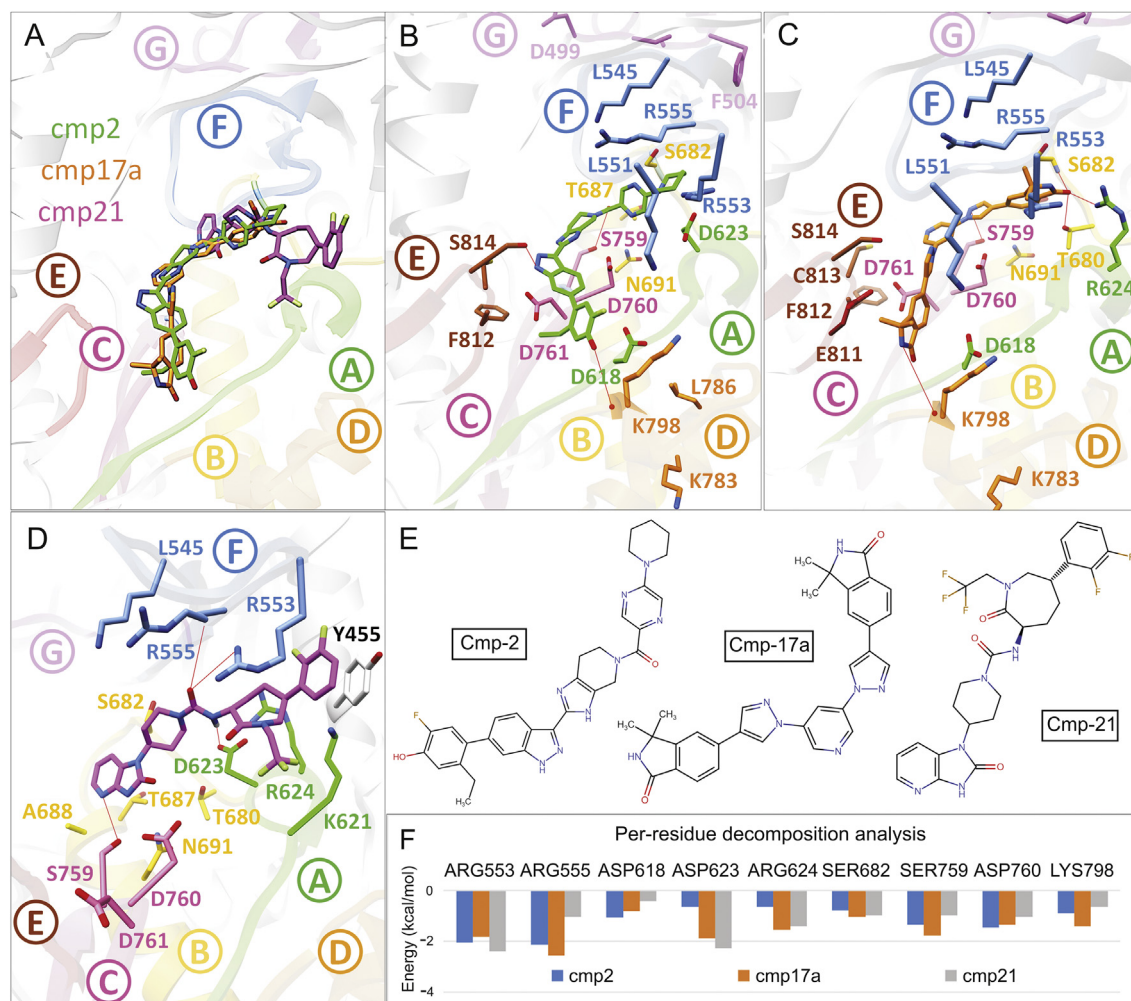


**Fig. 2.** Post-molecular dynamics (MD) analysis of SARS-CoV-2 Mpro hits. **(A)** Molecular surface representation of Mpro with MD simulated representative conformation of cmp3 (yellow), cmp12 (green), cmp14 (magenta), cmp17 (cyan) and cmp18 (dark blue) inside the substrate binding site zoomed with subsites S1 (orange), S2 (pink) and S4 (cyan) and residues are labelled accordingly. Molecular interactions representations of each complex with interacting residues are highlighted in blue sticks and catalytic dyad in green sticks. **(B)** Per-residue energy decomposition analysis of potential substrate-binding site residues. Terminal moieties of cmp12 and cmp14 revealed similar binding modes while cmp3, cmp17 and cmp18 showing different binding mode are displayed in **(C)** and **(D)** followed by chemical structure representations of these compounds in connection with interactions of structural moieties with binding pocket subsites.

