

Mechanism of ligand recognition by human ACE2 receptor

Apurba Bhattarai¹, Shristi Pawnikar¹ and Yinglong Miao^{1*}

¹Center for Computational Biology and Department of Molecular Biosciences,
University of Kansas, Lawrence, KS 66047, USA

* Corresponding email: miao@ku.edu

Running Title: Ligand recognition by ACE2 receptor

Abstract

Angiotensin converting enzyme 2 (ACE2) plays a key role in renin-angiotensin system regulation and amino acid homeostasis. Human ACE2 acts as the receptor for severe acute respiratory syndrome coronaviruses SARS-CoV and SARS-CoV-2. ACE2 is also widely expressed in epithelial cells of lungs, heart, kidney and pancreas. It is considered an important drug target for treating SARS-CoV-2, as well as pulmonary diseases, heart failure, hypertension, renal diseases and diabetes. Despite the critical importance, the mechanism of ligand binding to the human ACE2 receptor remains unknown. Here, we address this challenge through all-atom simulations using a novel ligand Gaussian accelerated molecular dynamics (LiGaMD) method. Microsecond LiGaMD simulations have successfully captured both binding and unbinding of the MLN-4760 inhibitor in the ACE2 receptor. In the ligand unbound state, the ACE2 receptor samples distinct Open, Partially Open and Closed conformations. Ligand binding biases the receptor conformational ensemble towards the Closed state. The LiGaMD simulations thus suggest a conformational selection mechanism for ligand recognition by the ACE2 receptor. Our simulation findings are expected to facilitate rational drug design of ACE2 against coronaviruses and other related human diseases.

Keywords: Angiotensin converting enzyme 2 (ACE2), coronaviruses, ligand binding, ligand Gaussian accelerated molecular dynamics (LiGaMD), conformational selection.

Introduction

Angiotensin converting enzyme 2 (ACE2) plays a key role in renin-angiotensin system regulation and amino acid homeostasis (Gheblawi et al., 2020;Gross et al., 2020). Human ACE2 acts as the receptor for severe acute respiratory syndrome coronaviruses SARS-CoV and SARS-CoV-2. ACE2 plays a vital role as a catalytic protease converting angiotensin II to angiotensin 1-7 and angiotensin I to angiotensin 1-9. Proteolytic reactions of these peptide hormones aid in the conversion of vasoconstrictors to vasodilators and thus help to maintain blood pressure (Gross et al., 2020). Likewise, ACE2 is widely expressed in epithelial cells of lungs, heart, kidney, and pancreas. It is an important drug target for treating pulmonary diseases, heart failure, hypertension, renal diseases and diabetes (Patel et al., 2016;Yan et al., 2020). ACE2 also stabilizes amino acid transporter B^oAT1 to regulate the gut microbiome and amino acid homeostasis (Perlot and Penninger, 2013).

Human ACE2 has been identified as the functional receptor for severe acute respiratory syndrome coronaviruses including SARS-CoV and SARS-CoV-2 (Hoffmann et al., 2020). SARS-CoV-2 is responsible for 2019 coronavirus pandemic (COVID-19). By mid-October 2020, ~40 million people have been infected by COVID-19 with ~1.1 million deaths around the world. With the unprecedented pandemic, it is of paramount importance to investigate virus infection and develop effective treatments of SARS-CoV-2. The entry of SARS-CoV-2 is mediated by interaction of the receptor binding domain (RBD) in the virus spike protein S1 subunit with the host ACE2 receptor. The transmembrane protease serine 2 (TMPRSS2) promotes priming of the spike protein and facilitates its S2 subunit to initiate the viral-cell membrane fusion. Hence, inhibiting the interaction between the viral RBD and host ACE2 presents a promising strategy for blocking SAR-COV-2 entry to the human cells.

ACE2 consists of the N-terminal catalytic peptidase domain (PD) on the extracellular side and the C-terminal transmembrane collectrin-like domain (CLD) with a cytoplasmic tail on the intracellular side. The enzyme PD domain can be inhibited by compounds like MLN-4760 (Towler et al., 2004), which bind to the protein active site and prevent substrate binding. MLN-4760 binding biases the receptor to adopt a “Closed” conformation, in which the protein active site formed by two subdomains of the PD is closed from the external environment (**Figure 1A**). Furthermore, the receptor undergoes conformational changes with hinge-bending movement of the dynamic N-terminal subdomain I relative to the stable subdomain II, e.g., $\sim 16^\circ$ bending upon inhibitor binding (Towler et al., 2004). In the absence of ligand binding, the two subdomains move apart from each other and the protein active site becomes exposed to solvent in an “Open” conformation (**Figure 1A**). In complex with RBD of the SARS-CoV or SARS-CoV-2, the ACE2 receptor also adopts a “Partially Open” conformation, in which the subdomain I lies between the “Open” and “Closed” conformations (Gross et al., 2020) (**Figure 1A**). Among over 20 experimental structures of the ACE2 receptor present in Protein Data Bank (PDB), most of them exhibit “Open” and “Partially Open” conformations but only one structure has been identified in the “Closed” conformation (PDB: 1R4L) (Towler et al., 2004). Despite tremendous efforts to determine these experimental structures (Huentelman et al., 2004; Towler et al., 2004; Li et al., 2005; Hernández Prada et al., 2008; Song et al., 2018; Shang et al., 2020; Wang et al., 2020; Yan et al., 2020), the dynamics and functional mechanism of the ACE receptor are still poorly understood (Nami et al., 2020).

MLN-4760 is a highly selective and potent (IC_{50} : 0.44 nM) small-molecule inhibitor of the ACE2 receptor (Dales et al., 2002). The inhibitor has two carboxylate groups contributing to -2 net charge of the molecule (**Figure S1**). One of the negatively charged carboxylate groups interacts

with the positively charged zinc ion, by which the ACE2 receptor functions as a metallopeptidase enzyme. Depending on ligand or viral RBD binding, the receptor adopts different conformations, but the pathways and mechanism of ligand binding in the ACE2 receptor remain unknown. In the context of SARS-CoV-2 and many other medical implications, it is important to understand the mechanism of ligand recognition by the ACE2 receptor in order to design effective drugs against the virus.

Ligand Gaussian accelerated molecular dynamics (LiGaMD) (Miao et al., 2020) is an enhanced sampling computational technique that is developed to efficiently simulate both binding and unbinding of ligand molecules. It is based on Gaussian accelerated molecular dynamics (GaMD), which works by adding a harmonic boost potential to smooth the biomolecular potential energy surface (Miao et al., 2015). GaMD greatly reduces energy barriers and accelerates biomolecular simulations by orders of magnitude (Miao, 2018). GaMD provides unconstrained enhanced sampling without the requirement of pre-defined collective variables or reaction coordinates. Moreover, because the boost potential exhibits a Gaussian distribution, biomolecular free energy profiles can be properly recovered through cumulant expansion to the second order (Miao et al., 2015). In LiGaMD (Miao et al., 2020), the ligand non-bonded interaction potential energy is selectively boosted to enable ligand dissociation. Another boost potential is applied to the remaining potential energy of the entire system in a dual-boost algorithm to facilitate ligand rebinding. LiGaMD has been demonstrated on host-guest and protein-ligand binding model systems. LiGaMD allows us to capture repetitive ligand binding and unbinding, and thus characterize both ligand thermodynamics and kinetics simultaneously. The calculated ligand binding free energy and kinetic rate constants compared very well with experimental data (Miao et al., 2020).

