

**RESEARCH ARTICLE**

# Fragment-based *in silico* design of SARS-CoV-2 main protease inhibitors

Sarfraz Ahmad<sup>1</sup> | Muhammad Usman Mirza<sup>2</sup> | Lee Yean Kee<sup>1</sup> | Mamoona Nazir<sup>3</sup> |  
Noorsaadah Abdul Rahman<sup>1</sup> | John F. Trant<sup>2</sup> | Iskandar Abdullah<sup>1</sup>

<sup>1</sup>Drug Design Development Research Group, Department of Chemistry, Faculty of Science, Universiti Malaya, Kuala Lumpur, Malaysia

<sup>2</sup>Department of Chemistry and Biochemistry, University of Windsor, Windsor, ON, Canada

<sup>3</sup>Department of Pharmacy, The University of Lahore, Lahore, Pakistan

**Correspondence**

Sarfraz Ahmad, Drug Design Development Research Group, Department of Chemistry, Faculty of Science, Universiti Malaya, Kuala Lumpur 50603, Malaysia.  
Email: sarfraz.ahmad@um.edu.my

**Funding information**

Universiti Malaya, Grant/Award Number: IIRG003A-2019; Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: 553704-20

**Abstract**

3CLpro is essential for SARS-CoV-2 replication and infection; its inhibition using small molecules is a potential therapeutic strategy. In this study, a comprehensive crystallography-guided fragment-based drug discovery approach was employed to design new inhibitors for SARS-CoV-2 3CLpro. All small molecules co-crystallized with SARS-CoV-2 3CLpro with structures deposited in the Protein Data Bank were used as inputs. Fragments sitting in the binding pocket (87) were grouped into eight geographical types. They were interactively coupled using various synthetically reasonable linkers to generate larger molecules with divalent binding modes taking advantage of two different fragments' interactions. In total, 1,251 compounds were proposed, and 7,158 stereoisomers were screened using Glide (standard precision and extra precision), AutoDock Vina, and Prime MMGBSA. The top 22 hits having conformations approaching the linear combination of their constituent fragments were selected for MD simulation on Desmond. MD simulation suggested 15 of these did adopt conformations very close to their constituent pieces with far higher binding affinity than either constituent domain alone. These structures could provide a starting point for the further design of SARS-CoV-2 3CLpro inhibitors with improved binding, and structures are provided.

**KEYWORDS**

coronavirus COVID-19, fragment-based drug discovery, main protease Mpro 3CLpro, multivalency, SARS-CoV-2, X-ray crystal structure

## 1 | INTRODUCTION

Coronavirus disease 2019 (COVID-19), an ongoing pandemic announced by the World Health Organisation in March 2020, remains, as of early 2021, an escalating worldwide public health emergency affecting 216 countries with over 64 million individuals were infected, claiming around 1.5 million lives by early December 2020 (Alamri et al., 2020a). As of May 11, 2021, a total of 32,571,814 cases of coronavirus disease 2019 (COVID-19) and 579,366 associated deaths had been reported in the United States ([https://covid.cdc.](https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days)

[gov/covid-data-tracker/#cases\\_casesper100klast7days](https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days)). It is caused by the novel coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Alamri et al., 2020a; Alamri et al., 2020b; Mirza et al., 2020). The elderly, and those afflicted with chronic illness or possessing a compromised immune system are more likely to be severely affected (Shah et al., 2020). Virions are readily transmitted through aerosols or droplets adsorbing via the mucous membranes of the eyes, mouth, lungs, and nose (Karia et al., 2020). At this point, the zoological origin of the disease remains uncertain, with competing hypotheses suggesting it might have been

transferred directly from bats, or perhaps it passed through pangolin hosts where it recombined with another virus to make the highly infective version present in humans (Lam et al., 2020). This high infectivity, coupled with an approximate 1% fatality rate, has stimulated an unprecedented global effort for the identification of both effective and safe vaccines to drive down infections, as well as antiviral drug treatments to lower fatality rates when the vaccines fail for an individual (Lurie et al., 2020; Zhang, Huang, et al., 2020; Zhang, Zeng, et al., 2020).

Vaccines are the frontline strategy to curb viruses; however, prior to 2020, no coronavirus vaccine has ever been clinically deployed; partially due to the uncertainties around the duration of the protection offered and the relatively mild nature of most endemic coronaviruses such as those responsible for the common cold. The remarkable efforts of pharmaceutical companies and government scientists worldwide to produce a slew of vaccines working through different mechanisms is nothing less than an Olympian accomplishment; however, even the best efficacy rates only reach approximately 95%, insufficient for complete protection of a world of 7.5 billion people. Furthermore, we need to worry about continually monitoring antibody levels in the vaccinated population. It is likely that the protection offered is not extremely long-lasting, and the emergence of new mutants constantly threatens the protection offered by vaccines.

Coronaviruses are enveloped, single-stranded positive-sense RNA viruses with the largest documented genome size of any virus, varying between 26 and 32 kb for different strains (Alamri et al., 2020 b; Mirza & Froeyen, 2020). They have a 5'-cap and a 3'- polyadenylate tail containing 6–12 open reading frames (ORFs) (Lu et al., 2020). The first ORF (ORF 1a/b) comprises about two-thirds of the genome length and undergoes direct translation to yield two polyproteins, pp1a and pp1ab, following an a-1 frameshift between ORF1a and ORF1b (Chen et al., 2020; Hussain et al., 2005). These polyproteins are further cleaved into 16 functional non-structural proteins (nsps) by 3CLpro (Liu & Wang, 2020; Yang et al., 2003). The coronavirus 3CLpro is an approximately 300-amino acid-long, three-domain, cysteine protease (Anand, 2003; Bacha et al., 2004). Active 3CLpro requires the homodimer with two promoters within (Lai et al., 2006; Xia & Kang, 2011). A non-classical catalytic dyad (Cys145, His41) is positioned between domain I and II (Alamri et al., 2020b; Fan et al., 2005; Mirza & Froeyen, 2020). It has the capability to precisely recognize the 11 cleavage sites required to release nsp4 to nsp16 (Fan et al., 2005). In addition, it was found to exhibit self-hydrolysis activity (Ramajayam et al., 2011; Yang et al., 2006). The nsp4-to-nsp16 gene products are essential proteins for viral propagation, conducting genome replication, transcription, protein translation, cleavage, modification, and nucleic acid synthesis (Anand, 2003; Berry et al., 2015). Since proteins analogous to 3CLpro are

not found in humans, 3CLpro is an ideal antiviral target, and many groups are exploring its inhibition (Liu et al., 2020).

Unfortunately, conventional de novo drug discovery takes years to provide clinical candidates. Faster approaches include drug repurposing, structure-based drug design, and fragment-based drug design (Alamri et al., 2020 a; Ikram et al., 2019; Khalid et al., 2020; Mirza & Ikram, 2016; Mirza et al., 2016, 2019; Rehman et al., 2020; Salo-Ahen et al., 2021). This work aims to provide a comprehensive starting point for rational fragment-based drug discovery of SARS-CoV-2 3CLpro inhibitors by analyzing all the current available crystal data on the PDB. A large body of metadata collated and developed in this study is also provided to benefit drug discovery work by others.

## 2 | METHODS AND MATERIALS

A block diagram of work-breakdown structure is presented in Figure 1.

### 2.1 | Designations

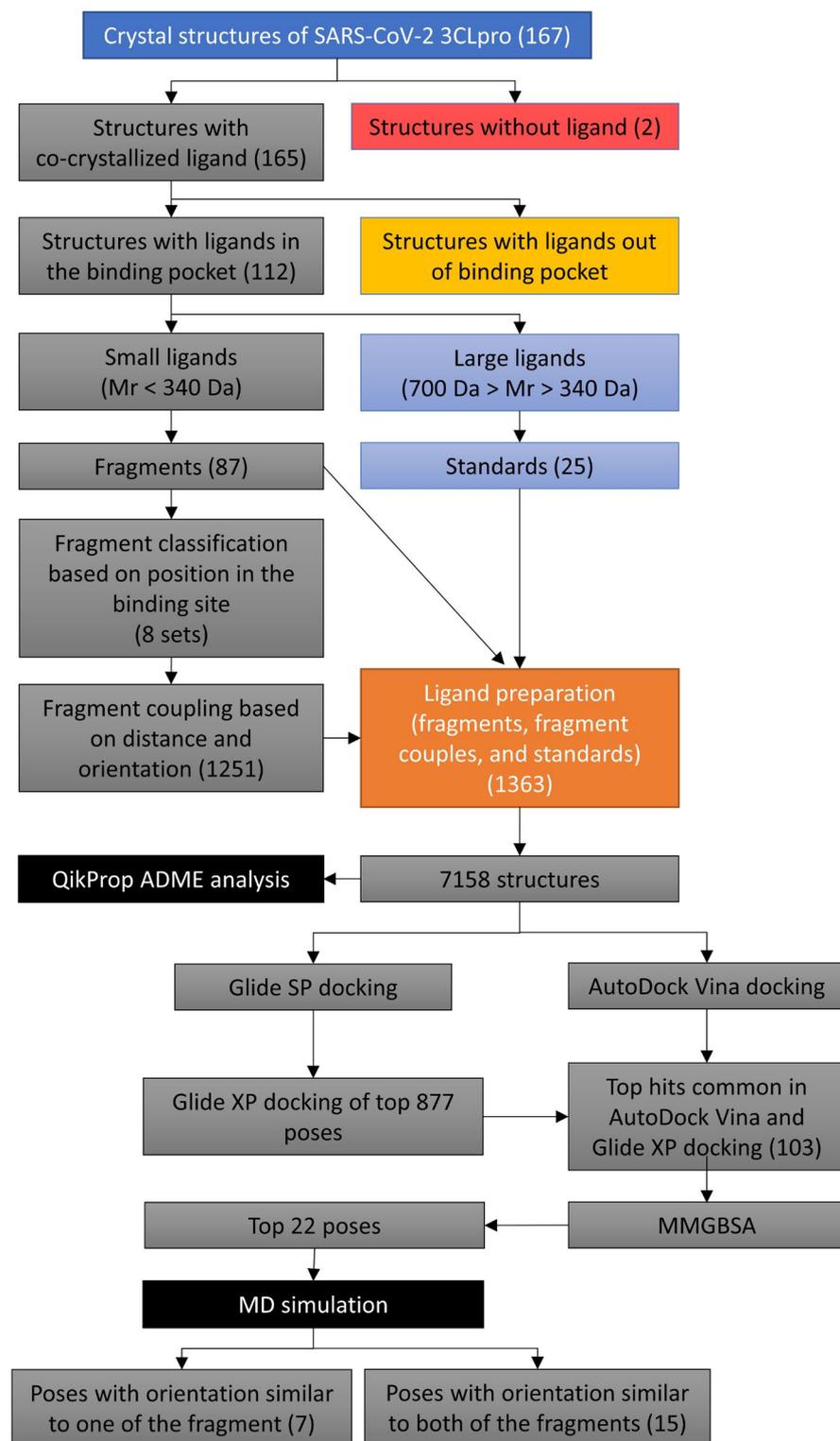
All the ligands and protein structures are presented according to their PDB identifiers. The coupled ligands were named according to their PDB identifiers separated by an underscore. The ligands with 700>Mr>340 Dalton were treated as standards, and their PDB identifiers were prefixed with L underscore (L\_PDB ID).