MMGBSA total energy, which indicated 3 H-bonds were established between Ser310, Glu375 and Lys288 and terminal region of cmp11, and interacted mainly with the side chains of Glu374, Met378, Ala312, Arg178, Ser535, Ser539 and Asp534 (Fig. 4B). These residues were also found conserved in Upf1-ADP complex except

Met378 and Arg178. Per-residue energy decomposition analysis indicated major contributions from Arg178 of domain 1B, Ala312 and Ala314 of domain 1A, which indicated  $< -1.5$  kcal/mol binding energy with all four compounds. Together, Ser288 and Ser310 also contributed in total free energy of binding (Fig. 4C).





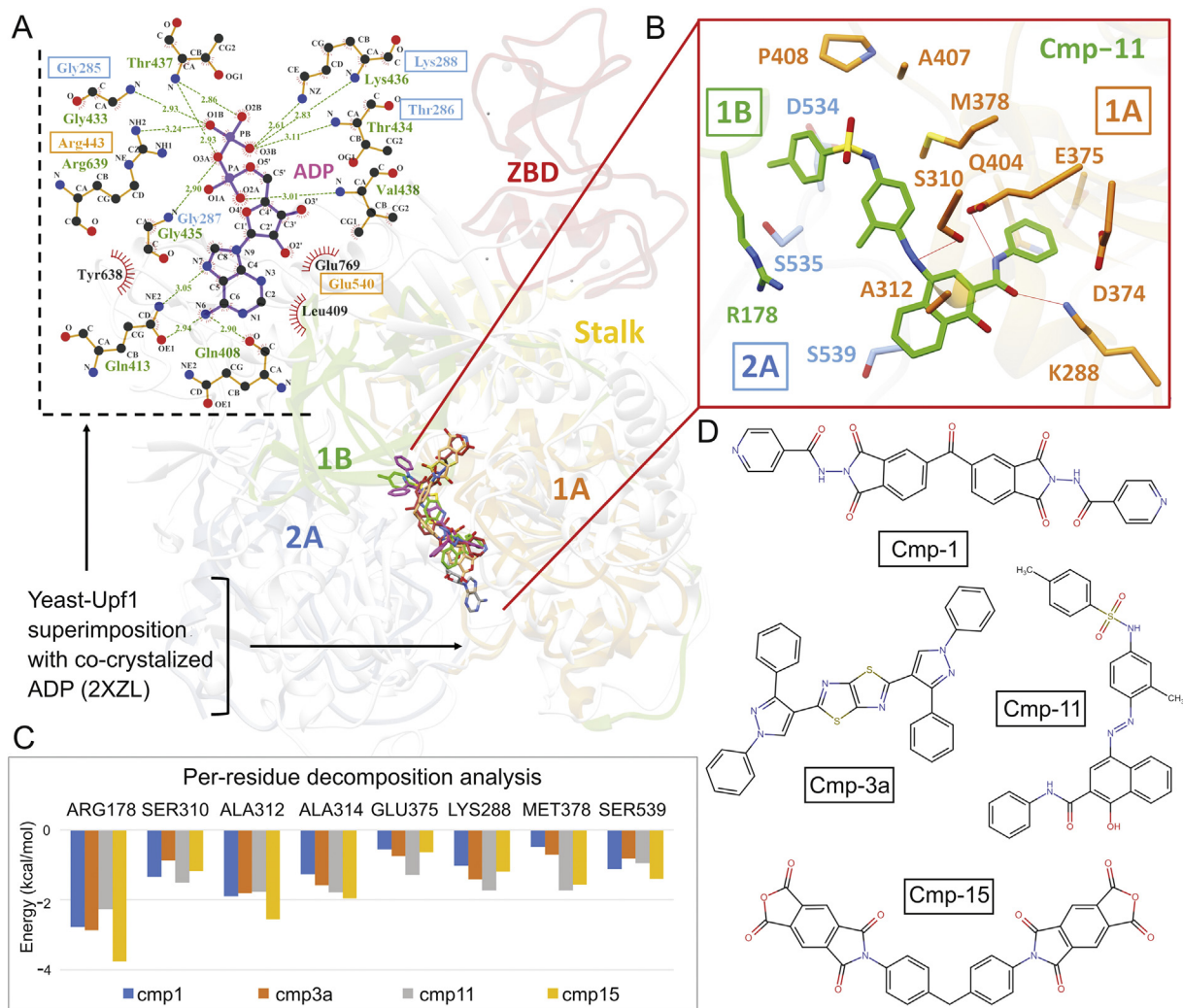
**Fig. 3.** Post-molecular dynamics (MD) analysis of SARS-CoV-2 Nsp12 RdRp complexes. (A) MD simulated conformations of cmp2 (green) (B), cmp17a (orange) (C), cmp21 (magenta) (D) inside the predicted binding pocket of Nsp12 polymerase. The arrangement of motifs and colors are the same as in Fig. 1B. Molecular interactions of individual compounds are displayed in B and C and D, and residues are labelled accordingly. (E) Representations of chemical structures of compounds, and (F) Per-residue energy decomposition analysis of potential binding pocket residues.