Here, we have applied all-atom LiGaMD simulations to investigate binding and unbinding of the MLN-4760 inhibitor and associated conformational changes of the ACE2 receptor. Microsecond LiGaMD simulations were able to capture both ligand binding and unbinding. The ACE2 receptor sampled Open, Partially Open and Closed conformational states, being consistent with previous experimental structures. Ligand binding could bias the receptor conformational ensemble to the Closed state, suggesting a conformational selection mechanism rather than induced fit. In summary, the LiGaMD simulations allowed us to understand the mechanism of ligand recognition by the ACE2 receptor.

Materials and Methods

Ligand Gaussian accelerated molecular dynamics (LiGaMD)

LiGaMD is an enhanced sampling computational technique that is developed to efficiently simulate ligand binding and unbinding based on the previous GaMD (Miao et al., 2015). Details of the LiGaMD method has been described in the previous study (Miao et al., 2020). A brief summary will be provided here.

We consider a system of ligand L binding to a protein P in a biological environment E . The system comprises of N atoms with their coordinates $r \equiv \{\vec{r}_1, \dots, \vec{r}_N\}$ and momenta $p \equiv \{\vec{p}_1, \dots, \vec{p}_N\}$. The system Hamiltonian can be expressed as:

$$H(r, p) = K(p) + V(r), \quad (1)$$

where $K(p)$ and $V(r)$ are the system kinetic and total potential energies, respectively. Next, we decompose the potential energy into the following terms:

$$\begin{aligned}
 V(r) = & V_{P,b}(r_P) + V_{L,b}(r_L) + V_{E,b}(r_E) \\
 & + V_{PP,nb}(r_P) + V_{LL,nb}(r_L) + V_{EE,nb}(r_E) \\
 & + V_{PL,nb}(r_{PL}) + V_{PE,nb}(r_{PE}) + V_{LE,nb}(r_{LE}). \tag{2}
 \end{aligned}$$

where $V_{P,b}$, $V_{L,b}$ and $V_{E,b}$ are the bonded potential energies in protein P , ligand L and environment E , respectively. $V_{PP,nb}$, $V_{LL,nb}$ and $V_{EE,nb}$ are the self non-bonded potential energies in protein P , ligand L and environment E , respectively. $V_{PL,nb}$, $V_{PE,nb}$ and $V_{LE,nb}$ are the non-bonded interaction energies between P - L , P - E and L - E , respectively. According to classical molecular mechanics force fields (Cornell et al., 1996; Duan et al., 2003; Vanommeslaeghe et al., 2010; Vanommeslaeghe and MacKerell, 2014), the non-bonded potential energies are usually calculated as:

$$V_{nb} = V_{elec} + V_{vdW}. \tag{3}$$

Where V_{elec} and V_{vdW} are the system electrostatic and van der Waals potential energies. Presumably, ligand binding mainly involves the non-bonded interaction energies of the ligand, $V_{L,nb}(r) = V_{LL,nb}(r_L) + V_{PL,nb}(r_{PL}) + V_{LE,nb}(r_{LE})$. In LiGaMD, we add boost potential selectively to the essential ligand non-bonded potential energy according to the GaMD algorithm:

$$\Delta V_{L,nb}(r) = \begin{cases} \frac{1}{2} k_{L,nb} (E_{L,nb} - V_{L,nb}(r))^2, & V_{L,nb}(r) < E_{L,nb} \\ 0, & V_{L,nb}(r) \geq E_{L,nb} \end{cases} \tag{4}$$

where $E_{L,nb}$ is the threshold energy for applying boost potential and $k_{L,nb}$ is the harmonic constant. For simplicity, the subscript of $\Delta V_{L,nb}(r)$, $E_{L,nb}$ and $k_{L,nb}$ is dropped in the following. When E is set to the lower bound $E=V_{max}$, k_0 can be calculated as:

$$k_0 = \min(1.0, k'_0) = \min\left(1.0, \frac{\sigma_0 V_{max} - V_{min}}{\sigma_V V_{max} - V_{avg}}\right). \tag{5}$$

Alternatively, when the threshold energy E is set to its upper bound $E = V_{min} + \frac{1}{k}$, k_0 is set to:

$$k_0 = k_0'' \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right) \frac{V_{max} - V_{min}}{V_{avg} - V_{min}}, \quad (6)$$

if k_0'' is found to be between 0 and 1. Otherwise, k_0 is calculated using Eqn. (5).

Next, one can add multiple ligand molecules in the solvent to facilitate ligand binding to proteins in MD simulations (Dror et al., 2011; Miao et al., 2018). This is based on the fact that the ligand binding rate constant k_{on} is inversely proportional to the ligand concentration. The higher the ligand concentration, the faster the ligand binds, provided that the ligand concentration is still within its solubility limit. In addition to selectively boosting the bound ligand, another boost potential could thus be applied on the unbound ligand molecules, protein and solvent to facilitate both ligand dissociation and rebinding. The second boost potential is calculated using the total system potential energy other than the non-bonded potential energy of the bound ligand in Eqn. (2) as:

$$\begin{aligned} V_D(r) &= V(r) - V_{L,nb}(r) \\ &= V_{P,b}(r_P) + V_{L,b}(r_L) + V_{E,b}(r_E) \\ &\quad + V_{PP,nb}(r_P) + V_{EE,nb}(r_E) + V_{PE,nb}(r_{PE}). \end{aligned} \quad (7)$$

$$\Delta V_D(r) = \begin{cases} \frac{1}{2} k_D (E_D - V_D(r))^2, & V_D(r) < E_D \\ 0, & V_D(r) \geq E_D \end{cases} \quad (8)$$

where E_D and k_D are the corresponding threshold energy for applying the second boost potential and the harmonic constant, respectively. This leads to dual-boost LiGaMD (LiGaMD_Dual) with the total boost potential $\Delta V(r) = \Delta V_{L,nb}(r) + \Delta V_D(r)$.