### 2.2 | Fragments

PDB (www.rcsb.org) comprised 167 crystal structures of SARS-CoV-2 3CLpro at the time of data accumulation (November 10, 2020). These structures were acquired and analyzed by an initial filtration for ligands with a molecular mass of less than 700 Da and residing in the active site binding pocket. This filtration leads to 112 viable crystal structures for our further study. Subsequently, 87 ligands with molecular mass less than 340 Da were selected as fragments for further development while the remaining, larger, 25 ligands were selected as controls for the docking process.

### 2.3 | Fragment coupling and linker selection

The selected 87 fragments were divided into eight groups based on their amino acid interactions and location in the binding site. The fragments were then coupled across different groups, as shown in Table 1. Coupling was ignored where more than two-atom overlap was found. While designing the linker, it was kept in mind that the joining two ligands should occupy similar



**FIGURE 1** A schematic block diagram of the workflow followed in the study [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

conformation in the binding pocket after the connection. In this regard, the distance between the joining atoms was considered to determine the number of new bonds and bond angles being formed. It is also plausible to find functional groups from other sets that occupy the empty space for the connection of any two sets. The type of amino acids in the vicinity of the putative linker was also considered while designing linkers.

## 2.4 | Virtual screening

### 2.4.1 | Ligand preparation

The SMILE strings of the 1,363 ligands (87 fragments, 25 large molecules with  $M_r > 340$  Da as standards, and 1,251 combinations) were placed in the first column of the MS

**TABLE 1** Tabulated representation of possible combinations of 8 sets of fragments. Green cells represent the potential coupling of two groups based on their positions. The digits above the red diagonal are the number of bonds that a linker could have. The numbers below the red diagonal represent the number of couples generated across two sets (sum 1,251). In the case of set 8, two couplings were possible within the set

	Set 1–24	Set 2–11	Set 3–17	Set 4–13	Set 5–8	Set 6–5	Set 7–5	Set 8–4
Set 1–24				1, 2	1	2	3	2
Set 2–11				1		2	4	1
Set 3–17						3	2	1
Set 4–13	333	143					2	
Set 5–8	158					3		1
Set 6–5	120	55	65		37		3	3
Set 7–5	120	55	30	2		25		3
Set 8–4	53	11	11		8	10	13	2

Excel and their corresponding unique identifiers were placed in the second column. Both columns were copied and pasted in the notepad, generating a smile, tab space, and the identifier in each line in the file. The contents of the notepad were converted into a single sdf file with the help of OpenBabel (v 3.1.1). The sdf file was loaded into Maestro, and the 3D structures of the compounds were prepared and optimized using the LigPrep module of Maestro (Schrödinger Release 2020–4: Maestro, LigPrep, Schrödinger, LLC, New York, NY, 2020) (Sastry et al., 2013). Possible tautomers of the compounds were produced using Epik (Roos et al., 2019) at the target pH of  $7.0 \pm 2.0$ . For stereoisomers, specified chiralities were retained, while the other stereogenic centers were varied to get a maximum of 32 isomers. Finally, a total of 7,158 structures were obtained for the docking purpose.

## 2.4.2 | Protein preparation

The crystal structure of SARS-CoV-2 3CLpro protease was retrieved in the form of the biological assembly from the Protein Data Bank (PDB ID 5R83, resolution 1.58 Å) (Di Pizio et al., 2017; Douangamath et al., 2020). 3CLpro exists as a homodimer. Thus, both chains were used in the screening process. The structure was loaded in UCSF Chimera (v 1.15) and saved as a pdb file to get both chains. The Protein Preparation Wizard of the Maestro molecular modeling software (Schrödinger Release 2020–4: Protein Preparation Wizard, Schrödinger, LLC, New York, NY, 2020) was used to prepare protein after grouping both chains of the dimer. The preparation steps involved the addition of hydrogens, optimization of the hydrogen bond networks, and assignment of protonation states of histidine residues. The water molecules were removed, and a restrained minimization was performed using the OPLS3e force field (Roos et al., 2019). Subsequently, the Receptor Grid Generation module of Maestro was employed to identify the docking site by generating a cubical grid box that was centered at (7.3, 0.9, 25.2) with 22 Å length.

## 2.5 | Docking

### 2.5.1 | Glide

Virtual screening in the specified search space of the SARS-CoV-2 3CLpro protease was performed by using the Glide docking tool of Maestro with SP (Standard Precision) mode. The top 877 poses were redocked using XP (Extra Precision) mode (Li et al., 2011).

### 2.5.2 | AutoDock Vina

For comparison purposes, 7,158 ligand poses prepared by LigPrep in the previous step were exported and docked on 3CLpro (prepared previously) using an automated Molecule server (Kiss et al., 2012). Same search box parameters were used as previously defined.

### 2.5.3 | MMGBSA

The results from Glide XP and AD Vina were compared with selected top 103 poses. The estimated binding free energy of the top ligand poses was calculated with the Prime/MMGBSA (molecular mechanics generalized Born surface area) module using the VSGB solvation model and the OPLS3e force field (Li et al., 2011).

## 2.6 | Molecular dynamics simulation

The Maestro System Builder module was used to prepare docked protein–ligand complexes for molecular dynamics (MD) simulation. Simple point-charge (SPC) water model was used with an orthorhombic box extending 10 Å from protein. The placement of ions was excluded within 20 Å of the ligand. The system was neutralized using sodium or chloride ions, and 0.15 M sodium chloride was added. OPLS3e force field was used to record 12 ns simulation with 3 ps

recording interval. NPT ensemble was used with 300 K, and 1.01325 bar, and the system was relaxed (100 ps) before simulation (Schrödinger Release 2020–4: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2020. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2020).

## 2.7 | ADME prediction

ADME (absorption, distribution, metabolism, and excretion) properties of the 7,158 ligands containing selected fragments, fragment couples, and standard molecules were predicted using QikProp tool of Maestro (Schrödinger Release 2020–4: QikProp, Schrödinger, LLC, New York, NY, 2020).

## 3 | RESULTS AND DISCUSSION

At the time of data accumulation (10 November 2020), 167 crystal structures of SARS-CoV-2 3CLpro were present in the PDB database. An in-depth analysis of these structures provided a set of 112 structures having ligands in the recognized catalytic binding pocket (Figure 2). These ligands were further divided into two subsets based on molecular mass. A small molecule-fragment subset with 87 ligands was analyzed for this study. The frequency of the protein's amino acids involvement in binding interactions was determined (Figure 3). This information was then used to group the extant ligands into 8 families of inhibitors (Figure 4).

### 3.1 | Binding site topology

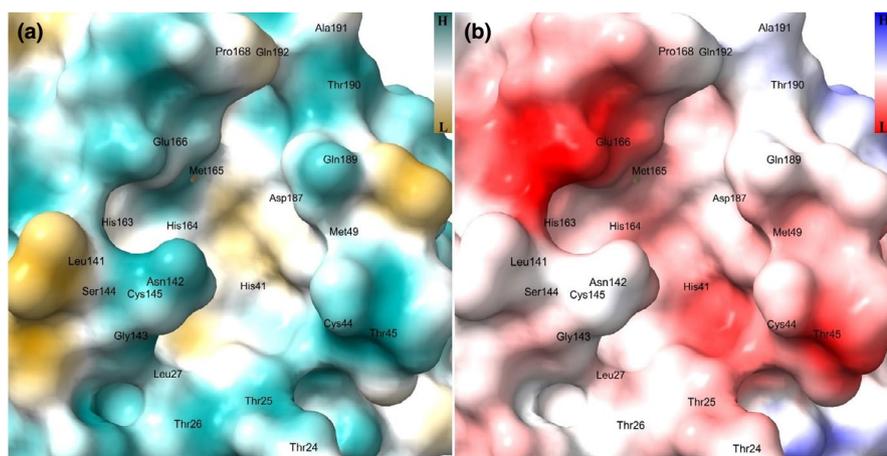
Based on the observed interactions, 23 residues are implicated in ligand contact, of which 9 residues were involved in only one interaction. The binding site is predominantly composed of polar amino acids (Figure 2), and Gly143, Ser144, and Cys145 were found making H-bonds with 52, 32, and 49

ligands, respectively (Figure 3). In contrast, Leu27, Met49, and Met165 are the lone prominent lipophilic residues involved in only 4, 3, and 13 hydrophobic interactions, respectively. The highest diversity of interactions was observed for His41, located at the base of the pocket, which participated in 25  $\pi$ -stacking, eight  $\pi$ -cationic interactions, 5 H-bonds, three hydrophobic interactions, three water bridges, and one salt bridge.