#### 4. Discussion

The computational methods offer an immediate and scientifically sound basis to design a highly specific inhibitor against important viral proteins and give aid in antiviral drug discovery [43,44]. A recent structural modeling effort predicted few repurposed drugs against SARS-CoV-2 vital proteins including antibacterial, anti-fungal, anti-HIV drugs, and flavonoids from different sources [45]. In the current study, a brief structural elucidation was performed together, which revealed strikingly similar features in SARS-CoV-2, when compared with crystal structures of SARS-CoV-RdRp (6NUR) and helicase (6JYT). These viral proteins play a crucial role in the viral life cycle and considered among the most popular strategies for antiviral therapeutics [44,46]. Structural analysis has revealed that the substrate-binding pockets of SARS-CoV-Mpro, active site of RdRp and NTPase binding pocket of helicase are highly conserved [47], which can lead to the concept of “wide-spectrum inhibitors” for targeting CoVs. Furthermore, extensive SBVS procedure led to the identification of potential hits against each target, which were assessed through MD simulations. MM-GBSA and per residue energy decomposition divulged the relative importance of amino acid involved in binding supported with the contribution of different components of binding free energy.

Overall, SARS-CoV-2-Mpro hits were found to interact with catalytic dyad (His41 or Cys145) together with residues located at S4, S2 and S1 subsites of substrate binding pocket and mutational studies of His41 or Cys145 resulted in complete loss of enzymatic activity, which confirmed the role of catalytic dyad [48,49]. In the substrate-binding pocket, catalytic dyad residues, highly conserved residues of S1, Gln189 and Met165 of S4 displayed relative importance in ligand binding, suggesting a possibility for inhibitor design targeting this region in the Mpro. In case of SARS-CoV-2 Nsp12, the in-depth interaction analysis inside the conserved motifs (Fig. 3) revealed functionally important aspartate residues of motif A (Asp623) and motif C (Asp760), together with conserved arginine residues of motif F (Arg553 and Arg555), and Ser759 (motif C) which were found to highly interact with the compounds. Among these, motif A and C are most strikingly conserved aspartate residues that bind divalent metal ions necessary for catalytic activity [30,31]. Moreover, the importance of aspartate residues in RdRps supported by various other mutational studies [37–39] suggested a conserved binding feature of cmp2, cmp17a and cmp21 along with side chains of Arg553, Arg555 and Ser759 which lie deep in the binding pocket of nCoV-RdRp, thus enhancing it as an anchor for inhibitory molecules.

In SARS-CoV-2-helicase, the structural information of NTPase active site obtained after superimposing SARS-helicase and yeast-



**Fig. 4.** Post-molecular dynamics (MD) analysis of SARS-CoV-2 Nsp13 helicase complexes. (A) Yeast Upf1-ADP complex (transparent white) superimposed on Nsp13 helicase to locate the ATP putative binding pocket and 2D-interaction plot is displayed. MD simulated cmp1 (brown), cmp3a (magenta), cmp11 (green) and cmp15 (sandy brown) are displayed inside the pocket. The arrangement of domains is coloured the same as in Fig. 1C. (B) MD simulated conformations of cmp11 (green) inside the predicted binding pocket of Nsp13 helicase interacted with key residues of domain 1A, 2A and 1B. (C) Per-residue energy decomposition analysis of potential binding pocket residues, and representations of chemical structures of compounds are displayed in (D).

