Accelerated simulations and computer-aided drug design of membrane proteins

By
Apurba Bhattarai

Submitted to the graduate degree program in the Center for Computational Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson: Yinglong Miao, Ph.D.

Ilya Vakser, Ph.D.

Joanna S.G. Slusky, Ph.D.

Krzysztof Kuczera, Ph.D.

Michael S. Wolfe, Ph.D.

Date Defended: 21 April 2022

The dissertation committee for Apurba Bhattarai certifies that this is the approved version of the following dissertation:

Accelerated simulations and computer-aided drug design of membrane proteins

Chairperson: Yinglong Miao, Ph.D.

Date Approved: 21 April 2022

Abstract

Membrane proteins represent ~30% of the proteome[1] and more than 60% of drug targets.[2] They play key roles in physiological functions, such as signal transduction, transport, ion regulation and enzymatic activities. Malfunction of these proteins result in deadly human diseases, such as paralysis, cancer, heart failure and Alzheimer's diseases (AD).[3] Despite the importance, the structural dynamics and functional mechanisms of many membrane proteins remain poorly understood, which has greatly hindered effective drug design. Gaussian accelerated Molecular Dynamics (GaMD) is a novel technique that provides simultaneous unconstrained enhanced sampling and free energy calculations.[4,5] GaMD has been successfully applied to study physical pathways and mechanisms of protein folding and ligand binding of important membrane proteins.[4,6,7] Based on the principles of GaMD, Ligand GaMD (LiGaMD) and Peptide GaMD (Pep-GaMD) methods were developed to efficiently simulate ligand and peptide binding/unbinding, respectively. Here, I have integrated GaMD, LiGaMD and Pep-GaMD with advanced complementary computational techniques including molecular docking, homology modeling and free energy calculations for accelerated simulations and computer-aided drug design of different membrane proteins. In Chapters 1 and 2, I focus on applying GaMD and Pep-GaMD to determine the mechanisms of γ-secretase activation and substrate processive proteolysis. In Chapters 3 and 4, I focus on investigating the dependence of GPCR-membrane interactions on the receptor activation state and retrospective docking of known positive allosteric modulators (PAMs) to GaMD simulation-derived structural ensembles of the adenosine A_1 receptor (A_1AR), respectively. Finally, in Chapter 5, I apply LiGaMD to understand the mechanism of ligand recognition by human Angiotensin Converting Enzyme-2 (ACE2) receptor.

Dedication

I dedicate this thesis to my parents, Arun and Anita Bhattarai, my grandparents, Ram Chandra and Sushila Acharya, my wife, Punam Rawal, and my brother Adarsha Bhattarai as they were the reason, I did this all, and they are the reason I will continue to do more.

Acknowledgements

First and foremost, I offer my sincerest gratitude to my mentor and PI, Dr. Yinglong Miao. Thank you for believing in me and accepting me as the first graduate student in the Miao lab. Thank you for teaching me how to do science and to be a hardworking-dedicated researcher. Not only professionally, but I am also grateful to Dr. Miao for inspiring me to be a humble, respectful, disciplined and a virtuous human being. And secondly, I am extremely grateful to Dr. Jinan Wang, who is like a second PI to me. Thank you, Jinan, for teaching me all the technical and intellectual aspects of being a computational scientist. Thank you for being a kind motivator and a good friend. I will forever be indebted towards both my teachers, Dr. Miao and Jinan, for their deep-rooted impact they have in my professional and personal life.

I would like to thank the members of my defense committee, Dr. Ilya Vakser, Dr. Michael S. Wolfe, Dr. Krzysztof Kuczera and Dr. Joanna S.G. Slusky for their constructive criticism, encouragement, and support in various ways. My sincere thanks to Dr. Christian Ray, one of the members of my comprehensive exam committee for being generous and helping me during the exam. I would also like to thank our collaborators, Dr. Michael S. Wolfe, Dr. Sujan Devkota and Dr. Sanjay Bhattarai, Dr. David Thal and Dr. Alisa Glukhova, Dr. Yuan Chen, and Dr. Giulia Palermo for the wonderful teamwork.

Finally, I want to express gratitude to my friends and family, lab members, departmental colleagues, professors, and all the KU Jayhawks who have given their heart whelming support in making my PhD journey a magnificent experience.

Table of Contents

Chapter 1

Mechanisms of γ-Secretase Activation and Substrate Processing

Apurba Bhattarai[1], Sujan Devkota[2], Sanjay Bhattarai[2], Michael S. Wolfe[2] and Yinglong Miao[1]

[1]Center for Computational Biology and Department of Molecular Biosciences, [2]Department of Medical Chemistry, School of Pharmacy, University of Kansas, Lawrence, KS 66047, USA

Reprinted (Adapted) with permission from Bhattarai, Apurba, Sujan Devkota, Sanjay Bhattarai, Michael S. Wolfe, and Yinglong Miao. "Mechanisms of γ-secretase activation and substrate processing." ACS central science 6, no. 6 (2020): 969-983[8].

Abstract

Amyloid β-peptide, the principal component of characteristic cerebral plaques of Alzheimer's disease (AD), is produced through intramembrane proteolysis of the amyloid precursor protein (APP) by γ-secretase. Despite the importance in pathogenesis of AD, the mechanisms of intramembrane proteolysis and substrate processing by γ-secretase remain poorly understood. Here, complementary all-atom simulations using a robust Gaussian accelerated molecular dynamics (GaMD) method and biochemical experiments were combined to investigate substrate processing of wildtype and mutant APP by γ-secretase. The GaMD simulations captured spontaneous activation of γ-secretase, with hydrogen bonded catalytic aspartates and water poised for proteolysis of APP at the ε cleavage site. Furthermore, GaMD simulations revealed that familial AD mutations I45F and T48P enhanced the initial ε cleavage between residues Leu49-Val50, while M51F mutation shifted the ε cleavage site to the amide bond between Thr48-Leu49. Detailed analysis of the GaMD simulations allowed us to identify distinct low-energy conformational states of γ-secretase, different secondary structures of the wildtype and mutant APP substrate, and important active-site sub-pockets for catalytic function of the enzyme. The simulation findings were highly consistent with experimental analyses of APP proteolytic products using mass spectrometry and western blotting. Taken together, the GaMD simulations and biochemical experiments have enabled us to elucidate the mechanisms of γ-secretase activation and substrate processing, which should facilitate rational computer-aided drug design targeting this functionally important enzyme.

Keywords: Alzheimer's disease (AD), amyloid precursor protein (APP), γ-secretase, Gaussian accelerated molecular dynamics (GaMD), mass spectrometry (MS), western blotting.

Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by cerebral atrophy, beginning with areas of the brain involved in learning and memory. Deposition of 42-residue amyloid β-peptide (Aβ42) in the form of plaques is a defining pathological feature of AD and begins many years before onset of symptoms.[9] For these reasons, Aβ42 has been a major target for the development of potential therapeutics[10] as well as a key biomarker for AD.[11] Aβ peptides are derived through proteolytic processing of the membrane-traversing amyloid precursor protein (APP), first by β-secretase outside the membrane, generating a membrane-bound 99-residue C-terminal fragment (C99), and then by γ-secretase within the membrane.[12] γ-Secretase is a membrane-embedded aspartyl protease complex, with presenilin (PS1) as the catalytic component that carries out intramembrane proteolysis of >90 substrates, including APP and the Notch family of cell-surface receptors.[13] Cleavage of the APP transmembrane (TM) domain by γ-secretase determines the length of Aβ peptides, the proportion of the hydrophobic TM domain retained in the Aβ product, and therefore the tendency of Aβ to aggregate into plaques.

Proteolysis of the APP TM domain by γ-secretase is complex.[14] Initial endoproteolysis of C99 at the ε site generates 48- or 49-residue Aβ (Aβ48 or Aβ49) and corresponding APP intracellular domains (AICD49-99 or AICD50-99) **(Figure S1).**[15] These initially formed Aβ peptides are then trimmed every 3-4 amino acids through a carboxypeptidase activity of γ-secretase along two pathways, Aβ48→Aβ45→Aβ42→Aβ38 and Aβ49→Aβ46→Aβ43→Aβ40,[16,17] and this trimming is dictated by three active-site pockets that recognize substrate residues P1', P2' and P3' (i.e., immediately C-terminal of the scissile amide bond).[18] Mutations in the APP TM domain associated with early-onset familial AD (FAD) can

skew ε cleavage in favor of Aβ48 (i.e., to the pathological Aβ42 pathway).[18,19] Alternatively, these mutations can be "pathway switchers", affecting carboxypeptidase activity to switch from the Aβ40 pathway to the Aβ42 pathway.[18]

Little is known about the mechanism by which γ-secretase accomplishes intramembrane proteolysis. A substantial advance in understanding substrate recognition came recently with reports of cryo-electron microscopic (cryo-EM) structure determination of the γ-secretase complex bound to the Notch and APP substrates **(Figure S1).**[20,21] The cryo-EM structures were consistent with expectations from previous studies using small-molecule probes and mutagenesis. In both structures, the substrate TM assumed a helical conformation starting from the extracellular side and was surrounded by TM2, TM3 and TM5 of PS1. The helix ended just before entry into the enzyme active site, becoming first partially unwound and then fully extended into a β-strand toward the intracellular side. The substrate β-strand interacted with an antiparallel β-strand in the intracellular side of PS1 TM7, which in turn interacted with another β-strand from the enzyme TM6. This β-sheet motif was suggested to be essential for substrate recognition by the γ-secretase.[21,22] While a *tour de force* for the field, stabilization of the substrate-enzyme complex required (1) mutation of one of the catalytic aspartates (Asp385) to alanine in PS1 (inactivating the enzyme) and (2) double cysteine mutagenesis and disulfide crosslinking between substrate and presenilin (with the potential for deviation from normal wildtype interactions).

Computational modeling, especially molecular dynamics (MD) simulation, has proven useful in understanding the structural dynamics of γ-secretase. Previous studies have provided valuable insights into the conformational changes[23-26], enzyme allosteric modulation[27], substrate binding[23,26,28-30], water distribution[23,24], lipid interactions[24] and ligand binding of γ-

secretase[31-33]. A putative active conformation was described for the substrate-free (apo) γ-secretase with the two catalytic aspartates moving to close proximity in few[23-25], but none has characterized the enzyme active state poised for proteolysis with both the water and peptide substrate. Hence, the dynamic mechanisms of enzyme activation and substrate processing by γ-secretase remained poorly understood.