Energetic Reweighting

To calculate potential of mean force (PMF) (Roux, 1995) from LiGaMD simulations, the probability distribution along a reaction coordinate is written as $p^*(A)$. Given the boost potential $\Delta V(\vec{r})$ of each frame, $p^*(A)$ can be reweighted to recover the canonical ensemble distribution, $p(A)$, as:

$$p(A_j) = p^*(A_j) \frac{\langle e^{\beta \Delta V(\vec{r})} \rangle_j}{\sum_{i=1}^M \langle p^*(A_i) e^{\beta \Delta V(\vec{r})} \rangle_i}, \quad j = 1, \dots, M, \quad (9)$$

where M is the number of bins, $\beta = k_B T$ and $\langle e^{\beta \Delta V(\vec{r})} \rangle_j$ is the ensemble-averaged Boltzmann factor of $\Delta V(\vec{r})$ for simulation frames found in the j^{th} bin. The ensemble-averaged reweighting factor can be approximated using cumulant expansion:

$$\langle e^{\beta \Delta V(\vec{r})} \rangle = \exp \left\{ \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k \right\}, \quad (10)$$

where the first two cumulants are given by

$$\begin{aligned} C_1 &= \langle \Delta V \rangle, \\ C_2 &= \langle \Delta V^2 \rangle - \langle \Delta V \rangle^2 = \sigma_v^2. \end{aligned} \quad (11)$$

The boost potential obtained from LiGaMD simulations usually follows near-Gaussian distribution. Cumulant expansion to the second order thus provides a good approximation for computing the reweighting factor (Miao et al., 2014; Miao et al., 2015). The reweighted free energy $F(A) = -k_B T \ln p(A)$ is calculated as:

$$F(A) = F^*(A) - \sum_{k=1}^2 \frac{\beta^k}{k!} C_k + F_c, \quad (12)$$

where $F^*(A) = -k_B T \ln p^*(A)$ is the modified free energy obtained from LiGaMD simulation and F_c is a constant.

Simulations of ligand binding in human ACE2 receptor

LiGaMD simulations using the dual-boost scheme were performed on ligand binding to the ACE2 receptor. The X-ray crystal structure of the MLN-4760 inhibitor-bound ACE2 (PDB: 1R4L (Towler et al., 2004)) was used to prepare the simulation system (**Figure S1**). The N- and C-termini of the peptides were capped with the acetyl (ACE) and N-methyl amide (NME) neutral groups, respectively. The missing hydrogen atoms were added using the *tleap* module in AMBER (Case et al., 2005a). The AMBER ff19SB force field (Tian et al., 2019) were used for the protein. The MLN-4760 ligand was modeled with GAFF-2 (Case et al., 2020) force field. Atomic partial charges of ligand were obtained through B3LYP/6-31G* quantum calculations of the electrostatic potential, for which the charges were fitted using the antechamber program (Case et al., 2005a; Case et al., 2005b). Each system was neutralized by adding counter ions and immersed in a cubic TIP3P (Jorgensen et al., 1983) water box, which was extended 10 Å from the receptor surface. A total of 10 ligand molecules (one in the X-ray bound conformation and another nine placed randomly in the solvent) were included in the system to facilitate ligand binding. This design was based on the fact that the ligand binding rate is inversely proportional to the ligand concentration. The higher the ligand concentration, the faster the ligand binds, provided that the ligand concentration is still within its solubility limit.

The built simulation system was energy minimized with 1 kcal/mol/Å² constraints on the heavy atoms of the protein and ligand, including the steepest descent minimization of 5,000 steps followed by a conjugate gradient minimization of 5,000 steps. The system was then heated from 0 K to 300 K for 200 ps. It was further equilibrated using the NVT ensemble at 300K for 800 ps and the NPT ensemble at 300 K and 1 bar for 1 ns with 1 kcal/mol/Å² constraints on the heavy atoms of the protein and ligand, followed by 2 ns short cMD without any constraint. The LiGaMD

simulations proceeded with 14 ns short cMD to collect the potential statistics, 64 ns GaMD equilibration after adding the boost potential and then three independent 1000 ns production runs (**Table S1**). Initial testing simulations showed that when the threshold energy for applying boost potential to the ligand non-bonded energy was set to the lower bound (i.e., $E = V_{\max}$), the bound ligand maintained the X-ray conformation and did not dissociate. In comparison, when the threshold energy was set to the upper bound (i.e., $E = V_{\min} + 1/k$), it enabled high enough boost potential to dissociate the ligand from the protein. In addition, boost applied not only to the ligand non-bonded energy but also to all the bonded energies can accelerate the molecular transitions for ligand dissociation and protein conformational change. Therefore, the threshold energy for applying the ligand boost potential was set to the upper bound in the LiGaMD_Dual simulations. Similarly, second boost potential was applied to the system total potential energy other than the ligand bonded/non-bonded potential energy to provide sufficient acceleration for sampling ligand rebinding. The threshold energy was set to the upper bound for rebinding part as well. LiGaMD_Dual production simulation frames were saved every 0.2 ps for analysis.

The VMD (Humphrey et al., 1996) and CPPTRAJ (Roe and Cheatham, 2013) tools were used for simulation analysis. The 2D PMF profiles were calculated through energetic reweighting of the LiGaMD_Dual simulations using *PyReweighting* toolkit (Miao et al., 2014). 2D PMF profiles of the interdomain Glu56:CA - Ser128:CA atom distance and root-mean-square deviation (RMSD) of the MLN-4760 inhibitor relative to X-ray conformation were calculated to analyze conformational changes of the protein upon ligand binding. Residues Glu56 and Ser128 lie on the tip of $\alpha 4$ (subdomain I) and $\alpha 6$ helices (subdomain II), respectively. The bin size was set to 2.0 Å for the atom distances. Similarly, 2D PMF profiles of the Glu56:CA - Ser128:CA atom distance and RMSD of subdomain I relative to X-ray conformation were calculated to characterize domain

flexibility in the ACE2 receptor. The cutoff of simulation frames in one bin for 2D PMF reweighting was set to 500. Structural clustering was performed on the LiGaMD simulations based on the RMSD of the receptor as well as the ligand snapshots using hierarchical agglomerative algorithm in CPPTRAJ (Roe and Cheatham, 2013). The RMSD cutoff was set to 2.0 Å. The top structural clusters were identified as the representative conformations of the receptor and ligand corresponding to low-energy states in the PMF profiles.

Results

Atomic trajectories were obtained from LiGaMD simulations on the MLN-4760 inhibitor-bound ACE2 receptor (PDB: 1R4L) (Towler et al., 2004) (**Figure 1**). During the LiGaMD equilibration, the bound ligand dissociated from the active site to the bulk solvent, accompanied by large conformational changes of the protein subdomain I (**Figure S2**). Upon ligand dissociation, subdomain I of the receptor changed from the “Closed” to “Open” conformation. Three independent 1000 ns LiGaMD production runs were further performed with randomized initial atomic velocities after the equilibration simulation.

During one of the three 1000 ns LiGaMD production simulations (“Sim 2” in **Table S1**), the MLN-4760 inhibitor bound to the active site of the ACE2 receptor during ~100-500 ns (**Figure 1B**), while no ligand binding was observed in the other two simulations (**Figure S3**). In “Sim 2”, RMSD of the ligand relative to the 1R4L X-ray structure reached a minimum of ~0.99 Å near ~420 ns (**Figure 1B**). The ligand then dissociated at ~500 ns into the bulk solvent and stayed unbound during the remainder of the 1000 ns LiGaMD simulation. Meanwhile, the receptor underwent large-scale conformational changes and sampled different conformations including the “Open”,

“Partially Open” and “Closed” (**Figure 1A**), which were consistent with the receptor experimental structures.