Upf1-ADP divulged the location of conserved residues. The identified hits were positioned similarly with reference to the interaction network reported for co-crystallized ADP (Fig. 4). In the active site, the side chains of Ser310, Lys288 and Glu375 actively participated in H-bond, while Arg178 and conserved alanine residues (Ala312 and Ala314) contributed substantially to total binding free energy. Moreover, the importance of these residues has already been highlighted in mutational studies [42].

Additionally, the identified compounds exhibited promising ADMET properties and none of the compounds were found to contain any high-risk chemical group. The drug candidates with poor ADMET might lead to toxic metabolites [50] or be unable to cross membranes [51], and the predicted solubilities were estimated within an acceptable range [52]. Most of the compounds were predicted as non-inhibitors for cytochrome P450 isozymes which play a vital role in drug metabolism and inhibition of these enzymes leads to toxicity and poor excretion from the liver [53]. Moreover, all compounds were found novel as they were not reported elsewhere against antiviral activities. Among various hits identified, cmp2 against SARS-CoV-2 RdRp was recently reported as a Pan-Janus Kinase Inhibitor Clinical Candidate (PF-06263276) for its potential role in the treatment of inflammatory diseases of

lungs [54]. Hence, the cmp2 could be ideal for reducing the inflammatory activity during pathogenic COVID-19 epidemic, as lung inflammation has been observed during the SARS and MERS outbreaks.

Conclusively, this comprehensive study presents an integrated computational approach to the identification of novel inhibitors into the area of anti-nCoV treatment as a starting point, which warrants in vitro testing to evaluate compound efficacy. The in-depth structural elucidation of interacting residues together with the dynamic conformations adopted over a period of 20 ns MD simulations of identified compounds offers the way to design wide-spectrum or selective inhibitors against HCoV.

#### Compliance with ethics requirement

This article does not contain any study with human or animal subjects.

#### Conflicts of interest

The authors declare that there are no conflicts of interest.



## Acknowledgments

We would like to thank the computational facilities provided by KU Leuven. This study was supported by IRO scholarship Ph.D. Grant.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jppha.2020.04.008>.