Here, we present the first report of MD computational modeling of activation of APP-bound γ-secretase using the latest cryo-EM structures of substrate-bound enzyme. The enzyme and substrate were computationally restored to the wildtype. Extensive all-atom simulations using a novel and robust Gaussian accelerated molecular dynamics (GaMD) method were employed to capture the extremely slow motions underlying activation of γ-secretase for proteolysis of substrate within the cell membrane (k_{cat} in proteoliposomes estimated at 1.9 h^{-1}).[34]

GaMD is an enhanced sampling computational technique that works by adding a harmonic boost potential to smooth the biomolecular potential energy surface.[4] GaMD greatly reduces energy barriers and accelerates biomolecular simulations by orders of magnitude.[35] GaMD does not require pre-defined collective variables or reaction coordinates. Compared with the enhanced sampling methods that rely on careful selection of the collective variables, GaMD is of particular advantage for studying complex biological processes[36] such as enzyme activation and substrate processing by γ-secretase. Moreover, because the boost potential follows a Gaussian distribution, biomolecular free energy profiles can be properly recovered through cumulant expansion to the second order.[4] GaMD builds on the previous accelerated MD (aMD) method[37,38], but solves its energetic reweighting problem[39] for free energy calculations of large biomolecules. GaMD has successfully revealed physical pathways and

mechanisms of protein folding and ligand binding, which are consistent with experiments and long-timescale conventional MD simulations.[4,6,7] It has also been applied to characterize protein-protein,[40,41] protein-membrane,[42] and protein-nucleic acid[43,44] interactions. Therefore, GaMD was applied in this study for enhanced sampling of the γ-secretase complex, a well-known slow enzyme.[45,46]

Furthermore, the GaMD simulations were highly consistent with parallel mass spectrometry (MS) and western blotting biochemical experiments on the processing of both wildtype and mutant APP substrates. Remarkably, one of the mutations (M51F) in APP shifted the substrate ε cleavage site to the amide bond between residue Thr48-Leu49, while another two mutations (I45F and T48P) enhanced the ε cleavage between Leu49-Val50 compared with the wildtype. The GaMD simulations and biochemical experiments together offered a deep atomic-level understanding of intramembrane proteolysis by γ-secretase.

Results

Activation of computationally restored wildtype γ-secretase is captured in GaMD simulations. Our initial testing GaMD simulations using the earlier published cryo-EM structure of Notch-bound γ-secretase **(Figure S1A)**—with Asp385 computationally restored—showed that Asp257 rather than Asp385 should be protonated in the active site, as in this case the two aspartates were able to approach each other for catalysis **(Table 1** and **Figure S2).** Further testing GaMD simulations using the cryo-EM structure of APP-bound γ-secretase **(Figure S1B** and **Table 1)** revealed an active conformation of the PS1 catalytic subunit with computationally restored Asp385 while the enzyme-substrate disulfide bond was kept **(Figure**

S3 and **Movie S1).** Building upon these testing results, we proceeded to remove the artificial enzyme-substrate disulfide bond to completely restore the wildtype γ-secretase for further simulations **(Table 1).** During three 2-μs GaMD enhanced simulations, spontaneous activation of APP-bound γ-secretase was observed starting from its inactive cryo-EM conformation **(Figures 1A** and **S4A** and **Movie S2).** The activation was characterized by coordinated hydrogen bonding interactions between the active-site aspartates, APP and a water molecule. Active site Asp257 and Asp385 moved closer to form a hydrogen bond between the protonated Asp257 and the carbonyl oxygen in Leu49 of the scissile amide bond in APP **(Figure 1B)**. The two aspartates were ~7 Å apart between their Cγ atoms. Water entered the enzyme active site from the intracellular side and formed hydrogen bonds with the aspartates. The hydrogen bonds with the catalytic aspartates activated the water needed for nucleophilic attack of the carbonyl carbon of the scissile amide bond in APP. The distance between the carbonyl carbon of Leu49 and water oxygen was ~3.8 Å. This active conformation is well poised for ε cleavage of the amide bond between residues Leu49 and Val50 of APP.

RMSFs were calculated from GaMD simulations of the enzyme-substrate complex **(Figure S5)**. In nicastrin, extracellular helices α1, α2, α4a, the C-terminal regions of α5, α12, α17, and TM domain exhibited high fluctuations with ~3 Å RMSF **(Figure S5A)**. The TM6 and Helix-8 of APH1 were also flexible during the simulations. In PS1, TM2 extracellular domain, TM6 and TM6a were flexible with ~2.5-3 Å RMSF. Through structural clustering of GaMD simulation snapshots **(see Methods),** the top cluster was obtained as the representative wildtype active conformation of the enzyme. Relative to the cryo-EM structure, the extracellular end of TM2 moved outwards by ~2.5 Å **(Figure 1C)** and TM6a moved upwards by ~2 Å **(Figure 1D)**. Conformational changes of these domains involved a significant number

of PS1 FAD mutation sites, including Gln127, Arg128, Ser132, Pro264, Pro267, Arg269, Leu271, Val272, Glu273 and Thr274 (www.alzforum.org). Interestingly, His131 from TM2 and Cys263 from TM6a flipped their side chains. The N-terminal helix region of APP moved outwards by ~10 Å during enzyme activation (**Figure 1A**), while the C-terminal β-strand of APP moved by ~6 Å to interact with the PS1 TM6a helix. In the process, APP residue Leu52 made new contacts with residues Val272 and Ala275 in TM6a of PS1 **(Figure 1E)**. The movement was consistent with previous finding that TM6a undergoes large conformational change upon substrate binding and plays a key role in activation of the enzyme.[20] In addition, the intracellular ends of TM8 and TM1 moved by ~4.5 Å and ~3.5 Å, respectively (**Figure 1F**). Residues Ser104, Phe105 and Tyr106 in the N-terminal region of PS1 HL1 changed into a helical conformation during enzyme activation **(Figures S6B** and **S6F)**. In summary, we have captured activation of computationally restored wildtype γ-secretase bound by wildtype APP in the GaMD simulations.

GaMD simulations correlated with biochemical experiments on cleavage of wildtype and mutant APP. MS experiments were carried out to analyze AICD species (AICD49-99 and AICD50-99) generated in proteolysis of the wildtype APP and three mutants (I45F, T48P and M51F) by γ-secretase assay **(Figures 2A-2D)**. For the wildtype APP, the MALDI-TOF analysis showed the presence of both AICD species, but the AICD50-99 species had relatively higher intensity than the AICD49-99 species **(Figure 2A).** The difference in the amount of AICD fragments suggested that the γ-secretase preferred ε cleavage between Leu49-Val50 to the cleavage between Thr48-Leu49 in the wildtype APP, as has been previously reported.[18] Such experimental data correlated well with GaMD simulations of γ-secretase with the wildtype APP

substrate, during which the activated enzyme was poised to cleave wildtype APP between Leu49-Val50 **(Figure 1B)**.

During activation, the wildtype APP-bound γ-secretase also sampled "inhibited' and "intermediate" low-energy states as identified from the GaMD reweighted free energy profile **(Figure 2E)**. In the inhibited state, the catalytic aspartates moved very close to each other, with only ~4 Å distance between the Cγ atoms, while the substrate was ~6 Å away from the active site. This conformation could not accommodate water between the aspartates to form hydrogen bonds. A similar inhibited state of the enzyme was also observed in the dipeptidic inhibitor N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester (DAPT)-bound cryo-EM structure (PDB: 5FN2).[47] With the enzyme active site in the inhibited state, APP substrate moved away from the catalytic aspartates in the GaMD simulations. The carbonyl oxygen in Leu49 of APP was ~6 Å from the protonated oxygen of Asp257 **(Figure 2E).** In the intermediate state, the Cγ atoms of the catalytic aspartates were ~6 Å apart, while the carbonyl oxygen of Leu49 of the APP substrate was ~7 Å away from the Asp257 oxygen **(Figure 2E)**.

For the I45F and T48P mutants of APP substrate, the MS analysis showed decreased amount of the AICD49-99 species from proteolysis of both mutants compared with the wildtype substrate, and AICD50-99 was the predominant AICD product **(Figures 2B and 2C)**. Thus ε cleavage between Leu49-Val50 was even more preferred for these two mutants than for the wildtype substrate. In parallel with the experiments, further GaMD simulations were performed on γ-secretase bound by the I45F and T48P mutant APP substrates. The I45F mutant substrate-bound γ-secretase became activated during 1.1 μs GaMD simulations **(Figure S4B).** A low-energy conformation was observed in the I45F active state for which the distance between the Cγ atoms of Asp257 and Asp385 was ~7 Å, while the Leu49 carbonyl oxygen and Asp257

protonated oxygen formed a hydrogen bond with ~3 Å distance **(Figure 2F)**. The boost potential was 10.30 ± 6.79 kcal/mol in the GaMD simulations of the I45F mutant substrate-bound enzyme, which was comparable to that of the wildtype system (10.45 ± 6.78 kcal/mol) **(Table 1)**. However, the I45F-mutant APP substrate-bound γ-secretase was activated within shorter simulation time compared with the wildtype APP substrate-bound enzyme, with higher probability of conformations for the ε cleavage between Leu49-Val50 in APP **(Figures S4A and S4B)**. The simulation findings agreed well with the experimental data. Analysis of AICD products by MALDI mass spectrometry revealed a higher peak intensity of AICD50-99 than AICD49-99 for the I45F mutant APP compared with the wildtype APP **(Figures 2A and 2B)**. Another low-energy conformation of I45F APP substrate-bound enzyme was observed in the inhibited state **(Figure 2F)**, being similar to the inhibitor DAPT-bound structure of γ-secretase (PDB: 5FN2)[47].

