Computational study on peptidomimetic inhibitors against SARS-CoV-2 main protease

Tuanjai Somboon [a], Panupong Mahalapbutr [b], Kamonpan Sanachai [a], Phornphimon Maitarad [c], Vannajan Sanghiran Lee [d], Supot Hannongbua [e], Thanyada Rungrotmongkol [a,f,*]

[a] Structural and Computational Biology Research Unit, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
[b] Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand
[c] Research Center of Nano Science and Technology, Shanghai University, Shanghai 200444, PR China
[d] Department of Chemistry, Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia
[e] Center of Excellence in Computational Chemistry (CECC), Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
[f] Program in Bioinformatics and Computational Biology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

ARTICLE INFO

ABSTRACT

The emergence outbreak caused by a novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has received significant attention on the global risks. Due to itscrucial role in viral replication, the main protease 3CL^{pro} is an important target for drug discovery and development to combat COVID-19. In this work, the structural and dynamic behaviors as well as binding efficiency of the four peptidomimetic inhibitors (N3, 11a, 13b, and 14b) recently co-crystalized with SARS-CoV-2 3CL^{pro} were studied and compared using all-atom molecular dynamics (MD) simulations and solvated interaction energy-based binding free energy calculations. The per-residue decomposition free energy results suggested that the key residues involved in inhibitors binding were H41, M49, L141–C145, H163–E166, P168, and Q189–T190 in the domains I and II. The van der Waals interaction yielded the main energy contribution stabilizing all the focused inhibitors. Besides, their hydrogen bond formations with F140, G143, C145, H164, E166, and Q189 residues in the substrate-binding pocket were also essential for strengthening the molecular complexation. The predicted binding affinity of the four peptidomimetic inhibitors agreed with the reported experimental data, and the 13b showed the most efficient binding to SARS-CoV-2 3CL^{pro}. From rational drug design strategies based on 13b, the polar moieties (e.g., benzamide) and the bulky N-terminal protecting groups (e.g., thiazole) should be introduced to P1' and P4 sites in order to enhance H-bonds and hydrophobic interactions, respectively. We hope that the obtained structural and energetic information could be beneficial for developing novel SARS-CoV-2 3CL^{pro} inhibitors with higher inhibitory potency to combat COVID-19.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The novel coronavirus (SARS-CoV-2) was first identified as the source of the illness with the infection trace in a seafood market in Wuhan city, Hubei province, China [1,2]. Nevertheless, infectious diseases of related coronaviruses at pandemic level occurred in 2002 and 2012 for SARS-CoV [3–5] and Middle East Respiratory Syndrome (MERS-CoV) [6,7], respectively. Coronaviruses are a large family of viruses that cause symptoms of the common cold and a serious respiratory illness that can lead to mortality [8]. More importantly, this current contagious disease is spreading significantly and there is still no vaccine or antiviral drug available.

SARS-CoVs are single stranded positive RNA viruses belonging to the *Coronaviridae* family. They consist of a 5′-untranslated region (UTR), a replicate complex (ORF1ab) encoding non-structural proteins (NSPs), a spike protein (S), an envelope protein (E), a membrane protein (M), a nucleocapsid protein (N), 3′-UTR, and several unidentified non-structural open reading frames [9–11]. In NSPs of ORF1ab, SARS-CoVs contain two proteases that cleave different sites of the replicase polyproteins. The 3 chymotrypsin-like cysteine protease (3CL^{pro}), also called main protease (M^{pro}) cleaves 11 sites, whereas the papain-like protease (PL^{pro}) cleaves three sites in the polyproteins [12,13]. Consequently, SARS-CoV 3CL^{pro} plays a key role in proteolytic processing of the replicate polyproteins and has been recognized as a potential target for anti-viral drugs.

The sequence identity between SARS-CoV 3CL^{pro} and SARS-CoV-2 3CL^{pro} are 96%, and the amino acid residues involved in the catalytic and substrate-binding sites are 100% conserved [13,14]. Catalytic dyad is composed of the residues H41 and C145. Besides, Q189 residue commonly found in the main protease of both viruses involves in hydrogen bond formations with inhibitors [14–16]. The 3CL^{pro} requires dimerization for its proteolytic activity [17]. The catalytic/substrate binding cleft of 3CL^{pro} is located between the antiparallel β-barrels of domains I and II (Fig. 1) at the P1 and P2 positions [18], while the domain III is notably required for enzymatic reaction [19].

Broad-spectrum inhibitors targeting SARS-CoV 3CL^{pro} have been designed and evaluated [20–22]. A recent discovery has shown that N3 is a potent Michael acceptor inhibitor to SARS-CoV-2 3CL^{pro} (EC$_{50}$ of 16.77 μM in SARS-CoV-2-infected Vero cells and $k_{obs}/[I]$ of 11,300 $\pm$ 880 M^{-1} s^{-1} for SARS-CoV-2 3CL^{pro}) [23]. Meanwhile, L. Zhang et al. have developed a potent α-ketoamide (13b) inhibitor with the IC$_{50}$ value of 0.67 $\pm$ 0.18 μM for anti-SARS-CoV-2 3CL^{pro} and it exhibits a higher plasma half-life (~3-fold) and a lower rapid clearance than its parent lead compound [24]. A loss of the Boc group in compound 14b weakens the inhibitory potency and becomes almost inactive. Lately, the crystal structure of SARS-CoV-2 3CL^{pro} in complex with the peptidomimetic aldehyde inhibitor 11a has been determined, indicating the achievement of a covalent bond formation between the aldehyde group and the catalytic residue C145 [25]. This compound shows excellent inhibitory potency with the IC$_{50}$ value of 0.053 $\pm$ 0.005 μM and also possesses good pharmacokinetic properties with low toxicity. Accordingly, these different types of peptidomimetic inhibitors are promising for anti-viral drug developments against SARS-CoV-2. In this work, the molecular complexation between these four recently reported inhibitors (N3, 11a, 13b, and 14b; Fig. 2) and SARS-CoV-2 3CL^{pro} was elucidated and compared by means of molecular dynamics (MD) simulations and binding free energy calculations.

2. Computational details

2.1. Structural preparation

The co-crystal structures of SARS-CoV-2 3CL^{pro} with the three inhibitors N3 (6LU7), 11a (6LZE), and 13b (6Y2F) were downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank [26]. All inhibitors were covalently bonded with C145 residue

of SARS-CoV-2 3CL^{pro} [23–25]. This covalent bond was removed using Accelrys Discovery Studio 2.5 [27]. The warhead of N3, 11a, and 13b at P1' position was then restored to the α,β double bond, aldehyde group, and α-keto group, respectively, while the vacancy side chain of C145 in all structures was protonated. The two missing amino acid residues E47 and D48 in the 13b/3CL^{pro} complex were built using SWISS-MODEL server [28]. The 13b/3CL^{pro} complex was used to prepare the 14b system by eliminating the butyloxycarbonyl moiety at P4 site of 13b using the Accelrys Discovery Studio 2.5.

According to the standard protocols [29–31], the electrostatic potential (ESP) charges of ligands were calculated with the HF/6-31G (d) level using the Gaussian 09 program [32]. The antechamber and parmchk modules in AMBER program version 16 [33] were respectively used to generate the restrained ESP charges and prepare the parameters of all inhibitors. The bonded and non-bonded parameters for the protein and ligand were treated with the AMBER ff14SB force field [34] and generalized AMBER force field version 2 (GAFF2) [35], respectively. The protonation states of all ionizable amino residues were assigned at pH 7.4 using PDB2PQR server version 2.0.0 [36]. Hydrogen atoms were added using the LEaP module. The systems were firstly minimized to remove any steric hindrance or inappropriate geometry. Prepared structures were placed in a 10 Å truncated box filled with TIP3P water molecules [37]. Sodium ions were randomly added to neutralize the system. The added hydrogen atoms and water molecules were minimized using 500 steps of steepest descent (SD) followed by 1500 steps of conjugated gradient (CG) methods before running the MD simulations, while the remaining parts of the system were held fixed. The protein and ligand were further minimized by SD (500 iterations) and CG (1500 iterations) methods with constrained solvent molecules. Finally, the whole complex was fully minimized using the same procedure.