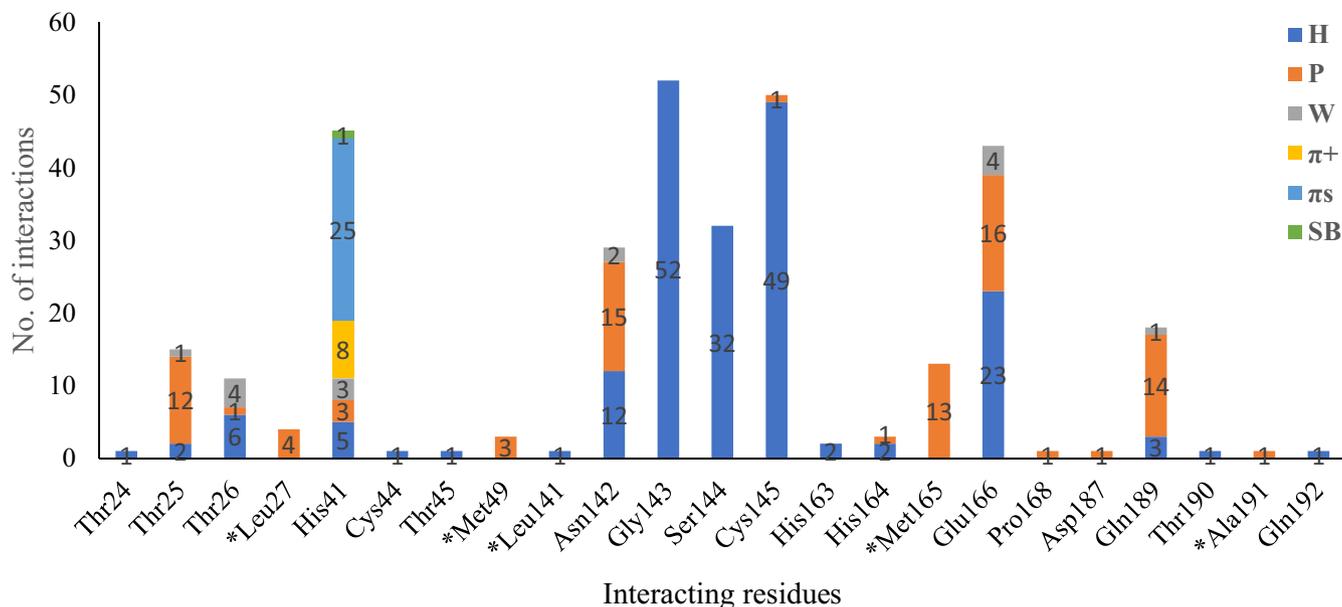
### 3.2 | Fragments

The selected fragment complexes were meticulously analyzed for ligand–protein interactions, including the frequency of binding site residues across the 87 selected complexes (Figure 3). The ligands were categorized based on their occupancy geometry and ligand–amino acid interactions in the binding site and divided into eight sets (Figure 4).

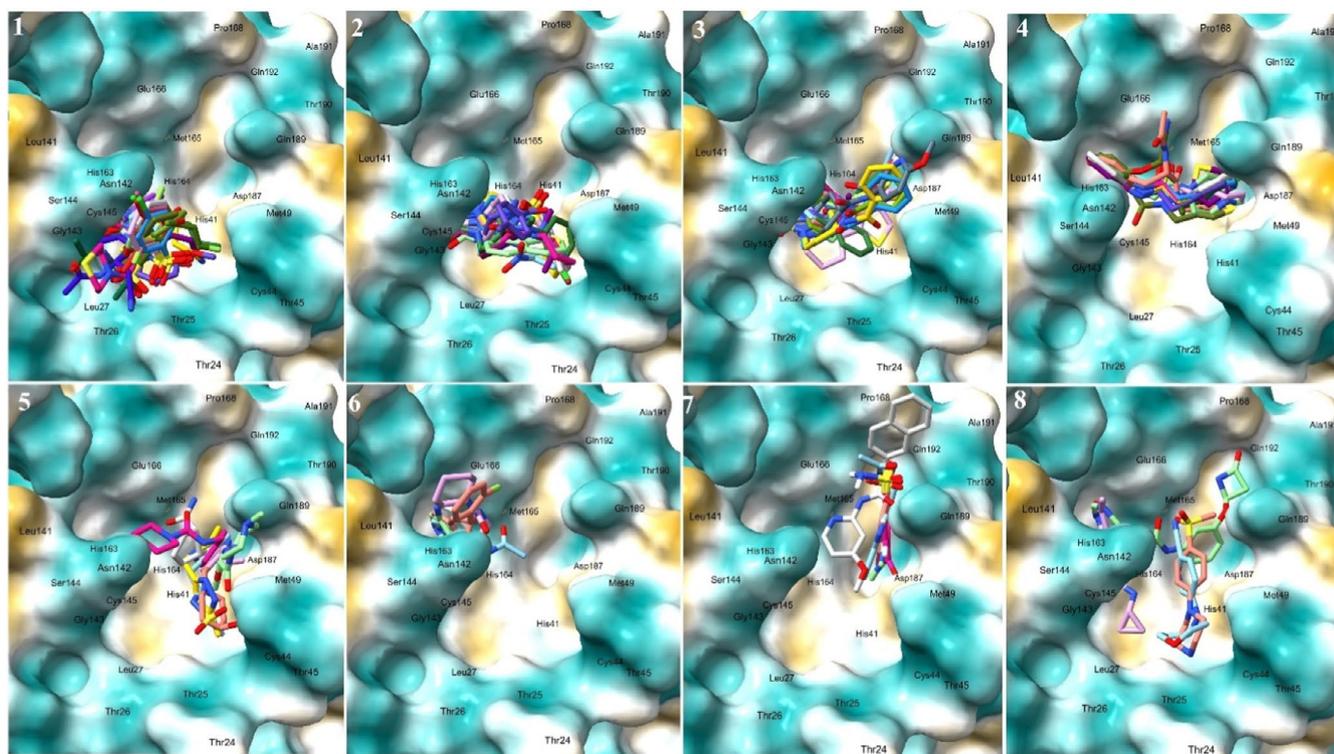
Set 1 is the largest category containing 24 compounds. These ligands were found residing in the bottom left of the binding pocket and interact with Thr25, Thr26, Leu27, His41, Asn142, Gly143, Ser144, and Cys145 (Figure 5). Set 2 consists of 11 compounds, residing slightly higher in the binding pocket than Set 1. These compounds favor interacting with Gly143, Ser144, and Cys145; their lack of interactions with Thr26 and Asn142 differentiate them from Set 1. Set 3 consists of 17 compounds. The primary difference between set 2 and set 3 is that the latter tend to interact strongly with His41. The ligand T8 M was represented by two crystal structures 5RFX and 5RHA, nearly identical except that an H-bond with Ser144 in 5RFX is replaced with a  $\pi$ -stacking interaction with His41 in 5RHA. Set 4 consists of 13 compounds. Here, Glu166 was interacting with most ligands via both H-bond and hydrophobic interactions. Set 5 contains eight compounds, which primarily reside in the center of the pocket and interact with His41. Set 6 contains the only five compounds residing in the left gorge of the binding pocket and stabilized by interactions with Asn142 and Glu166. Set 7 comprises five compounds that predominantly interact with the top gorge



**FIGURE 2** The surface location of the 23 residues that of 3CLpro that participated with the 87 observed ligands (PDB ID 5R83): (a) surface plot of atomic lipophilicity (b) surface plot of amino acid lipophilicity, the Kyte-Doolittle scale. H and L in the top right of both images represent hydrophilicity and lipophilicity, respectively. The images were generated using UCSF ChimeraX (v 1.1) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 3** Number and type of interactions presented by the 3CLpro binding site amino acid residues with the 87 selected fragments in terms of H (H-bond), P (hydrophobic interaction), W (water bridge),  $\pi_s$  ( $\pi$ -stacking),  $\pi^+$  ( $\pi$ -cationic), and SB (salt bridge); \* represents lipophilic amino acids [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

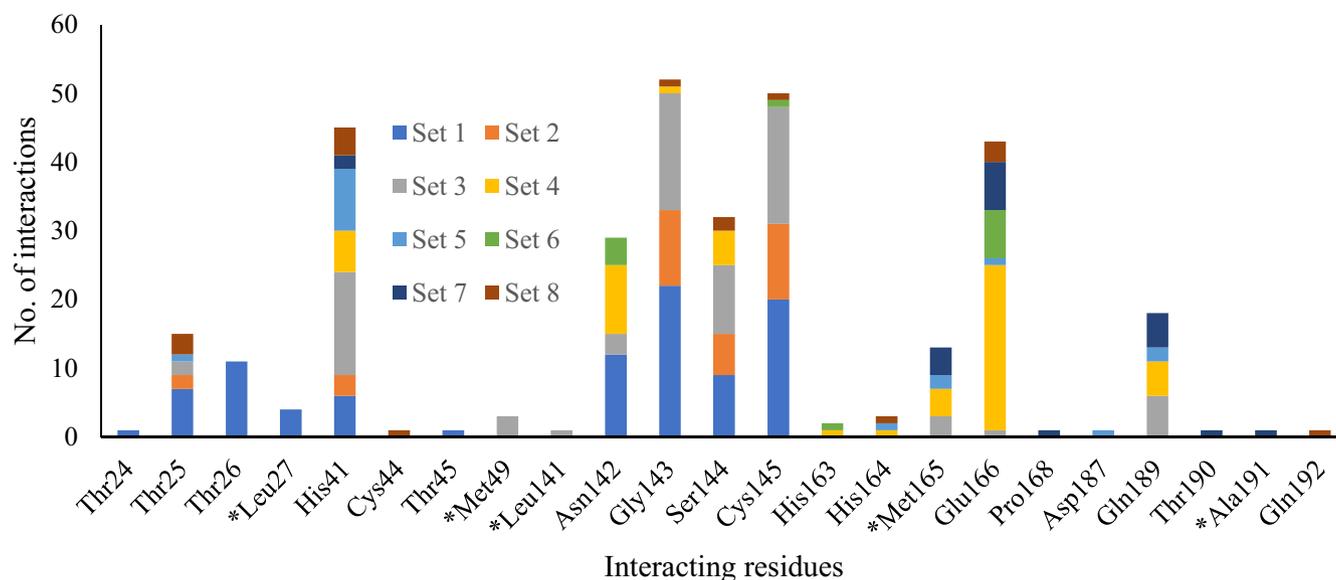


**FIGURE 4** Pictorial representation of the eight families of the co-crystallized fragments in the binding site of 3CLpro. In each case, all members of the family are mutually superimposed in a static binding pocket [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

amino acids (Met165, Glu166, and Gln189). Set 8 represents the remaining “miscellaneous” four compounds with diverse binding modes; however, even here, NTG and US7 share a similar binding site orientation; but UOP and UGD occupy bottom left and top left gorges, respectively (Figure 6).

### 3.3 | Fragment coupling and linker selection

Fragment-based drug design is widely employed in drug discovery (Erlanson et al., 2019; Kashyap et al., 2018). Commercial fragment libraries are available as online



**FIGURE 5** Graphical representation of the eight families of co-crystallized fragments based on their interactions with the binding site residues of 3CLpro. Lipophilic residues are labeled with \* [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

databases comprising large bodies of drug-like fragments that can be incorporated into larger molecules for in silico analysis against the target protein (Keserü et al., 2016; Rudling et al., 2017). The newly designed hits are synthesized then screened for their in vitro efficacy, and selected leads are further explored using X-ray crystallography to confirm the predicted mode (Aitipamula & Vangala, 2017). The current study is unique in terms of fragment selection by inverting this process: crystallographic binding poses of the selected fragments in the designated active sites of SARS-CoV-2 3CLpro were used as a starting point to design larger molecules. It was hypothesized that using appropriate linkers to join the fragment poses could provide a larger molecule whose divalent binding would encompass the interactions formed by its component domains. This approach has been used before for this protein, but only on a far more limited scale where three co-crystallized fragments were used to design 19 ligands (Luan & Huynh, 2020).