For the T48P-mutant APP substrate-bound γ-secretase, activation was observed during one of three 1.5 μs GaMD simulations **(Figure S4C).** Low-energy conformations were identified from the free energy profile in the active, inhibited and intermediate states **(Figure 2G)**. In the T48P active state, the catalytic aspartates were positioned ~7 Å apart (Cγ to Cγ), and the substrate Leu49 carbonyl oxygen aligned with the protonated oxygen of Asp257 to form a hydrogen bond. The boost potential was 10.87 ± 6.87 kcal/mol, which was also comparable to that of wildtype APP simulations **(Table 1)**. The T48P-mutant APP substrate-bound γ-secretase transitioned into the active state within shorter simulation time compared to the wildtype system **(Figure S4A and S4C)**. The T48P APP substrate mutant had a higher probability than the wildtype APP substrate of aligning the aspartates and water with the scissile amide bond between Leu49-Val50 in APP. This computational finding was again consistent

with MALDI mass spectrometric analysis of AICD products: AICD50-99 intensity is higher than AICD49-99 for the T48P-mutant substrate compared to that of wildtype system **(Figures 2A and 2C)**. The observed inhibited state (**Figure 2G**) was similar to that seen in the wildtype and I45F systems **(Figures 2E and 2F)** as well as the inhibitor DAPT-bound cryo-EM structure of γ-secretase (PDB: 5FN2).[47] In the T48P intermediate state, the Cγ atoms of the catalytic aspartates were ~6 Å apart, while the Leu49 carbonyl oxygen of APP substrate and the protonated oxygen of Asp257 were ~7 Å apart **(Figure 2G)**. The I45F and T48P-mutant APP substrate-bound γ-secretase showed similar structural flexibility as the wildtype system in the RMSFs calculated from GaMD simulations **(Figures S5A).** In both systems, extracellular helices α1, α2, α4a, the C-terminal regions of α5, α12, α17, and TM domain of nicastrin exhibited high fluctuations. The TM6 and Helix-8 of APH-1, and the TM2 extracellular domain, TM6 and TM6a of PS1 were also flexible during the GaMD simulations **(Figures S5C and S5D)**.

Shifted ε cleavage site of APP in the M51F mutant

MS analysis of AICD products from the M51F-mutant system revealed AICD49-99 as the major product, suggesting that the predominant ε cleavage site of M51F APP was between residues Thr48-Leu49 **(Figure 2D).** A low level of AICD48-99 was also detected, revealing that the M51F APP substrate was cleaved to a limited degree between Ile47-Thr48 **(Figure 2D).** This was consistent with previous studies that a Phe residue is not tolerated in the P2' position of substrate or transition-state analogue inhibitors of γ-secretase.[18,48] Thus, M51F mutation of the APP substrate shifted the ε cleavage site from Leu49-Val50 to Thr48-Leu49. Such shift of ε

cleavage was consistently observed in 1.5 μs GaMD simulations of the M51F-mutant APP substrate bound to γ-secretase **(Figure S4D** and **Movie S3).** The protonated oxygen of PS1 Asp257 was hydrogen bonded to the carbonyl oxygen of Thr48, and the activated water molecule targeted the scissile amide bond between Thr48 and Leu49 in the M51F APP mutant for ε cleavage. In comparison, residue Thr48 in the wildtype APP maintained a distance of ~8-9 Å between its carbonyl oxygen and the protonated oxygen of the PS1 Asp257 **(Figure S4E)**. A distinct low-energy state was identified for the "shifted" conformation in the free energy profile of the M51F APP system **(Figure 2H)**. In the shifted state, the Cγ atoms of the catalytic aspartates were ~7 Å apart, and the carbonyl oxygen of APP substrate Thr48 and protonated oxygen of PS1 Asp257 formed a hydrogen bond with ~3 Å distance. Moreover, in one of the three GaMD simulations, the ε cleavage site of M51F-mutant APP was further shifted to the amide bond between Ile47-Thr48. The distance between the carbonyl oxygen of APP substrate residue Ile47 and the protonated oxygen of PS1 Asp257 became ~3 Å **(Figure S7)**. The Cγ atoms of the catalytic aspartates were ~7 Å apart. This observation was consistent with the low level of AICD48-99 fragment detected by MS of AICD products of the M51F APP mutant **(Figure 2D)**.

In addition to the MS experiments, the effect of APP mutations was investigated by detecting the total amount of Flag-tagged AICD species in *in-vitro* γ-secretase assays by western blotting using anti-Flag antibodies (**Figure 2I)**. The AICD production increased substantially for the M51F mutant substrate compared to wildtype APP substrate. In contrast, ε proteolysis of the I45F and T48P mutants showed no drastic change in the total AICD level compared with wildtype APP substrate **(Figure 2I)**. This was highly consistent with the GaMD simulations. In the systems with wildtype, I45F and T48P mutant APP substrate, the low-energy

inhibited state was observed in the free energy profiles of γ-secretase, but not for M51F-mutant system **(Figures 2E-2H)**.

Structural clustering was performed on GaMD simulations of M51F APP-bound γ-secretase, and the top cluster was identified as the shifted conformational state of the enzyme. Compared with the wildtype active conformation **(Figure 3A)**, the extracellular end of TM2 in PS1 moved outwards by ~2.5 Å **(Figure 3B)**. The helix involving residues Thr124, Val125, Gly126 and Gln127 became disordered in this region **(Figure 3B).** The APP substrate moved downwards by ~4 Å in the substrate binding channel of the enzyme **(Figures 3A and 3C)**. In comparison, the catalytic aspartates and flanking regions of TM6 and TM7 moved less than APP **(Figures 3C and 3E)**. Upon shifting of the ε cleavage site, local rearrangements of APP residues were required to establish the coordinated hydrogen bonding interactions at the active site **(Figure 3C** and **Movie S3)**. Sidechain flipping of the APP Thr48 residue led to formation of a hydrogen bond between its carbonyl oxygen and the PS1 Asp257 protonated oxygen (**Figure S4D**). Residue Leu49 initially facing the activated water between these catalytic aspartates flipped the sidechain and moved downwards by ~4 Å. The PS1 TM6 helix in the M51F shifted state moved towards the active site by ~4 Å relative to the wildtype active conformation **(Figure 3A)**. Moreover, the TM6a helix tilted by ~60° relative to the wildtype active conformation **(Figures 3A and 3D)**. Meanwhile, the β-strand at the C-terminus of APP deformed to a turn as it moved away from the TM6a helix. APP Leu52, interacting with the non-polar residues of TM6a in the wildtype active conformation, flipped its side chain and moved away from these residues by ~6 Å **(Figure 3E)**. The intracellular domains of TM1 and TM8 in the M51F shifted state moved by ~2.5 Å compared with the wildtype active conformation **(Fig 3F)**. PS1 FAD mutation sites Ala79, Val82, Ile83, Met84, Leu85, Pro88,

Leu424 and Ala426 from TM1 and TM8 showed similar movements of their sidechains. In addition, RMSF of the M51F-mutant APP-bound γ-secretase calculated from GaMD simulations showed higher flexibility in TM2, TM6 and TM6a regions of PS1 **(Figures S5B)**. This extra flexibility is consistent with the ability of the M51F-mutant system to readjust the positioning of the substrate in the active site in shifting the ε cleavage site.

Changes in secondary structures of APP substrate mutants

Changes in secondary structures of the wildtype and mutant APP substrate in γ-secretase were monitored during the GaMD simulations **(Figures 4, S8** and **S9)**. Secondary structures of APP substrate in the active conformations of the wildtype, I45F and T48P mutant systems and the shifted active conformation of the M51F mutant system were compared using their top ranked structural clusters obtained from the corresponding simulations **(Figure S10)**. For wildtype APP substrate, residues Gly29 to Val46 formed a helical conformation throughout the simulations except between residues 42 and 43 **(Figure 4A)**. Residues Asn27-Lys28 fluctuated between turn and coil conformations during last ~700 ns for activation whereas Ile47-Leu49 fluctuated between helix and turn conformations throughout the simulation. The N-terminal region of APP substrate was very flexible and sampled turn and coil conformations. The C-terminal residues Leu52 to Lys55 primarily maintained an antiparallel β-sheet conformation. Residues Val50-Met51, immediately after the Leu49-Val50 ε cleavage site, formed a turn for a number of times that exposed this APP scissile amide bond to the enzyme aspartates and coordinated water for proteolysis.

The I45F and T48P mutants of APP substrate, which maintained ε cleavage between Leu49-Val50, showed similar secondary structures as the wildtype substrate, although unique features were also observed in each mutant. Both mutant substrates formed turns at residues Val50-Met51 despite fluctuations **(Figures 4B** and **4C)** and adopted a β-sheet conformation at the C-terminus during simulations. However, only residues Ile31-Val46 formed a helix in the I45F mutant, with ~2-3 residues towards the N-terminus losing the helical conformation **(Figure 4B)**. The N-terminal region of I45F-mutant APP substrate was thus more flexible than the wildtype substrate and bent over the HL1 loop of PS1 **(Figures S6C, S6F** and **S11)**. For the I45F APP mutant substrate, multiple hydrogen bonds were formed between the N-terminal residues of APP and PS1 HL1. Residue Gln112 in the PS1 HL1 loop—mutated to Cys112 in the cryo-EM structure to generate a disulfide bond and restored in our simulation—formed three hydrogen bonds with residues Ser26, Asn27 and Lys28 of the APP N-terminus **(Figure S11A)**. Two of these hydrogen bonds involved backbone atoms. In addition, the backbone N atom of Ile114 in PS1 HL1 formed a hydrogen bond with the backbone O atom of Lys28 of APP. These hydrogen bonds contributed to a parallel β-sheet between the PS1 HL1 and APP N-terminus as reflected in the secondary structure plots **(Figures 4B, S8, S11A)**. In contrast, the N-terminal loop of wildtype APP was observed flexible without bending over the PS1 HL1 **(Figures S6B** and **S6F).** For the T48P-mutant APP, residues Gly29-Ala42 formed a helical conformation, whereas residues Thr43-Ile47 fluctuated between the α-helix, 3-10 helix and turn conformations **(Figure 4C)**. Residues Ser104, Phe105, Tyr106 and Thr107 of the PS1 HL1 loop formed a helix in the T48P active conformation, similar to what was observed in the wildtype active state **(Figure S6D** and **S6F)**.