LiGaMD simulations captured both ligand binding and dissociation of ACE2

During the “Sim 2” LiGaMD trajectory, starting from the bulk solvent, one of the MLN-4760 inhibitor molecules first attached to the interface between the receptor 3_{10} H4 and $\alpha 5$ helices within ~100 ns, moved up into the space between the two protein subdomains and entered the active site of the ACE2 receptor between ~100-160 ns (**Figure 1C**). The ligand bound at the active site of the receptor during ~100-500 ns. At ~500 ns, the ligand dissociated from the active site to bulk solvent (**Figure 1D**). The dissociation pathway was observed to be different from that of binding. Ligand dissociated from the opening between the receptor $\alpha 2$ and $\alpha 4$ helices as the subdomain I transitioned from “Closed” to “Open” conformation (**Figure 1D**). On the other hand, the ligand bound through the space just above the 3_{10} H4 and $\alpha 5$ helices (**Figure 1C**). During ligand binding and dissociation in the “Sim 2” LiGaMD trajectory, subdomain I of the receptor sampled different conformations. However, such conformational changes were also observed in other two simulations (“Sim1” and “Sim 2”) regardless of ligand binding/dissociation (**Figure S4**).

Free energy profiles of ligand binding in the human ACE2 receptor

A 2D potential of mean force (PMF) free energy profile was calculated with the ligand RMSD relative to X-ray conformation and the interdomain distance by combining the three independent 1000 ns LiGaMD production trajectories (**Figures 2A**). Six low-energy conformational states of the receptor were identified from the PMF profile, including the “Bound (B)”, “Intermediate-1 (I-

1)”, “Intermediate-2 (I-2)”, “Intermediate-3 (I-3)”, “Unbound-1 (U-1)” and “Unbound-2 (U-2)”. Particularly, the system adopted the “Bound” state with ligand RMSD $< 5 \text{ \AA}$, the “Unbound” state with ligand RMSD $> 35 \text{ \AA}$ and intermediate states with $5 - 35 \text{ \AA}$ ligand RMSD relative to the 1R4L X-ray structure.

In the “Bound” state, the ligand bound at the protein active site and the protein interdomain distance was $\sim 14 \text{ \AA}$. The ligand exhibited a minimum RMSD of 0.99 \AA compared with the X-ray structure (**Figures 2A and 2B**). The system sampled three different intermediate states during ligand binding, *i.e.*, “Intermediate-1 (I-1)”, “Intermediate-2 (I-2)” and “Intermediate-3 (I-3)”. In the I-1 state, the ligand RMSD was $\sim 9.6 \text{ \AA}$ and the interdomain distance was $\sim 18-20 \text{ \AA}$ (**Figures 2A and 2C**). The ligand was located near the active site making interactions with residues of the $\alpha 4$ helix, $\alpha 5$ helix, $\alpha 11$ helix, $3_{10}H8$ helix, $\beta 4-\beta 5$ loop, and $\alpha 8-3_{10}H4$ loop in the two protein subdomains. In the “I-2” state, the ligand RMSD was $\sim 32.1 \text{ \AA}$ and the interdomain distance was $\sim 12 \text{ \AA}$ (**Figures 2A and 2C**). The ligand interacted with the $\alpha 2$ helix of subdomain I and the $\alpha 4$ helix of subdomain II in the receptor. In the “I-3” state, the ligand RMSD was $\sim 29.6 \text{ \AA}$ and the interdomain distance was $\sim 23 \text{ \AA}$ (**Figures 2A and 2C**). The ligand interacted with the $\alpha 8$ helix, $\alpha 20$ helix, $\alpha 8-\beta 3$ loop, and the C-terminal loop after the $\alpha 20$ helix in the protein subdomain II (**Figure 2C**). The system sampled two “Unbound (U-1)” and “Unbound (U-2)” states, where the ligand RMSD was $\sim 60-70 \text{ \AA}$ and the interdomain distances were $\sim 12 \text{ \AA}$ and $\sim 23 \text{ \AA}$, respectively. In these states, the ligand was found far away from the receptor in the bulk solvent and the receptor could change between the “Closed” and “Open” conformations (**Figures 2A and 2D**).

Low-energy intermediate conformational states of ligand binding to human ACE2 receptor

Low-energy intermediate conformational states “I-1”, “I-2” and “I-3” of the MLN-4760 inhibitor binding to the human ACE2 receptor identified from the LiGaMD simulation free energy profiles are shown in **Figures 3A, 3B** and **3C**, respectively. Polar and charged groups present in different parts of the receptor made favorable interactions with the charged carboxylate groups and the polar chloride and nitrogen atoms in the ligand molecule. These interactions played important role in recognition and binding of the MLN-4760 inhibitor to the receptor.

In the intermediate “I-1” state, the receptor adopted a “Partially Open” conformation with ~18 Å interdomain distance. The ligand molecule was located near the receptor active site with ~9.6 Å RMSD relative to the X-ray structure and formed interactions with residues from both subdomains of ACE2. One of the ligand carboxylate groups formed ionic interaction with the positively charged protein residue His345 (**Figure 3A**). The carboxylate group also interacted with protein residue Thr371. Similarly, one of the nitrogen atoms in the ligand’s central ring formed ionic interaction with the negatively charged protein residue Glu145. The ligand chloride group formed polar interactions with protein residue Asn149.

In the intermediate “I-2” state, the receptor adopted the “Closed” conformation and the ligand interacted with the α 2 helix of subdomain I and the α 4 helix of subdomain II in the receptor (**Figure 3B**). One of the ligand chloride atoms formed polar interactions with protein residues Gln60 and Asn63, while the other chloride atom formed polar interactions with protein residues Thr118 and Asn121. One of the ligand’s negatively charged carboxylate groups formed ionic interaction with the protein positively charged residue Lys114. The ligand terminal benzene ring formed hydrophobic interactions with protein residue Val59 (**Figure 3B**).

In the intermediate “I-3” state, one of the ligand chloride atoms formed polar interactions with protein residues Arg245, Ser602 and Met249 of subdomain II (**Figure 3C**). One of the ligand

charged carboxylate groups formed polar interactions with protein residue Asn601. The ligand terminal benzene ring formed hydrophobic interactions with Pro258 in the receptor.

During one of the LiGaMD trajectories (“Sim3”), the MLN-4760 ligand was observed to interact with residue Gly354 of the ACE2 receptor (**Figure S5**). The ligand terminal methyl group was able to form close hydrophobic interactions with the Gly354 residue in subdomain I of the ACE2 receptor. Notably, residue Gly354 is considered an important residue making direct interactions with the RBD of SARS-CoV and SARS-CoV-2 as seen in the experimental structures (Wang et al., 2020).

Conformational plasticity of human ACE2 receptor

In the LiGaMD simulations, while the protein subdomain II was stable maintaining the 1R4L X-ray conformation with $\sim 2\text{-}4$ Å RMSD (**Figures S2 and 4B**), subdomain I in the human ACE2 receptor exhibited high flexibility and underwent large conformational changes with $\sim 3\text{-}10$ Å RMSD compared with the X-ray conformation (**Figure 4A**). We calculated 2D PMF regarding the interdomain distance and subdomain I RMSD relative to the 1R4L X-ray conformation. Three low-energy conformational states were identified in the PMF profile, including the “Open”, “Partially Open” and “Closed” (**Figure 3C**).