2.2. Molecular dynamics simulations

All molecular dynamics (MD) simulations were carried out using AMBER16 [33]. The prepared inhibitors/3CL^{pro} complexes were heated from 10 to 310 K for 200 ps with harmonic positional restraint of 30.0 kcal/mol Å^2 to Cα atoms of protein. Each complex was then subjected to four steps of restrained MD simulations at the same temperature with harmonic restraints of 30, 20, 10, and 5 kcal/mol Å^2 for 1300 ps in total and another 200 ps was performed without any restraint.
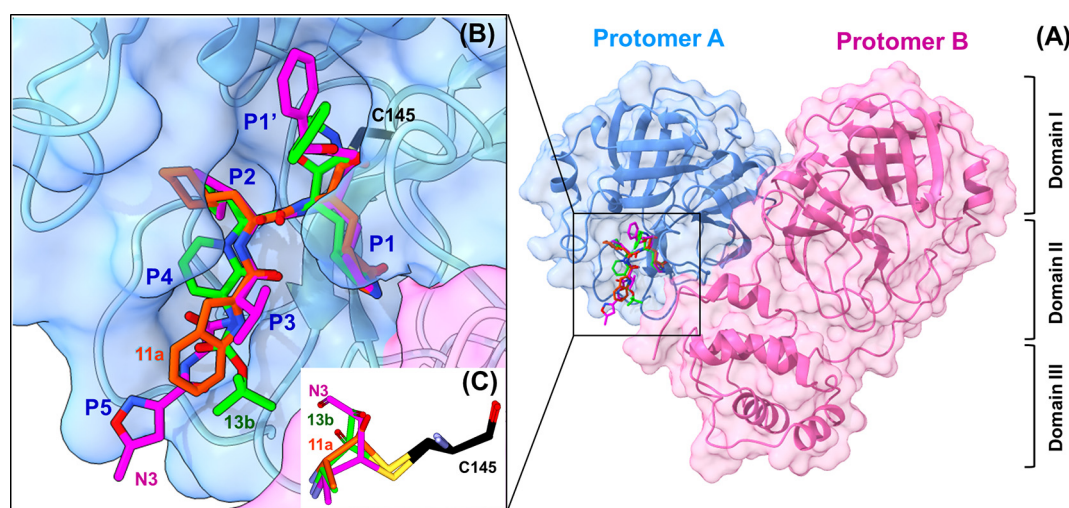
Fig. 1. Superimposition of SARS-CoV-2 3CL^{pro} co-crystal structures. (A) Protomer A (blue) in complex with N3 (magenta stick, 6 LU7.PDB), 11a (red orange stick, 6LZE.PDB), and 13b (green stick, 6Y2F.PDB) and protomer B (pink) in unbound form. Domains I, II, and III are demonstrated. (B) A close-up view of the binding orientation of N3, 11a, and 13b inhibitors in the substrate-binding cleft. (C) The covalent bond (C–S) between peptidomimetic inhibitors and C145.

Fig. 2. 2D structure of the four studied compounds (Michael acceptor inhibitor N3, aldehyde inhibitor 11a, and two α-ketoamide inhibitors 13b and 14b). The canonical binding fragments are identif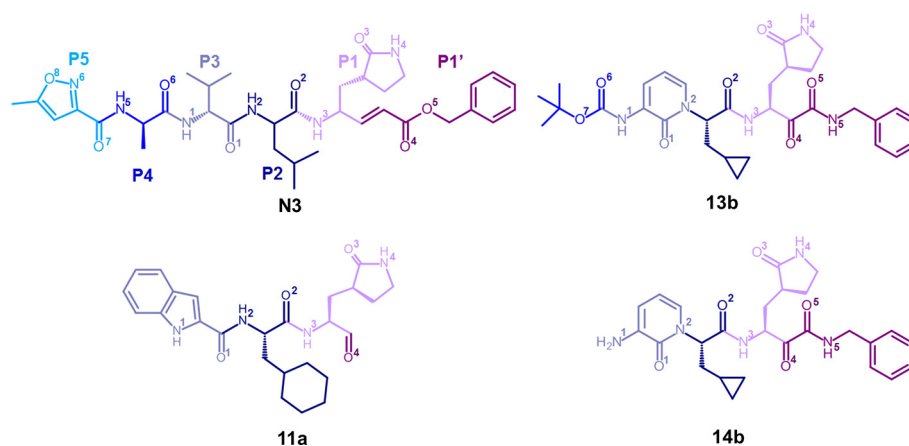ied demonstrating binding pocket moieties (P5, P4, P3, P2, P1 and P1'). Atomic labels of each inhibitor are clarified for further discussion.

Finally, the entire system was simulated under the *NPT* ensemble (310 K, 1 atm) until reaching 100 ns. Non-bonded interactions were set to 10-Å cutoff distance, whereas the particle mesh Ewald summation method [38] was employed to treat the electrostatic interactions. The SHAKE algorithm [39] was used to constrain all covalent bonds involving hydrogen atoms. A 2-fs simulation time step was used throughout the MD simulations. Temperature and pressure were controlled by the Langevin thermostat [40] with a collision frequency of 2 ps^{-1} and the

Berendsen barostat [41] with a pressure-relaxation time of 1 ps, respectively. The MD trajectories were saved every 10 ps. The results from MD simulations were analyzed using the same criteria as previously described [14]. The CPPTRAJ [42] was used to calculate the structural dynamics properties and intermolecular hydrogen bonding of the inhibitors/3CL^{pro} complexes. The MM/GBSA per-residue decomposition free energy ($\Delta G_{bind}^{residue}$) calculations [43] were carried out on the 100 snapshots extracted from the last 20-ns MD simulations using the
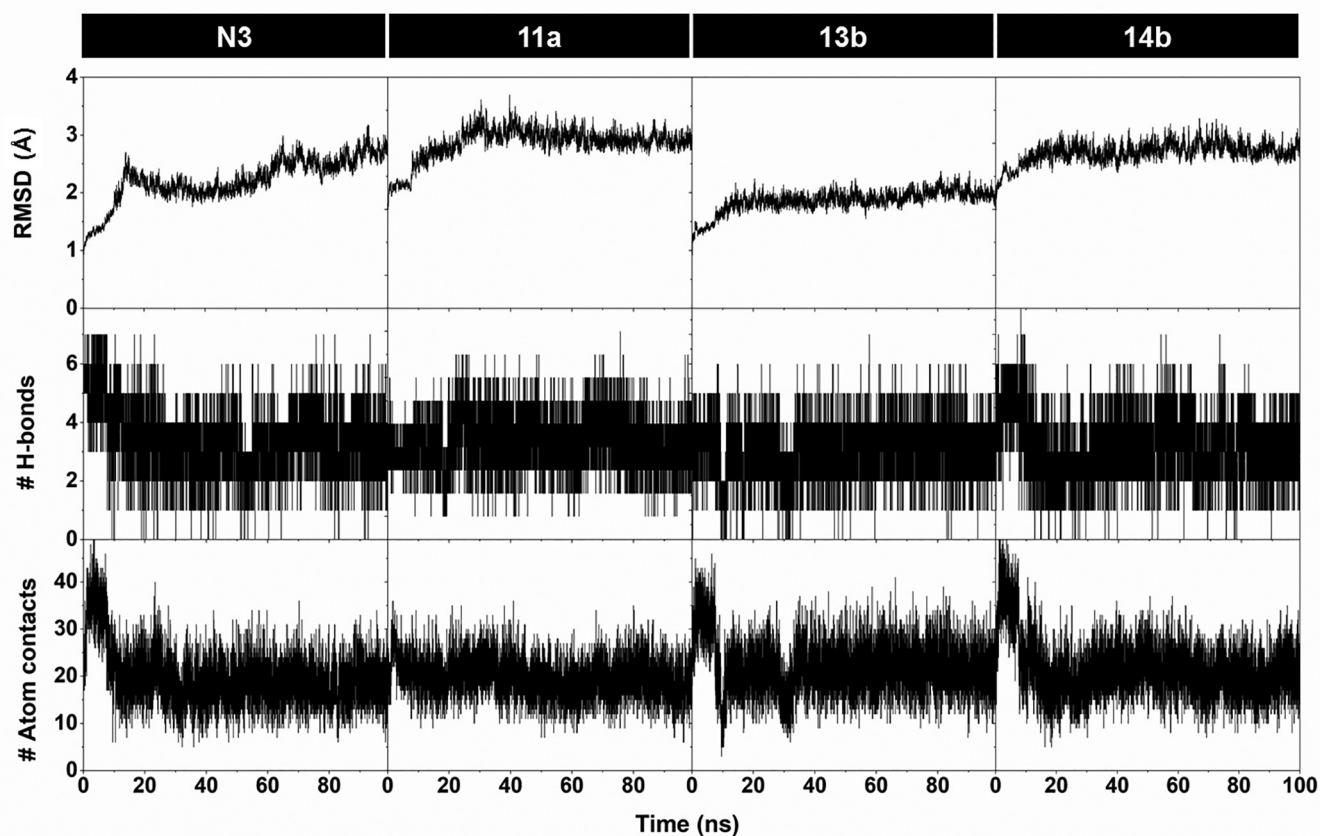

Fig. 3. All-atom RMSD, # H-bonds, and # atom contacts of N3, 11a, 13b, and 14b in complex with the SARS-CoV-2 3CL^{pro} plotted along the 100-ns MD simulations.