For the M51F mutant APP, residues Ala30-Val46 formed a helical conformation during the GaMD simulations **(Figure 4D).** A longer turn appeared starting from residue Leu49 to Leu52 in M51F APP during the simulations **(Figures 4D, 3E** and **S10)**. In comparison, a turn was formed for only residues Val50-Met51 in the wildtype APP that exposed the Leu49-Val50 scissile amide bond for ε cleavage (**Figures 4A** and **S10**). The shift of this turn correlated with the shift of the ε cleavage site. The C-terminal β-strand became shorter in the M51F APP substrate mutant **(Figure 4D)** and even completely disappeared in the representative M51F shifted active conformational state **(Figures S10** and **3E)**. As the M51F-mutant APP substrate moved downwards relative to PS1, its N-terminus formed more interactions with PS1 HL1 **(Figures S6E** and **S6F).** The backbone O and N atoms of Gly111 in HL1 often formed hydrogen bonds with the backbone atoms N of Glu22 and O of Phe20 in the APP substrate, respectively **(Figure S6E, S6F** and **S11B)**. These hydrogen bonds resulted in a parallel β-sheet conformation **(Figure 4D)**. The N-terminus of the T48P-mutant APP was also found in proximity with the PS1 HL1 loop **(Figure S6D** and **S6F).** The PS1 HL1 loop—with high flexibility and multiple interactions with APP substrate—make it one of the most important regions of PS1 in the context of enzymatic function and Alzheimer's disease pathogenesis[49,50]. The PS1 HL1 has a large number of FAD mutation sites, including Phe105, Gly111, Leu113, Tyr115, and Gln127. Hence, our simulation findings were consistent with the literature regarding importance of the HL1 loop.

The C-terminus of APP substrate-bound to the active wildtype conformation moved towards PS1 TM6a region by ~6 Å during enzyme activation **(Figure 1E)**. The C-terminus of APP maintained β-sheet conformations with the N-terminus of PS1 TM7 throughout the simulations. Hence, with the movement of C-terminus of APP, the N-terminus of PS1 TM7 also

moved along by ~6 Å **(Figure S12)**. In contrast, I45F and T48P mutant APPs maintained the β-sheet conformations with the N-terminus of PS1 TM7 without the movement of C-terminus. M51F mutant APP lost its interaction with the PS1 TM7 and hence losing the β-sheet conformations **(Figures 4D** and **3E).**

Comparison of the S1', S2' and S3' active-site subpockets in the wildtype and mutant APP substrate-bound γ-secretase. Representative active conformations of PS1 were identified as the top ranked structural clusters from the GaMD simulations of the wildtype, I45F and T48P mutant systems and the shifted active conformation from the M51F system simulations. These conformations were aligned and compared for the enzyme active-site S1', S2' and S3' subpockets that were occupied by APP substrate residues P1', P2' and P3', respectively[18] **(Figure 5)**. In the active wildtype conformation, the S1' subpocket occupied by P1' residue (Val50) constituted residues mostly from TM6 and TM6a as listed in **Table S1.** The S3'subpocket occupied by P3' residue (Leu52) constituted residues mostly from TM6a and the C-terminus of PS1-NTF. The S1' and S3' subpockets were located on the same side with respect to APP **(Figures 5A).** In contrast, the S2' subpocket occupied by P2' residue (Met51) constituted residues mostly from TM8, TM8-TM9 loop and the β-strand region of TM7 **(Table S3).**

In the I45F and T48P active conformations **(Figures 5A, 5B and 5C),** the S1' and S3' subpockets occupied by the P1' (Val50) and P3' (Leu52) residues, respectively, embodied the same S1' and S3' sub pockets of the wildtype active conformation. The S2' pocket occupied by P2' (Met51) of the I45F and T48P mutant APP substrate comprised residues from TM8, the

TM8-TM9 loop, the β-strand region of TM7 and part of TM1. Notably, both the S1' and S2'subpockets involved the PAL motif (P333-A434-L435) in the TM9 N-terminal region that is considered important for substrate binding[51]. This S2' subpocket occupied by the APP mutants was located on the same side of substrate but ~4 Å above the extended S2' subpocket in the wildtype active conformation.

For the M51F APP mutant, the presence of a bulky residue Phe at the P2' position induced local rearrangements and shifted the ε cleavage site. With the shift, Leu49, Val50 and Phe51 became the new P1', P2' and P3' residues, respectively. The new P1' residue occupied a distinct subpocket near to the S1' subpocket in the wildtype active conformation **(Figures 5A, 5D** and **5E).** In contrast, the new P2' residue occupied a new subpocket in the space between the S1' and S3' subpockets of the wildtype active conformation of PS1. The new P3' residue occupied the same extended pocket as S2' subpocket in the I45F active and T48P active conformations (**Figures 5B, 5C** and **5E**). Hence, the new subpocket occupied by the P3' residue (F51) is termed "shifted S3' subpocket" here and also involved the PAL motif **(Figures 5D** and **5E)**. Moreover, the L52 (P4') residue and K53 (P5') residue in the M51F shifted active conformation occupied the S2' and S3' subpockets as in the wildtype active conformation of PS1, respectively **(Figures 5D** and **5E).**

The location of the S2' subpocket differed among the active conformations of the wildtype active, I45F active and T48P active conformations. As the C-terminus of I45F and T48P mutant APP moved by ~6 Å compared with the wildtype APP **(Figure S12)**, the P2' residue (Met51) of these mutants occupied a different S2' subpocket **(Figures 5)**. Due to shift in the ε cleavage site, the C-terminus of M51F mutant APP lost interactions with the N-terminus of PS1 TM7 and PS1 TM6a. This resulted in large conformational tilting of PS1 TM6a helix in

the M51F shifted conformation **(Figure 3).** Therefore, the conformational changes and molecular interactions of the APP with the γ-secretase provided important insights into the mechanisms of activation and substrate processing by the enzyme.

Discussion

The PS1-containing γ-secretase complex is a founding member of intramembrane-cleaving proteases (I-CLiPs) which carry out hydrolysis of substrate TM domains within the hydrophobic environment of the lipid bilayer.[52] I-CLiPs also include the S2P metalloproteases, rhomboid serine proteases, and presenilin-like aspartyl proteases. Although microbial representatives of each of these other I-CLiP classes have been crystallized for high-resolution structure determination [53-56], visualizing the active state and elucidating the molecular mechanism of intramembrane proteolysis has been challenging. Only very recently has a rhomboid protease been studied through time-resolved x-ray crystallography to reveal how this serine protease hydrolyzes transmembrane substrates.[57] Most recently, structures of the γ-secretase complex bound to Notch and APP substrates have been reported, providing critical insights into substrate recognition of γ-secretase.[20,58] Nevertheless, mutations in the enzyme and substrate needed for stabilization of the substrate-enzyme complex precluded visualization of the active protease and raised the possibility of unnatural substrate interactions.

Using the latest cryo-EM structures, we have, for the first time, developed an all-atom MD model for activation of the APP substrate-bound γ-secretase poised for intramembrane proteolysis that is in excellent agreement with mass spectrometry and western blotting biochemical experiments. Extensive simulations using a novel GaMD enhanced sampling

method have captured spontaneous activation of γ-secretase in the presence of APP and water (**Figure 6**). The catalytic aspartates moved into close proximity, similar to previous simulation findings,[23-25] although, these studies were performed without the APP substrate bound to the γ-secretase active site. Previous studies suggested a putative active conformation of the apo γ-secretase but was unable to fully characterize the enzyme activation involving additional coordinated hydrogen bond interactions with the substrate. In the GaMD simulations, water molecules entered the active site, one of which coordinated with the two aspartates (**Figure 6B** and **Movies S1** and **S2).** Moreover, Asp257 formed a hydrogen bond with the carbonyl oxygen of the scissile amide bond between APP residues Leu49-Val50. The activated water molecule was poised for nucleophilic attack on the backbone carbon atom of this activated Leu49-Val50 amide bond. While a number of regions of nicastrin, Aph-1, and Pen-2 displayed flexibility during simulations of the activated enzyme-substrate complex, the PS1 TM6a was the most noteworthy, as this region interacted directly with substrate near the cleavage site and appeared to play a role in enzyme activation. The wildtype enzyme-substrate complex additionally sampled the inhibited and intermediate conformational states, the former closely resembling the conformation of the DAPT inhibitor-bound γ-secretase.[47] The current ~2-μs GaMD simulation of γ-secretase with wildtype APP captured the enzyme activation for ε cleavage of APP between Leu49-Val50. The ε cleavage of wildtype APP between Thr48-Leu49 with lower probability, as detected by MS, would likely require longer simulation time and more sufficiently sampling.

GaMD simulations on I45F and T48P APP substrate-bound γ-secretase revealed faster activation of PS1 for proteolysis at the ε cleavage site between Leu49-Val50 with these two FAD mutations compared to the complex with wildtype APP substrate. These observations were consistent with MS analysis of AICD proteolytic products: the two FAD-mutant substrates

were cleaved by γ-secretase with a greater AICD50-99/AICD49-99 ratio than was the wildtype substrate. Moreover, the M51F mutation resulted in dramatic conformational changes of APP **(Figure 6C** and **Movie S3)**, setting up ε cleavage between Thr48-Leu49. These results were entirely consistent with the known incompatibility of Phe in the P2' position.[18] MS experimental results also showed the major AICD product generated by γ-secretase from the M51F-mutant APP substrate was due to cleavage between Thr48-Leu49. Little or no cleavage occurred between Leu49-Val50. In addition, western blotting revealed a substantial increase in the total AICD production in the *in-vitro* γ-secretase assay for the M51F-mutant APP substrate compared to the wildtype APP substrate. In contrast, I45F and T48P mutant APP-bound γ-secretase showed similar amount of AICD production as the wildtype APP bound γ-secretase. This was in exceptional agreement with the GaMD simulation: the low-energy inhibited state was observed in the free energy profiles of the wildtype, I45F and T48P mutant APP bound γ-secretase, but absent in the M51F mutant APP system. These strong correlations between the GaMD simulations and biochemical experiments provided substantial validity to our dynamic model of γ-secretase.