After identification of the small drug-like fragments, the second step in the rational fragment-based drug discovery approach involves the coupling of these fragments to build large molecules with improved affinity for the selected target protein. In this regard, the choice of the linker group connecting two fragments and the connection sites of the fragments is of critical importance. In this study, the sole criterion for linker selection was to ensure that the resulting two-domain molecule retained the binding modes that both domains showed independently in their crystal structures. A rigorous campaign of ligand coupling was carried out. Some couplings were rejected due to the overlap of fragments, and some couples were linked through multiple linkers, resulting in a total of 1,251 structures. In cases where the fragments are in close proximity, direct connection through an amine,

or a methylene or ethanyl bridge was used. Amine and amide containing linkers were used when three to four bond-lengths were desired. In some Set 2—Set 4 couplings, cyclizations were considered to form a ring between two adjacent piperazine carbons in Set 2 and urea functionalities in Set 4 to make a five-membered ring. A SMILE file containing these compound structures, along with the 87 starting fragments and 25 control compounds is provided in the supplementary information.

### 3.4 | Virtual screening

#### 3.4.1 | Ligand preparation

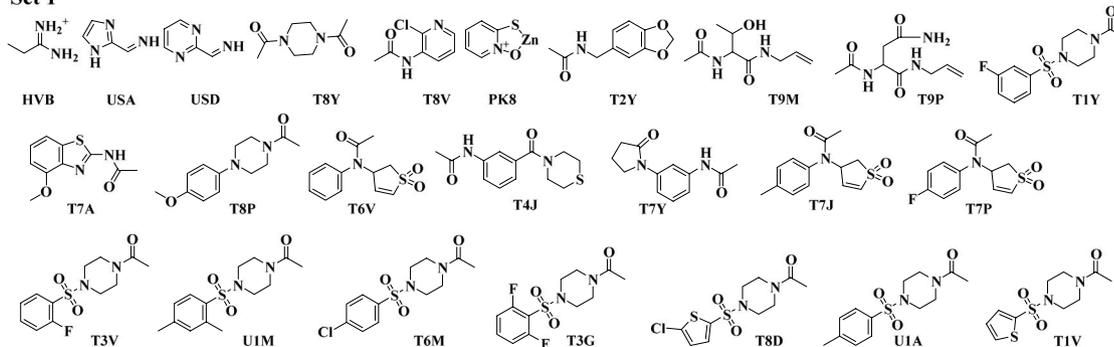
A reasonably high number of stereoisomers were sampled (maximum of 32 for each structure) to obtain a larger number of conformations for screening and maximize the probability of identifying the stereoisomers best able to imitate the original bound fragments. The twenty-five co-crystallized ligands with MW >340 Da were used as standards during docking. A SMILE file with 7,158 stereoisomers derived from the 1,251 compounds is provided in the supplementary information.

### 3.5 | Glide and AutoDock Vina

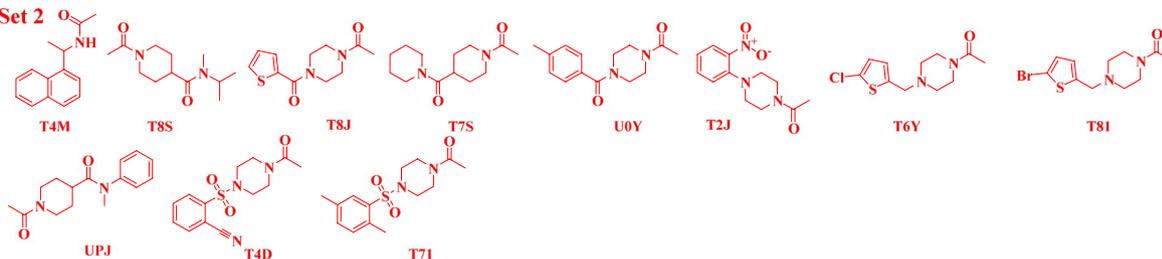
#### 3.5.1 | MMGBSA

Molecular mechanics generalized Born surface area (MMGBSA) was used to predict binding free energy ( $\Delta G_{\text{bind}}$ ) of the ligand–receptor complex. The top 103 ligands were ranked according to their MMGBSA values, and the top

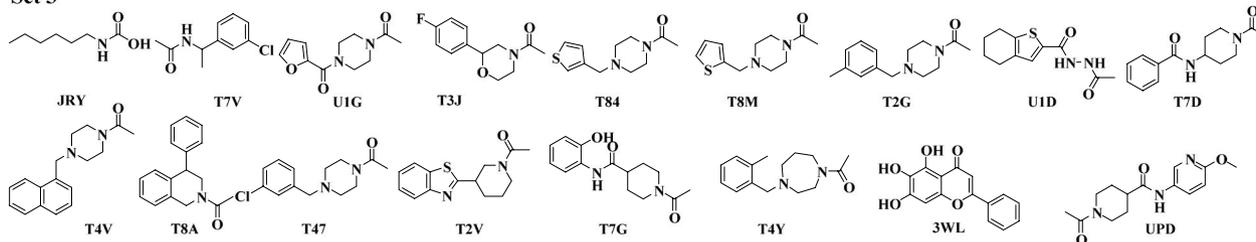
## Set 1



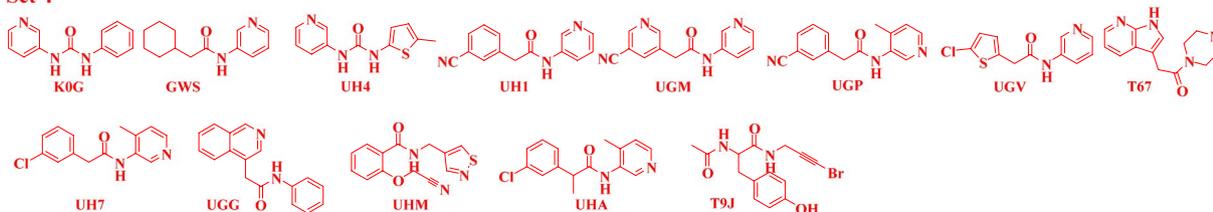
## Set 2



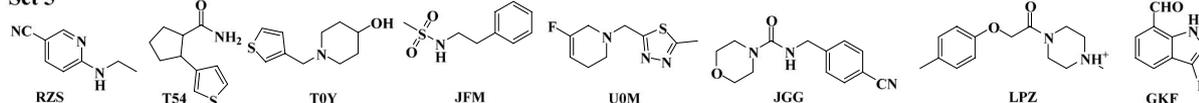
## Set 3



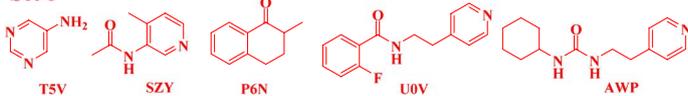
## Set 4



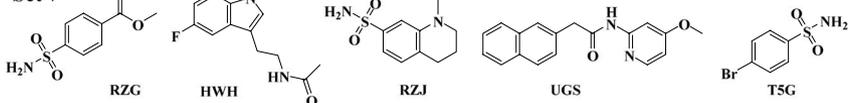
## Set 5



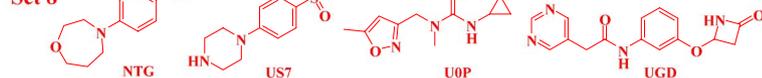
## Set 6



## Set 7



## Set 8



**FIGURE 6** Structures of the 87 3CLpro ligands divided into eight sets based on their binding modes. The PDB IDs of the ligands are given to access related data (<https://www.rcsb.org/>) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

22 ligands (Table 2) were selected for molecular dynamics simulation.

### 3.5.2 | Molecular dynamics simulation

Due to the binding site's polar nature, as it opens to solvent, the predominant interactions are H-bonds and water bridges. Hydrophobic interactions are far less pronounced, as noted above. Consequently, we ranked the top 22 compounds by H-bond occupancy during a 12 ns simulation (Figure 7) as an additional proxy for the strength of their interaction with the pocket and also provided the fraction of the hydrophobic interactions, extant water bridges, and  $\pi$ - $\pi$  and  $\pi$ -cation interactions over the simulation.

RMSD graphs of 22 ligands with respect to the reference conformation (first frame at time  $t = 0$ ) are given in Figure 8. These demonstrate the extent of ligand deviation from the initial pose (at  $t = 0$ ) required before reaching a stable conformation. They also illustrate relative ligand fluctuation,