In the “Closed” conformation, subdomain I moved near subdomain II closing the active site. The receptor interdomain distance was ~ 14 Å and RMSD of subdomain I was ~ 4 Å compared with the 1R4L X-ray structure (**Figure 3C**). In the “Partially Open” conformation, the receptor interdomain distance increased to ~ 18 Å and RMSD of subdomain I was ~ 4 Å compared with the 1R4L X-ray structure (**Figure 3C**). Finally, the receptor interdomain distance could increase

further to ~ 20 Å and the subdomain I RMSD relative to the 1R4L X-ray structure increased to ~ 6 Å in the “Open” conformation. Notably, conformations of the ACE2 receptor in the “Partially Open” and “Open” low-energy states were closely similar to the experimental 6ACK cryo-EM and 6LZG X-ray structures, respectively (**Figure 3C**). Therefore, the different low-energy states of ACE2 receptor revealed from our LiGaMD simulations highlighted the receptor conformational plasticity during its function for ligand binding and interactions with other proteins (e.g., the coronavirus spike protein).

Discussion

Since its discovery in 2000 (Donoghue et al., 2000), the ACE2 receptor has been recognized as a critical protease enzyme with multiple physiological roles in renin-angiotensin system, amino acid transport, gut microbiome ecology and innate immunity. The ACE2 receptor has also been identified as the functional receptor for SARS-CoV and SARS-CoV-2. The COVID-19 pandemic caused by SARS-CoV-2 has been recognized as a serious global health threat as it has no proper treatment and continues to spread across the world. With the infection cases rising daily, it is critical to develop therapeutics against SARS-CoV-2. Here, we have applied all-atom simulations using a novel LiGaMD method to investigate the mechanism of ligand binding to the human ACE2 receptor.

Through LiGaMD enhanced sampling simulations, we have, for the first time, successfully captured both binding and dissociation of a ligand in the human ACE2 receptor. During the simulations, the receptor could sample distinct conformational states, revealing remarkable conformational plasticity of the receptor. Ligand binding biased the receptor conformational

ensemble to the Closed state, suggesting a conformational selection mechanism rather than induced fit. Furthermore, the MLN-4760 ligand molecule with -2 net charge, formed polar interactions with charged and polar residues in the low energy intermediate states. This finding suggested that electrostatic interactions played an important role in the recognition and binding/dissociation of the MLN-4760 inhibitor to the ACE2 receptor, being consistent with previous findings of “electrostatic steering” in recognition of charged ligands by proteins (Wade et al., 1998;Miao et al., 2020).

Despite our encouraging simulation findings, it is important to note that the number of ligand binding and unbinding events captured in the presented LiGaMD simulations was too small and the simulation free energy profiles were not converged. More sufficient sampling would be needed in order to obtain converged simulations and calculate the ligand binding thermodynamics and kinetics. This can be potentially achieved through additional and longer simulations, as well as further method developments combining LiGaMD with other enhanced sampling algorithms such as replica exchange (Huang et al., 2018;Oshima et al., 2019) and Markov state models (Plattner and Noe, 2015).

In this context, the MLN-4760 inhibitor binds to the human ACE2 receptor with high affinity (IC_{50} : 0.44 nM) (Towler et al., 2004). It is extremely difficult to simulate the ligand dissociation and binding with long-timescale cMD and even the enhanced sampling methods. Very few computational studies have been carried out on ligand binding to the ACE2 receptor. A recent study (Nami et al., 2020) showed that the MLN-4760 ligand could dissociate from the ACE2 receptor upon binding of the viral RBD. However, this study was not able to characterize ligand binding to the ACE2 receptor. In comparison, our LiGaMD simulations could capture both ligand binding and dissociation in the human ACE2 receptor.

In case of SARS-CoV-2 infection, RBD of the viral spike protein attaches to the top region of subdomain I in the ACE2 receptor, which is distant from the receptor peptidase active site. Thus, the MLN-4760 inhibitor does not compete directly with the RBD of SARS-CoV-2. However, there have been evidences of allosteric signaling between the RBD binding site and active site in the ACE2 receptor. Huentelman et. al. (Huentelman et al., 2004) performed *in silico* molecular docking studies to identify a novel ACE2 inhibitor that could also block SARS-CoV spike protein-mediated receptor attachment. Our LiGaMD simulations further highlighted the dynamic nature of the receptor in terms of the large-scale movement of subdomain I and ligand binding. Similarly, LiGaMD simulations showed that MLN-4760 could form hydrophobic interactions with RBD binding site in subdomain I of the receptor. This information could be helpful in designing drugs that could block RBD binding to the receptor. Likewise, since the human ACE2 receptor shows conformational selection for ligand binding, virtual screening using ensemble docking (Lin et al., 2002; Amaro et al., 2008; Amaro et al., 2018) with receptor structural ensembles generated from the LiGaMD simulations will be a promising approach to designing potent drug molecules of the ACE2 receptor.

In summary, we have successfully simulated both ligand binding and dissociation in the human ACE2 receptor using the novel LiGaMD enhanced sampling method. During the LiGaMD simulations, the receptor could sample distinct Closed, Partially Open and Open conformational states. Ligand binding biased the receptor towards the Closed conformation, suggesting a conformational selection mechanism. Our simulation findings are expected to facilitate rational drug design targeting ACE2 for the therapeutic treatments of COVID-19 and other related human diseases.

Author Contributions

Conceived the study: Y.M., A.B. and S.P.; Performed simulations: A.B.; Analyzed data: A.B. and S.P.; Wrote the manuscript: A.B., S.P. and Y.M.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

This work used supercomputing resources with allocation award TG-MCB180049 through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562, and the Research Computing Cluster at the University of Kansas. This work was supported in part by the National Institutes of Health (R01GM132572) and the startup funding in the College of Liberal Arts and Sciences at the University of Kansas.

Supplementary Material

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/>.