The active-site S1', S2' and S3' subpockets were visualized in the wildtype active, I45F active, T48P active and M51F shifted active conformations of PS1 obtained from GaMD simulations. The protein residues **(Figure 5** and **Table S3)** found in the S1' and S3' subpockets of the active wildtype, I45F and T48P conformations were the same as those identified in a recent computational study by Hitzenberger et.al.[33] However, the S2' pocket of the wildtype active PS1 was identified in a distinct location that shifted by ~4 Å towards the APP C-terminus from the previously described S2' pocket.[33] The subpocket described by Hitzenberger et.al.[33], on the other hand, appeared to be the S2' pocket in the I45F and T48P active conformations

and the shifted S3' subpocket for the M51F APP **(Figure 5E)**. Shift of the S2' subpocket from the wildtype active conformation to the I45F and T48P active conformations resulted from the simultaneous movements of the APP C-terminus and PS1 TM7 N-terminus towards the PS1 TM6a in order to maintain the β-sheet structure of this domain in the GaMD simulations **(Figure S12).** Therefore, the GaMD simulations revealed a newly identified S2' subpocket for wildtype APP, while the previously described S2' subpocket[33] was used as the S2' for I45F and T48P APP as well as the shifted S3' for the M51F APP **(Figure 5).**

In summary, we have combined all-atom GaMD simulations with MS and western blotting experiments to probe the mechanisms of γ-secretase activation and its ε cleavage of the wildtype and mutant APP substrates. Extensive GaMD simulations using the latest cryo-EM structures of γ-secretase have captured spontaneous activation of the enzyme, for which the active-site Asp385 has been restored and the artificial enzyme-substrate disulfide bond has been removed. The active conformation is characterized by water-bridged hydrogen bonds between the two catalytic aspartates, one of which formed another hydrogen bond with the carbonyl oxygen of the target scissile amide bond for the ε cleavage of APP. Free energy calculations of the GaMD simulations also allowed us to identify distinct intermediate, inhibited and shifted active conformational states of γ-secretase. The simulations predicted ε cleavage preferences of the wildtype and three mutants of APP that were highly consistent with MS and western blotting experimental findings of the AICD species. The validated GaMD simulations were then used to interpret the experimental data at an atomistic level. Remarkably, the M51F mutation shifted the ε cleavage site of APP from the amide bond between Leu49-Val50 to the Thr48-Leu49 bond, generating predominantly the AICD49-99 fragment instead of the AICD50-99 as detected by MS. Finally, the GaMD simulations have systematically revealed the active-site S1', S2' and

S3’ subpockets of γ-secretase that interacted with the P1’, P2’ and P3’ residues in the wildtype and mutant APP. This provides an in-depth picture of the ε proteolytic cleavage of different APP substrates by γ-secretase. The GaMD method is apparently very well suited for the study of this extremely slow-acting membrane protease complex. In order to fully understand the functional mechanisms of γ-secretase, further simulation and experimental studies have been planned on the tripeptidase activity of the enzyme and effects of FAD mutations in both the APP substrate and γ-secretase. These studies are expected to greatly facilitate rational drug design targeting γ-secretase for the AD therapeutic treatments.

Materials and Methods

Cloning

All mutations in C100 FLAG were introduced by site-directed mutagenesis (QuickChange Lightning Site Directed Mutagenesis kit, Agilent) in pET 22b vector. All constructs were verified by sequencing by ACGT.

C100-FLAG substrates purification

E. coli BL21 cells were grown in LB media until OD_{600} reached 0.6. Cells were induced with 0.5 mM IPTG and were grown post induction for 4 hours. The cells were then pelleted by centrifugation and resuspended in 50 mM HEPES pH 8, 1% Triton X-100. The cells were lysed by French press and the lysate was incubated with anti-FLAG M2-agarose beads from SIGMA.

Bound substrates were then eluted from the beads with 100 mM Glycine pH 2.5, 0.25% NP-40 detergent and then neutralized with Tris HCl prior to being stored at -80^{0}C.

γ-secretase expression and purification

γ-secretase was expressed in HEK 395F cells by transfection with pMLINK vector containing all four components (Presenilin-1, Pen-2, Aph-1, Nicastrin) of g-secretase complex (provided by Yigong Shi). For transfection, HEK 395F cells were grown in unsupplemented Freestyle 293 media (Life Technologies, 12338-018) until cell density reached 2x10^6 cells/ml. 150 mg of vector was mixed with 450 mg of 25 kDa linear polyethylemimines (PEI) and incubated for 30 minutes at room temperature. The DNA-PEI mixtures were added to HEK cells and cells were grown for 60 hours. The cells were harvested, and g-secretase was purified as described previously.[18]

In vitro γ-secretase assay and detection of AICD species

γ -secretase purification and assays were carried out as described previously.[18] Briefly, 30 nM purified γ-secretase was dissolved into total brain lipid extract (Avanti) in 50 mM HEPES pH 7.0, 150 mM NaCl, 0.25% CHAPSO. The detergent/lipid/enzyme solution was mixed with SM-2 bio-beads (Bio-Rad) for 2 h at 4 ^{0}C to remove the detergent. After removal of the bio beads, the proteoliposome solution was mixed with 3 mM recombinant C100 substrates to initiate the cleavage reaction. The reaction was carried out for 16 h at 37 ^{0}C. After 16 h, AICD-Flag products were isolated by immunoprecipitation with anti-FLAG M2 beads (SIGMA) in 10 mM

MES pH 6.5, 10 mM NaCl, 0.05% DDM detergent overnight at 4 ^{0}C. AICD products were then eluted from the anti-FLAG beads with acetonitrile:water (1:1) with 0.1% trifluoroacetic acid. The elutes were run on a Bruker MALDI-TOF mass spectrometer.

Western blotting

Samples from γ-secretase assays were run on 4-12% bis-tris gel and transferred into PVDF membrane. The membrane was treated with 5% dry milk in PBS Tween-20 for 1 h at ambient temperature. The membrane was then incubated with the anti-Flag M2 antibodies at 4 ^{o}C overnight. The membrane was washed 3 times with PBS Tween-20 and was incubated with anti-mouse secondary antibodies for 1 h. The membrane was washed and imaged for chemiluminescence.

Gaussian Accelerated Molecular Dynamics (GaMD)

GaMD is an enhanced sampling technique, in which a harmonic boost potential is added to smooth the potential energy surface and reduce the system energy barriers.[4] GaMD is able to accelerate biomolecular simulations by orders of magnitude.[7,59] GaMD does not need predefined collective variables. Moreover, because GaMD boost potential follows a gaussian distribution, biomolecular free energy profiles can be properly recovered through cumulant expansion to the second order.[4] GaMD has successfully overcome the energetic reweighting problem in free energy calculations that was encountered in the previous accelerated molecular dynamics (aMD)method[37,39] for free energy calculations of large molecules. GaMD has been

implemented in widely used software packages including AMBER[4,60], NAMD[6] and GENESIS[61]. A brief summary of GaMD is provided here.

Consider a system with *N* atoms at positions $\vec{r} = \{\vec{r}_1, \cdots, \vec{r}_N\}$. When the system potential $V(\vec{r})$ is lower than a reference energy *E*, the modified potential $V^*(\vec{r})$ of the system is calculated as:

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}),$$

$$\Delta V(\vec{r}) = \begin{cases} \frac{1}{2}k\big(E - V(\vec{r})\big)^2, & V(\vec{r}) < E \\ 0, & V(\vec{r}) \geq E \end{cases} \qquad (1)$$

where *k* is the harmonic force constant. The two adjustable parameters *E* and *k* are automatically determined based on three enhanced sampling principles.[4] The reference energy needs to be set in the following range:

$$V_{max} \leq E \leq V_{min} + \frac{1}{k}, \qquad (2)$$

where V_{max} and V_{min} are the system minimum and maximum potential energies. To ensure that Eqn. (2) is valid, *k* has to satisfy: $k \leq \frac{1}{V_{max} - V_{min}}$ Let us define $\equiv k_0 \frac{1}{V_{max} - V_{min}}$, then $0 < k_0 \leq 1$. The standard deviation of ΔV needs to be small enough (i.e., narrow distribution) to ensure proper energetic reweighting[62]: $\sigma_{\Delta V} = k\big(E - V_{avg}\big)\sigma_V \leq \sigma_0$ where V_{avg} and σ_V are the average and standard deviation of the system potential energies, $\sigma_{\Delta V}$ is the standard deviation of ΔV with σ_0 as a user-specified upper limit (e.g., $10k_B$T) for proper reweighting. When *E* is set to the lower bound $E=V_{max}$, k_0 can be calculated as:

$$k_0 = \min(1.0, k_0') = \min\left(1.0, \frac{\sigma_0}{\sigma_V}\frac{V_{max} - V_{min}}{V_{max} - V_{avg}}\right). \qquad (3)$$

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

$$k_0 = k_0^{"} \equiv (1 - \frac{\sigma_0}{\sigma_V}) \frac{V_{max} - V_{min}}{V_{avg} - V_{min}}, \qquad (4)$$

if $k_0^{"}$ is found to be between *0* and *1*. Otherwise, k_0 is calculated using Eqn. (3).

Similar to aMD, GaMD provides schemes to add only the total potential boost ΔV_P, only dihedral potential boost ΔV_D, or the dual potential boost (both ΔV_P and ΔV_D). The dual-boost simulation generally provides higher acceleration than the other two types of simulations[63]. The simulation parameters comprise of the threshold energy E for applying boost potential and the effective harmonic force constants, k_{0P} and k_{0D} for the total and dihedral potential boost, respectively.

