MDPI

Article

Identifying the Hot Spot Residues of the SARS-CoV-2 Main Protease Using MM-PBSA and Multiple Force Fields

Jinyoung Byun and Juyong Lee *

Department of Chemistry, Division of Chemistry and Biochemistry, Kangwon National University, Chuncheon 24341, Korea; byunjy0614@kangwon.ac.kr
* Correspondence: juyong.lee@kangwon.ac.kr

Abstract: In this study, we investigated the binding affinities between the main protease of SARS-CoV-2 virus (M^{pro}) and its various ligands to identify the hot spot residues of the protease. To benchmark the influence of various force fields on hot spot residue identification and binding free energy calculation, we performed MD simulations followed by MM-PBSA analysis with three different force fields: CHARMM36, AMBER99SB, and GROMOS54a7. We performed MD simulations with 100 ns for 11 protein–ligand complexes. From the series of MD simulations and MM-PBSA calculations, it is identified that the MM-PBSA estimations using different force fields are weakly correlated to each other. From a comparison between the force fields, AMBER99SB and GROMOS54a7 results are fairly correlated while CHARMM36 results show weak or almost no correlations with the others. Our results suggest that MM-PBSA analysis results strongly depend on force fields and should be interpreted carefully. Additionally, we identified the hot spot residues of M^{pro}, which play critical roles in ligand binding through energy decomposition analysis. It is identified that the residues of the S4 subsite of the binding site, N142, M165, and R188, contribute strongly to ligand binding. In addition, the terminal residues, D295, R298, and Q299 are identified to have attractive interactions with ligands via electrostatic and solvation energy. We believe that our findings will help facilitate developing the novel inhibitors of SARS-CoV-2.

Keywords: SARS-CoV-2; main protease; molecular dynamics; multiple force fields; MM-PBSA; ligand binding; absolute binding free energy

1. Introduction

The new type of coronavirus, the SARS-CoV-2 virus, outbroke in China at the end of 2019. Since its first outbreak, it has led to more than 265 million infections and 5.25 million deaths as of November 2021 [1,2]. Although many vaccines are approved and being developed, no small molecule drug, which has a molecular weight under 500 amu, has been approved by FDA yet as of November 2021 [3,4]. The SARS-CoV-2 genome encodes more than 20 proteins, which include the main protease (M^{pro}) that shares 96.1% similarity with 3CLP of SARS-CoV [5]. M^{pro} plays an important role in SARS virus replication and transcription. When the mRNA of the virus is translated into a polyprotein chain, M^{pro} is first autocleaved to become a mature enzyme, which in turn cleaves all of the 11 remaining downstream nonstructural proteins of the polyproteins to polypeptides, which are required for the replication process of the virus [5].

M^{pro} of SARS-CoV-2 is essential for viral replication and is considered a promising target for drug development [6–8]. Based on this assumption, various studies have been performed to find potent inhibitors targeting M^{pro} [7,9]. Thus, it is critical to evaluate the binding free energy between a ligand and its receptor both accurately and efficiently to accelerate COVID-19 drug discovery.

Because the shape of the binding site of M^{pro} was well characterized [6,10–12], docking studies between ligands and the M^{pro} receptor have been carried out to find out possible hits or leads [13,14]. Alamri et al. used the covalent docking tool, Schrödinger's Covalent

Docking (CoVDock) [15], to discover potential covalent inhibitors from the AFCL library consisting of 1000 covalent molecules and 116 FDA-approved molecules. As a result, paritaprevir and simeprevir from FDA-approved molecules were identified as potential covalent inhibitors of M^{pro} of SARS-CoV-2 [13]. Peele et al. screened 24 natural plant-derived molecules with antiviral property, 22 FDA-approved antiviral drugs, and 16 antimalarial drugs with the docking approach [14]. They found several molecules, lopinavir, amodiaquine, and theaflavin digallate that showed a good binding affinity with M^{pro}.

To further refine docking results, more rigorous absolute binding free energy calculations are performed in general. There are many approaches to obtain the absolute binding free energies of protein–ligand complexes: attach-pulling of a ligand from its receptor, linear interaction energy (LIE), molecular mechanics/Poisson–Boltzmann surface area and generalized Born surface area (MM-PBSA and MM-GBSA), free energy perturbation (FEP), thermodynamic integration (TI), and nonequilibrium molecular dynamics simulation [16,17]. All of these computational methods have a trade-off between computational cost and accuracy. Both FEP and TI approaches are considered most rigorous and accurate than the other methods. However, they demand much more computational resources to obtained well-converged binding free energies [18–21]. The ligand attach-pulling approach pulls ligand from its binding site and calculates a binding free energy based on the Jarzynski or Crooks theorem [22]. This nonequilibrium approach depends on a pulling pathway, and it is still not clear how much computational resources are required to obtain a reliable binding free energy. As an alternative approach to these perturbation methods, the MM-PB/GBSA approach, also known as an end-state method, is widely used to estimate absolute binding free energies of protein–ligand complexes due to its relatively lower computational burden. However, the approach suffers from rather large computational errors, lower accuracy, than the perturbation-based methods. Despite its lower accuracy, the MM-PB/GBSA method is becoming more popular to rank the binding free energies of different ligands with a single receptor quickly [23–25].

Here, we briefly review previous computational studies that used free energy calculation methods to discover possible drug candidates that bind to M^{pro}. Ngo et al. used three computational methods, molecular docking, fast pulling of ligand (FPL), and FEP, to discovery a potential SARS-CoV-2 M^{pro} from the existing drugs and natural compounds [26]. They screened potential inhibitor candidates from the large library and obtained the two natural compounds that showed high binding affinity in comparison with 13b ligand, which was from the recently reported experimental work [9]. Moreover, they suggested the E166 residue of M^{pro} is critical in ligand binding through alanine mutation calculation. Sk et al. have evaluated the binding free energies of two molecules (α-ketoamide and Z31792168) with M^{pro} by using the MM-PB/GBSA approach [27]. They performed per-residue energy decomposition analysis and showed that the ΔE_{vdw}, ΔE_{elec}, and ΔE_{nonpol} terms contributed favorably to ligand binding, and the ΔE_{vdw} contributed most. Based on the per-residue decomposition analysis, they identified hotspot residues, which have higher interaction energy than -1.0 kcal/mol. For both ligands, H163, H164, and M165 were identified as hot spot residues. Nutho et al. selected lopinavir and ritonavir as potential inhibitors, which were found to be effective against HIV protease [28]. From the MM-PB/GBSA analysis, they found that the ΔE_{vdw} is the most favored force to binding free energy more twice than ΔE_{elec} with both ligands. Moreover, they also identified important residues for ligand binding whose decomposed binding free energies are lower than -1.0 kcal/mol. Four residues, MET49, M165, P168, and Q189, were identified for the M^{pro}-lopinavir complex and nine residues, L27, H41, M49, F140, N142, G143, H164, M165, and E166, were identified for the M^{pro}-ritonavir complex. Wang et al. also used the MM-PBSA with a flexible docking method to identify potential inhibitors for SARS-CoV-2 M^{pro} protease [29]. They suggested that the five molecules, carfilzomib, ervacycline, valrubicin, lopinavir, and elbasvir, as the potential inhibitors against M^{pro}. Among them, carfizomib was identified as the strongest binder.

The main objectives of this study are (1) to identify the hot spots of M[pro] of SARS-CoV-2 via MM-PBSA calculations using multiple force fields and (2) to assess the force field dependency of MM-PBSA calculation results.

In this study, we performed the MD simulations of known 11 M[pro]–ligand complexes and calculated the binding affinities using MM-PBSA calculations to identify the hot spot residues of M[pro]. Six ligands are from Diamond Light Source (diamond.ac.uk/COVID-19/for-scientists/Main-protease-structure-and-XChem/Downloads.html), three HIV protease inhibitors (darunavir, lopinavir, and indinavir), and three more recently reported ligands (11r, 13a, and 13b) [9], respectively. For all 11 M[pro]–ligand complexes, we performed three MD simulations using different force fields to investigate the force field dependency of binding free energy estimates. Three different force fields were used: GROMOS54a7 (g54a7) [30,31], AMBER99SB (ff99sb) [32], and CHARMM36 force field (c36) [33]. For each complex, 100 ns of MD simulations were performed. The MM-PBSA results obtained with three force fields are compared and the hot spot residues are identified via energy decomposition analyses.

2. Materials and Methods

2.1. Molecular Docking Calculation

When we used GalaxyDock3 [34] to generate the initial M[pro]–ligand configurations, we selected the center of mass of the 13b (PDB ID: 6Y2F) as the center of a grid box with the SARS-CoV-2 M[pro] structure whose PDB ID is 6YB7. The ligand data were downloaded from diamond.ac.uk/COVID-19. The atomic partial charges of each ligand were calculated with the Gasteiger approach [35,36]. The lowest energy docking pose was selected as the initial configuration for MD simulations of 11r, 13a, darunavir, lopinavir, and indinavir whose native structure are not identified yet.

2.2. Molecular Dynamics Simulations

In this study, we used two MD programs (GROMACS-2019.4 [37] and AMBER18 [38]) and three force fields: g54a7 [30,31], ff99sb [32], and c36 [33].

2.2.1. MD Simulations with GROMACS

The g54a7 was used to perform MD simulations with GROMACS-2019.4 [37]. The ligand topology files were obtained from the ATB (Automated Topology Builder) server [39], which generates ligand parameters compatible with the protein force field. We used the original geometry output of ATB. The systems were solvated with the SPC water model [40] with a 12 Å padding region. We added Na^+ ions to neutralize the system. The time step of MD simulation was set to 2 fs with the LINCS algorithm to constraint hydrogen-connected covalent bonds [41]. The long-range electrostatic interaction was calculated with the fast smooth particle-mesh Ewald electrostatics method [42,43]. The solvated system was minimized by using the steepest descent method for 10,000 steps. After minimization, a series of equilibrium MD simulations were performed: the 200 ps of NVT, 1 ns of NPT, and 50 ns of NVT with weak harmonic position restraints. At each step, the simulations were performed with the Nosé–Hoover thermostat [44] and the Parrinello–Rahman barostat [45] at 300 K. Last, the production step was performed with NPT condition during 100 ns.