References

- Amaro, R.E., Baron, R., and Mccammon, J.A. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design* 22, 693-705.
- Amaro, R.E., Baudry, J., Chodera, J., Demir, Ö., Mccammon, J.A., Miao, Y., and Smith, J.C. (2018). Ensemble Docking in Drug Discovery. *Biophysical Journal* 114, 2271-2278.
- Case, D., Belfon, K., Ben-Shalom, I., Brozell, S., Cerutti, D., Cheatham, T., Cruzeiro Iii, V., Darden, T., Duke, R., and Giambasi, G. (2020). AMBER 2020. *University of California, San Francisco*.
- Case, D.A., Cheatham Iii, T.E., Darden, T., Gohlke, H., Luo, R., Merz Jr, K.M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005a). The Amber biomolecular simulation programs. *Journal of computational chemistry* 26, 1668-1688.
- Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005b). The Amber biomolecular simulation programs. *Journal of Computational Chemistry* 26, 1668-1688.
- Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1996). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *Journal of the American Chemical Society* 118, 2309-2309.
- Dales, N.A., Gould, A.E., Brown, J.A., Calderwood, E.F., Guan, B., Minor, C.A., Gavin, J.M., Hales, P., Kaushik, V.K., and Stewart, M. (2002). Substrate-based design of the first class of angiotensin-converting enzyme-related carboxypeptidase (ACE2) inhibitors. *Journal of the American Chemical Society* 124, 11852-11853.
- Donoghue, M., Hsieh, F., Baronas, E., Godbout, K., Gosselin, M., Stagliano, N., Donovan, M., Woolf, B., Robison, K., and Jeyaseelan, R. (2000). A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circulation research* 87, e1-e9.
- Dror, R.O., Pan, A.C., Arlow, D.H., Borhani, D.W., Maragakis, P., Shan, Y., Xu, H., and Shaw, D.E. (2011). Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci U S A* 108, 13118-13123.
- Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G.M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J.M., and Kollman, P. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* 24, 1999-2012.
- Gheblawi, M., Wang, K., Viveiros, A., Nguyen, Q., Zhong, J.-C., Turner, A.J., Raizada, M.K., Grant, M.B., and Oudit, G.Y. (2020). Angiotensin-converting enzyme 2: SARS-CoV-2 receptor and regulator of the renin-angiotensin system: celebrating the 20th anniversary of the discovery of ACE2. *Circulation research* 126, 1456-1474.
- Gross, L.Z.F., Sacerdoti, M., Piiper, A., Zeuzem, S., Leroux, A.E., and Biondi, R.M. (2020). ACE2, the Receptor that Enables the Infection by SARS-CoV-2: Biochemistry, Structure, Allostery and Evaluation of the Potential Development of ACE2 Modulators.
- Hernández Prada, J.A., Ferreira, A.J., Katovich, M.J., Shenoy, V., Qi, Y., Santos, R.A., Castellano, R.K., Lampkins, A.J., Gubala, V., and Ostrov, D.A. (2008). Structure-based identification of small-molecule angiotensin-converting enzyme 2 activators as novel antihypertensive agents. *Hypertension* 51, 1312-1317.

- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., and Nitsche, A. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*.
- Huang, Y.M., Mccammon, J.A., and Miao, Y. (2018). Replica Exchange Gaussian Accelerated Molecular Dynamics: Improved Enhanced Sampling and Free Energy Calculation. *J Chem Theory Comput* 14, 1853-1864.
- Huentelman, M.J., Zubcevic, J., Hernandez Prada, J.A., Xiao, X., Dimitrov, D.S., Raizada, M.K., and Ostrov, D.A. (2004). Structure-based discovery of a novel angiotensin-converting enzyme 2 inhibitor. *Hypertension* 44, 903-906.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* 14, 33-38.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* 79, 926-935.
- Li, F., Li, W., Farzan, M., and Harrison, S.C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309, 1864-1868.
- Lin, J.H., Perryman, A.L., Schames, J.R., and Mccammon, J.A. (2002). Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *Journal of the American Chemical Society* 124, 5632-5633.
- Miao, Y. (2018). Acceleration of Biomolecular Kinetics in Gaussian Accelerated Molecular Dynamics. *Journal of Chemical Physics* 149, 072308.
- Miao, Y., Bhattarai, A., Nguyen, A.T.N., Christopoulos, A., and May, L.T. (2018). Structural Basis for Binding of Allosteric Drug Leads in the Adenosine A1 Receptor. *Sci. Rep.* 8, 16836.
- Miao, Y., Bhattarai, A., and Wang, J. (2020). Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. *Journal of Chemical Theory and Computation* 16, 5526-5547.
- Miao, Y., Feher, V.A., and Mccammon, J.A. (2015). Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* 11, 3584-3595.
- Miao, Y., Sinko, W., Pierce, L., Bucher, D., and Mccammon, J.A. (2014). Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation* 10, 2677-2689.
- Nami, B., Ghanaeian, A., Ghanaeian, K., and Nami, N. (2020). The effect of ACE2 inhibitor MLN-4760 on the interaction of SARS-CoV-2 spike protein with human ACE2: a molecular dynamics study.
- Oshima, H., Re, S., and Sugita, Y. (2019). Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules. *J Chem Theory Comput* 15, 5199-5208.
- Patel, V.B., Zhong, J.-C., Grant, M.B., and Oudit, G.Y. (2016). Role of the ACE2/angiotensin 1-7 axis of the renin-angiotensin system in heart failure. *Circulation research* 118, 1313-1326.
- Perlot, T., and Penninger, J.M. (2013). ACE2-From the renin-angiotensin system to gut microbiota and malnutrition. *Microbes and infection* 15, 866-873.

- Plattner, N., and Noe, F. (2015). Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun* 6, 7653.
- Roe, D.R., and Cheatham, T.E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* 9, 3084-3095.
- Roux, B. (1995). The Calculation of the Potential of Mean Force Using Computer-Simulations. *Computer Physics Communications* 91, 275-282.
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221-224.
- Song, W., Gui, M., Wang, X., and Xiang, Y. (2018). Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS pathogens* 14, e1007236.
- Tian, C., Kasavajhala, K., Belfon, K.A., Raguette, L., Huang, H., Migués, A.N., Bickel, J., Wang, Y., Pincay, J., and Wu, Q. (2019). ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* 16, 528-552.
- Towler, P., Staker, B., Prasad, S.G., Menon, S., Tang, J., Parsons, T., Ryan, D., Fisher, M., Williams, D., and Dales, N.A. (2004). ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *Journal of Biological Chemistry* 279, 17996-18007.
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell, A.D. (2010). CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* 31, 671-690.
- Vanommeslaeghe, K., and Mackerell, A.D., Jr. (2014). CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim Biophys Acta*.
- Wade, R.C., Gabdouliline, R.R., Ludemann, S.K., and Lounnas, V. (1998). Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations. *Proc Natl Acad Sci U S A* 95, 5942-5949.
- Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., and Yuen, K.-Y. (2020). Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*.
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444-1448.

Figure Captions

Figure 1. (A) X-ray and cryo-EM structures of the ACE2 receptor with subdomain I in the “Open” (cyan, PDB: 6LZG), “Partially Open” (magenta, PDB: 6ACK) and “Closed” (orange, PDB: 1R4L) conformations. Subdomain II is stable and colored in gray. In the “Closed” conformation, the receptor is bound by the MLN-4760 inhibitor. The protein is shown as ribbons and the ligand as sticks (orange). (B) Root-mean-square deviations (RMSDs) of ten MLN-4760 inhibitor molecules relative to the bound X-ray conformation (PDB: 1R4L) are calculated from the 1000 ns “Sim 2” LiGaMD trajectory, in which the ligand bound to the ACE2 active site during ~100-400 ns and then dissociated into the bulk solvent. (C) Two views of the observed ligand binding pathway, for which the center ring of MLN-4760 is represented by lines and colored by simulation time in a blue-white-red (BWR) color scale. (D) Two views of the observed ligand dissociation pathway, for which the center ring of MLN-4760 is represented by lines and colored by simulation time in a blue-white-red (BWR) color scale.