Energetic Reweighting of GaMD Simulations

To calculate potential of mean force (PMF)[64] from GaMD 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:

$$p(A_j) = p^*(A_j) \frac{\langle e^{\beta \Delta V(\vec{r})} \rangle_j}{\sum_{i=1}^{M} \langle p^*(A_i) e^{\beta \Delta V(\vec{r})} \rangle_i}, \; j = 1, \ldots, M, \qquad (5)$$

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:

$$\langle e^{\beta \Delta V(\vec{r})} \rangle = exp\left\{\sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k\right\}, \qquad (6)$$

where the first two cumulants are given by

$$\begin{aligned} C_1 &= \langle \Delta V \rangle, \\ C_2 &= \langle \Delta V^2 \rangle - \langle \Delta V \rangle^2 = \sigma_v^2. \end{aligned} \qquad (7)$$

The boost potential obtained from GaMD simulations usually follows near-Gaussian distribution. Cumulant expansion to the second order thus provides a good approximation for computing the reweighting factor[4,62]. The reweighted free energy $F(A) = -k_B T \ln p(A)$ is calculated as:

$$F(A) = F^*(A) - \sum_{k=1}^{2} \frac{\beta^k}{k!} C_k + F_c, \qquad (8)$$

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

System Setup

The earlier published cryo-EM structure of γ-secretase bound by Notch (PDB: 6IDF)[58] was used for initial GaMD simulations. This system was used to optimize our simulation protocol, especially the protonation state of aspartates in the active site. Another cryo-EM structure of γ-secretase bound by APP (PDB: 6IYC)[20] was used to perform further GaMD simulations as per the optimized protocol. For the wildtype enzyme, residue Asp385 that was mutated to Ala at the active site in the cryo-EM structure was restored for setting up the simulation system. Similarly, the disulfide bond between Cys112 of PS1-Q112C and Cys24 of APP-V24C were

removed, and the wildtype residues were restored for simulation setup. Five unresolved residues at the N-terminus of APP substrate C83 were added through homology modelling using SWISS-MODEL.[65] All chain termini were capped with neutral groups, i.e. the acetyl group (ACE) for the N-terminus and methyl amide group (CT3) for C terminus. Protein residues were set to the standard CHARMM protonation states at neutral pH with the *psfgen* plugin in VMD.[66] Then the complex was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer with all overlapping lipid molecules removed using the *Membrane* plugin in VMD[66] **(Figure S1)**. The system charges were then neutralized at 0.15 M NaCl using the *Solvate* plugin in VMD.[66] Periodic boundary conditions were applied on the simulation systems. The simulation systems of γ-secretase bound by APP are summarized in **Table 1**.

For APP-mutant simulations systems, isoleucine, threonine and methionine were mutated to phenylalanine, proline and phenylalanine computationally at the 29th, 32nd and 35th residue of APP substrate, respectively. These corresponded to I45F, T48P and M51F mutations as per the numbering based on C99, the substrate that was cleaved to Aβ, although the actual substrate in the model was C83.

Simulation Protocol

The CHARMM36 parameter set[67] was used for the protein and POPC lipids. Initial energy minimization and thermalization of the γ-secretase complex followed the same protocol as used in the previous GaMD simulations of membrane proteins.[7,68] The simulation proceeded with equilibration of lipid tails. With all the other atoms fixed, the lipid tails were energy minimized for 1000 steps using the conjugate gradient algorithm and melted with constant number,

volume, and temperature (NVT) run for 0.5 ns at 310 K. Each system was further equilibrated using constant number, pressure, and temperature (NPT) run at 1 atm and 310 ^{o}K for 10 ns with 5 kcal (mol Å^{2})$^{-1}$ harmonic position restraints applied to the protein. Further equilibration of the systems was performed using an NPT run at 1 atm and 310 ^{o}K for 0.5 ns with all atoms unrestrained. Conventional MD simulation was performed on each system for 10 ns at 1 atm pressure and 310 ^{o}K with a constant ratio constraint applied on the lipid bilayer in the X-Y plane. The GaMD simulations were carried out using AMBER 18.[4,60] Dual-boost GaMD simulations were performed to study the substrate-bound γ-secretase complex (**Table 1**). In the GaMD simulations, the threshold energy E for adding boost potential was set to the lower bound, i.e. $E = V_{max}$.[4,6] The simulations included 50 ns equilibration after adding the boost potential and then multiple independent production runs lasting 1-2 μs with randomized initial atomic velocities. GaMD production simulation frames were saved every 0.2 ps for analysis.

Simulation analysis

The VMD[66] and CPPTRAJ[69] tools were used for trajectory analysis. In particular, distance was calculated between the Cγ atoms of catalytic aspartate residues. Hydrogen bond distance was calculated between donor protonated oxygen atom of PS1 Asp257 and the acceptor carbonyl oxygen atom of APP substrate residue Leu49, Thr48 or Ile47. Root-mean-square fluctuations (RMSFs) were calculated for the protein residues, averaged over three independent GaMD simulations and color coded for schematic representation of each complex system. The CPPTRAJ was used to calculate the protein secondary structure plots. The *PyReweighting* toolkit[62] was applied to reweight GaMD simulations for free energy calculations by combining

all simulation trajectories for each system. A bin size of 1 Å was used for the PMF calculation of distances. The cutoff was set to 500 frames in each bin for calculating the 2D PMF profiles. Protein snapshots were taken every 1 ps for structural clustering. Clustering was performed on the GaMD simulations of wildtype, I45F, T48P and M51F-mutant APP bound γ-secretase based on the RMSD of PS1 using hierarchical agglomerative algorithm in CPPTRAJ[69] generating ~10 representative structural clusters for each system. The top structural cluster was identified as the representative active (wildtype) and shifted active conformational states of the wildtype and M51F mutant APP bound γ-secretase systems, respectively. The top structural cluster was also identified as the active (I45F and T48P) conformational state of the I45F and T48P mutant APP bound γ-secretase.

Restoration of wildtype γ-secretase for molecular dynamics simulations

We performed initial GaMD simulations on the earlier published cryo-EM structure of γ-secretase bound by the Notch substrate (PDB: 6IDF[58]). Residue Ala385 at the active site was mutated back to aspartate, whereas the disulfide bond between the N-terminus of Notch substrate and hydrophobic loop 1 (HL1) loop was kept intact. In aspartyl proteases, proximity between the two active site Asp residues necessitates protonation of one of them, preventing charge repulsion. Testing GaMD simulations were performed to determine which of the two Asp residues was protonated in γ-secretase. We performed multiple 300 ns GaMD simulations **(Table 1)** on three different systems: the original cryo-EM structure, Asp257 protonated, and with Asp385 protonated. The distance time course plots **(Figure S2)** revealed that the system with protonated Asp257 in the N-terminal fragment (NTF) subunit of PS1 facilitated the

activation of γ-secretase, as the two active-site aspartates approached each other to a distance of ~6-7 Å between the Cγ atoms. In contrast, simulations of the original cryo-EM structure did not show significant change from the starting Asp257:Cγ-Ala385:Cβ distance of ~10-11 Å. In the system with protonated Asp385 in the C-terminal fragment (CTF) of PS1, the two aspartates maintained a distance of ~10-11 Å between the Cγ atoms of the two aspartates. Therefore, Asp257 was protonated in subsequent GaMD simulations being similar to the setup of a previous computational study[24].

Next, we proceeded to simulate γ-secretase bound by APP (PDB: 6IYC[20]). Residue Ala385 was similarly mutated back to Asp385 in the wildtype γ-secretase, which was compared to the original cryo-EM system in 300 ns GaMD simulations **(Table 1)**. The disulfide bond between substrate and enzyme was still kept. Free energy calculations showed that the active-site residues Asp257 and Ala385 maintained ~10 Å distance between their sidechain terminal C atoms in the cryo-EM system even though water molecules were observed entering the active site **(Figure S3A)**. The substrate remained distant from the active site residues, with ~5-6 Å between the Cγ atom of protonated Asp257 and the carbonyl oxygen of Leu49 in APP. In contrast, activation of γ-secretase was observed during three independent 300 ns GaMD simulations of the computationally restored wildtype enzyme **(Figure S3B, Movie S1)**. The protonated Asp257 formed a hydrogen bond with the carbonyl oxygen in Leu49 of the scissile amide bond in APP. Water molecules entered the PS1 active site. One water molecule was trapped between the two catalytic Asp residues through stable hydrogen bonds. This would induce nucleophilic attack of the carbonyl carbon of Leu49 by the activated water molecule, which is a key step for substrate proteolysis. The enzyme active site was thus well poised for proteolysis of APP for the ε cleavage between residues Leu49 and Val50. The two aspartates

were ~7 Å apart between their Cγ atoms **(Figure S3B)**. The distance between the carbonyl carbon of Leu49 and the water oxygen was ~3.8 Å. Therefore, our GaMD simulations successfully captured activation of the APP-bound γ-secretase in the presence of the enzyme-substrate disulfide bond. Next, the artificial disulfide bond between the N-terminus of APP substrate and the HL1 loop of PS1 was also removed by computationally restoring the wildtype residues for further simulations as summarized in **Table 1.**

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 project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231, and the Research Computing Cluster at the University of Kansas. This work was supported in part by the startup funding in the College of Liberal Arts and Sciences at the University of Kansas (Y.M.) and GM122894 from the National Institutes of Health (M.S.W.).

Table 1: Summary of GaMD simulations performed on different systems of γ-secretase bound by the Notch and APP substrates.

Enzyme	Substrate	[a]Disulfide Bond	[b]N_{atoms}	Dimension (Å^3)	Simulation (ns)	[c]ΔV_{avg} (kcal/mol)	[d]$\sigma_{\Delta v}$ (kcal/mol)
D385A (Cryo-EM)	Notch	Present	240,021	141 x 124 x 146	300 x 1	12.91	7.91
D385-protonated	Notch	Present	240,358	141 x 124 x 146	300 x 3	9.97	6.76
D257-protonated	Notch	Present	240,358	141 x 124 x 146	300 x 3	10.36	6.46
D385A (Cryo-EM)	APP	Present	253,650	141 x 124 x 147	300 x 1	12.53	6.58
Wildtype	APP	Present	253,647	141 x 124 x 147	300 x 3	10.46	6.91s
Wildtype	APP	Absent	241,351	141 x 124 x 147	2000 x 3	10.45	6.78
Wildtype	I45F APP	Absent	241,355	141 x 124 x 147	1100 x 3	10.30	6.79
Wildtype	T48P APP	Absent	241,348	141 x 124 x 147	1400 x 3	10.87	6.87
Wildtype	M51F APP	Absent	241,360	141 x 124 x 147	1500 x 3	10.08	7.38

[a]The artificial disulfide bond between the N-terminus of APP and PS1 HL1 loop of the γ-secretase is kept ("Present") or removed ("Absent").
[b]N_{atoms} is the number of atoms in the simulation systems.
[c]ΔV_{avg} and [d]$\sigma_{\Delta V}$ are the average and standard deviation of the GaMD boost potential, respectively.