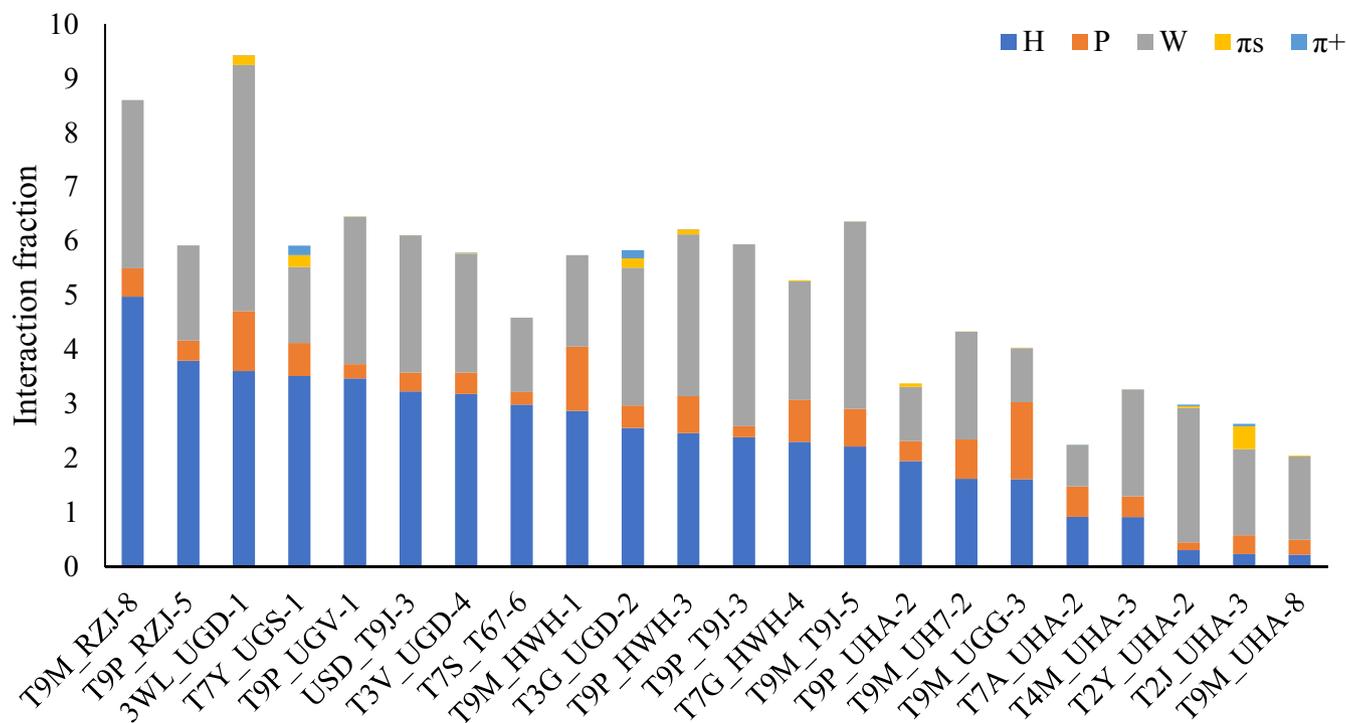
should there be multiple low-lying and low-barrier states; or stability, should there be one dominant energy well. For example, T7A\_UHA-2 was stable between 3 and 3.5 Å, and T9P\_RZJ-5 was stable between 0.8 and 1 Å, indicating that the former sampled several similar binding modes, while the latter had predominantly one stable form. To better contextualize this data, we have chosen three of these compound ligands, T9P\_RZJ-5, T7Y\_UGS-1, and T7S\_T67-6. The binding interaction profile of all the remaining compounds is provided in supplementary information from Figure S1 to S19. In order to analyze the binding interaction profile, the most representative conformation of binding pose of each compound was selected through MD clustering. MD clustering represents the clusters of conformations with a deviation of  $<1$  Å, and the largest cluster displays the most representative of the ligand. Therefore, binding pose of each compound from the largest cluster was used to generate 2D and 3D interaction plots.

T9P\_RZJ-5 provided a  $-6.94$  kcal/mol XP score and a  $\Delta G_{\text{bind}}$  of  $-64.02$  kcal/mol, according to MMGBSA (Table

**TABLE 2** The top 22 hits, as per MMGBSA scores, of compound molecules. Glide extra precision (XP) and AutoDock (AD) Vina scores are presented along with the number of fragments whose binding mode is conserved from their crystal structures. The three compounds discussed in greater detail below are highlighted in red. In the "Compounds" column, two fragments involved in making a compound are separated by an underscore, and the number after the dash represents pose number generated by LigPrep

Sr. No.	Compound	Glide XP score	Conserved fragments	AD Vina score	Conserved fragments	MMGBSA $\Delta G_{\text{bind}}$
1	T7Y_UGS-1	-7.42	2	-9	2	-76.06
2	T3V_UGD-4	-8.11	1	-8.1	1	-66.79
3	T3G_UGD-2	-7.02	1	-8.3	2	-67.35
4	T7G_HWH-4	-7.72	1	-9.0	1	-64.76
5	T9M_UH7-2	-7.89	2	-7.0	2	-63.48
6	T9P_UGV-1	-7.14	1	-7.3	2	-64.78
7	T9M_RZJ-8	-7.02	2	-7.4	1	-64.86
8	T9P_UHA-2	-8.01	2	-7.6	1	-62.55
9	T9P_RZJ-5	-6.94	2	-7.5	2	-64.02
10	T2Y_UHA-2	-7.43	1	-7.6	2	-61.52
11	T9M_HWH-1	-6.84	2	-7.9	2	-59.25
12	T9M_UHA-8	-7.40	2	-7.6	1	-58.02
13	3WL_UGD-1	-7.07	2	-10.1	2	-58.58
14	T9M_UGG-3	-7.07	2	-8.6	0	-57.41
15	T4M_UHA-3	-7.93	2	-7.8	2	-53.61
16	USD_T9J-3	-6.88	2	-7.9	2	-55.41
17	T2J_UHA-3	-6.95	2	-8.2	1	-55.26
18	T7S_T67-6	-6.98	1	-9.0	1	-55.18
19	T7A_UHA-2	-7.01	2	-7.9	2	-53.73
20	T9P_HWH-3	-7.20	2	-7.5	1	-52.09
21	T9P_T9J-3	-7.30	1	-7.7	0	-50.68
22	T9M_T9J-5	-7.22	2	-7.2	1	-46.63

Note: The energy values are in kcal/mol.



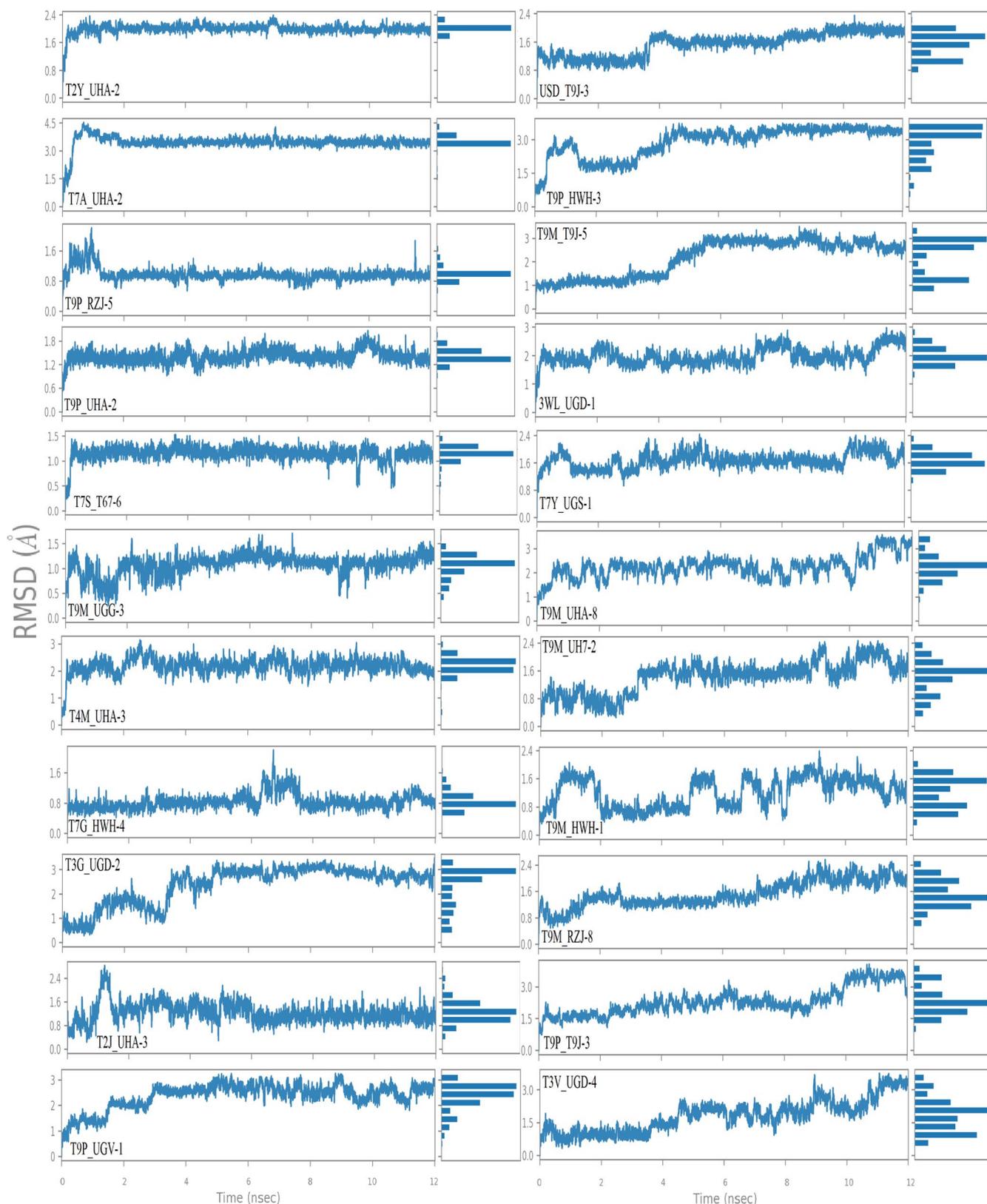
**FIGURE 7** Interaction fraction (the number of interactions normalized over the 12 ns simulation period) of the selected ligands for hydrogen bond (H), hydrophobic interaction (P), water bridge (W),  $\pi$  stacking ( $\pi_s$ ), and  $\pi$  cationic ( $\pi^+$ ) interactions during MD simulation. An interaction fraction of 3.1 for water bridges of T9M\_RZJ-8 stands for 3 water bridges 100% of the simulation time and a 4th one 10% of the simulation time [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

2). An aromatic sulphonamide, it has the 2nd highest H-bond occupancy ranking among the studied drugs. Thr25 and Gln189 both make stable single H-bonds, while His41 forms two very stable H-bonds with the ligand (Figure 9). All the ligand atoms have an RMSF of around 1 Å except one sulphonamide oxygen (2 Å) and the CH<sub>2</sub> of the terminal alkene (1.5 Å) (Figure 10), which occupies a larger empty pocket between Cys145 and Met165 (Figure 9). It presented a stable RMSD of around 0.8 Å compared to the reference frame ( $t = 0$ ), indicating that the pose obtained from XP docking remained largely static during the simulation (Figure 8). Figure 9 shows a comparison of co-crystallized poses of T9P and RZJ with the MD simulated pose of T9P\_RZJ-5. RZJ and the same domain in T9P\_RZJ-5 adopt similar conformations with only a slight tilt; however, the T9P component flips its orientation in its binding site. This molecule appears to be an entirely novel prophetic compound, and no identical and similar hits were found in ChEMBL, PubChem, and SciFinder repositories. ChEMBL3467038, an N-acetyl derivative of the fragment RZJ, has been found to be nontoxic toward HepG2 cells (Singh et al., 2011), and a related compound ChEMBL275919 is a proposed estrone sulfatase inhibitor (Di Pizio & Niv, 2015).