Figure 2. (A) The 2D potential of mean force (PMF) free energy profile of the ligand RMSD and interdomain distance calculated by combining the three independent 1000 ns LiGaMD production simulations of human ACE2 receptor. Six low energy conformational states were identified, including the “Bound (B)”, “Intermediate-1 (I-1)”, “Intermediate-2 (I-2)”, “Intermediate-3 (I-3)”, “Unbound-1 (U-1)” and “Unbound-2 (U-2)”. (B) Conformations of the ACE2 receptor and MLN-4760 ligand in the “Bound (B)” state (green) compared with the X-ray conformation (Orange, PDB: 1R4L). (C) Conformations of the ACE2 receptor during binding of MLN-4760 in the “Intermediate-1 (I-1)” (blue) and “Intermediate-2 (I-2)” (yellow) and “Intermediate-3 (I-3)” (pink) states. (D) Conformations of the ACE2 receptor in the “Unbound-1 (U-1)” (ice blue) and “Unbound-2 (U-2)” (lime) states. The protein is shown as ribbons and ligand as sticks.

Figure 3. (A) The “Intermediate-1 (I-1)” conformational state of the MLN-4760 ligand (blue balls and sticks) bound to the ACE2 receptor (ribbons). Residues of subdomain I (blue) and subdomain II (gray) including H345, T371, E145 and N149 formed polar interactions with the ligand. (B) The “Intermediate-2 (I-2)” conformational state of MLN-4760 (yellow) bound to the ACE2 receptor (ribbons). Residues of subdomain I (yellow) and subdomain II (gray) including Q60, N63, N171 and N121 formed polar interactions with the ligand. (C) The “Intermediate-3 (I-3)” conformational state of MLN-4760 (pink) bound to the ACE2 receptor (ribbons). Residues of subdomain I (pink) and subdomain II (gray) including R245, M249, S602 and N601 formed polar interactions with the ligand.

Figure 4. RMSDs of (A) subdomain I and (B) subdomain II of the ACE2 receptor relative to the closed X-ray conformation (PDB:1R4L) are calculated from three independent LiGaMD production simulations. (C) 2D potential of mean force (PMF) of the subdomain I RMSD and interdomain distance calculated by combining the three LiGaMD simulations. Three low energy conformational states of the receptor are identified in the PMF profile, including the “Closed”, “Partially Open” and “Open”, which are similar to the 6LZG, 6ACK and 1R4L PDB structures, respectively. (D) Low-energy conformations of the ACE2 receptor with subdomain I found in the “Open” (red), “Partially Open” (blue) and “Closed” (green) states in the LiGaMD simulations. Subdomain II is stable and colored in gray.