2.2.2. MD Simulations with AMBER

We performed MD simulations with the two force fields, ff99sb [32] and c36 [33] using AMBER18 [38].

- Preparing systems with an AMBER99SB force field (ff99sb)
  The M[pro] and ligand topology files were generated using tleap [46] and Antechamber [47] programs implemented in AMBER18 with ff99sb and generalized AMBER force field (GAFF) [48]. The systems were solvated with the TIP3P water model with a padding region of 12 Å. Energy minimization calculations of the solvated systems were performed by using the steepest descent method. After energy minimization,

equilibrium MD simulations were performed as follows: 200 ps of NVT, 1 ns of NPT, and 50 ns of NVT with weak harmonic positional restraints. The time step for MD simulations was set to 2 fs. To constrain all bonds involving hydrogen atoms, the SHAKE algorithm was used [49]. The particle mesh Ewald (PME) method also calculated the long-range electrostatic interactions [50]. The temperature was controlled with Langevin thermostat [51]. After equilibrium simulations were finished, we performed production simulations for 100 ns with NPT condition and the trajectory of each simulation was recorded every 2 ps.

- Preparing systems with CHARMM36 force field (c36)
  The parameter and topology files of complexes were generated from CHARMM-GUI [52] with c36 [33]. The parameters for ligands were generated by CHARMM generalized force field (CGenFF) [53,54]. The subsequent NVT 125 ps, NPT 1 ns, NVT 50 ns equilibration simulations were carried out at 300 K, followed by production runs for 100 ns under NPT ensemble condition. The TIP3P water model was used to represent explicit water molecules. We added the sodium ions in order to neutralize the system. The time step for MD simulation was set to 2 fs with the SHAKE algorithm. After the energy of the solvated system was minimized by using the steepest descent method, the subsequent steps were carried out using the identical parameters with the AMBER setting described above.

2.3. MM-PBSA Analysis

In this study, the interactions between the M^{pro} receptor and ligands were calculated with the MM-PBSA analysis. We utilized the two MM-PBSA tools; one is `g_mmpbsa` [55] for the GROMACS MD results and the other is `MMPBSA.py` [56] from the AMBER18 package, respectively.

- `g_mmpbsa` setting
  Each topology and parameter files required for MM-PBSA calculations are obtained from the MD simulation with GROMACS section. For a trajectory of 100 ns, 10,000 frames were selected for MM-PBSA calculations as input. For nonpolar solvation energy calculations, we selected the SASA model [57–59] which is one of the most used nonpoloar models. For the other parameters, the default values of GROMACS were used.
- `MMPBSA.py` setting
  Each parameter and system topology files for a solvated complex, nonsolvated complex, receptor, and ligand were used from the preparation steps of MD simulation with AMBER section. Additionally, we performed the solvated M^{pro} receptor simulation and solvated ligand simulation, respectively, to estimate the receptor and ligand contribution in implementing MM-PBSA analysis. Similar to the `g_mmpbsa` setting, 10,000 frames also were used for MM-PBSA analyses with `MMPBSA.py` from trajectories of 100 ns. For PB calculations, the SASA (solvent accesible surface area) for the nonpolar solvation energy used the LCPO model [60] and the ionic strength of solution was set to 0.1 M and the mbondi [61] radius set was used. The rest of the options were set to the default values of `MMPBSA.py` in AMBER 18 package.

2.4. Linear Interaction Energy (LIE) Calculation

ΔG_{LIE} was computed as the mean of vdW and electrostatic interaction difference between bound state and unbound state $\Delta G_{LIE} = \alpha \langle \Delta E_{elec} \rangle + \beta \langle \Delta E_{vdW} \rangle + \gamma$. The coefficient γ reflects the hydrophobicity of various species of inhibitors to the binding cleft conceding. The coefficient α and β are weight parameters for nonpolar and polar interactions, respectively. Based on the ligands whose experimental IC_{50} value was identified, we estimated the α, β, and γ coefficient values, followed by recalculating the other ligands' energies to evaluate ΔG_{LIE}.

3. Results

The structure of M^pro consists of three domains: domain I (residues 10–99), domain II (residues 100–184), domain III (residues 201–303), and a long loop region (residues 185–200) (Figure 1a). Domain III is connected to domain II via the long loop structure. The substrate-binding site is located in a cleft between domain I, II with the loop region involved. It is composed of four subsites (S1, S1', S2 and S4) (Figure 1b,c). Each subsite is spatially located and composed of consecutive amino acids in the substrate-binding site. The S1 subsite consists of F140, N142, C145, H163, E166, and H172 residues. (Figure 1b red region). The S1' subsite is composed of T26 and L27 (Figure 1b lime region). The side chains of H41, M49, Y54, M165, D187 are involved in the formation of the S2 subsite (Figure 1b light blue). The L167, F185, Q189, Q192 consists of S4 subsite (the orange region in Figure 1b yellow).
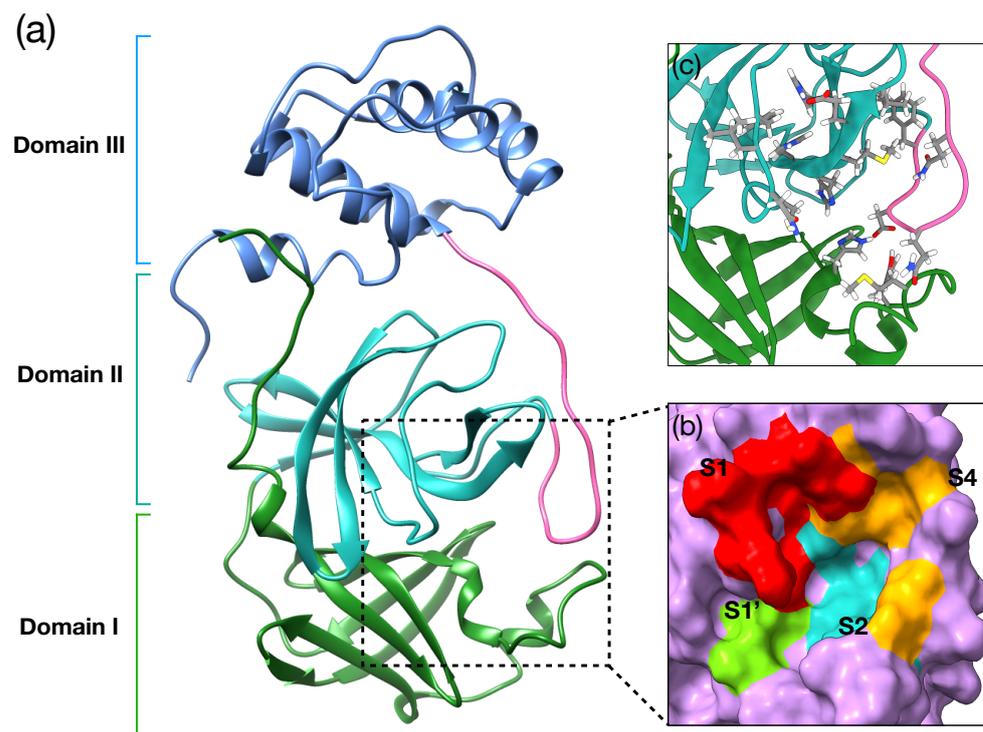
Figure 1. The structure of a single M^pro receptor (PDB ID: 6Y2F). (**a**) The three domains are colored and the pink colored region represents a long loop region. Domain I (green, residues 10–99), II (sea green, residues 100–184), III (cornflower blue, residues 201–305), and the pinked long loop, which is located between domains II and III (residues 185–200). The binding site is located within the dashed square box, composed of Domain I, Domain II, and the loop region. (**b**) The surface was represented by the catalytic binding center. (**c**) The surface removed view of the binding site. The residues involved in each subsite are represented.

3.1. MD Results of Complexes

Most MD simulations of the M^pro complexes remained bound stably, except the M^pro-X0072 and -X0691 complexes. In the simulations of the M^pro-X0072 complex with both ff99sb and c36, the X0072 ligand was detached from M^pro. The X0691 ligand with c36 also was detached from M^pro during simulations. The trajectories showed that both the X0072 and X0691 ligands occupied a region between E166 and Q189 in S4 subsite for a while and finally were detached from M^pro. These results suggest that both residues, E166 and Q189, may play the important roles of gateways of ligand binding.

The spatial occupancy patterns of ligands sampled during simulations are visually investigated to identify which residues participate in ligand binding of M^pro. To perform this investigation, the 'grid' command of the CPPTRAJ [62] program of the AMBER18 package was used to bin the distribution of a ligand into a 3D grid. Before performing the

analyses, periodic boundary conditions were removed and centered the system to put a ligand at the origin of each trajectory.