## References

- [1] D. Benvenuto, M. Giovanetti, A. Ciccozzi, et al., The 2019-new coronavirus epidemic: Evidence for virus evolution, *bioRxiv*, 2020.
- [2] R.-f. Zhu, R.-l. Gao, S.-H. Robert, et al., Systematic review of the registered clinical trials of coronavirus diseases 2019 (COVID-19), *medRxiv*, 2020.
- [3] N. Yamamoto, S. Matsuyama, T. Hoshino, et al., Nelfinavir inhibits replication of severe acute respiratory syndrome coronavirus 2 in vitro, *bioRxiv*, 2020.
- [4] M. Wang, R. Cao, L. Zhang, et al., Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro, *Cell Res.* 30 (2020) 269–271.
- [5] M.L. Holshue, C. DeBolt, S. Lindquist, et al., First case of 2019 novel coronavirus in the United States, *N. Engl. J. Med.* 382 (2020) 929–936.
- [6] L. Caly, J.D. Druce, M.G. Catton, et al., The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro, *Antivir. Res.* 178 (2020) 104787.
- [7] P. Colson, J.-M. Rolain, J.-C. Lagier, et al., Chloroquine and hydroxychloroquine as available weapons to fight COVID-19, *Int. J. Antimicrob. Agents* 55 (2020) 105932.
- [8] P. Gautret, J.-C. Lagier, P. Parola, et al., Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial, *Int. J. Antimicrob. Agents* (2020) 105949.
- [9] B. Damle, M. Vourvahis, E. Wang, et al., Clinical pharmacology perspectives on the antiviral activity of azithromycin and use in COVID-19, *Clin. Pharmacol. Ther.* 108 (2020) 201–211.
- [10] Y. Chen, Q. Liu, D. Guo, Coronaviruses: genome structure, replication, and pathogenesis, *J. Med. Virol.* 92 (4) (2020) 418–423.
- [11] L.E. Gralinski, V.D. Menachery, Return of the coronavirus: 2019-nCoV, *Viruses* 12 (2020) 135.
- [12] E. De Clercq, Potential antiviral and antiviral strategies against SARS coronavirus infections, *Expert Rev. Anti-infect. Ther.* 4 (2006) 291–302.
- [13] H. Yang, W. Xie, X. Xue, et al., Design of wide-spectrum inhibitors targeting coronavirus main proteases, *PLoS Biol.* 3 (2005) e324.
- [14] M.A. Marra, S.J. Jones, C.R. Astell, et al., The genome sequence of the SARS-associated coronavirus, *Science* 300 (2003) 1399–1404.
- [15] P.A. Rota, M.S. Oberste, S.S. Monroe, et al., Characterization of a novel coronavirus associated with severe acute respiratory syndrome, *Science* 300 (2003) 1394–1399.
- [16] P.C. Woo, S.K. Lau, C.-m. Chu, et al., Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia, *J. Virol.* 79 (2005) 884–895.
- [17] U. Bacha, J. Barrila, A. Velazquez-Campoy, et al., Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro, *Biochemistry* 43 (2004) 4906–4912.
- [18] T. Schwede, J. Kopp, N. Guex, et al., SWISS-MODEL: an automated protein homology-modeling server, *Nucleic Acids Res.* 31 (2003) 3381–3385.
- [19] J.A. Maier, C. Martinez, K. Kasavajhala, et al., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, *J. Chem. Theor. Comput.* 11 (2015) 3696–3713.
- [20] J.J. Irwin, B.K. Shoichet, modeling, ZINC – a free database of commercially available compounds for virtual screening, *J. Chem. Infor.* 45 (2005) 177–182.
- [21] R. Kiss, M. Sandor, F.A. Szalai, A public web service for drug discovery, *J. Cheminf.* 4 (2012) P17. <http://Mcule.com>.
- [22] O. Trott, A.J. Olson, AutoDock Vina, Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi-threading, *J. Comput. Chem.* 31 (2010) 455–461.
- [23] N. Ikram, M.U. Mirza, M. Vanmeert, et al., Inhibition of oncogenic kinases: an in vitro validated computational approach identified potential multi-target anticancer compounds, *Biomolecules* 9 (2019) 124.
- [24] M.U. Mirza, M. Vanmeert, M. Froeyen, et al., In silico structural elucidation of RNA-dependent RNA polymerase towards the identification of potential Crimean-Congo hemorrhagic fever virus inhibitors, *Sci. Rep.* 9 (2020) 1–18.
- [25] T. Hou, J. Wang, Y. Li, et al., Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, *J. Chem. Inf. Model.* 51 (2011) 69–82.
- [26] H.K. Srivastava, G.N. Sastry, Molecular dynamics investigation on a series of HIV protease inhibitors: assessing the performance of MM-PBSA and MM-GBSA approaches, *J. Chem. Inf. Model.* 52 (2012) 3088–3098.
- [27] J.J. Tan, W.Z. Chen, C.X. Wang, Investigating interactions between HIV-1 gp41 and inhibitors by molecular dynamics simulation and MM–PBSA/GBSA calculations, *J. Mol. Struct.* 766 (2006) 77–82.