Table 1

Energy components (kcal/mol) of focused inhibitors/3CL^{pro} complexes calculated with solvated interaction energy (SIE) method.

Energy component	N3	11a	13b	14b
SIE-based binding free energy				
E_{vdW}	-60.37 ± 0.34	-57.63 ± 0.55	-62.64 ± 0.36	-56.18 ± 0.28
E_{ele}	-14.33 ± 0.30	-13.95 ± 0.24	-17.85 ± 0.28	-15.04 ± 0.31
ΔG_{RF}	18.42 ± 0.21	15.90 ± 0.18	19.14 ± 0.20	15.66 ± 0.18
ΔG_{cavity}	-10.81 ± 0.06	-9.10 ± 0.09	-9.85 ± 0.06	-8.93 ± 0.03
ΔG_{bind}	-9.92 ± 0.04	-9.68 ± 0.07	-10.35 ± 0.04	-9.64 ± 0.03
Experimental binding free energy				
ΔG_{Exp}^a	-6.83^* [21]	-9.92 [25]	-8.42 [24]	N/A

Data are shown as mean $\pm$ standard error of the mean (SEM). $^a\Delta G_{Exp}$ was converted from the reported IC$_{50}$ values using the Cheng–Prusoff equation of $\Delta G_{Exp} = RT \cdot \ln(IC_{50})$ [48]. $^*\Delta G_{Exp}$ of N3 was calculated from K_i.
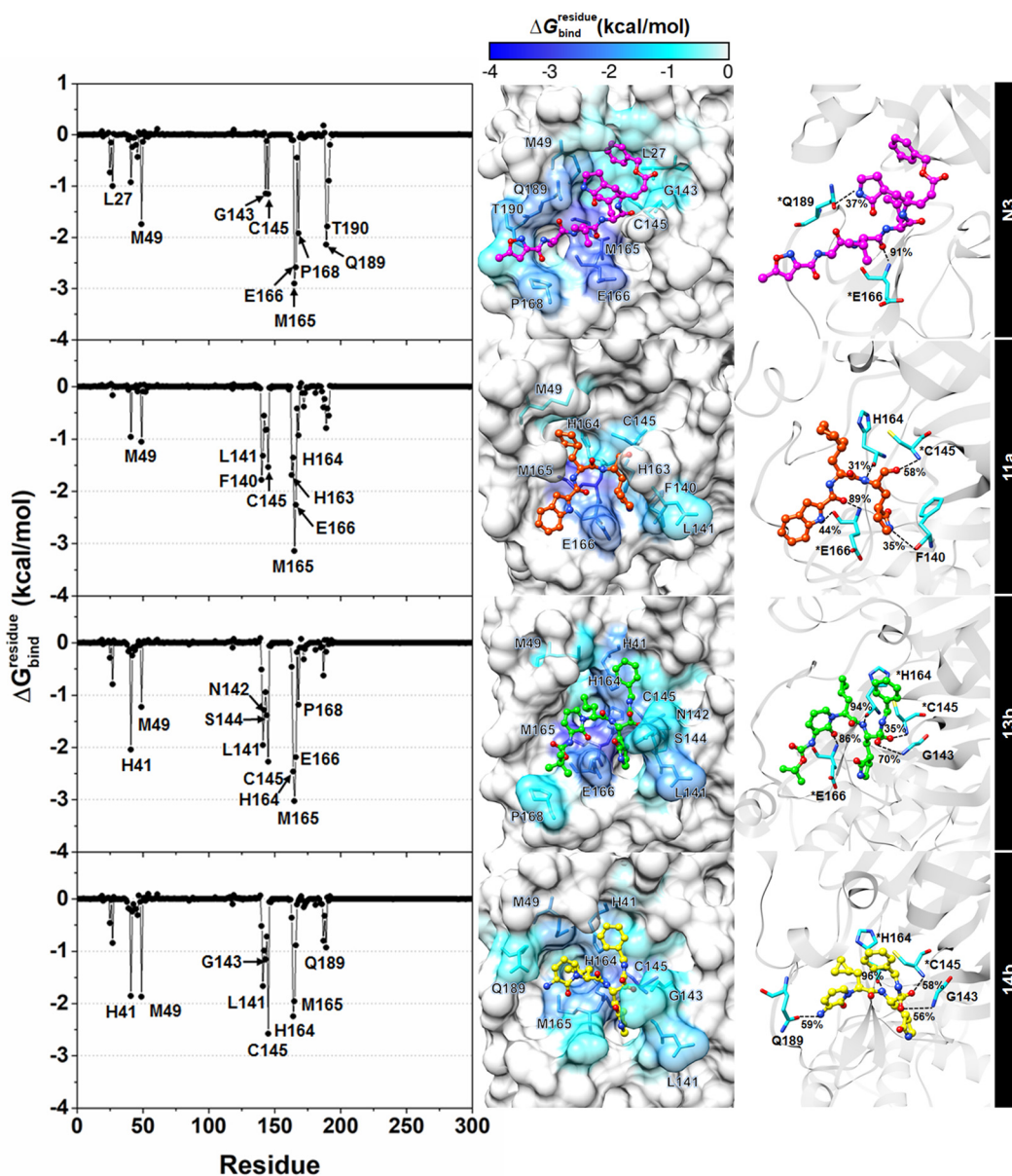
Fig. 4. $\Delta G_{bind}^{residue}$ of inhibitors/3CL^{pro} complexes (left). The contributing residues involved in ligand binding are colored according to their $\Delta G_{bind}^{residue}$ values, where the highest to lowest free energies is shaded from white to blue, respectively (middle). Binding pattern of inhibitors/3CL^{pro} complexes is drawn from the last MD snapshot (right). Black dashed lines represent H-bond formation. H-bonds observed in the crystal structures and remained after MD simulations are denoted as "*".

MMPBSA.py [44] implemented in AMBER16. The binding free energy of protein–ligand binding (ΔG_{bind}) was estimated from the solvated interaction energy (SIE) approach [45] as the summation of the van der Waals (E_{vdW}), electrostatic (E_{ele}), reaction field (G_{RF}), cavity ($\gamma\Delta SA(\rho)$), and constant (C) as the following equation:

$$\Delta G_{bind}(\rho, D_{in}, \alpha, \gamma, C) = \alpha[E_{vdW} + E_{ele}(D_{in}) + \Delta G_{RF}(\rho, D_{in}) + \gamma\Delta SA(\rho)] + C$$

where D_{in} is the solute interior dielectric constant and coefficients used in the calculations are $\alpha = 0.105$, $\gamma = 0.013$, and $C = -2.89$. E_{vdW} and E_{ele} are the intermolecular van der Waals and Coulomb interaction energies in the bound state, respectively. ΔG_{RF} is the electrostatic contribution of the solvation free energy to binding, and ΔG_{cavity} ($\gamma\Delta SA$) is the nonpolar contribution of the solvation free energy to binding.

3. Results and discussion

3.1. System stability

The system stability of all four inhibitors/SARS-CoV-2 3CL^{pro} complexes was determined using all-atom RMSD upon the simulation time. Fig. 3 shows that the RMSD values of all systems are increased in the first 40 ns and then maintained at the fluctuation of ~2.0–3.0 Å until the end of simulation. Thus, the number of intermolecular H-bonds (# H-bonds) and the number of atom contacts (# atom contacts) of inhibitor plotted along the simulation were also taken into account in a consideration of equilibrium phase. It can be seen that all systems reach equilibrium at 40 ns; however, in this work the MD trajectories from 80 to 100 ns were extracted for further analysis. Over the last 20 ns, the # H-bonds of N3, 11a, 13b, and 14b are 3.33 ± 1.04, 2.97 ± 1.11, 3.15 ± 0.93, and 3.13 ± 0.98, respectively. The # atom contacts of these inhibitors are observed in the range of ~18–21, and among them, the α-ketoamide inhibitor 13b shows the highest # atom contacts. This result suggests that 13b could interact more favorable to the SARS-CoV-2 3CL^{pro} than the remaining inhibitors (discussed in more detail later).