The MD result of darunavir showed a significantly localized distribution than those of lopinavir and indinavir, which are the inhibitors of HIV protease. The sulfonate group of darunavir stayed at the S4 subsite and the benzene ring also was located in S1 subsite and did not fluctuate much (Figures 2 and 3d). However, in the case of lopinavir, the functional groups of lopinavir fluctuated much, except that the dimethyl substituted benzene group stayed at the S4 subsite. The branches of indinavir showed that the rings fluctuated, and did not occupy the S4 subsite similar to other HIV protease inhibitors. The other ligands, 11r and 13a showed different binding behaviors from 13b. The ligands, 11r and 13a, showed more concentrated spatial distribution consistently than the 13b ligand in simulations performed with all force fields. On the contrary, the trajectory of 13b showed more widely distributed 3D histograms because of the fluctuation of its P1, P3 branches (Figures 2 and 3a–c). The P1 branch of both 11r and 13a remained stably at the S1 subsite during simulations. The P3 branch of 11r and the P1′ part of 13a with ff99sb, and the cyclohexyl ring of 13a with c36 also remained steadily at the S4 subsite, where most strong binders are occupied significantly during simulations. However, the P1 branch of 13b was located at a different region compared with 11r and 13a and fluctuated more than them, and 13b did not occupy the S4 subsite between E166 and Q189 during the simulations. It appears that this dynamical behaviors of the ligands are consistent with experimental data, the stronger binding free energy of 13b than 11r and 13a, which will be discussed further in the next section.
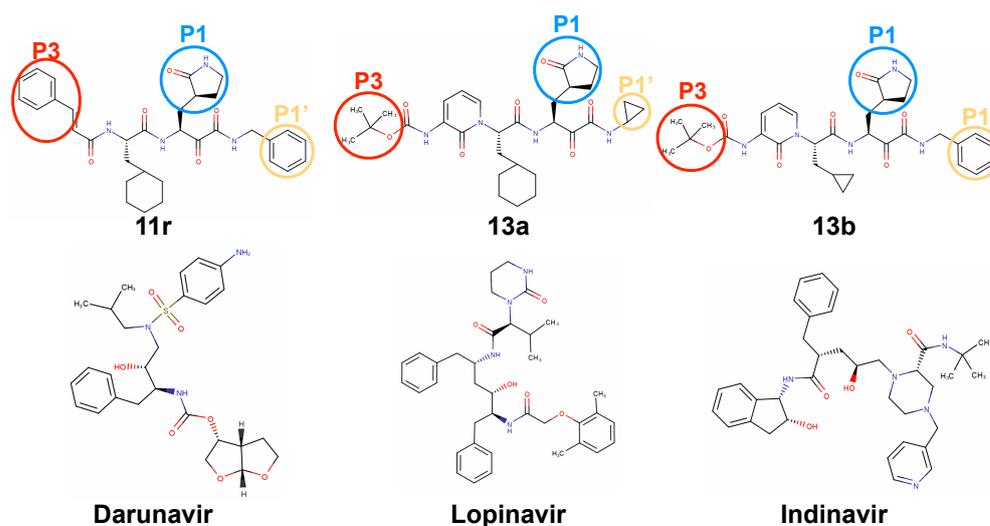
Figure 2. The ligands analyzed in Results section. 11r, 13a, and 13b ligands are at the upper row. The branches of 11r, 13a, and 13b ligands were represented in colored-circle. Darunavir, lopinavir, and indinavir are in the lower row. All ligands anaylzed in this study are in Figures S1–S5.

The fact that the S4 subsite was occupied by many strong binders was employed in a previous study that developed a new drug candidate from a weak hit molecule, perampanel [63]. The authors performed a structure-based substrate scope analysis, which considers the interaction and environments between M^{pro} and hit molecules, and rationally changed the branches of perampanel to fill the void of the S4 subsite. Most modified ligands interacted strongly with the E166, Q189, and Q192 residues. Especially, most modified ligands whose chloride atom was substituted for various alkoxy groups such as propoxy, butoxy group, or more bulky benzyloxy group to occupy the S4 subsite region, became more than 10 times stronger than the hit molecule in terms of IC_{50} values. This tendency is consistent with the MM-PBSA results, which will be discussed in the next section and may be attributed to hydrogen bonds between ligands and M^{pro}. The complete results of all the 3D histogram results are illustrated in Supplementary Information (Figures S15–S17).
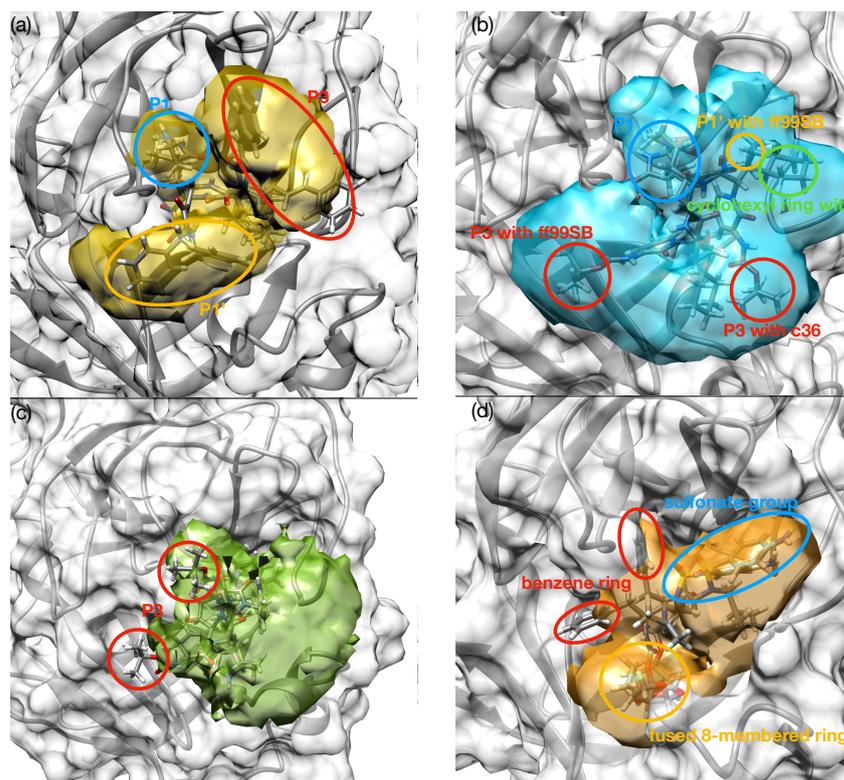
Figure 3. The examples of 3D histograms of ligand occupancy, comparing MS simulation results using ff99sb and c36. The colored region is a 3D histogram. The 3D histogram is spatial distribution of ligand atom sampled during simulation with ff99sb. The representative conformations from both ff99sb and c36 are imposed. (**a**) 11r (**b**) 13a (**c**) 13b (**d**) darunavir ligand.

3.2. Comparison of MD Simulation Results of Complexes Using Different Force Fields

In this section, we compared the MM-PBSA calculation results with three different force fields to investigate the effect of force field parameters on binding free energy estimates (Figures 4 and S6–S11). We estimated the entropic contribution of each ligand's binding free energy by using normal mode analysis (Table 1). The estimated entropy values were similar in the range of ±1 kcal/mol. Thus, entropic contributions are not considered for analysis.

3.2.1. A Comparison of ff99sb and c36 Force Field Results

The MM-PBSA results obtained with ff99sb and c36 showed weak correlations between them, Kendall $\tau = 0.067$, Pearson r = 0.022, and Spearman $\rho = 0.042$, respectively (Figure 4a). Overall, the MM-PBSA binding free energy estimates obtained with ff99sb were more favorable than those with c36 simulations. Most binding energy results obtained with MM-PBSA calculations using ff99sb had lower bind energy by more than about 10 kcal/mol than the c36 results.

The strongest binders of each force field are 11r (-54.990 kcal/mol) with ff99sb and 13a (-28.523 kcal/mol) with c36, respectively. 13a also identified as the top3 strong binder with ff99sb (-37.971 kcal/mol). However, 11r was not identified as a strong binder with c36. It suggested that the two force fields did show consistent results with each other. From experiments [9,64], the binding free energies of 11r, 13a, 13b, and darunavir ligands were -9.23, -7.70, -8.45, and -6.14 kcal/mol, respectively. In terms of consistency with experimental data, MM-PBSA with ff99sb correctly identified 11r as the strongest binder. However, c36 results predicted that 13a would be a stronger binder than 11r, which is inconsistent with the experiment. Both ff99sb (-29.535 kcal/mol) and c36 (5.579 kcal/mol) estimated that 13b is the weakest binder among the 11r, 13a, and 13b despite 13b is the second strong binder based on the experiment. We assumed the fact that 13b did not occupy the S4 subsite unlike 11r and 13a may explain why the MM-PBSA results were

not consistent with the experiment among 11r, 13a, and 13b ligands. Moreover, although darunavir is the weakest binder from the experimental data, the two force field failed to estimate the fact that darunavir is the weakest binder.
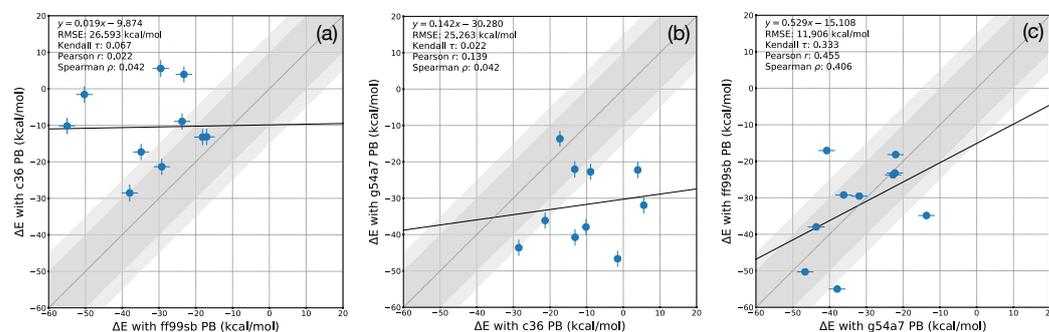
Figure 4. A comparison of the ΔE_{MM} obtained with MM-PBSA calculations using ff99sb, c36, and g54a7. (**a**) A comparison of MM-PBSA results between ff99sb and c36 (**b**) A comparison of MM-PBSA results between c36 and g54a7 (**c**) A comparison of MM-PBSA results between g54a7 and ff99sb The dashed diagonal line corresponds to the perfect agreement between the two results. The dotted lines correspond to a range of an absolute error of 10 kcal/mol. The error bars are presented with points. Because the X0072 and X0691 ligand were detached from M^{pro}, they were omitted.