T7Y\_UGS-1 has a  $-7.42$  kcal/mol XP score, a  $-9$  kcal/mole AD Vina score, and the highest MMGBSA  $\Delta G_{\text{bind}}$  of  $-76.06$  kcal/mol in the top 22 subset (Table 2). It is a naphthalene derivative having pyridine and pyrrolidone residues.

It was found making two H-bonds with Gln166 and one H-bond with His41 (Figure 10f). RMSF plot of ligand fit on protein showed that naphthalene, methoxy N-acetyl residues fluctuated more than 1 Å, while the rest of the atoms were close to 1 Å (Figure 10b). Furthermore, Figure 8 showed that the ligand mostly stayed at an RMSD between 1.2 and 2 Å compared to the initial pose (at  $t = 0$ ). It is evident from Figure 10b that UGS part of T7Y\_UGS-1 is very closely aligned with its co-crystallized posture, and T7Y part of T7Y\_UGS-1 is drifted toward His41 as compare to its co-crystallized pose that is closer to Ans142. T7Y\_UGS-1 or its closely related derivatives have not been reported in ChEMBL, PubChem, and SciFinder. In terms of fragments, a derivative of UGS having benzene in place of pyridine (ChEMBL1384462) has been reported for multiple biological activities, particularly against hepatitis C virus and influenza NS1 (Pydi et al., 2015). T7Y (ChEMBL1565601) has been reported against DNA polymerase  $\alpha$ , geminin, lysosomal  $\alpha$ -glucosidase, and apurinic/aprimidinic endonuclease-1 (Floriano et al., 2006).

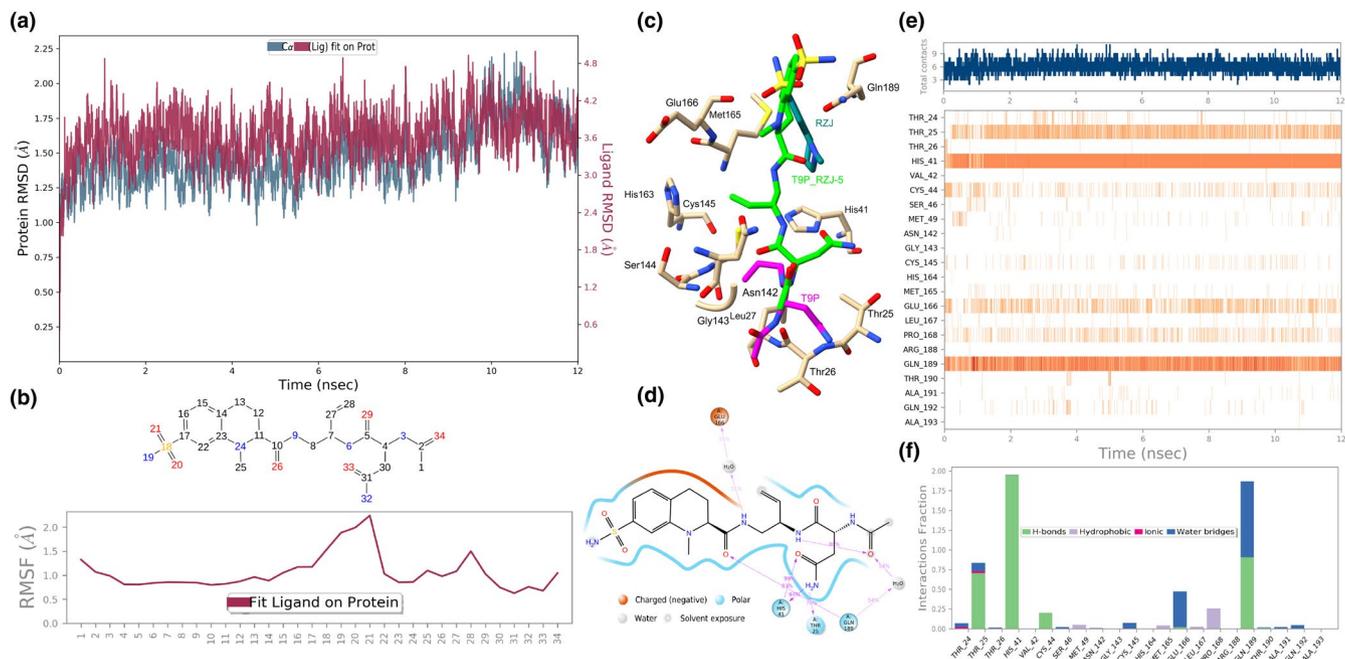
T7S\_T67-6 is a 7-azaindole derivative with one N-methylpiperazine and two piperidine residues. It has a Glide XP score of  $-6.98$  kcal/mol, an AD Vina score of  $-9$  kcal/mol, and an MMGBSA  $\Delta G_{\text{bind}}$  of  $-55.18$  kcal/mol (Table 2). It forms two H-bonds with Gln166 and one H-bond with Gln192 during most of the simulation period. A conserved hydrophobic interaction was also observed with Gln189 (Figure 11e,f). The RMSFs of all atoms of the ligand are  $<1$  Å except for carbon 11 of the



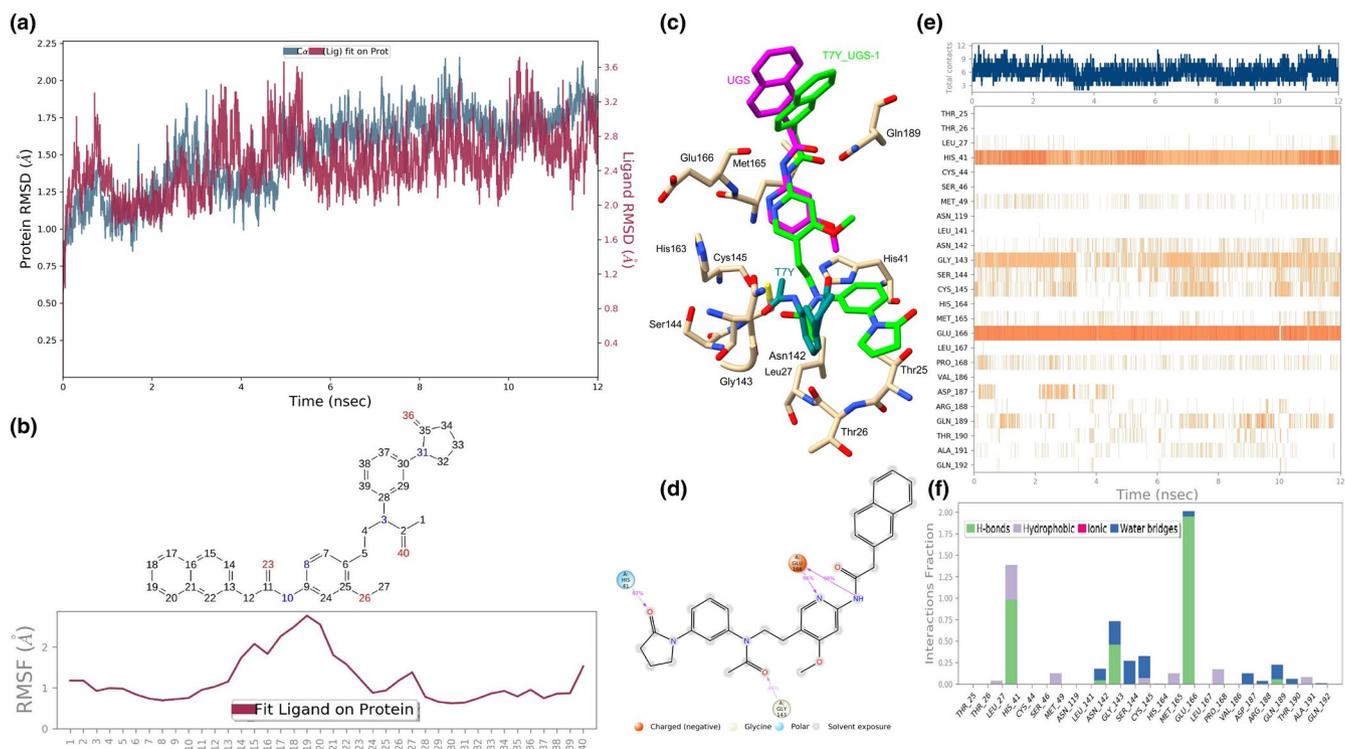
**FIGURE 8** Root mean square deviation (RMSD) of 22 ligands with respect to the reference conformation (first frame at time  $t = 0$ ). In the bar graphs, each bar's length represents the fraction of time at a particular RMSD out of 12 ns [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

terminal piperidine, which was slightly higher than 1 Å (Figure 11b). The RMSD showed dynamic stability, remaining between 1 and 1.4 Å when compared to the first frame ( $t = 0$ ) (Figure

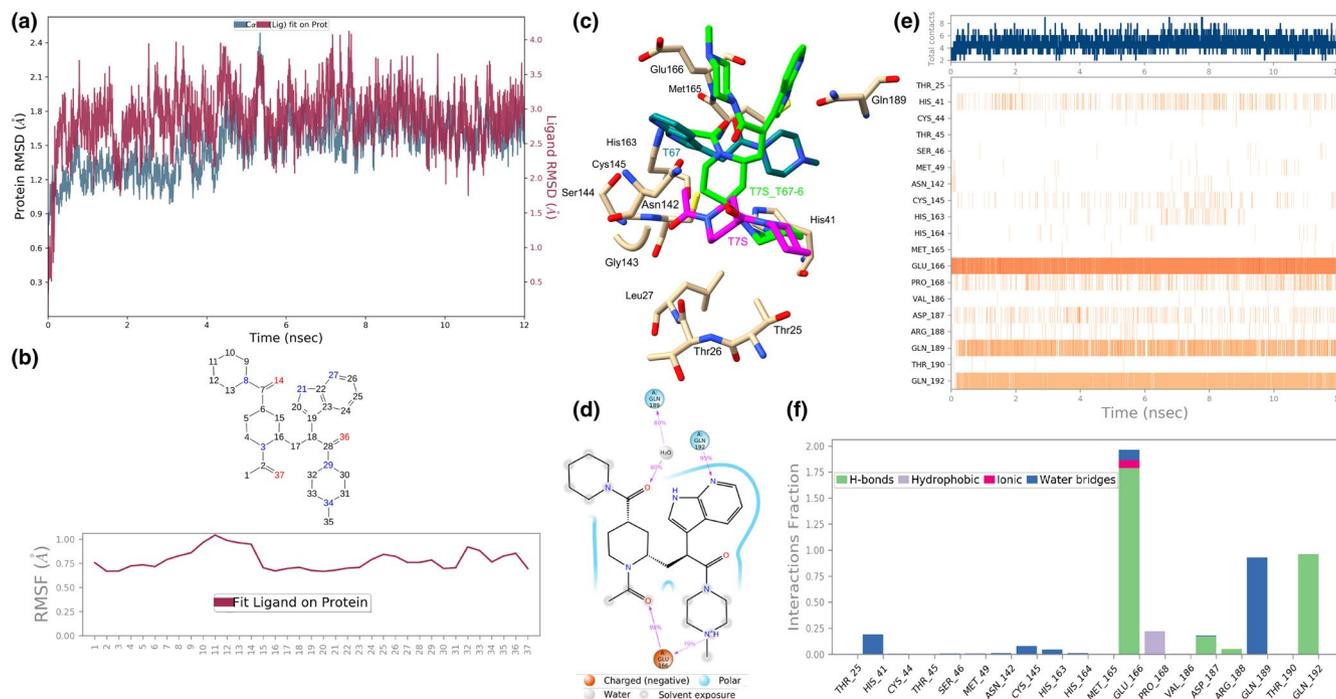
8). The T7S domain of T7S\_T67-6 remained nearly coincident with the co-crystallized pose of the isolated fragment; but the length of the linker does not allow the T67 domain to adopt