- [28] F. Wu, S. Zhao, B. Yu, et al., Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. <https://www.ncbi.nlm.nih.gov/nuccore/MN908947>, 2020.
- [29] H.-P. Chang, C.-Y. Chou, G.-G. Chang, Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride, *Biophys. J.* 92 (2007) 1374–1383.
- [30] O.B. Peersen, Picornaviral polymerase structure, function, and fidelity modulation, *Virus Res.* 234 (2017) 4–20.
- [31] O. Poch, I. Sauvaget, M. Delarue, et al., Identification of four conserved motifs among the RNA-dependent polymerase encoding elements, *EMBO J.* 8 (1989) 3867–3874.
- [32] P. Gong, O.B. Peersen, Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase, *Proc. Natl. Acad. Sci. USA* 107 (2010) 22505–22510.
- [33] P. Surana, V. Satchidanandam, D.T. Nair, RNA-dependent RNA polymerase of Japanese encephalitis virus binds the initiator nucleotide GTP to form a mechanically important pre-initiation state, *Nucleic Acids Res.* 42 (2013) 2758–2773.
- [34] M.U. Mirza, N.I. Noor-Ul-Huda Ghorri, A.R. Adil, et al., Pharmacoinformatics approach for investigation of alternative potential hepatitis C virus nonstructural protein 5B inhibitors, *Drug Des. Devel. Ther.* 9 (2015) 1825–1841.
- [35] M.U. Mirza, N. Ikram, Integrated computational approach for virtual hit identification against ebola viral proteins VP35 and VP40, *Int. J. Mol. Sci.* 17 (2016) 1748.
- [36] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, *Sci. Rep.* 7 (2017) 42717.
- [37] S.K. Biswas, D.P. Nayak, Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1, *J. Virol.* 68 (1994) 1819–1826.
- [38] A.L. Vázquez, J.M.M. Alonso, F. Parra, Mutation analysis of the GDD sequence motif of a calicivirus RNA-dependent RNA polymerase, *J. Virol.* 74 (2000) 3888–3891.
- [39] Y. Zhou, H. Zheng, F. Gao, et al., Mutational analysis of the SDD sequence motif of a PRRSV RNA-dependent RNA polymerase, *Sci. China Life Sci.* 54 (2011) 870–879.
- [40] Y. Gao, L. Yan, Y. Huang, et al., Structure of the RNA-dependent RNA Polymerase from COVID-19 Virus, *Science* 368 (2020) 779–782.
- [41] T.C. Appleby, J.K. Perry, E. Murakami, et al., Structural basis for RNA replication by the hepatitis C virus polymerase, *Science* 347 (2015) 771–775.
- [42] Z. Jia, L. Yan, Z. Ren, et al., Delicate structural coordination of the severe acute respiratory syndrome coronavirus Nsp13 upon ATP hydrolysis, *Nucleic Acids Res.* 47 (2019) 6538–6550.
- [43] J. Bajorath, *Cheminformatics and Computational Chemical Biology*, Springer, 2011.
- [44] M.U. Mirza, M. Vanmeert, A. Ali, et al., Perspectives towards antiviral drug discovery against Ebola virus, *J. Med. Virol.* 91 (2019) 2029–2048.
- [45] C. Wu, Y. Liu, Y. Yang, et al., Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharm. Sin.* 10 (2020) 766–788.
- [46] A. Sampath, R. Padmanabhan, Molecular targets for flavivirus drug discovery, *Antivir. Res.* 81 (2009) 6–15.
- [47] W. Liu, J.S. Morse, T. Lalonde, et al., Learning from the past: Possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV, *Chembiochem* 21 (2020) 730–738.
- [48] C. Huang, P. Wei, K. Fan, et al., 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism, *Biochemistry* 43 (2004) 4568–4574.
- [49] J. Tan, K.H. Verschuere, K. Anand, et al., pH-dependent conformational flexibility of the SARS-CoV main proteinase (Mpro) dimer: molecular dynamics simulations and multiple X-ray structure analyses, *J. Mol. Biol.* 354 (2005) 25–40.
- [50] A. Krüger, V.G. Maltarollo, C. Wrenger, et al., ADME Profiling in Drug Discovery and a New Path Paved on Silica, *Drug Discovery and Development–New Advances*, 2019. IntechOpen.
- [51] D.E. Pires, L.M. Kaminskas, D.B. Ascher, Prediction and Optimization of Pharmacokinetic and Toxicity Properties of the Ligand, *Computational Drug Discovery and Design*, Springer, 2018, pp. 271–284.
- [52] H.D. Williams, N.L. Trevasakis, S.A. Charman, et al., Strategies to address low drug solubility in discovery and development 65 (2013) 315–499.
- [53] U.M. Zanger, M. Schwab, Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities and impact of genetic variation, *Pharmacol. Ther.* 138 (2013) 103–141.
- [54] P. Jones, R.I. Storer, Y.A. Sabnis, et al., Design and synthesis of a pan-Janus kinase inhibitor clinical candidate (PF-06263276) suitable for inhaled and topical delivery for the treatment of inflammatory diseases of the lungs and skin, *J. Med. Chem.* 60 (2017) 767–786.