3.2.2. A Comparison of c36 and g54a7 Force Field Results

Next, we compared the MM-PBSA results obtained with c36 and g54a7 (Figure 4b). A weak correlation was also observed with a correlation coefficient, Kendall $\tau = 0.022$, Pearson $r = 0.139$, and Spearman $\rho = 0.042$, respectively. Most binding energies obtained with g54a7 showed more favorable values than those of c36. MM-PBSA results with g54a7 estimated that darunavir would be the strongest binder with a binding energy of -46.640 kcal/mol. The top contributing residues for the binding of darunavir with g54a7 were M49, E166, M165, and Q189 with decomposed energies (-3.625, -2.740, -1.972, and -1.491 kcal/mol, respectively). Because the M165, E166, and Q189 residues are in the S4 subsite, g54a7 also suggested that the S4 subsite was important for M^{pro}-ligand binding. Although c36 showed less favorable energies than g54a7, it is valuable to compare the binding energy of darunavir, which is the strongest binding with g54a7, between g54a7 and c36. With c36, both E166 and Q189 had the unfavorable decomposed energies—4.748 and 1.314 kcal/mol, respectively. However, these energies were reversed with g54a7— -2.740 and -1.491, respectively. We found that both residues have with highly unfavorable electrostatic interactions with the ligand, E$_{elec}$ of 19.815, 4.431 kcal/mol, respectively.

3.2.3. A Comparison of g54a7 and ff99sb Force Field Results

The MM-PBSA binding energies obtained with ff99sb and g54a7 were relatively well-correlated with Kendall $\tau = 0.333$, Pearson $r = 0.455$, and Spearman $\rho = 0.406$, respectively (Figure 4a,c). g54a7 also estimated that 13b is the weakest binder among 11r, 13a, 13b, and darunavir. These results were consistent with all three force fields. The MM-PBSA results with ff99sb showed a similar trend with the MM-PBSA results with g54a7 except for 11r and lopinavir cases. If the three outliers are excluded, a correlation coefficient between g54a7 and ff99sb was $\rho = 0.933$. This result suggests that g54a7 yields similar binding free energies with the all-atom model from ff99sb, although g54a7 used united-atom model parameters. The strongest binder without the outlier ligand was darunavir: -50.288 kcal/mol with ff99sb and -46.640 kcal/mol with g54a7. Darunavir was also the second strongest binder with the outlier ligands with ff99sb.

To investigate why the results of ff99sb and g54a7 are correlated, we calculated the correlation coefficients between individual energy terms of MM-PBSA results: ΔE_{vdw}, ΔE_{elec}, ΔE_{PB}, and ΔE_{nonpol}, respectively (Figure 5). Among the energy terms, the ΔE_{vdw} term showed a correlation between two force fields while the other energetic terms did not show any correlation. This shows that a correlation between ff99sb and g54a7 can

be attributed to the van der Waals term. This may be due to the fact that electrostatic or solvation interaction terms depend on more delicate parameters, such as partial charges and atomic radii for Poisson–Boltzmann calculations
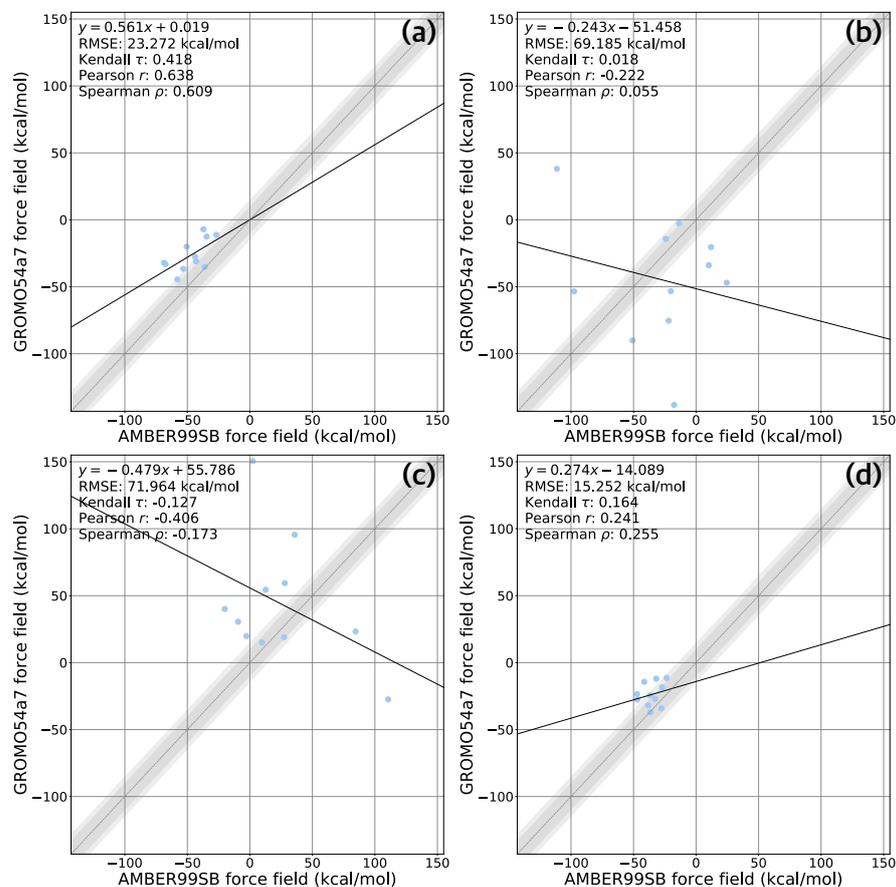
Figure 5. A comparison of the ΔE obtained with MM-PBSA calculations between ff99sb and g54a7. (**a**) ΔE_{vdw}, (**b**) ΔE_{elec}, (**c**) ΔE_{PB}, and (**d**) ΔE_{nonpol} All the subfigures are same in *x*- and *y*-axis scale to compare easily. The dashed diagonal line corresponds to the perfect agreement between the two results. The dotted lines correspond to a range of an absolute error of 10 kcal/mol. Because the X0072 and X0691 ligand were detached from M^{pro}, they were omitted.

3.3. Linear Interaction Energy Analysis

To improve the accuracy of binding affinity calculations, LIE analysis is commonly used to tune the balance between the electrostatic and van der Waals interactions. To check whether the LIE analysis improves the correlations between the calculated results and experiments, we performed the LIE analysis using the four ligands whose experimental IC$_{50}$ were known. (Figure 6).

It is identified that the LIE analysis did not improve the correlations between results obtained with different force fields. The correlation coefficient between the results of ff99sb and g54a7 deteriorated from 0.45 to 0.11. Meanwhile, the correlation coefficient between c36 and g54a7 elevated to 0.31 from 0.14, which is the highest among the LIE results. The correlation between c36 and ff99sb remained low after the LIE analysis with a Pearson correlation coefficient of 0.09. These results demonstrate that, even with the LIE analysis, the MM-PBSA results strongly depend on the choice of force field.
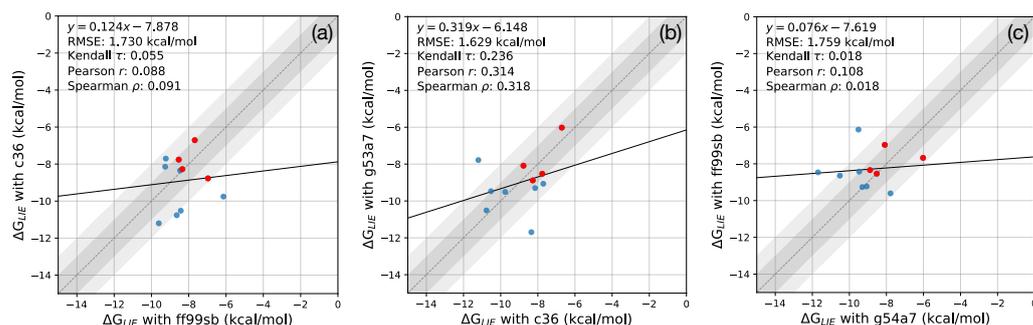
Figure 6. A comparison between MM-PBSA calculation results fitted by the linear interaction energy analysis. The left plot corresponds to the correlation between c36 and ff99sb. (**a**) A comparison between ff99sb and c36 results fitted by the LIE analysis. (**b**) A comparison between c36 and g54a7 results fitted by the LIE analysis. (**c**) A comparison between g54a7 and ff99sb results fitted by the LIE analysis. The red dots correspond to the ligand whose experimental IC_{50} values are identified. The dark shaded and light shaded regions correspond to an absolute error of 1 and 2 kcal/mol, respectively.

3.4. A Comparison of Polar Solvation Energy

In this section, we compared ΔE_{PB} from the MM-PBSA calculation results with three different force fields to investigate the effect of solvation on ligand binding shown in Figure 7. Usually, the PB energies contributed unfavorably to ligand binding. Specifically, the c36 force field showed the most unfavorable contribution for the binding of 13b and darunavir. 13b with c36 showed that E166 and R188 have ΔE_{PB} of 19.1 and 15.3 kcal/mol, respectively. Moreover, darunavir with c36 also showed that E166 and R188 have ΔE_{PB} of 19.8 and 11.3 kcal/mol, respectively, while 13a contributed to ΔE_{PB} of 6.2 and 4.9 kcal/mol for E166 and R188, respectively.
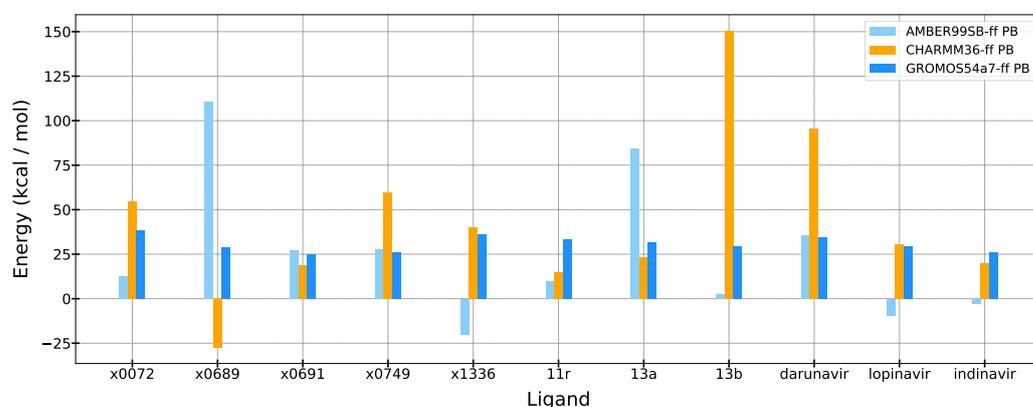
Figure 7. A comparison of ΔE_{PB} with ff99, c36, and g54a7 force fields from MM-PBSA analysis.