Figure 1

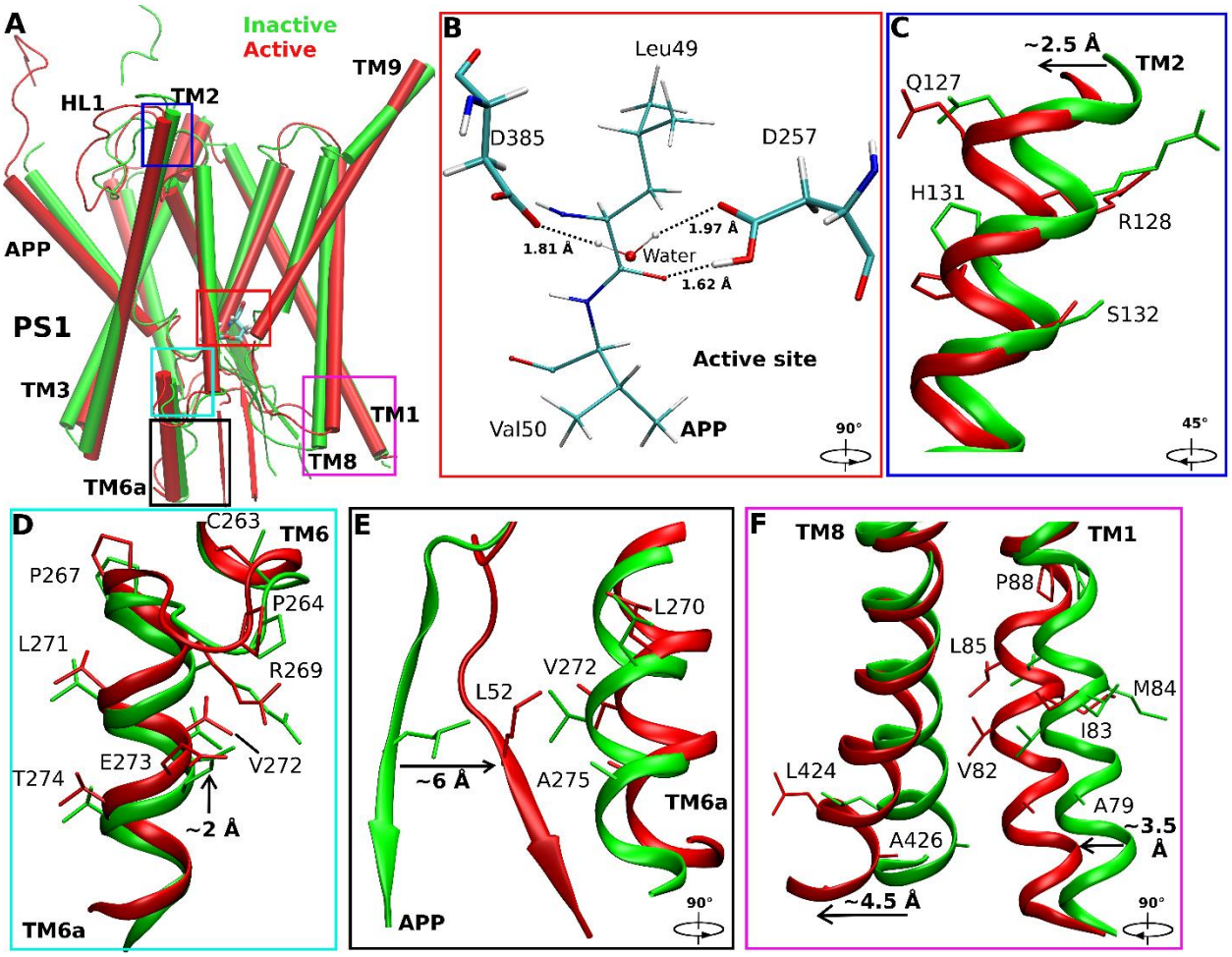

Figure 1: Conformational changes of the catalytic subunit presenilin (PS1) and APP substrate during activation of the computationally restored wildtype γ-secretase. (A) Comparison of the inactive cryo-EM structure (green) and wildtype active conformation of APP-bound PS1 (red). (B) The active site poised for proteolysis. Water entered the active site and formed hydrogen bonds with the catalytic aspartates, being ready for nucleophilic attack on the scissile amide bond between residues Leu49 and Val50 of APP for ε cleavage. (C-F) Conformational changes of (C) PS1 TM2, (D) PS1 TM6a, (E) the C-terminus of APP, (F) PS1 TM1 and PS1 TM8 during activation of γ-secretase. The extracellular end of TM2 moved outwards by ~2.5 Å in the active PS1 relative to the inactive cryo-EM structure. The PS1 TM6a moved upwards by ~2 Å compared to the cryo-EM structure. The C-terminal β-strand region of APP moved closer to interact with the PS1 TM6a helix. Residue Leu52 of APP moved by ~6 Å towards non-polar residues Val272, Leu270 and Ala275 in the enzyme TM6a. The intracellular ends of TM8 and TM1 moved from the cryo-EM structure by ~4.5 Å and ~3.5 Å, respectively.

Figure 2

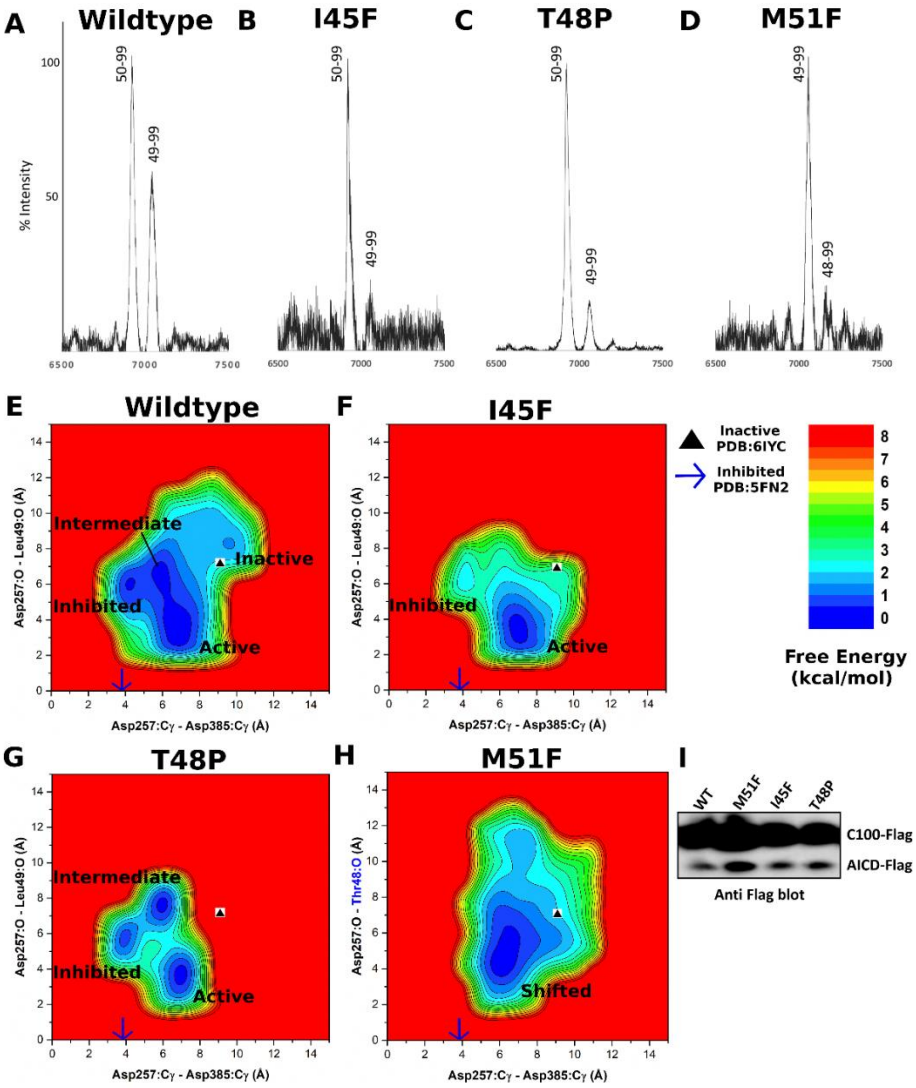

Figure 2: Mass spectrometry and western blotting of the APP intracellular domain (AICD) fragments and GaMD free energy profiles of wildtype and mutant APP-bound γ-secretase. (A-D) The intensity of different AICD fragments detected by mass spectrometry for (A) wildtype (AICD 50-99, expected mass: 6905.6 g/mol, observed mass: 6907.4 g/mol; AICD 49-99, expected mass: 7018.8 g/mol, observed mass: 7019.6 g/mol), (B) I45F (AICD 50-99, expected mass: 6905.6 g/mol, observed mass: 6905.4 g/mol; AICD 49-99, expected mass: 7018.8 g/mol, observed mass: 7019.8 g/mol), (C) T48P (AICD 50-99, expected mass: 6905.6 g/mol, observed mass: 6907.4 g/mol; AICD 49-99, expected mass: 7018.8 g/mol, observed mass: 7041.8 g/mol) and (D) M51F (AICD 49-99, expected mass: 7034.8 g/mol, observed mass: 7031.4 g/mol; AICD 48-99, expected mass: 7135.8 g/mol, observed mass: 7132.2 g/mol) APP substrate as cleaved by γ-secretase. (E-G) 2D free energy profiles of the Asp257:Cγ - Asp385:Cγ and Asp257:protonated O - Leu49:O distances calculated from GaMD simulations of (E) wildtype,

(F) I45F and (G) T48P APP substrate. (H) 2D free energy profile of the Asp257:Cγ - Asp385:Cγ and Asp257:protonated O – Thr48:O distances calculated from GaMD simulations of the M51F APP substrate. (I) The total amount of AICD species in γ-secretase determined *in vitro* by western blotting using anti-Flag antibodies of γ-secretase.