**FIGURE 9** A typical MD simulation workup for a compound drug: data for T9P\_RZJ-5. (a) Protein C $\alpha$  RMSD on the left y-axis and ligand RMSD fit on protein on the right y-axis. (b) RMSF of ligand atoms. (c) MD simulated pose of the compound ligand in lime, superimposed on the co-crystallized poses of its fragments in teal and magenta. (d) Highlight of the key protein–ligand interactions that exist during more than 30.0% of the simulation time. (e) A timeline representation of the interactions and contacts (H-bonds, hydrophobic, ionic, and water bridges) between the 23 implicated residues and the compound ligand. (f) Interaction fraction (the number of interactions normalized throughout the trajectory) of the ligand for hydrogen bond, hydrophobic interaction, ionic interaction, and water bridges with each of the key residues [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 10** A typical MD simulation workup for a compound drug: T7Y\_UGS-1. (a) Protein C $\alpha$  RMSD on the left y-axis and ligand RMSD fit on protein on the right y-axis. (b) RMSF of ligand atoms. (c) MD simulated pose of the compound ligand in lime, superimposed on the co-crystallized poses of its fragments in teal and magenta. (d) Highlight of the key protein–ligand interactions that exist during more than 30.0% of the simulation time. (e) A timeline representation of the interactions and contacts (H-bonds, hydrophobic, ionic, and water bridges) between the 23 implicated residues and the compound ligand. (f) Interaction fraction (the number of interactions normalized throughout the trajectory) of the ligand for hydrogen bond, hydrophobic interaction, ionic interaction, and water bridges with each of the key residues [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 11** A typical MD simulation workup for a compound drug: T7S\_T67-6. (a) Protein C $\alpha$  RMSD on the left y-axis and ligand RMSD fit on protein on the right y-axis. (b) RMSF of ligand atoms. (c) MD simulated pose of the compound ligand in lime, superimposed on the co-crystallized poses of its fragments in teal and magenta. (d) Highlight of the key protein–ligand interactions that exist during more than 30.0% of the simulation time. (e) A timeline representation of the interactions and contacts (H-bonds, hydrophobic, ionic, and water bridges) between the 23 implicated residues and the compound ligand. (f) Interaction fraction (the number of interactions normalized throughout the trajectory) of the ligand for hydrogen bond, hydrophobic interaction, ionic interaction, and water bridges with each of the key residues [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

a conformation consistent with that of the isolated fragment (Figure 11c). Two closely relatives of T7S (ChEMBL1574087 and ChEMBL1541069) are potent inhibitors of human TrxR1, SMN2, and IMPase proteins, and influenza A NS1 protein in the low nanomolar range (Biar $\acute{n}$ es et al., 2010; Miguet et al., 2006). Moreover, ChEMBL1574087 is an inhibitor of hepatitis C virus in a cell-based assay (5,012 nM). Furthermore, close analogues of T67 (ChEMBL231023 and ChEMBL3754439) are weak anti-inflammatory agents in the rat model (Tan et al., 2012). Of course, again, the new ligand has no known similar compounds according to a Scifinder search.

### 3.5.3 | QikProp

ADME property assessment of the top-22 hits are provided in the supplementary information.

## 4 | CONCLUSION AND FUTURE DIRECTIONS

This comprehensive study of known small molecule fragments co-crystallized with 3CLpro employed a crystal

structure-guided fragment-based approach to identify potential new inhibitors. The 87 fragments matching the criteria were connected *via* linkers reasonably expected to provide conformations of the compound drugs that reflect those preferred by its component domains. The 1,251 resulting structures were screened using docking, MMGBSA calculations, and MD simulations. The top 22 hits identified, after MD simulation, generally had one or both of their component domains adopt conformations consistent with the isolated fragments. A closer examination of three promising hits with diverse structures (T7Y\_UGS-1, T9P\_UHA-2, and T7S\_T67-6), suggest they might demonstrate significant activity that warrants their synthesis and biological testing. This two-domain approach could be extended to a three-domain strategy using mutually non-overlapping sets, such as set1+set5+set6 or set1+set6+set7, that could offer further enhanced affinity.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the Universiti Malaya (grant no.: IIRG003A-2019) for the necessary resources to complete this work, and the Natural Sciences and Engineering Research Council of Canada COVID-Alliance Funding Stream (grant no.: 553704-20 to JFT).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Sarfraz Ahmad contributed to study design, experimental work, and data collection; Sarfraz Ahmad, Iskandar Abdullah, and Muhammad Usman Mirza contributed to in silico analysis and interpretation of the data; Lee Yean Kee, Sarfraz Ahmad, and Mamoona Nazir contributed to drafting the manuscript and literature survey; John F. Trant and Noorsaadah Abdul Rahman contributed to critical revision of the manuscript.

## DATA AVAILABILITY STATEMENT

Freeware OpenBabel (v 3.1.1), UCSF Chimera (v 1.15), UCSF ChimeraX (v 1.1), and mcule Autodock Vina platform, and commercial Schrödinger drug discovery suite (Release 2020–4) were used in the study. A significant volume of data produced in the study is provided in the supporting information to help scientific community to continue the further development.

## REFERENCES

- Aitipamula, S., & Vangala, V. R. (2017). X-ray crystallography and its role in understanding the physicochemical properties of pharmaceutical cocrystals. *Journal of the Indian Institute of Science*, *97*(2), 227–243. <https://doi.org/10.1007/s41745-017-0026-4>
- Alamri, M. A., Tahir ul Qamar, M., Mirza, M. U., Alqahtani, S. M., Froeyen, M., & Chen, L.-L. (2020). Discovery of human coronaviruses pan-papain-like protease inhibitors using computational approaches. *Journal of Pharmaceutical Analysis*, *10*(6), 546–559. <https://doi.org/10.1016/j.jpha.2020.08.012>
- Alamri, M. A., Tahir ul Qamar, M., Mirza, M. U., Bhadane, R., Alqahtani, S. M., Muneer, I., Froeyen, M., & Salo-Ahen, O. M. H. (2020). Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro. *Journal of Biomolecular Structure and Dynamics*, *39*, 1–13.
- Anand, K. (2003). Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-SARS drugs. *Science*, *300*(5626), 1763–1767. <https://doi.org/10.1126/science.1085658>
- Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A., & Freire, E. (2004). Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry*, *43*(17), 4906–4912.
- Berry, M., Fielding, B. C., & Gamielien, J. (2015). Potential broad spectrum inhibitors of the coronavirus 3CLpro: A virtual screening and structure-based drug design study. *Viruses*, *7*(12), 6642–6660. <https://doi.org/10.3390/v7122963>
- Biarnés, X., Marchiori, A., Giorgetti, A., Lanzara, C., Gasparini, P., Carloni, P., Born, S., Brockhoff, A., Behrens, M., & Meyerhof, W. (2010). Insights into the binding of Phenyltiocarbamide (PTC) agonist to its target human TAS2R38 bitter receptor. *PLoS One*, *5*(8), e12394. <https://doi.org/10.1371/journal.pone.0012394>
- Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*, *92*(4), 418–423. <https://doi.org/10.1002/jmv.25681>
- Di Pizio, A., Kruetzfeldt, L.-M., Cheled-Shoval, S., Meyerhof, W., Behrens, M., & Niv, M. Y. (2017). Ligand binding modes from low resolution GPCR models and mutagenesis: Chicken bitter taste receptor as a test-case. *Scientific REPORTS*, *7*(1), 1–11. <https://doi.org/10.1038/s41598-017-08344-9>
- Di Pizio, A., & Niv, M. Y. (2015). Promiscuity and selectivity of bitter molecules and their receptors. *Bioorganic & Medicinal Chemistry*, *23*(14), 4082–4091. <https://doi.org/10.1016/j.bmc.2015.04.025>
- Douangamath, A., Fearon, D., Gehrtz, P., Krojer, T., Lukacik, P., Owen, C. D., Resnick, E., Strain-Damerell, C., Aimon, A., Ábrányi-Balogh, P., Brandão-Neto, J., Carbery, A., Davison, G., Dias, A., Downes, T. D., Dunnett, L., Fairhead, M., Firth, J. D., Jones, S. P., ... Walsh, M. A. (2020). Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nature Communications*, *11*(1), 1–11. <https://doi.org/10.1038/s41467-020-18709-w>
- Erlanson, D. A., Davis, B. J., & Jahnke, W. (2019). Fragment-based drug discovery: Advancing fragments in the absence of crystal structures. *Cell Chemical Biology*, *26*(1), 9–15. <https://doi.org/10.1016/j.chembiol.2018.10.001>
- Fan, K., Ma, L., Han, X., Liang, H., Wei, P., Liu, Y., & Lai, L. (2005). The substrate specificity of SARS coronavirus 3C-like proteinase. *Biochemical and Biophysical Research Communications*, *329*(3), 934–940. <https://doi.org/10.1016/j.bbrc.2005.02.061>
- Floriano, W. B., Hall, S., Vaidehi, N., Kim, U., Drayna, D., & Goddard, W. A. (2006). Modeling the human PTC bitter-taste receptor interactions with bitter tastants. *Journal of Molecular Modeling*, *12*(6), 931–941. <https://doi.org/10.1007/s00894-006-0102-6>
- Hussain, S., Pan, J., Chen, Y., Yang, Y., Xu, J., Peng, Y., Wu, Y., Li, Z., Zhu, Y., Tien, P., & Guo, D. (2005). Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. *Journal of Virology*, *79*(9), 5288–5295. <https://doi.org/10.1128/JVI.79.9.5288-5295.2005>
- Ikram, N., Mirza, M. U., Vanmeert, M., Froeyen, M., Salo-Ahen, O. M. H., Tahir, M., Qazi, A., & Ahmad, S. (2019). Inhibition of oncogenic kinases: An in vitro validated computational approach identified potential multi-target anticancer compounds. *Biomolecules*, *9*(4), 124. <https://doi.org/10.3390/biom9040124>
- Karia, R., Gupta, I., Khandait, H., Yadav, A., & Yadav, A. (2020). COVID-19 and its modes of transmission. *SN Comprehensive Clinical Medicine*, *2*(10), 1798–1801. <https://doi.org/10.1007/s42399-020-00498-4>
- Kashyap, A., Singh, P. K., & Silakari, O. (2018). Counting on fragment based drug design approach for drug discovery. *Current Topics in Medicinal Chemistry*, *18*(27), 2284–2293. <https://doi.org/10.2174/1568026619666181130134250>
- Keserű, G. M., Erlanson, D. A., Ferenczy, G. G., Hann, M. M., Murray, C. W., & Pickett, S. D. (2016). Design principles for fragment libraries: Maximizing the value of learnings from pharma fragment-based drug discovery (FBDD) programs for use in academia. *Journal of Medicinal Chemistry*, *59*(18), 8189–8206. <https://doi.org/10.1021/acs.jmedchem.6b00197>
- Khalid, H., Landry, K. B., Ijaz, B., Ashfaq, U. A., Ahmed, M., Kanwal, A., Froeyen, M., & Mirza, M. U. (2020). Discovery of novel Hepatitis C virus inhibitor targeting multiple allosteric sites of NS5B polymerase. *Infection, Genetics and Evolution*, *84*, 104371. <https://doi.org/10.1016/j.meegid.2020.104371>
- Kiss, R., Sandor, M., & Szalai, F. A. (2012). <http://Mcule.com>: A public web service for drug discovery. *Journal of Cheminformatics*, *4*(S1), P17.