On the other hand, in the cases of lopinavir and indinavir, which are moderate binders with ff99sb, the ΔE_{PB} showed favorable energy. The exposed residues showed highly favorable ΔE_{PB}. The D263 residue had highly favorable ΔE_{PB} of -20.6 kcal/mol with lopinavir and -19.7 kcal/mol with indinavir, respectively, whereas that of 11r is only -5.6 kcal/mol. Additionally, E178 also showed highly favorable ΔE_{PB} of -16.6 kcal/mol with lopinavir and -19.0 kcal/mol with indinavir, while that of 11r is only -11.3 kcal/mol.

3.5. Summary of MM-PBSA Results

In summary, the MM-PBSA estimates obtained with c36 is showing a different trend with those obtained with ff99sb and g54a7 (Table 1 and Figure 8). The absolute values of binding energies obtained with c36 are smaller than those with ff99sb and g54a7. Unfortunately, all three force fields failed to predict the rankings of known experimental binding affinity data. All three force fields predicted that 13b ligand was the weakest binder among

11r, 13a, and 13b; however, 13b was the second strong binder based on the experimental data. Only ff99sb correctly identified the 11r ligand as the strongest binder. However, more experimental data are necessary to evaluate the accuracy of force fields more rigorously in obtaining correct absolute binding free energies of M^{pro} complexes.

These results show that, based on the MM-PBSA calculations, no force field is accurate enough to obtain the absolute binding affinities of M^{pro}-ligand complexes yet, and the calculation results may show large variations according to the tested force field. Thus, our results also show that a care should be taken to interpret the absolute binding free energy estimates obtained with MM-PBSA and different force fields. These results also emphasizes the necessity of using more rigorous absolute binding free energies approaches [65–68]. The complete results of all energies are Supplementary Information Table S1.
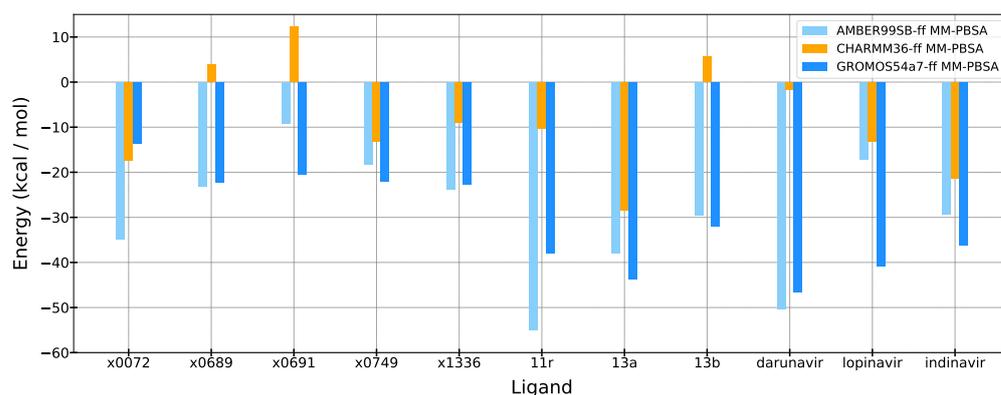
Figure 8. Overall binding free energy from MM-PBSA. All values has kcal/mol energy unit.

Table 1. Overall MM-PBSA difference energy results. All values in a kcal/mol energy unit. The superscript ^{a} means that all the energies are from the AMBER18 MD engine with ff99sb, while the superscript ^{b} indicates the results from the AMBER18 MD engine with c36. The superscript ^{c} indicates that all the energy values are from GROMACS-2019.4 MD engine with g54a7. NMA ΔS is that the entropy is calculated with normal mode analysis and QH ΔS means that the entropy is estimated by quasiharmonic (QH) approximation with only c36 system because `MMPBSA.py` package only supports normal mode analysis with ff99sb, not c36. All entropy values are computed by `MMPBSA.py` in AMBER18 MD package.

Ligand	MM-PBSA ^{a}	MM-PBSA ^{b}	MM-PBSA ^{c}	NMA ΔS	QH ΔS
X0072	−34.857	−17.302	−13.659	−16.455	−12.441
X0689	−23.223	3.943	−22.241	−16.899	−12.723
X0691	−9.075	12.182	−20.549	−18.119	−12.761
X0749	−18.167	−13.242	−22.040	−17.327	−12.677
X1336	−23.727	−8.932	−22.734	−19.280	−12.791
11r	−54.990	−10.177	−37.905	−31.088	−13.369
13a	−37.971	−28.523	−43.613	−27.803	−13.389
13b	−29.535	5.579	−31.927	−25.499	−13.404
darunavir	−50.288	−1.556	−46.640	−29.247	−13.330
indinavir	−29.238	−21.343	−36.142	−30.431	−13.430
lopinavir	−17.064	−13.166	−40.744	−27.693	−13.451

3.6. Hot Spot Residues for Ligand Binding

To identify hot spot residues playing important roles in ligand binding, we performed a per-residue energy decomposition analysis and calculated the average binding energy contribution of all ligands. The top 10 residues that contribute most to ligand binding of M^{pro} based on the average per-residue interaction energy is illustrated in Figure 9. All the top 10 contributing residues have ΔE^{MM} lower than −1.0 kcal/mol on average.
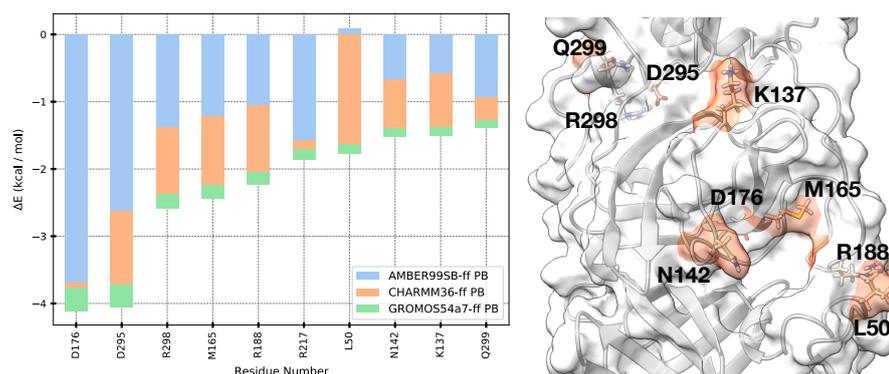
Figure 9. Averaged contribution energy from MM-PBSA and corresponding hot spot residues (**left**) are illustrated. Hot spot residues are highlighted on the surface of M^{pro} (**right**). Because the X0072 and X0691 ligands were detached from M^{pro}, they were omitted when generating the figure. Because the terminal residues are in the backside of the figure, their orange-colored regions were hidden.

The M165 residue is identified as the top contributing residue consistently across all tested force fields. M165 showed $\Delta E_{\text{avg}}^{\text{contrib}} = -5.0$ kcal/mol which is 2∼3 fold stronger than other residues' ΔE^{MM}. This result is consistent with the fact that most strongly binding ligands remained stably around the S4 subsite. The decomposition analysis revealed that M165 interacts favorably with ligands via both electrostatic and van der Waals interactions. With the strongest binder 11r with ff99sb and 13a with c36, M165 also was the most favorable residue. For the 11r case, E_{vdw}, E_{elec}, and E_{PB} were −2.508, −1.977, and 1.462 kcal/mol with ff99sb and −1.564, −1.614, and 1.714 kcal/mol with c36, respectively. For the 13a case, E_{vdw}, E_{elec}, and E_{PB} were −1.304, 0.842, and 1.003 kcal/mol with ff99sb and −1.721, −0.910, and 1.343 kcal/mol with c36, respectively. These results show that E_{vdw} mostly contribute to ligand binding.

From the atomic contact figures (Figures S18 and S19) generated by ligPlot+ [69], we identified that there are many short-range interactions between M^{pro} and the ligands, represented by red semicircles on protein residues or ligand atoms. Many atoms in the rings of ligands, mostly aromatic rings, are in close contact with the binding site residues. For example, most nonpolar atoms of three six-membered rings (two benzene rings and one cyclohexyl ring) of 11r showed that they form close contacts with binding site residues: S46 with the P1' benzene ring, E166 with the P3 benzene ring, and H164, M165, R188, and Q189 with the cyclohexyl ring.

Additionally, indinavir, 13a, and 13b have a large nonpolar group, a *tert*-butyl group, participating in protein–ligand binding (Figure S18). The *tert*-butyl groups of the ligands have many close contacts with surrounding residues: N119 with 13a, and N119 with 13b, and L141, S144, C145, and H164 with indinavir. Specifically, N119, which is away from the center of the binding site, largely interacts with ligands. Thus, we assume that the *tert*-butyl groups interact strongly with N119 via van der Waals interaction due to its bulky volume. Therefore, nonpolar groups, such as ring structure and *tert*-butyl group, play important roles in M^{pro}–ligand binding with many residues at diverse locations. The presence of many contacts between nonpolar atoms makes E_{vdw} the most contributing energetic term than the other terms.

Based on the fact that 11r with ff99sb and 13a with c36 are the strongest binders, we performed a residuewise binding free energy decomposition analysis to identify the hot spot residues of M^{pro}. The top contributing residues of 11r with ff99sb were D295, D187, D176, and H163 with binding energies of −5.538, −4.983, −4.350 and −3.624 kcal/mol, respectively. Except for D176 (4.185 kcal/mol), their E_{elec} were lower than −5.0 kcal/mol. The top contribution residues of 13a with c36 were D295, Y161, R298, L58, L50 with binding energies of −3.457, −3.308, −3.195, and −3.025 kcal/mol, respectively. They also had low E_{elec} except for Y161 (1.435 kcal/mol).