3.2. Predicted inhibitory efficiency

The binding affinity of all studied inhibitors/3CL^{pro} systems was estimated using solvated interaction energy (SIE) method on 100 snapshots extracted from the last 20 ns of simulations. The obtained results are listed in Table 1, whereas the experimental binding free energies (ΔG_{Exp}) were converted from the reported IC$_{50}$ for 11a [25] and 13b [24] and K_i for N3 [21]. The molecular mechanics energy (ΔE_{MM}) shows that van der Waals (vdW) interaction plays a vital role in molecular complexation with SARS-CoV-2 3CL^{pro} (ΔE_{vdW} of -60.37 ± 0.34, -57.63 ± 0.55, -62.64 ± 0.36, and -56.18 ± 0.28 kcal/mol for N3, 11a, 13b, and 14b, respectively), whereas the electrostatic attractions are observed in the range of ~-14 to -18 kcal/mol. It is noticed that vdW interaction is ~3.5–4.2-fold stronger than the electrostatic attraction, in line with several reported SARS-CoV-2 3CL^{pro} inhibitors, including ritonavir and lopinavir [14], Cmp-17 [46], and α-ketoamide 13b [47]. From SIE-based ΔG_{bind} calculations, we found that, among all studied compounds, the α-ketoamide inhibitor 13b shows the highest binding affinity (ΔG_{bind} of -10.35 ± 0.04 kcal/mol), which is to some extent greater than α-ketoamide inhibitor 14b (-9.64 ± 0.03 kcal/mol) and the two remaining inhibitors (-9.92 ± 0.04, and -9.68 ± 0.07 kcal/mol for N3 and 11a, respectively). The predicted ΔG_{bind} of these inhibitors are somewhat in correspondence with the reported experimental data. However, it is worth noting that experimental values of IC$_{50}$ and K_i were obtained from different laboratories. Although all studied inhibitors are not covalently attached to SARS-CoV-2 3CL^{pro} but the results show that these inhibitors are

strongly bound in the complex form, which are comparable well with X-ray crystallographic data.

3.3. Key residues upon inhibitors binding

The key residues involved in the binding of ligands to the SARS-CoV-2 3CL^{pro} were investigated using $\Delta G_{bind}^{residue}$ calculation based on MM/GBSA method. The obtained results are shown in Fig. 4, whereas the binding orientation of each inhibitor is depicted in the middle panel, in which the hot-spot residues are colored according to their $\Delta G_{bind}^{residue}$ values. It should be noted that only residues that exhibit the energy stabilization of ≤ -1.0 kcal/mol at least in one system are labeled and discussed.

In consistent with the ΔG_{bind} results (Table 1), the potent α-ketoamide inhibitor 13b shows the highest hot-spot residues (10 residues, including H41, M49, L141, N142, S144, C145, H164, M165, E166, and P168), whereas there are nine (L27, M49, G143, C145, M165, E166, P168, Q189, and T190) and eight (M49, F140, L141, C145, H163, H164, M165, and E166) important residues found in the Michael acceptor inhibitor N3 and the aldehyde inhibitor 11a, respectively. Note that these three potent compounds have the same binding residues as follows: M49, C145, M165, and E166. Although the eight residues (H41, M49, L141, G143, C145, H164, M165, and Q189) are found to be contributed for binding of the inactive inhibitor 14b, the interactions with the residues E166 and P168 positioned closed to the P4 site are dramatically reduced due to eradication of the Boc group.

Since the canonical fragments at P4 and P5 positions were designed to favor the S4 hydrophobic site [21], N3 is able to occupy the binding pocket in this region better than the other inhibitors and shows high affinity with Q189 ($\Delta G_{bind}^{residue}$ of -2.15 kcal/mol) and T190 ($\Delta G_{bind}^{residue}$ of -1.79 kcal/mol) residues. Notably, the hydrophobic contribution from M165 is found to be essential in all the studied inhibitors with the $\Delta G_{bind}^{residue}$ of -2.9, -3.1, -3.0, and -2.0 kcal/mol for N3, 11a, 13b, and
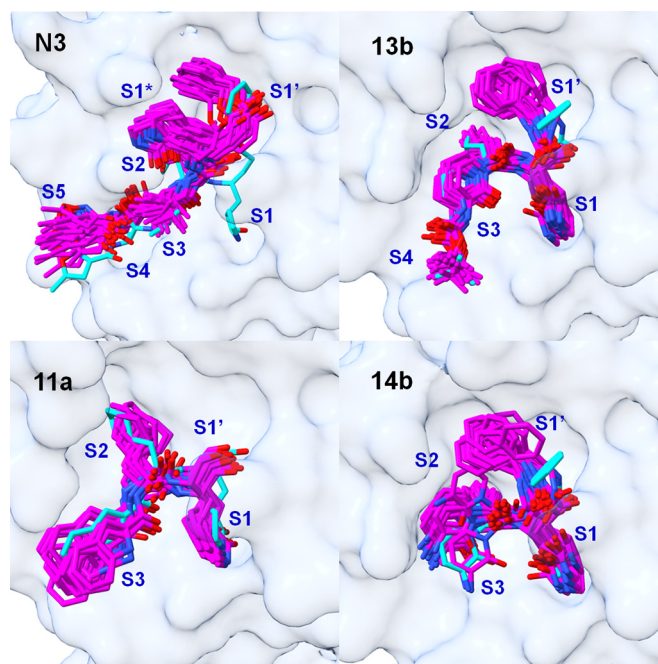

Fig. 5. Superimposition of N3, 11a, 13b, and 14b inhibitors at the substrate-binding pocket of SARS-CoV-2 3CL^{pro} derived from the last 20-ns MD simulations (magenta stick representation) compared with X-ray crystal structures (cyan stick representation). The canonical sites are labeled in blue text.

14b, respectively. This finding is in accordance with previous studies demonstrating that M49 and M165 stabilized inhibitors through hydrophobic interactions [14,49]. The G143 and C145 residues interacted with 13b ($\Delta G_{bind}^{residue}$ of -0.95 and -2.27 kcal/mol, respectively) and 14b ($\Delta G_{bind}^{residue}$ of -1.16 and -2.58 kcal/mol, respectively) better than N3 and 11a inhibitors, since this oxyanion hole stabilization is commonly found in α-ketoamide inhibitors [22,24]. The high contribution from the residue P168 was detected in the inhibitors containing the P4 site systems (of -1.9 and -1.2 kcal/mol for N3 and 13b), whereas this contribution is dramatically decreased in 11a and 14b. This result could be the reason why removing the Boc group (14b) weakens its inhibitory efficiency [24].

Superimposition over the last 20 ns of MD simulations (Fig. 5) shows that the active inhibitors N3, 11a, and 13b are well-aligned in the substrate-binding pocket, and among them, 13b is the most stable one. It should be noted that the S1 site is dramatically shifted to S1* for N3 and moves the benzyl ester portion up together (S1' site) due to lacking of H-bond formations at the catalytic center residues. Additionally, 14b also has large fluctuations at S1' and S3 sites, in good agreement with the reported low inhibitory activity toward SARS-CoV-2 CL^{pro} [24].