- Lai, L., Han, X., Chen, H., Wei, P., Huang, C., Liu, S., Fan, K., Zhou, L., Liu, Z., Pei, J., & Liu, Y. (2006). Quaternary structure, substrate selectivity and inhibitor design for SARS 3C-like proteinase. *Current Pharmaceutical Design*, *12*(35), 4555–4564.
- Lam, T.-T.-Y., Jia, N., & Cao, W.-C. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, *583*, 282–285.
- Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., & Friesner, R. A. (2011). The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins*, *79*(10), 2794–2812.
- Liu, X., & Wang, X.-J. (2020). Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *Journal of Genetics and Genomics*, *47*(2), 119. <https://doi.org/10.1016/j.jgg.2020.02.001>
- Liu, Y., Liang, C., Xin, L., Ren, X., Tian, L., Ju, X., Li, H., Wang, Y., Zhao, Q., Liu, H., Cao, W., Xie, X., Zhang, D., Wang, Y., & Jian, Y. (2020). The development of Coronavirus 3C-Like protease (3CLpro) inhibitors from 2010 to 2020. *European Journal of Medicinal Chemistry*, *206*, 112711. <https://doi.org/10.1016/j.ejmech.2020.112711>
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., ... Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *The Lancet*, *395*(10224), 565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
- Luan, B., & Huynh, T. (2020). *Crystal-structures-guided design of fragment-based drugs for inhibiting the main protease of SARS-CoV-2*.
- Lurie, N., Saville, M., Hatchett, R., & Halton, J. (2020). Developing Covid-19 vaccines at pandemic speed. *New England Journal of Medicine*, *382*(21), 1969–1973. <https://doi.org/10.1056/NEJMp2005630>
- Miguet, L., Zhang, Z., & Grigorov, M. G. (2006). Computational studies of ligand-receptor interactions in bitter taste receptors. *Journal of Receptors and Signal Transduction*, *26*(5–6), 611–630. <https://doi.org/10.1080/10799890600928210>
- Mirza, M. U., Ahmad, S., Abdullah, I., & Froeyena, M. (2020). Identification of novel human USP2 inhibitor and its putative role in treatment of COVID-19 by inhibiting SARS-CoV-2 papain-like (PLpro) protease. *Computational Biology and Chemistry*, *89*, 107376.
- Mirza, M. U., & Froeyen, M. (2020). Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. *Journal of Pharmaceutical Analysis*, *10*(4), 320–328.
- Mirza, M. U., & Ikram, N. (2016). Integrated computational approach for virtual hit identification against Ebola viral proteins VP35 and VP40. *International Journal of Molecular Sciences*, *17*(11), 1748. <https://doi.org/10.3390/ijms17111748>
- Mirza, M. U., Rafique, S., Ali, A., Munir, M., Ikram, N., Manan, A., Salo-Ahen, O. M., & Idrees, M. (2016). Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Scientific Reports*, *6*, 37313.
- Mirza, M. U., Vanmeert, M., Ali, A., Iman, K., Froeyen, M., & Idrees, M. (2019). Perspectives towards antiviral drug discovery against Ebola virus. *Journal of Medical Virology*, *91*(12), 2029–2048. <https://doi.org/10.1002/jmv.25357>
- Pydi, S. P., Jaggupilli, A., Nelson, K. M., Abrams, S. R., Bhullar, R. P., Loewen, M. C., & Chelikani, P. (2015). Abscisic acid acts as a blocker of the bitter taste G protein-coupled receptor T2R4. *Biochemistry*, *54*(16), 2622–2631. <https://doi.org/10.1021/acs.biochem.5b00265>
- Ramajayam, R., Tan, K.-P., & Liang, P.-H. (2011). Recent development of 3C and 3CL protease inhibitors for anti-coronavirus and anti-picornavirus drug discovery. *Biochemical Society Transactions*, *39*(5), 1371–1375. <https://doi.org/10.1042/BST0391371>
- Rehman, H. M., Mirza, M. U., Ahmad, M. A., Saleem, M., Froeyen, M., Ahmad, S., Gul, R., Alghamdi, H. A., Aslam, M. S., Sajjad, M., & Bhinder, M. A. (2020). Bhinder a putative prophylactic solution for COVID-19: Development of novel multiepitope vaccine candidate against SARS-COV-2 by comprehensive immunoinformatic and molecular modelling approach. *Biology*, *9*(9), 296.
- Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J. M., Lu, C., Dahlgren, M. K., Mondal, S., Chen, W., Wang, L., Abel, R., Friesner, R. A., & Harder, E. D. (2019). OPLS3e: Extending force field coverage for drug-like small molecules. *Journal of Chemical Theory and Computation*, *15*(3), 1863–1874. <https://doi.org/10.1021/acs.jctc.8b01026>
- Rudling, A., Gustafsson, R., Almlöf, I., Homan, E., Scobie, M., Warpman Berglund, U., Helleday, T., Stenmark, P., & Carlsson, J. (2017). Fragment-based discovery and optimization of enzyme inhibitors by docking of commercial chemical space. *Journal of Medicinal Chemistry*, *60*(19), 8160–8169. <https://doi.org/10.1021/acs.jmedchem.7b01006>
- Salo-Ahen, O. M. H., Alanko, I., Bhadane, R., Bonvin, A. M. J. J., Honorato, R. V., Hossain, S., Juffer, A. H., Kabedev, A., Lahtela-Kakkonen, M., Larsen, A. S., Lescrinier, E., Marimuthu, P., Mirza, M. U., Mustafa, G., Nunes-Alves, A., Pansar, T., Saadabadi, A., Singaravelu, K., & Vanmeert, M. (2021). Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes*, *9*(1), 71. <https://doi.org/10.3390/pr9010071>
- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, *27*(3), 221–234. <https://doi.org/10.1007/s10822-013-9644-8>
- Shah, A. U. M., Safri, S. N. A., Thevadas, R., Noordin, N. K., Rahman, A. A., Sekawi, Z., Ideris, A., & Sultan, M. T. H. (2020). COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government. *International Journal of Infectious Diseases*, *97*, 108–116. <https://doi.org/10.1016/j.ijid.2020.05.093>
- Singh, N., Pydi, S. P., Upadhyaya, J., & Chelikani, P. (2011). Structural basis of activation of bitter taste receptor T2R1 and comparison with Class A G-protein-coupled receptors (GPCRs). *Journal of Biological Chemistry*, *286*(41), 36032–36041. <https://doi.org/10.1074/jbc.M111.246983>
- Tan, J., Abrol, R., Trzaskowski, B., & Goddard, W. A. (2012). 3D structure prediction of TAS2R38 bitter receptors bound to agonists phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP). *Journal of Chemical Information and Modeling*, *52*(7), 1875–1885. <https://doi.org/10.1021/ci300133a>
- Xia, B., & Kang, X. (2011). Activation and maturation of SARS-CoV main protease. *Protein & Cell*, *2*(4), 282–290. <https://doi.org/10.1007/s13238-011-1034-1>
- Yang, H., Bartlam, M., & Rao, Z. (2006). Drug design targeting the main protease, the Achilles' heel of coronaviruses. *Current Pharmaceutical Design*, *12*(35), 4573–4590.
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R., & Rao, Z. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor.

- Proceedings of the National Academy of Sciences*, 100(23), 13190–13195. <https://doi.org/10.1073/pnas.1835675100>
- Zhang, C., Huang, S., Zheng, F., & Dai, Y. (2020). Controversial treatments: An updated understanding of the Coronavirus Disease 2019. *Journal of Medical Virology*, 92(9), 1441–1448. <https://doi.org/10.1002/jmv.25788>
- Zhang, J., Zeng, H., Gu, J., Li, H., Zheng, L., & Zou, Q. (2020). Progress and prospects on vaccine development against SARS-CoV-2. *Vaccines*, 8(2), 153. <https://doi.org/10.3390/vaccines8020153>

**How to cite this article:** Ahmad, S., Usman Mirza, M., Yean Kee, L., Nazir, M., Abdul Rahman, N., Trant, J. F., & Abdullah, I. (2021). Fragment-based *in silico* design of SARS-CoV-2 main protease inhibitors. *Chemical Biology & Drug Design*, 98, 604–619. <https://doi.org/10.1111/cbdd.13914>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.