Interestingly, the terminal charged residues, D295, R298, and Q299, which are far from the binding site interacted strongly with ligands. Although they were far from the binding site over about 30 Å during MD simulation, the energy decomposition analysis showed that these residues have strong electrostatic or polar solvation attractive interaction for ligand's binding. Recently, T. Sztain et al. reported that the C-terminal region was examined as a potential allosteric site [70], which is consistent with our per-residue decomposition analysis result.

It is observed that the residues consisting of the S1' subsite contributed less to ligand binding than the other subsites, which is consistent with the 3D histogram analysis. Except for the terminal D295, R298, and Q299 residues that are far from the binding site, the hot spot residues (N142, M165, D176 and, R188) are identified in the S1 or S4 subsites. This result suggests that the residues of the S4 and S1 subsites may interact stronger than the residues of the S1' and S2 subsites with ligands.

3.7. Hydrogen Bonds Analysis

We carried out a hydrogen bond analysis and investigated its percentage to find which residues are forming hydrogen bonds strongly with the ligands. The complete hydrogen bond analysis results are in Supplementary Information Figures S12–S14. It is identified that if a specific hydrogen bond lasts long during a simulation, it contributed strongly to binding free energy. For the cases of 11r and 13a, which were the strongest binders with ff99sb and c36, respectively, the hydrogen bonds that existed for more than 60% of the entire trajectory were identified to be the most strongly contributing residues H142, S144, 163, and E166. Noticeably, the 11r ligand with ff99sb formed two stable hydrogen bonds, which existed more than 70% of the entire trajectory (Figure 10).

For the simulation of 11r with ff99sb, the hydrogen atom attached at the NE atom of H163 formed hydrogen bonds with 11r for 79% of the trajectory. The hydrogen atom at the N atom of E166 formed hydrogen bonds with 11r for 73%. For the simulation of darunavir with ff99sb, the oxygen atom at H164 and the hydrogen atom at the N atom of E166 also formed the hydrogen bonds with darunavir with 84% and 66%, respectively. For the case of 13a with c36, which is the strongest binder with c36, the hydrogen atom at N atom of E166 also existed for 61%, while other hydrogen bonds existed lower than 10% of the trajectory. Except for the ligands discussed above (11r, 13a, and darunavir), the other ligands showed that the M^{pro}–ligand complexes formed hydrogen bonds under 30% with both ff99sb and c36 (Table 2). For example, X0749, which is the weakest binder with ff99sb, showed that all hydrogen bonds lasted lower than 10%. These results show that hydrogen bonds play important roles in strong binding affinity between M^{pro}–ligand complexes.
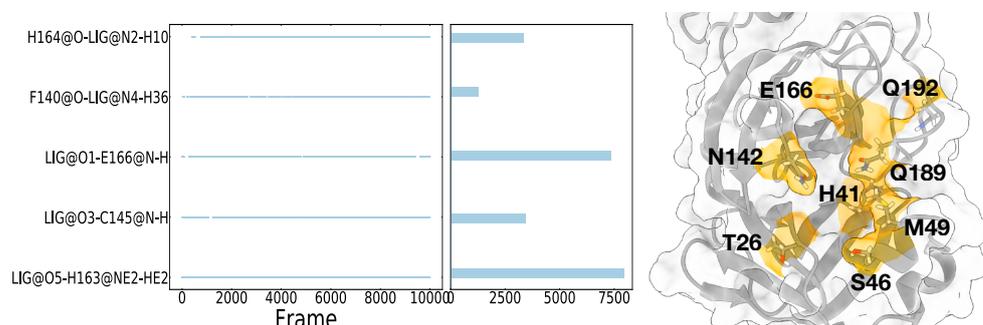
Figure 10. The hydrogen bond analysis of 11r, which was strongest binder with ff99sb at left. The residues that were frequently captured in hydrogen bond analysis at right.

The residues consisting of the subsites formed diverse hydrogen bonds during MD simulation with all of ligands (Tables 3 and 4). N142, E166, and Q189 formed highly stable hydrogen bonds with the ligands. The atoms that were frequently involved in forming hydrogen bonds were similar both with ff99sb and c36. Based on these results, the stable hydrogen bonds appeared to be independent of the force field. Most stable hydrogen bonds

were formed by the atoms in side chain. Except for the H-N atom in E166, all the frequencies of each backbone atom involved in hydrogen bonds were less than 10 times. This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, and the experimental conclusions that can be drawn.

Table 2. The number of hydrogen bonds of each ligand, which satisfies the thresholds with both ff99sb and c36. The total 10,000 number of frames was used to perform hydrogen bond analysis.

Threshold	0.1		0.2		0.3		0.5		0.7	
Ligand	ff99sb	c36	ff99sb	c36	ff99sb	c36	ff99sb	c36	ff99sb	c36
X0072	0	0	0	0	0	0	0	0	0	0
X0689	2	2	1	2	1	1	0	1	0	1
X0691	1	0	0	0	0	0	0	0	0	0
X0749	0	0	0	0	0	0	0	0	0	0
X1336	3	1	1	0	0	0	0	0	0	0
11r	7	2	4	1	4	1	2	0	2	0
13a	6	1	6	1	2	1	0	1	0	0
13b	2	3	0	1	0	1	0	0	0	0
darunavir	4	4	3	3	3	2	3	1	1	0
lopinavir	5	2	1	1	1	0	0	0	0	0
indinavir	3	1	2	1	1	0	1	0	0	0

Table 3. The M^{pro} backbone atoms on residues that frequently formed hydrogen bonds with ff99sb and c36 during simulations. The atoms were based on the hydrogen bonds in the top 5 existence percentage of each ligand MD simulation by hydrogen bond analysis. The number in parentheses indicates how many atoms formed hydrogen bonds in all ligand cases.

Residue	Backbone	
Force Field	ff99sb	c36
T26	H-N(5), O(2)	H-N(7)
H41	O(1)	O(1)
S46	O(3), H-N(5)	O(1), H-N(5)
N142	H-N(1), O(1)	H-N(1)
E166	H-N(10), O(4)	H-N(7)
Q189	O(3), H-N(2)	O(3), H-N(1)

Table 4. The M^{pro} side chain atoms on residues that frequently formed hydrogen bonds with ff99sb and c36 during simulations. The atoms were based on the hydrogen bonds in the top 5 existence percentage of each ligand MD simulation by hydrogen bond analysis. The number in parentheses means that how many atoms formed hydrogen bonds in all ligand cases.

Residue	Side Chain	
Force Field	ff99sb	c36
T26	HG1-OG1(2), OG1(1)	HG1-OG1(3)
H41	HE2-NE2(7), ND1(2), NE2(1)	HD1-ND1(5), NE2(2)
S46	HG-OG(10), OG(7)	HG-OG(7), OG(4)
N142	HD21-ND2(10), HD22-ND2(8), OD1(12), ND2(3)	HD22-ND2(8), HD21-ND2(7), OD1(6), ND2(3)
E166	OE1(6), OE2(7)	OE1(6), OE2(4)
Q189	HE22-NE2(10), HE21-NE2(10), OE1(1), NE2(1)	HE21-NE2(11), HE22-NE2(11), OE1(15), NE2(1)

4. Discussion

The ligand binding behaviors depended on which functional group (named branch in our manuscript) is interacting with the binding site and how a ligand posed in the binding site. Thus, we investigated their binding behaviors where the branches of ligands were located at the binding site and how they fluctuated during the MD simulations. In terms of binding pose, the 11r, 13a, and 13b ligands' branches are located at different subsites for 11r, 13a, and 13b (Figures 3, S15 and S16).

First, the P1 branch, which is a five-membered ring, is a common functional group for 11r, 13a, and 13b. The P1 branch of 11r and 13a consistently occupied the S1 subsite with both ff99sb and c36 (blue circle in Figure 3a,b). With the ff99sb force field, the RMSF values of the P1 branch of 11r and 13a are 11.1 Å and 10.7 Å, respectively. However, the P1 of 13b did not occupy the S1 subsite with an RMSF value of 14.9 Å. These results suggest that the occupancy of the S1 subsite of the P1 branch may play an important role in M^{pro}–ligand binding.

Second, the P3 branch of 13a (red circle in Figure 3b, RMSF; 15.9 Å with ff99sb and 14.6 Å with c36, respectively) fluctuated more than the P1' branch of 11r (yellow circle in Figure 3a, RMSF; 13.2 Å with ff99sb and 15.4 Å with c36, respectively) at the S1' subsite.

Additionally, the exposed P3 branch of 13b (red circle in Figure 3c) fluctuated most during simulations in terms of RMSF of 18.6 Å with ff99sb and 15.2 Å with c36, respectively. Although the different branch are located at the S1' subsite, the RMSF values indicate that occupancy of the S1' subsite is less important in ligand binding.

Furthermore, hydrogen bond patterns between the ligands and the binding site residues also showed different patterns by ligands. With ff99sb, 13b formed hydrogen bonds with M^{pro} less frequently than 11r and 13a (Figure S12). 11r formed hydrogen bonds with H163 and E166 for more than 75% of the trajectory (Figure S12f). 13a formed hydrogen bonds with N142 and S144 for more than 40% (Figure S12g). However, 13b did not form any hydrogen bond with a frequency of more than 40% (Figure S12h). The maximum frequency of hydrogen bond of 13b is about 20%, which is less than the halves of those of 11r and 13a. These indicate that 11r and 13a stay longer in a specific binding pose or location than 13b because the corresponding binding pose is more stable than the others. This pattern is also consistent with the MM-PBSA results; ΔG of 11r and 13a are lower than that of 13b with ff99sb in Figure 8.