The electrostatic ($\Delta E_{ele} + \Delta G_{polar}$) and vdW ($\Delta E_{vdW} + \Delta G_{nonpolar}$) energy contributions of the residues important for inhibitor binding are depicted in Fig. 6. We found that the main contribution for stabilizing the inhibitors is vdW interactions ($\Delta E_{vdW} + \Delta G_{nonpolar}$) up to -4.7 kcal/mol as seen by negative values (green grids), whereas there are rather low electrostatic attractions observed (~ -1.0 to 2.6 kcal/mol shaded by light green to red grids), correlating well with the ΔE_{MM} calculations based on SIE method (Table 1). The hydrophobic contributions from residues H41, M49, N142, C145, M165, and E166 are found to stabilize all the studied inhibitors. For contribution from H41, one of the catalytic dyad, the canonical moiety at P2 of all inhibitors stacks with the imidazole ring of H41 at the S2 site (Fig. 4). Interestingly, this catalytic residue shows high vdW contribution (< -2.4 kcal/mol) to the α-ketoamides (13b and 14b) over the remaining types of inhibitors, suggesting that cyclopropyl ring is preferentially occupied in this site, which commonly presents in preclinical/clinical drugs such as efavirenz, abacavir, and roflumilast [50]. It should be noted that the highest vdW contributions from the M165 (-3.11 kcal/mol) and E166 (-4.74 kcal/mol) residues are detected in the potent α-ketoamide inhibitor 13b, as previously reported [47]. However, 13b lacks the P5 site, resulting in a lower vdW contribution with P168 and Q189 residues

	N3 Elec	N3 vdW	11a Elec	11a vdW	13b Elec	13b vdW	14b Elec	14b vdW
T25	0.50	-1.23	0.06	-0.03	0.26	-0.55	0.21	-0.67
L27	0.12	-1.12	0.06	-0.22	0.14	-0.93	0.11	-0.95
H41	0.89	-1.82	0.78	-1.74	0.65	-2.69	0.56	-2.41
S46	1.30	-1.73	0.26	-0.35	0.18	-0.21	0.44	-0.75
M49	0.45	-2.20	0.17	-1.22	0.31	-1.54	0.35	-2.22
F140	0.05	-0.05	-0.75	-1.02	0.45	-0.97	0.44	-0.96
L141	0.10	-0.11	0.25	-1.57	-0.46	-1.49	-0.30	-1.36
N142	1.75	-1.71	1.44	-1.99	0.66	-1.94	0.91	-1.90
G143	0.04	-1.19	-0.20	-0.63	0.16	-1.10	0.07	-1.22
S144	0.14	-0.26	0.26	-1.08	-0.25	-1.14	0.39	-1.11
C145	0.09	-1.25	-0.60	-0.94	-0.85	-1.42	-1.09	-1.49
H164	0.60	-0.71	0.50	-1.86	-0.80	-1.66	-0.59	-1.66
M165	-0.14	-2.76	-0.24	-2.90	0.08	-3.11	0.76	-2.72
E166	-0.72	-1.86	1.41	-3.67	2.55	-4.74	1.36	-2.25
L167	0.71	-1.16	0.20	-0.62	0.77	-0.95	0.04	-0.14
P168	0.12	-2.05	0.01	-0.94	0.13	-1.31	0.00	-0.04
R188	0.52	-0.49	1.04	-1.28	0.51	-0.46	0.44	-0.77
Q189	1.34	-3.49	1.58	-2.37	0.51	-0.68	0.05	-0.98
T190	-0.30	-1.49	0.37	-0.78	0.21	-0.16	0.19	-0.16
A191	0.33	-1.22	0.10	-0.66	0.11	-0.14	0.11	-0.14

Energy contribution (kcal/mol)

-5 -4 -3 -2 -1 0 1 2 3 4 5

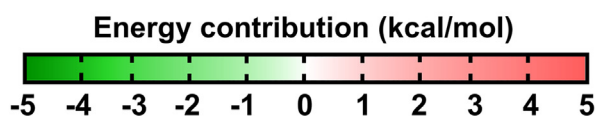**Fig. 6.** Electrostatic ($\Delta E_{ele} + \Delta G_{polar}$, red) and vdW ($\Delta E_{vdW} + \Delta G_{nonpolar}$, green) energy contributions from individual residues in the binding pocket of SARS-CoV-2 3CL^{pro} toward the inhibitors N3, 11a, 13b, and 14b.
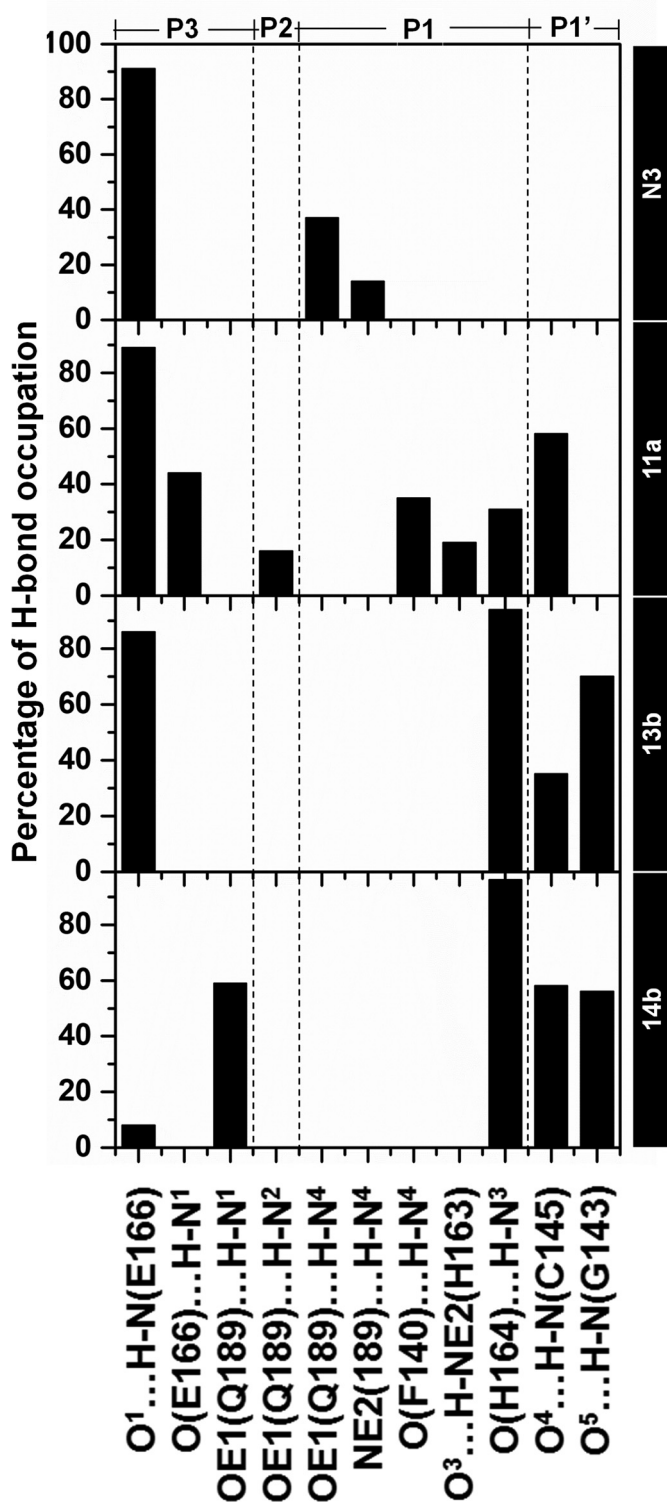
Fig. 7. Percentage of H-bond occupation of the SARS-CoV-2 3CL^{pro} residues contributed to N3, 11a, 13b, and 14b binding.

(-1.31 and -0.68 kcal/mol) compared to that of the Michael acceptor inhibitor N3 (-2.05 and -3.49 kcal/mol).

3.4. Protein-drug hydrogen bonding

In biological system, H-bonds play important role in many aspects, e.g., helping a complexation between ligand and protein. The percentages of the intermolecular H-bonds formed between the inhibitors and SARS-CoV-2 3CL^{pro} residues are plotted in Fig. 7, while the representative structures taken from the last snapshot of MD simulations are depicted in Fig. 4 (right panel). As expected, there are not many strong hydrogen bonds (> 80%) stabilizing the inhibitor binding, i.e., the E166 backbone nitrogen only interacts with the O^1 atom of all three potent inhibitors at P3 site, while the N^3 atom of the two α-ketoamides is highly stabilized by the H164 backbone oxygen.