Figure 1

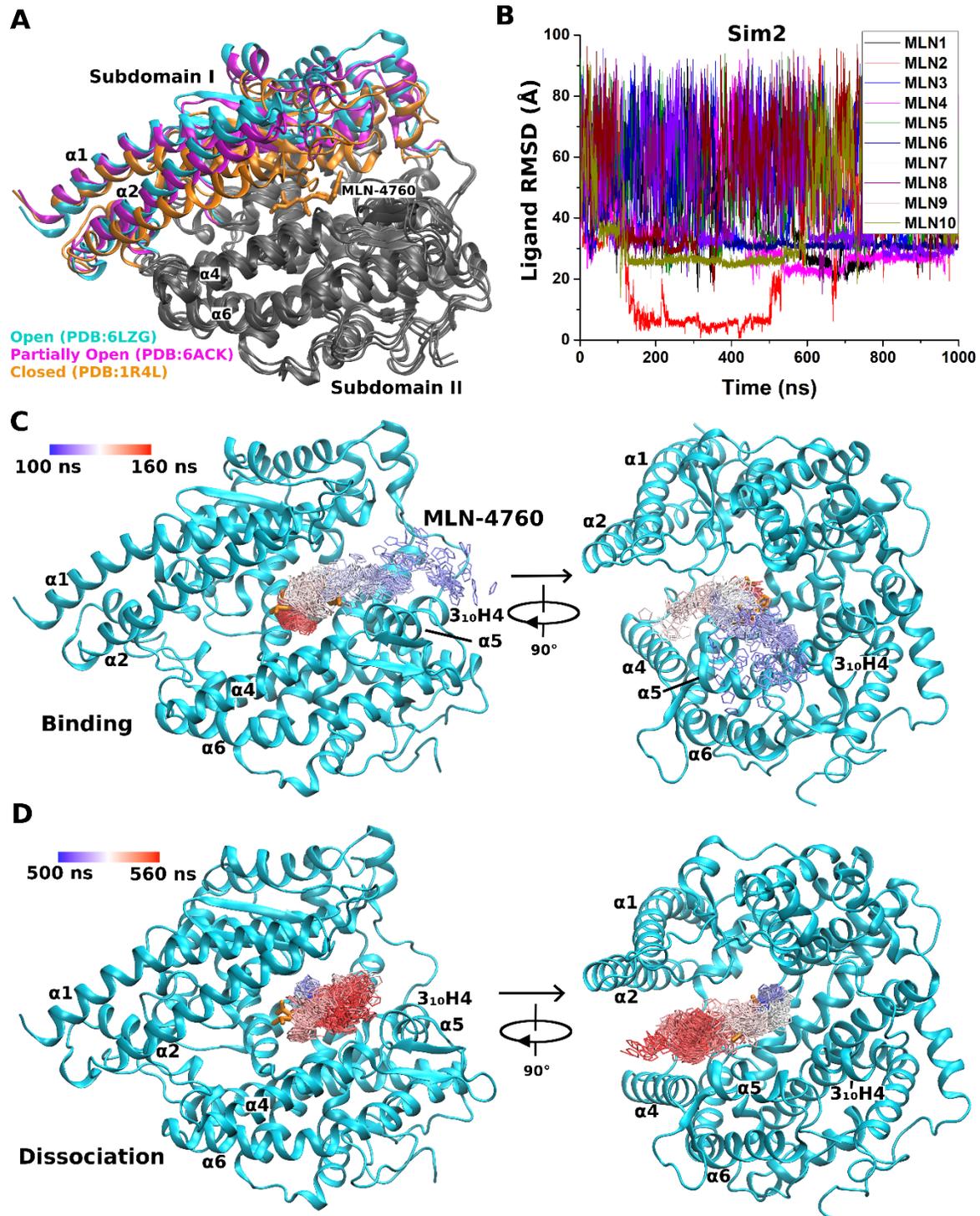


Figure2

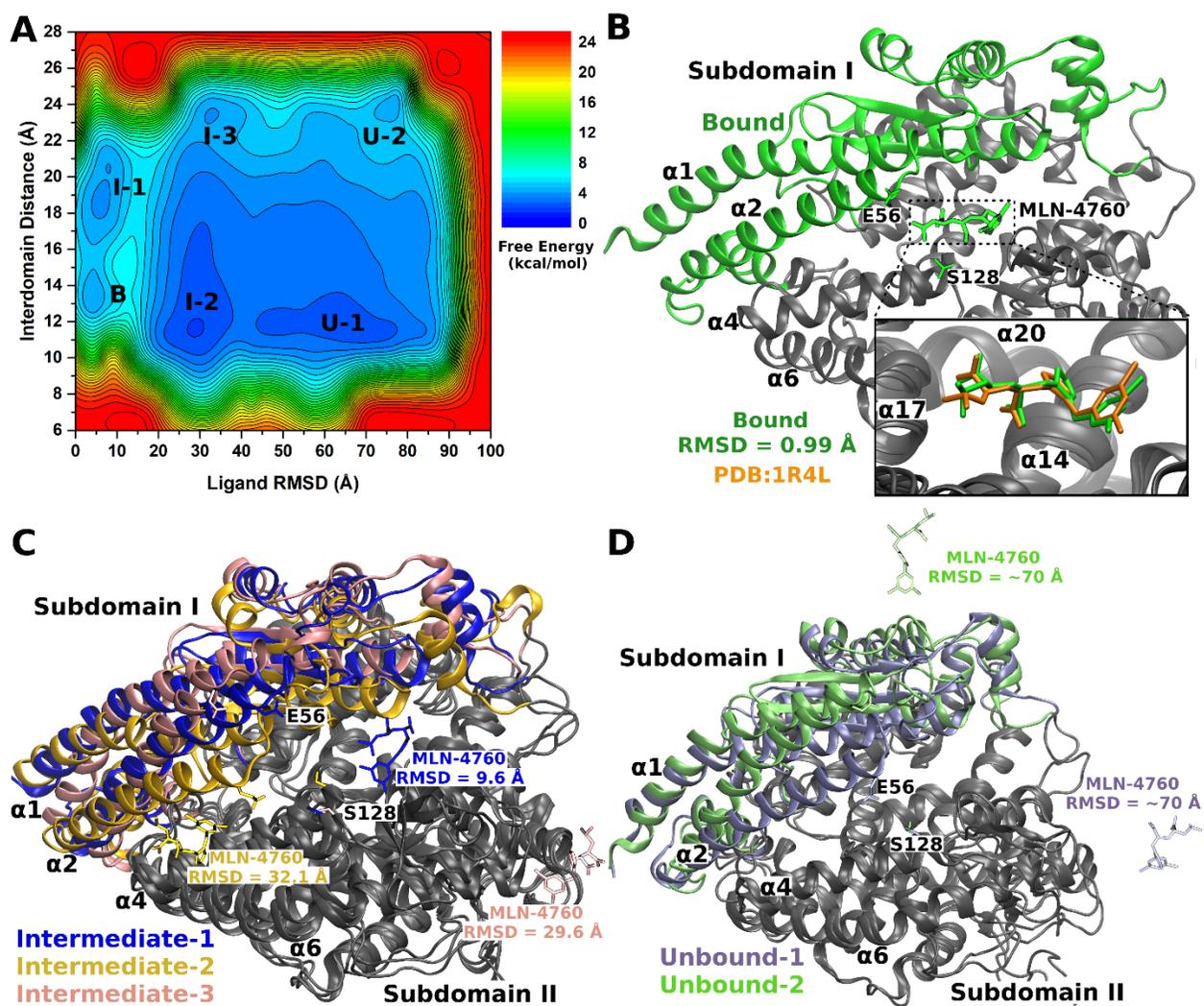


Figure 3

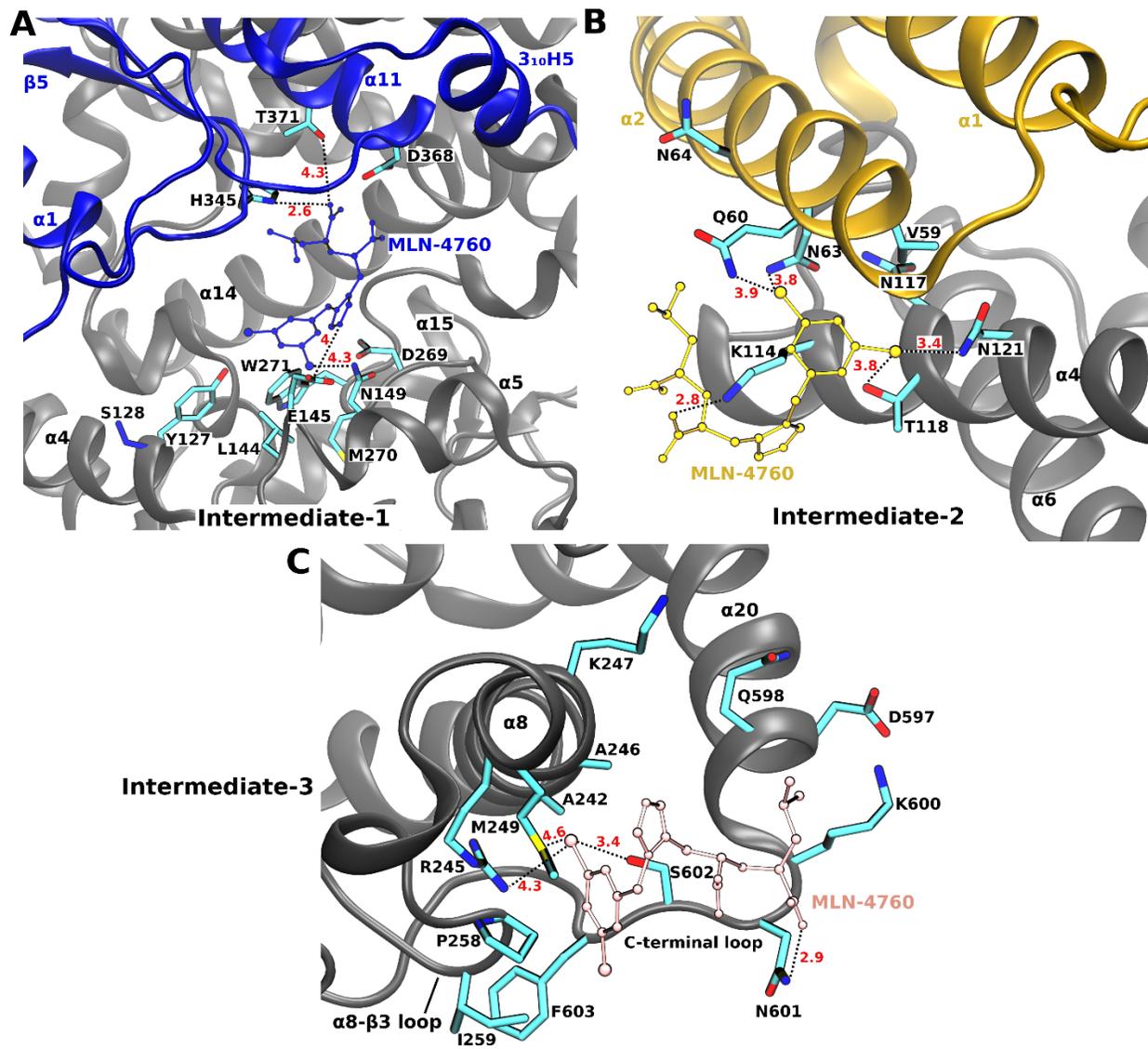


Figure 4