In addition to the hydrogen bonds mentioned above, most other hydrogen bonds of 11r and 13a formed for more than 20 % of the trajectory, whereas 13b did not. This suggests that 11r and 13a strongly interact with residues at the S1 (F140, N142, C145, H163, E166, and H172) and the S4 (L167, F185, Q189, Q192) subsite (Figure S12f,g). Thus, this supports the fact that MM-PBSA analysis can identify the hot spot residues discussed in Section 3.6. Moreover, only 13b formed hydrogen bonds with the S46 and the T26 residues, which were not observed in the hydrogen bond patterns of 11r and 13a. T26 and S46 residues are located at the S1' and S2 subsites, respectively. It suggests that the 13b can more interact with residues in S1' andS2 subsite, resulting in different ligand behaviors compared to 11r and 13a. This result is also consistent with the c36 force field results. 13a, the strongest binder with c36, showed the most stable hydrogen bond among 11r, 13a, and 13b. Furthermore, 13b more frequently formed a hydrogen bond with T26 in Figure S13.

In this study, we utilized the MM-PBSA approach to investigate the hot spot residues of SARS-CoV-2 M^{pro} with three different force fields: ff99sb, c36, and g54a7. There was no study that performed a comparison between different force fields. Here, we used the force fields ff99sb, c36, and g54a7 and compared their MM-PBSA analysis results. If the MM-PBSA methodology and all three force fields are well matured, all binding affinity estimates would have been be similar and show a reasonable correlation.

However, the estimated absolute binding energies of M^{pro} and known ligands calculated with MM-PBSA were not consistent between three different force fields. By comparing the MM-PBSA results from the tested force fields, we found that only the ff99sb and g54a7 results have a weak correlation. On the contrary, the c36 results showed almost no corre-

lation with the g54a7 and ff99sb results. The binding energies from c36 were consistently estimated higher than those of both ff99sb and g54a7. These results suggest that MM-PBSA results heavily depend on the choice of force field and that no single force field is able to predict the correct relative ranking of experimental binding affinities of the M^{pro} complexes. In other words, one should be careful in comparing binding free energy estimates obtained with different force fields and the MM-PBSA method, suggesting that the accuracy of MM-PBSA calculation is limited. From a practical perspective of drug design, our results indicate that the MM-PBSA analysis should be used only in the early stage of drug discovery and more rigorous absolute binding free energy methods, such as alchemical relative/absolute binding free energy calculations [26,71], should be used in later stages for more accurate lead optimization.

Based on the residuewise decomposed energies, the terminal residues (D295, R298, and Q299) were identified to contribute strongly to ligand binding. Because they were far from the known binding site, they were not considered as a residue that contributes the binding energy at first glance. However, the terminal residues had averaged decomposed energy in our ligands: −4.065, −2.582, and −1.383 kcal/mol, respectively, which has been suggested by the other study that the terminal residues were important in ligand binding [70]. The decomposed residue energies showed that the residues of the S1 or S4 subsites contributed more strongly to binding energy than the residues in the S1′ or S2 subsites, supporting that the ligand's atoms or branches located in S1′s subsite fluctuated more freely than those residues in the S1 or S4 subsites.

5. Conclusions

In growing threat of COVID-19, SARS-CoV-2 main protease (M^{pro}) is one of the potential pharmaceutical target protein because M^{pro} enzyme has no human analogue and is conserved among corona viruses [5].

There have been computational studies to predict the binding affinities of M^{pro}–ligand complexes [13,28,29]. Previous studies only performed MM-GBSA [13,29] or investigated only two ligands [28], using only one force field, which may affect the prediction quality heavily. Compared to previous computational studies, we performed MD simulations of 11 M^{pro}–ligand complexes with three different force fields and followed by the MM-PBSA analysis, which yields more accurate binding free energy estimates in general than MM-GBSA. There was no study that performed a comparison between different force fields. Here, we used the force fields ff99sb, c36, and g54a7 and compared their MM-PBSA analysis results. If the MM-PBSA methodology and all three force fields are well matured, all binding affinity estimates would have been be similar and show a reasonable correlation.

In this study, we investigated the binding affinity of SARS-CoV-2 M^{pro} with its potential inhibitors to offer insight about M^{pro}–ligand binding and accelerate M^{pro} drug discovery. We benchmarked the absolute binding free energy estimated obtained with three widely used force fields, ff99sb, c36, and g54a7. Unexpectedly, we identified that MM-PBSA results with different force fields are not consistent with each other. Only the results with ff99sb and g54a7 are weakly correlated. These results suggests that MM-PBSA results highly depend on the choice of force field. This also shows that one should be careful in comparing binding free energy estimates of M^{pro}–ligand complexes obtained with different force fields and the MM-PBSA method. Through energy decomposition analysis, we found that the residues of M^{pro} at the S1, S4 subsites or terminal regions play essential roles in ligand binding. This suggests that these residues may play important roles in developing drugs. We hope that our results could provide insight into the design and development of more potency inhibitor against SARS-CoV-2 M^{pro} and facilitate the drug development for the treatment of COVID-19.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/life12010054/s1, Figure S1: The used ligands in this work. Figure S2: The partial charge represented ligand figures generated by gaff. Ligands are x0072, x0689, x0691, x0749, and x1336. Figure S3: The partial charge represented ligand figures generated by gaff. Ligands are 11r, 13a, 13b, darunavir, lopinavir, and indinavir. Figure S4: The partial charge represented ligand figures generated by charmm-gui. Ligands are x0072, x0689, x0691, x0749, and x1336. Figure S5: The partial charge represented ligand figures generated by charmm-gui. Ligands are 11r, 13a, 13b, darunavir, lopinavir, and indinavir. Figure S6: MM-PBSA decomposed energy per-reisdue with ff99SB (a) X0072 (b) X0689 (c) X0691 (d) X0749 (e) X1336. Figure S7: MM-PBSA decomposed energy per-residue with ff99SB (a) 11r (b) 13a (c) 13b (d) darunavir (e) lopinavir (f) indinavir. Figure S8: MM-PBSA decomposed energy per-residue with c36 (a) X0072 (b) X0689 (c) X0691 (d) X0749 (e) X1336. Figure S9: MM-PBSA decomposed energy per-residue with c36 (a) 11r (b) 13a (c) 13b (d) darunavir (e) lopinavir (f) indinavir. Figure S10: MM-PBSA Decomposed energy per-residue with g54a7 (a) X0072 (b) X06889 (c) X0691 (d) X0749 (e) X1336. Figure S11: MM-PBSA Decomposed energy per-residue with g54a7 (a) 11r (b) 13a (c) 13b (d) darunavir (e) lopinavir (f) indinavir. Figure S12: Hydrogen bond existence map with ff99SB. The scatter plots show a formed hydrogen bond at corresponding frame. Figure S13: Hydrogen bond existence map with c36. The scatter plots show a formed hydrogen bond at corresponding frame. Figure S14: Hydrogen bond existene map with g54a7. The scatter plots show a formed hydrogen bond at corresponding frame. Figure S15: 3D histogram from the production trajectories with ff99sb. Figure S16: 3D histogram from the production trajectories with c36. Figure S17: 3D histogram from the production trajectories with g54a7. Figure S18: The protein-ligand interaction plot generated by ligPlot+ program with ff99sb force field. Figure S19: The protein-ligand interaction plot generated by ligPlot+ program with c36 force field. Table S1: Each energetic components of ΔG component (kcal/mol) of each ligand-M pro calculated by the MM-PBSA analysis.

Author Contributions: J.B. and J.L. conceived, designed, performed, and analyzed the experiments and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FDA	Food and Drug Administration
3CLP	3C-like proteinase
MD simulation	Molecular Dynamics simulation
MM-PBSA	Molecular Mechanics Poisson–Boltzmann Surface Area
ff14sb	Amber14sb force field
c36	Charmm36 force field
g54a7	Gromos54a7 force field

References

1. The Johns Hopkins Coronavirus Resource Center. 2021. Available online: https://coronavirus.jhu.edu/ (accessed on 1 November 2021).
2. Ioannidis, J.P. Global perspective of COVID-19 epidemiology for a full-cycle pandemic. *Eur. J. Clin. Investig.* **2020**, *50*, e13423. [CrossRef] [PubMed]
3. Altay, O.; Mohammadi, E.; Lam, S.; Turkez, H.; Boren, J.; Nielsen, J.; Uhlen, M.; Mardinoglu, A. Current Status of COVID-19 Therapies and Drug Repositioning Applications. *iScience* **2020**, *23*, 101303. [CrossRef]

4. Hughes, D.L. Quest for a Cure: Potential Small-Molecule Treatments for COVID-19, Part 2. *Org. Process. Res. Dev.* **2021**, *25*, 1089–1111. [CrossRef]

5. Fauquet, C.; Fargette, D. International Committee on Taxonomy of Viruses and the 3142 unassigned species. *Virol. J.* **2005**, *2*, 1–10.

6. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293. [CrossRef] [PubMed]

7. Ullrich, S.; Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorganic Med. Chem. Lett.* **2020**, *30*, 127377. [CrossRef]

8. Mengist, H.M.; Dilnessa, T.; Jin, T. Structural basis of potential inhibitors targeting SARS-CoV-2 main protease. *Front. Chem.* **2021**, *9*, 622898. [CrossRef] [PubMed]

9. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020**, *368*, 409–412. [CrossRef]

10. Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C.D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Ábrányi-Balogh, P.; et al. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **2020**, *11*, 1–11. [CrossRef]

11. Jin, Z.; Zhao, Y.; Sun, Y.; Zhang, B.; Wang, H.; Wu, Y.; Zhu, Y.; Zhu, C.; Hu, T.; Du, X.; et al. Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nat. Struct. Mol. Biol.* **2020**, *27*, 529–532. [CrossRef] [PubMed]

12. Bhardwaj, V.K.; Singh, R.; Das, P.; Purohit, R. Evaluation of acridinedione analogs as potential SARS-CoV-2 main protease inhibitors and their comparison with repurposed anti-viral drugs. *Comput. Biol. Med.* **2021**, *128*, 104117. [CrossRef]

13. Alamri, M.A.; Tahir Ul Qamar, M.; Mirza, M.U.; Bhadane, R.; Alqahtani, S.M.; Muneer, I.; Froeyen, M.; Salo-Ahen, O.M. Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro. *J. Biomol. Struct. Dyn.* **2020**, *39*, 4936–4948. [CrossRef]

14. Peele, K.A.; Durthi, C.P.; Srihansa, T.; Krupanidhi, S.; Ayyagari, V.S.; Babu, D.J.; Indira, M.; Reddy, A.R.; Venkateswarulu, T. Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: A computational study. *INformatics Med. Unlocked* **2020**, *19*, 100345. [CrossRef]

15. Zhu, K.; Borrelli, K.W.; Greenwood, J.R.; Day, T.; Abel, R.; Farid, R.S.; Harder, E. Docking covalent inhibitors: A parameter free approach to pose prediction and scoring. *J. Chem. Inf. Model.* **2014**, *54*, 1932–1940. [CrossRef] [PubMed]

16. Jarzynski, C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E* **1997**, *56*, 5018. [CrossRef]

17. Macchiagodena, M.; Pagliai, M.; Karrenbrock, M.; Guarnieri, G.; Iannone, F.; Procacci, P. Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. *J. Chem. Theory Comput.* **2020**, *16*, 7160–7172. [CrossRef] [PubMed]

18. Zwanzig, R.W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426. [CrossRef]

19. Beveridge, D.L.; DiCapua, F. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492. [CrossRef]

20. Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417. [CrossRef]

21. Heinzelmann, G.; Chen, P.C.; Kuyucak, S. Computation of standard binding free energies of polar and charged ligands to the glutamate receptor GluA2. *J. Phys. Chem. B* **2014**, *118*, 1813–1824. [CrossRef]

22. Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S.B.; Tinoco, I.; Bustamante, C. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* **2005**, *437*, 231–234. [CrossRef] [PubMed]

23. Srivastava, H.K.; Sastry, G.N. Molecular dynamics investigation on a series of HIV protease inhibitors: Assessing the performance of MM-PBSA and MM-GBSA approaches. *J. Chem. Inf. Model.* **2012**, *52*, 3088–3098. [CrossRef] [PubMed]

24. Williams, A.H.; Zhan, C.G. Fast Prediction of Binding Affinities of the SARS-CoV-2 Spike Protein Mutant N501Y (UK Variant) with ACE2 and Miniprotein Drug Candidates. *J. Phys. Chem. B* **2021**, *125*, 4330–4336. [CrossRef] [PubMed]

25. King, E.; Qi, R.; Li, H.; Luo, R.; Aitchison, E. Estimating the Roles of Protonation and Electronic Polarization in Absolute Binding Affinity Simulations. *J. Chem. Theory Comput.* **2021**, *17*, 2541–2555. [CrossRef]

26. Ngo, S.T.; Quynh Anh Pham, N.; Thi Le, L.; Pham, D.H.; Vu, V.V. Computational determination of potential inhibitors of SARS-CoV-2 main protease. *J. Chem. Inf. Model.* **2020**, *60*, 5771–5780. [CrossRef]

27. Sk, M.F.; Roy, R.; Jonniya, N.A.; Poddar, S.; Kar, P. Elucidating biophysical basis of binding of inhibitors to SARS-CoV-2 main protease by using molecular dynamics simulations and free energy calculations. *J. Biomol. Struct. Dyn.* **2020**, *39*, 3649–3661. [CrossRef]

28. Nutho, B.; Mahalapbutr, P.; Hengphasatporn, K.; Pattaranggoon, N.C.; Simanon, N.; Shigeta, Y.; Hannongbua, S.; Rungrotmongkol, T. Why are lopinavir and ritonavir effective against the newly emerged coronavirus 2019? Atomistic insights into the inhibitory mechanisms. *Biochemistry* **2020**, *59*, 1769–1779. [CrossRef]

29. Wang, J. Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study. *J. Chem. Inf. Model.* **2020**, *60*, 3277–3286. [CrossRef]

30. Poger, D.; Van Gunsteren, W.F.; Mark, A.E. A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem.* **2010**, *31*, 1117–1125. [CrossRef]

31. Schmid, N.; Eichenberger, A.P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A.E.; van Gunsteren, W.F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856. [CrossRef]

32. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinform.* **2006**, *65*, 712–725. [CrossRef]

33. Huang, J.; MacKerell, A.D., Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135–2145. [CrossRef] [PubMed]

34. Yang, J.; Baek, M.; Seok, C. GalaxyDock3: Protein–ligand docking that considers the full ligand conformational flexibility. *J. Comput. Chem.* **2019**, *40*, 2739–2748. [CrossRef]

35. Gasteiger, J.; Marsili, M. A new model for calculating atomic charges in molecules. *Tetrahedron Lett.* **1978**, *19*, 3181–3184. [CrossRef]

36. Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228. [CrossRef]

37. Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; de Groot, B.L.; Grubmüller, H. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J. Comput. Chem.* **2019**, *40*, 2418–2431. [CrossRef] [PubMed]

38. Lee, T.S.; Cerutti, D.S.; Mermelstein, D.; Lin, C.; Legrand, S.; Giese, T.J.; Roitberg, A.; Case, D.A.; Walker, R.C.; York, D.M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58*, 2043–2050. [CrossRef] [PubMed]

39. Malde, A.K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P.C.; Oostenbrink, C.; Mark, A.E. An automated force field topology builder (ATB) and repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026–4037. [CrossRef]

40. Berendsen, H.J.; Postma, J.P.; van Gunsteren, W.F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Springer: Berlin/Heidelberg, Germany, 1981; pp. 331–342.

41. Hess, B.; Bekker, H.; Berendsen, H.J.; Fraaije, J.G. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472. [CrossRef]

42. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]

43. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [CrossRef]

44. Hoover, W.G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695. [CrossRef]

45. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190. [CrossRef]

46. Salomon-Ferrer, R.; Case, D.A.; Walker, R.C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198–210. [CrossRef]

47. Wang, J.; Wang, W.; Kollman, P.A.; Case, D.A. Antechamber: An accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* **2001**, *222*, U403.

48. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [CrossRef] [PubMed]

49. Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phyics* **1977**, *23*, 327–341. [CrossRef]

50. Darden, T.; Perera, L.; Li, L.; Pedersen, L. New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **1999**, *7*, R55–R60. [CrossRef]

51. Uberuaga, B.P.; Anghel, M.; Voter, A.F. Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation. *J. Chem. Phys.* **2004**, *120*, 6363–6374. [CrossRef]

52. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865. [CrossRef]

53. Vanommeslaeghe, K.; MacKerell, A.D., Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154. [CrossRef]

54. Vanommeslaeghe, K.; Raman, E.P.; MacKerell, A.D., Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168. [CrossRef]

55. Kumari, R.; Kumar, R.; Consortium, O.S.D.D.; Lynn, A. g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951–1962. [CrossRef]

56. Miller, B.R., III; McGee, T.D., Jr.; Swails, J.M.; Homeyer, N.; Gohlke, H.; Roitberg, A.E. MMPBSA. py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321. [CrossRef] [PubMed]

57. Sitkoff, D.; Sharp, K.A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978–1988. [CrossRef]

58. Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129. [CrossRef]

59. Kuhn, B.; Kollman, P.A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786–3791. [CrossRef]

60. Weiser, J.; Shenkin, P.S.; Still, W.C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217–230. [CrossRef]

61. Tsui, V.; Case, D.A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolym. Orig. Res. Biomol.* **2000**, *56*, 275–291. [CrossRef]

62. Roe, D.R.; Cheatham, T.E., III. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *JOurnal Chem. Theory Comput.* **2013**, *9*, 3084–3095. [CrossRef]

63. Zhang, C.H.; Stone, E.A.; Deshmukh, M.; Ippolito, J.A.; Ghahremanpour, M.M.; Tirado-Rives, J.; Spasov, K.A.; Zhang, S.; Takeo, Y.; Kudalkar, S.N.; et al. Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent. Sci.* **2021**, *7*, 467–475. [CrossRef]

64. Mahdi, M.; Mótyán, J.A.; Szojka, Z.I.; Golda, M.; Miczi, M.; Tőzsér, J. Analysis of the efficacy of HIV protease inhibitors against SARS-CoV-2 main protease. *Virol. J.* **2020**, *17*, 1–8. [CrossRef]

65. He, X.; Liu, S.; Lee, T.S.; Ji, B.; Man, V.H.; York, D.M.; Wang, J. Fast, accurate, and reliable protocols for routine calculations of protein–ligand binding affinities in drug design projects using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **2020**, *5*, 4611–4619. [CrossRef]

66. Lee, J.; Tofoleanu, F.; Pickard, F.C.; König, G.; Huang, J.; Damjanović, A.; Baek, M.; Seok, C.; Brooks, B.R. Absolute binding free energy calculations of CBClip host–guest systems in the SAMPL5 blind challenge. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 71–85. [CrossRef]

67. Aldeghi, M.; Heifetz, A.; Bodkin, M.J.; Knapp, S.; Biggin, P.C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **2016**, *7*, 207–218. [CrossRef] [PubMed]

68. Aldeghi, M.; Heifetz, A.; Bodkin, M.J.; Knapp, S.; Biggin, P.C. Predictions of ligand selectivity from absolute binding free energy calculations. *J. Am. Chem. Soc.* **2017**, *139*, 946–957. [CrossRef]

69. Laskowski, R.A.; Swindells, M.B. LigPlot+: Multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **2011**, *51*, 2778–2786. [CrossRef] [PubMed]

70. Sztain, T.; Amaro, R.; McCammon, J.A. Elucidation of cryptic and allosteric pockets within the SARS-CoV-2 protease. *Biorxiv* **2020**. [CrossRef] [PubMed]

71. Ngo, S.T.; Tam, N.M.; Pham, M.Q.; Nguyen, T.H. Benchmark of Popular Free Energy Approaches Revealing the Inhibitors Binding to SARS-CoV-2 Mpro. *J. Chem. Inf. Model.* **2021**, *61*, 2302–2312. [CrossRef] [PubMed]