We found that there are three H-bonds formations detected in the Michael acceptor inhibitor N3 at P1 and P3 sites, including OE1 (Q189)···N-N^4 (P1) at 37.3%, NE2 (Q189)···N-N^4 (P1) at 13.7%, and O^1 (P3)···H-N (E166) at 90.7%, which are similarly found in the initial structure from X-ray crystallographic data [23]. In the case of aldehyde inhibitor 11a, it mainly formed H-bonds with O1 (P3)···H-N (E166), O4 (P1')···H-N (C145), and O (E166)···H-N1 (P3) at 88.7%, 58.4%, and 44.1%, respectively. For the α-ketoamide inhibitors, 13b and 14b share H-bond formation pattern as follows: O (H164)···H-N^3 (P1) at 93.8% and 95.8%, O^4 (P1')···H-N (C145) at 35.3% and 57.7%, and O^5 (P1')···H-N (G143) at 70.1% and 55.8%. The stabilization of oxyanion hole from residues G143 and C145 via two H-bonds is commonly found in α-ketoamide inhibitors [22], whereas the warhead of the aldehyde [51] and Michael [52] inhibitors formed only one H-bond with the catalytic center of the target protease. It should be noted that the inactive 14b inhibitor has low level of H-bond formation (8.2%) with E166 compared to the other inhibitors.

3.5. Solvent accessibility of binding pocket

The ability of water accessibility at the active site of SARS-CoV-2 3CL^{pro} with and without inhibitor bound (protomers A and B in Fig. 8A) was estimated by means of solvent-accessible surface area (SASA) calculation. The obtained results are shown in Fig. 8B, while the SASA averaged from the last 20 ns of MD simulations is listed in Table 2. Note that the SASA calculations were performed on the residues located within 4 Å of the inhibitor. The obtained results showed that the SASAs for protomer B are 1289 ± 84, 1193 ± 76, 1087 ± 81, and 1033 ± 84 Å^2 for the N3, 11a, 13b, and 14b systems, respectively. Once the inhibitor bound to the active site of protomer A, the water accessibility of this protomer is diminished to 734 ± 60, 879 ± 71, 919 ± 63, and 880 ± 60 Å^2 for N3, 11a, 13b, and 14b, respectively. A reduction of water accessibility upon the binding of inhibitors to the enzyme active site is also found in the reported anti-HIV drugs, lopinavir and ritonavir, bound to the SARS-CoV-2 3CL^{pro} [14] as well as in other protein-ligand complexations [53–55]. Since N3 is the largest molecule among studied inhibitors, accommodating well at the S1' and S1-S5 sites, a lower water accessibility to the substrate-binding cleft is observed. Although the α-ketoamide inhibitor 13b shows the greatest binding affinity, the SASA value is not lower than the inactive compound 14b. This suggested that further structural modification on 13b should be made, e.g., enlarging the functional group at P4 site to better occupy the hydrophobic S4-S5 sites of SARS-CoV-2 3CL^{pro}.

3.6. Rational drug design

According to the aforementioned results, the 13b α-ketoamide inhibitor was selected as a template to propose newly designed protease inhibitors (Fig. 9) owing to the following reasons: (i) it has the highest binding efficiency to SARS-CoV-2 CL^{pro} (Table 1), (ii) it shows the highest stability in the binding pocket (Fig. 5), and (iii) it exhibits the highest vdW contribution with the residues M165 and E166 (Fig. 6). However, 13b has lower interactions with S4-S5 sites, and its molecular complexation is mainly driven by vdW forces. Additionally, the phenyl ring of 13b at S1' site is relatively unstable compared to the other moieties (Fig. 5).

To enhance the binding efficiency, especially electrostatic contributions, to SARS-CoV-2 CL^{pro}, the structural modifications of P1' and P4-
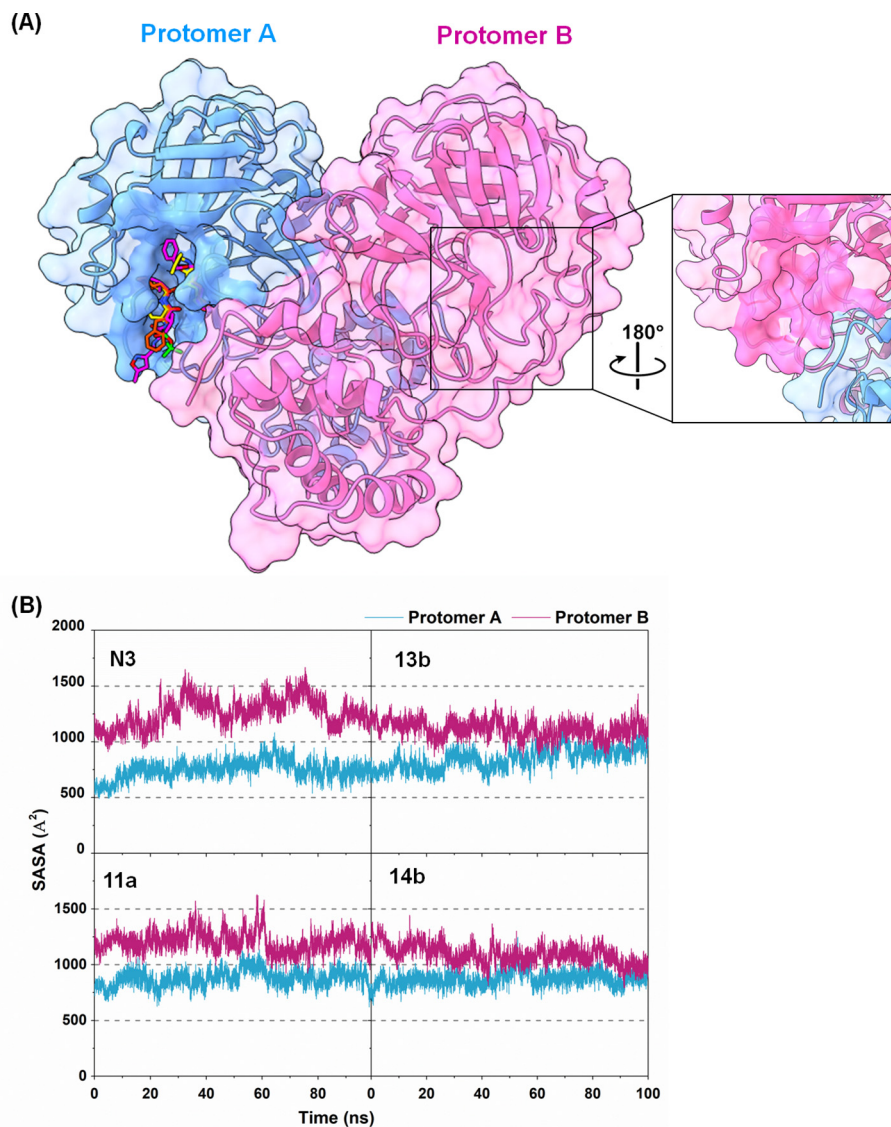
Fig. 8. (A) SARS-CoV-2 3CL^{pro} homodimer, in which (A) each inhibitor is bound at protomer A binding pocket (blue) and protomer B is without an inhibitor (red pink). The surface of residues within 4 Å from an inhibitor of each protomer is shown. (B) SASA plots along the simulations time of N3, 11a, 13b, and 14b systems.

P5 sites of 13b are suggested as follows: (i) the polar moieties (e.g. (1) benzamide and (2) benzoate fragments) should be added to the phenyl ring at P1' region to facilitate H-bond formations with the surrounding residues T25, H41, and C44 and (ii) the bulky N-terminal protecting groups (e.g., (1) thiazole, (2) isoxazole, and (3) MPC moieties) should be introduced to the P4 position to enhance hydrophobic interactions with P168 and T190 residues. However, the cyclopropyl ring at P2 position should be conserved as it strongly stacks with the catalytic H41 via hydrophobic interaction [50] (Fig. 6).

Six new designed 13b derivatives were proposed and taken to perform MD simulations for 100 ns. The binding affinity of them toward 3CL^{pro} was estimated using SIE method on 100 snapshots extracted from the last 20 ns of simulations. The obtained results are shown in Fig. 9B. There are four out of six compounds which show higher binding affinity than 13b (-10.35 ± 0.04 kcal/mol), including D1−1, D1−2, D2−2, and D2−3 with the ΔG_{bind} of -11.61 ± 0.08, -10.72 ± 0.06, -11.28 ± 0.06, and -10.73 ± 0.06 kcal/mol, respectively. Among these new design compounds, the modification of P1' to benzamide and P4 to thiazole (D1−1) shows the highest binding efficiency. The $\Delta G_{bind}^{residue}$ of D1−1 is improved for N142, S144, M165, E166, and P168 compared to 13b. The binding pattern of D1−1/3CL^{pro} complex is shown in Fig. 9E. In comparison with 13b, H-bond formation of O (H164)$\cdots$H-N^3 (P1) of D1−1 is mostly preserved (93.8%). Interestingly, the H-bond formation of benzamide fragment with C44 (H−N (P1')$\cdots$O (C44)) is magnicently enhanced (69% occupation) as proposed.

Table 2
Average SASA (Å^2) during the last 20 ns MD simulations of N3, 11a, 13b, and 14b inhibitors. Data are shown as mean ± standard deviation.

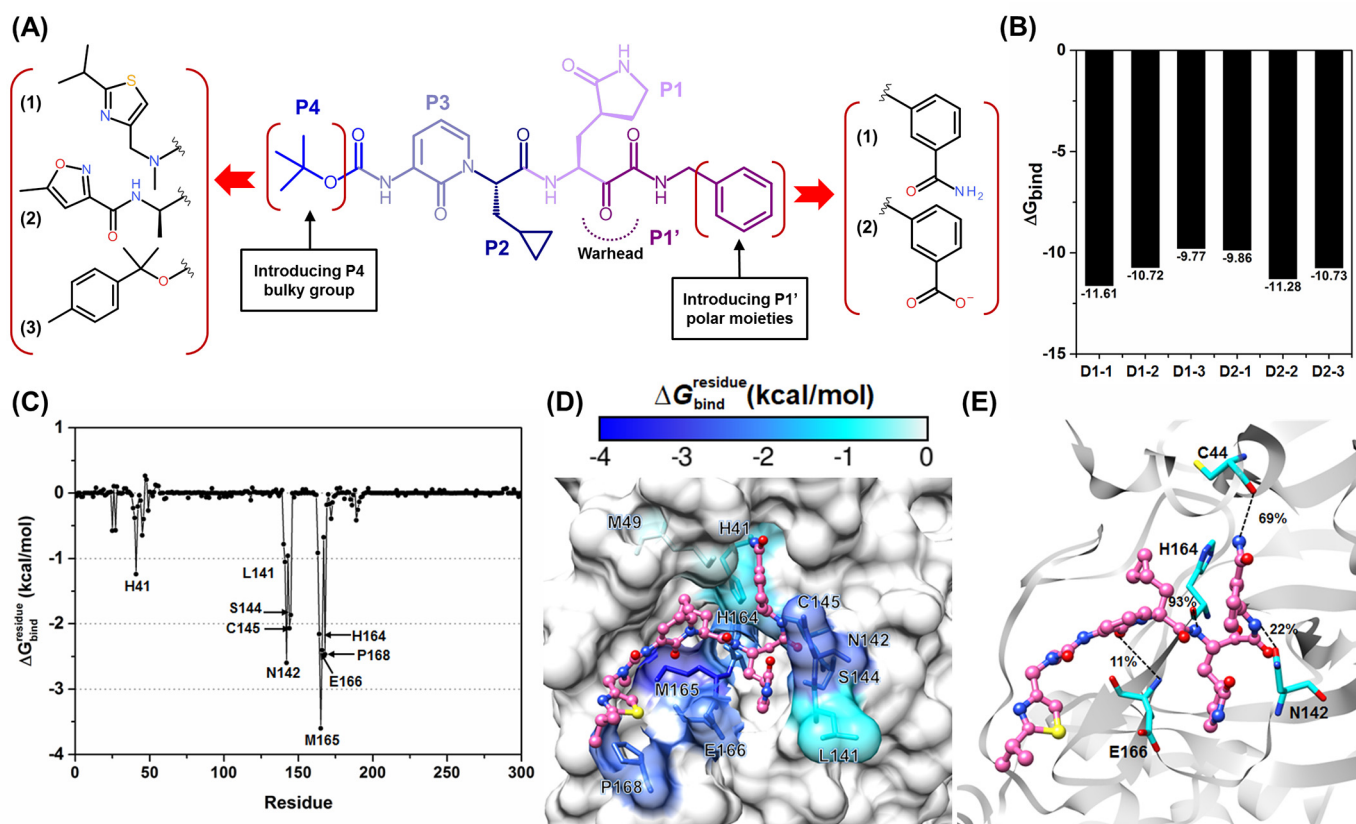| System	Average SASA (Å^2)	
	Protomer A	Protomer B
N3	734 ± 60	1289 ± 84
11a	879 ± 71	1193 ± 76
13b	919 ± 63	1087 ± 81
14b	880 ± 60	1033 ± 84

Fig. 9. Rational drug design of the SARS-CoV-2 3CL^{pro} inhibitors. (A) 2D structure of 13b with possible modified fragments. (B) Predicted ΔG_{bind} of modified 13b compounds from SIE method. (C) $\Delta G_{bind}^{residue}$ of D1–1/3CL^{pro} complex. (D) The contributing residues involved in D1–1 compound binding are colored according to their $\Delta G_{bind}^{residue}$ values, where the highest to lowest free energies is shaded from grey to blue, respectively. (E) Binding pattern of D1–1/3CL^{pro} complexes drawn from the last MD snapshot. Black dashed lines represent H-bond formation.

4. Conclusions

In the present study, the key binding and susceptibility of the reported peptidomimetic inhibitors (N3, 11a, 13, and 14b) against SARS-CoV-2 3CL^{pro} have been investigated using all-atom MD simulations. The $\Delta G_{bind}^{residue}$ calculations revealed that residues (i) H41, M49, L141, N142, S144, C145, H164, M165, E166, and P168, (ii) L27, M49, G143, C145, M165, E166, P168, Q189, and T190, and (iii) M49, F140, L141, C145, H163, H164, M165, and E166 are important for the binding of 13b, N3, and 11a, respectively. In the case of an inactive inhibitor 14b, residues H41, M49, L141, G143, C145, H164, M165, and Q189 are contributed. The molecular complexation between SARS-CoV-2 3CL^{pro} and all the studied inhibitors is driven mainly by vdW interactions. Besides, hydrogen bond formations at F140, G143, C145, H164, E166, and Q189 residues were also important for their binding efficiency. The predicted ΔG_{bind} somewhat agreed with the ΔG_{Exp}, and among them, the 13b exhibited the most efficient binding to SARS-CoV-2 3CL^{pro}. The bulky N-terminal protecting groups (e.g., thiazole, isoxazole, and MPC moieties) should be introduced at the P4 position in order to improve binding interactions with P168 and T190 hydrophobic residues, whereas the polar moieties (e.g. benzamide and benzoate fragments) should be added to the phenyl ring at P1' site to facilitate H-bond formations with the surrounding residues T25, H41, and C44. This information can be helpful for further drug development of novel peptidomimetic inhibitors targeting SARS-CoV-2 3CL^{pro}.

Declaration of Competing Interest

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

References

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, Z. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Lancet 395 (2020) 497.
[2] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, N. Engl. J. Med. 382 (2020) 727.
[3] C. Drosten, S. Günther, W. Preiser, S. van der Werf, H.-R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R.A.M. Fouchier, A. Berger, A.-M. Burguière, J. Cinatl, M. Eickmann, N. Escriou, K. Grywna, S. Kramme, J.-C. Manuguerra, S. Müller, V. Rickerts, M. Stürmer, S. Vieth, H.-D. Klenk, A.D.M.E. Osterhaus, H. Schmitz, H.W. Doerr, N. Engl. J. Med. 348 (2003) 1967.
[4] T.G. Ksiazek, D. Erdman, C.S. Goldsmith, S.R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J.A. Comer, W. Lim, P.E. Rollin, S.F. Dowell, A.-E. Ling, C.D. Humphrey, W.-J. Shieh, J. Guarner, C.D. Paddock, P. Rota, B. Fields, J. DeRisi, J.-Y. Yang, N. Cox, J.M. Hughes, J.W. LeDuc, W.J. Bellini, L.J. Anderson, N. Engl. J. Med. 348 (2003) 1953.
[5] R.H. Xu, J.F. He, M.R. Evans, G.W. Peng, H.E. Field, D.W. Yu, C.K. Lee, H.M. Luo, W.S. Lin, P. Lin, L.H. Li, W.J. Liang, J.Y. Lin, A. Schnur, Emerg. Infect. Dis. 10 (2004) 1030.
[6] R.J. de Groot, S.C. Baker, R.S. Baric, C.S. Brown, C. Drosten, L. Enjuanes, R.A.M. Fouchier, M. Galiano, A.E. Gorbalenya, Z.A. Memish, S. Perlman, L.L.M. Poon, E.J. Snijder, G.M. Stephens, P.C.Y. Woo, A.M. Zaki, M. Zambon, J. Ziebuhr, Journal of virology 87 (2013) 7790.
[7] A.M. Zaki, S. van Boheemen, T.M. Bestebroer, A.D.M.E. Osterhaus, R.A.M. Fouchier, N. Engl. J. Med. 367 (2012) 1814.
[8] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J.A. Xia, T. Yu, X. Zhang, L. Zhang, Lancet 395 (2020) 507.
[9] Y. Chen, Q. Liu, D. Guo, J. Med. Virol. 92 (2020) 418.

[10] S. Hussain, J. Pan, Y. Chen, Y. Yang, J. Xu, Y. Peng, Y. Wu, Z. Li, Y. Zhu, P. Tien, D. Guo, J. Virol. 79 (2005) 5288.

[11] R. Ramajayam, K.-P. Tan, P.-H. Liang, Biochem. Soc. Trans. 39 (2011) 1371.

[12] R. Hilgenfeld, FEBS J. 281 (2014) 4085.

[13] E.J. Snijder, P.J. Bredenbeek, J.C. Dobbe, V. Thiel, J. Ziebuhr, L.L.M. Poon, Y. Guan, M. Rozanov, W.J.M. Spaan, A.E. Gorbalenya, J. Mol. Biol. 331 (2003) 991.

[14] B. Nutho, P. Mahalapbutr, K. Hengphasatporn, N.C. Pattaranggoon, N. Simanon, Y. Shigeta, S. Hannongbua, T. Rungrotmongkol, Biochemistry 59 (2020) 1769.

[15] L. Mittal, A. Kumari, M. Srivastava, M. Singh, S. Asthana, J. Biomol. Struct. Dyn. (2020) 1.

[16] V. Nukoolkarn, V.S. Lee, M. Malaisree, O. Aruksakulwong, S. Hannongbua, J. Theor. Biol. 254 (2008) 861.

[17] X. Xue, H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam, Z. Rao, J. Mol. Biol. 366 (2007) 965.

[18] K. Fan, P. Wei, Q. Feng, S. Chen, C. Huang, L. Ma, B. Lai, J. Pei, Y. Liu, J. Chen, L. Lai, J. Biol. Chem. 279 (2004) 1637.

[19] U. Bacha, J. Barrila, A. Velazquez-Campoy, S.A. Leavitt, E. Freire, Biochemistry 43 (2004) 4906.

[20] H.A. Elshabrawy, J. Fan, C.S. Haddad, K. Ratia, C.C. Broder, M. Caffrey, B.S. Prabhakar, J. Virol. 88 (2014) 4353.

[21] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, PLoS Biol. 3 (2005), e324, .

[22] L. Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E.J. Snijder, H. Liu, R. Hilgenfeld, J. Med. Chem. 63 (2020) 4562.

[23] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Nature 582 (2020) 289.

[24] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Science 368 (2020) 409.

[25] W. Dai, B. Zhang, X.-M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.-K. Zhang, Y. Xu, H. Yang, H. Liu, Science 368 (2020) 1331.

[26] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic acids research 28 (2000) 235.

[27] Accelrys, Inc, San Diego, CA, USA, 2009.

[28] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, Nucleic Acids Research 46 (2018) W296.

[29] P. Mahalapbutr, K. Thitinanthavet, T. Kedkham, H. Nguyen, L.T.H. Theu, S. Dokmaisrijan, L. Huynh, N. Kungwan, T. Rungrotmongkol, J. Mol. Struct. 1180 (2019) 480.

[30] J. Kammarabutr, P. Mahalapbutr, B. Nutho, N. Kungwan, T. Rungrotmongkol, J. Mol. Graph. Model. 89 (2019) 122.

[31] P. Mahalapbutr, P. Wonganan, T. Charoenwongpaiboon, M. Prousoontorn, W. Chavasiri, T. Rungrotmongkol, Biomolecules 9 (2019) 545.

[32] G.W.T.M.J. Frisch, H.B. Schlegel, G.E. Scuseria, J.R.C.M.A. Robb, G. Scalmani, V. Barone, B. Mennucci, H.N.G.A. Petersson, M. Caricato, X. Li, H.P. Hratchian, J.B.A.F. Izmaylov, G. Zheng, J.L. Sonnenberg, M. Hada, K.T.M. Ehara, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, O.K.Y. Honda, H. Nakai, T. Vreven, J.A. Montgomery Jr., F.O.J.E.

Peralta, M. Bearpark, J.J. Heyd, E. Brothers, V.N.S.K.N. Kudin, T. Keith, R. Kobayashi, J. Normand, A.R.K. Raghavachari, J.C. Burant, S.S. Iyengar, J. Tomasi, N.R.M. Cossi, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, C.A.V. Bakken, J. Jaramillo, R. Gomperts, R.E. Stratmann, A.J.A.O. Yazyev, R. Cammi, C. Pomelli, J.W. Ochterski, K.M.R.L. Martin, V.G. Zakrzewski, G.A. Voth, J.J.D.P. Salvador, S. Dapprich, A.D. Daniels, J.B.F.O. Farkas, J.V. Ortiz, J. Cioslowski, A.D.J. Fox, Gaussian, Inc., Wallingford CT, 2010.

[33] D.S.C.D.A. Case, T.E. Cheatham, T.A. Darden III, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York, P.A. Kollman, University of California, San Francisco, 2017.

[34] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, J. Chem. Theory Comput. 11 (2015) 3696.

[35] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, J. Comput. Chem. 25 (2004) 1157.

[36] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, Nucleic Acids Res. 32 (2004) W665.

[37] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.

[38] T. Darden, D. York, L. Pedersen, J. Chem. Phys. 98 (1993) 10089.

[39] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, J. Comput. Phys. 23 (1977) 327.

[40] B.P. Uberuaga, M. Anghel, A.F. Voter, J. Chem. Phys. 120 (2004) 6363.

[41] H.J.C. Berendsen, J.P.M. Postma, W.F.V. Gunsteren, A. DiNola, J.R. Haak, The Journal of Chemical Physics 81 (1984) 3684.

[42] D.R. Roe, T.E. Cheatham, J. Chem. Theory Comput. 9 (2013) 3084.

[43] S. Genheden, U. Ryde, Expert Opin Drug Discov 10 (2015) 449.

[44] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, J. Chem. Theory Comput. 8 (2012) 3314.

[45] M. Naïm, S. Bhat, K.N. Rankin, S. Dennis, S.F. Chowdhury, I. Siddiqi, P. Drabik, T. Sulea, C.I. Bayly, A. Jakalian, E.O. Purisima, J. Chem. Inf. Model. 47 (2007) 122.

[46] M.U. Mirza, M. Froeyen, Journal of Pharmaceutical Analysis 10 (2020) 320.

[47] J. Liang, E. Pitsillou, C. Karagiannis, K.K. Darmawan, K. Ng, A. Hung, T.C. Karagiannis, Comput. Biol. Chem. 87 (2020) 107292.

[48] H.C. Cheng, J. Pharmacol. Toxicol. Methods 46 (2001) 61.

[49] P. Mukherjee, F. Shah, P. Desai, M. Avery, J. Chem. Inf. Model. 51 (2011) 1376.

[50] T.T. Talele, J. Med. Chem. 59 (2016) 8712.

[51] L. Zhu, S. George, M.F. Schmidt, S.I. Al-Gharabli, J. Rademann, R. Hilgenfeld, Antivir. Res. 92 (2011) 204.

[52] J. Tan, S. George, Y. Kusov, M. Perbandt, S. Anemüller, J.R. Mesters, H. Norder, B. Coutard, C. Lacroix, P. Leyssen, J. Neyts, R. Hilgenfeld, Journal of virology 87 (2013) 4339.

[53] K. Cao, N. Li, H. Wang, X. Cao, J. He, B. Zhang, Q.Y. He, G. Zhang, X. Sun, J. Biol. Chem. 293 (2018) 6075.

[54] P. Mahalapbutr, P. Wonganan, W. Chavasiri, T. Rungrotmongkol, Cancers (Basel) 11, 2019.

[55] P. Mahalapbutr, N. Darai, W. Panman, A. Opasmahakul, N. Kungwan, S. Hannongbua, T. Rungrotmongkol, Sci. Rep. 9 (2019) 10205.