Figure 3

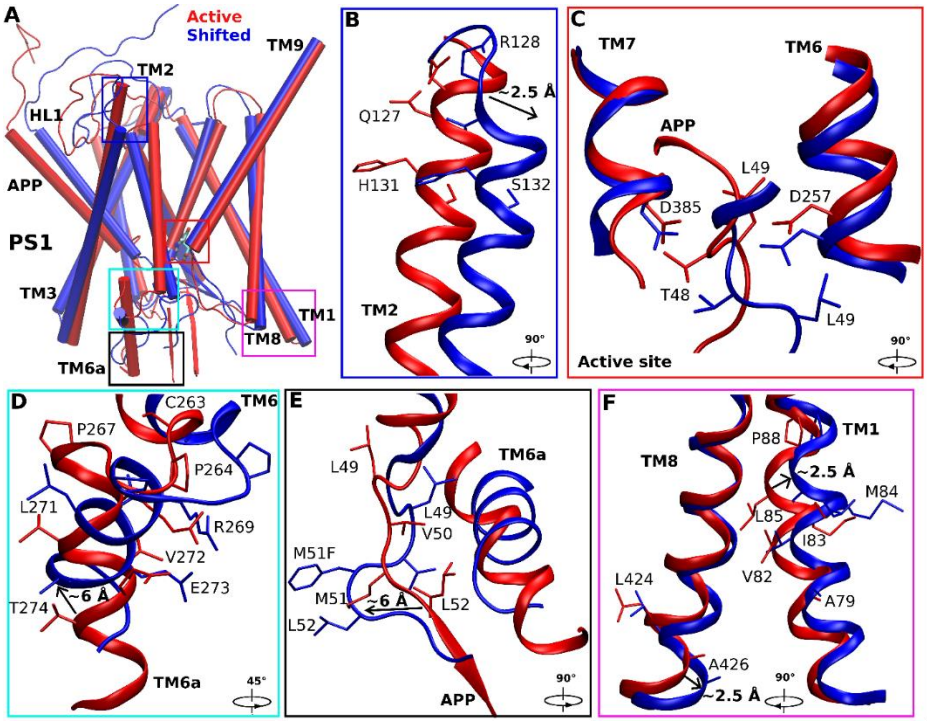

Figure 3: Conformational changes of the catalytic subunit presenilin (PS1) and APP in the shifted active (M51F) states of γ-secretase compared with the active (wildtype) state. (A) Overview of the active (red) and shifted active (blue) conformations of APP-bound PS1. (B) The extracellular end of the PS1 TM2 moved outwards by ~2.5 Å in the shifted active (M51F) conformation relative to the active (wildtype) structure. Residues Thr124, Val125, Gly126 and Gln127 in this region lost the helical conformation in the shifted active state. (C) The active site poised to attack the scissile amide bond between residues Leu49 and Val50 in the active state (red) and between residues Thr48 and Leu49 in the shifted active state (blue) of APP for ε cleavage. Sidechain flipping of the APP Thr48 residue led to formation of a hydrogen bond between its carbonyl oxygen and the PS1 Asp257 protonated oxygen. Residue Leu49 initially facing the center of two aspartates flipped to the other side with downward movement of ~4 Å. (D) The N-terminus of PS1 TM6 moved towards the active site by ~4 Å and the TM6a helix tilted by ~60°. (E) The β-strand at the C-terminus of APP substrate deformed to a turn as it moved away from TM6a in PS1. The APP Leu52 interacting with non-polar residues in PS1 TM6a in the active conformation flipped its side chain and moved in the opposite direction by ~6 Å. (F) The intracellular ends of TM1 and TM8 moved by ~2.5 Å and ~2.5 Å in PS1, respectively.

Figure 4

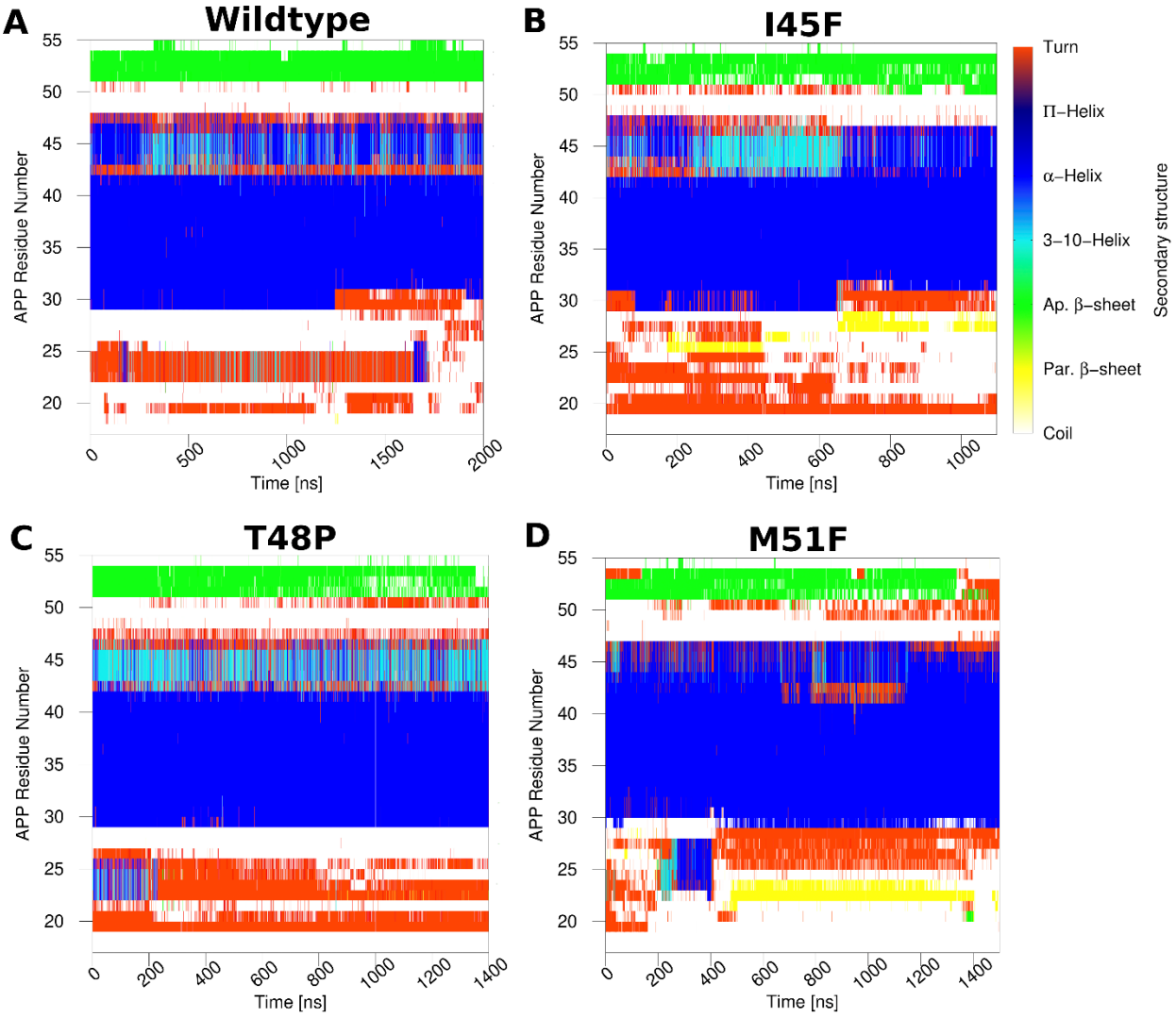

Figure 4: Time courses of the APP secondary structures in the (A) wildtype, (B) I45F, (C) T48P and (D) M51F forms as bound to γ-secretase calculated from their representative GaMD simulations. Results of the other simulations are plotted in **Figures S8** and **S9**.

Figure 5

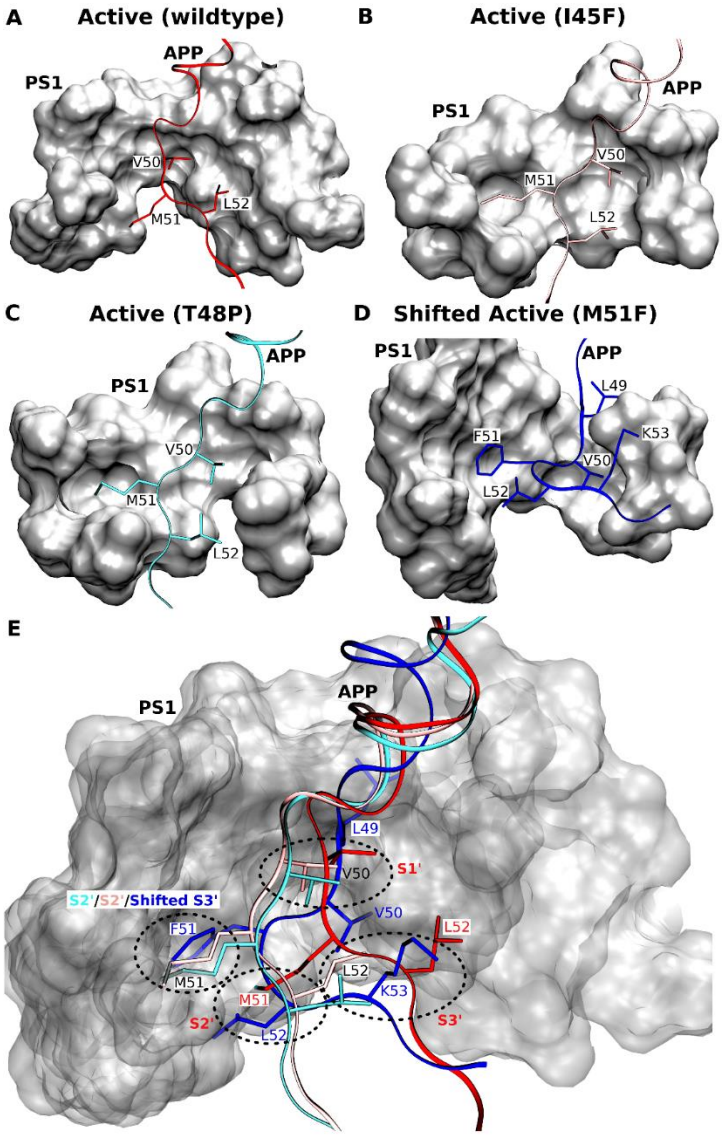

Figure 5: (A-D) Comparison of the locations of APP substrate residues P1', P2' and P3' in the (A) wildtype active, (B) I45F active, (C) T48P active and (D) shifted active M51F APP substrate-bound conformations of γ-secretase. (E) Comparison of the corresponding PS1 active-site S1', S2' and S3' pockets in these different conformational states of γ-secretase.

Figure 6:

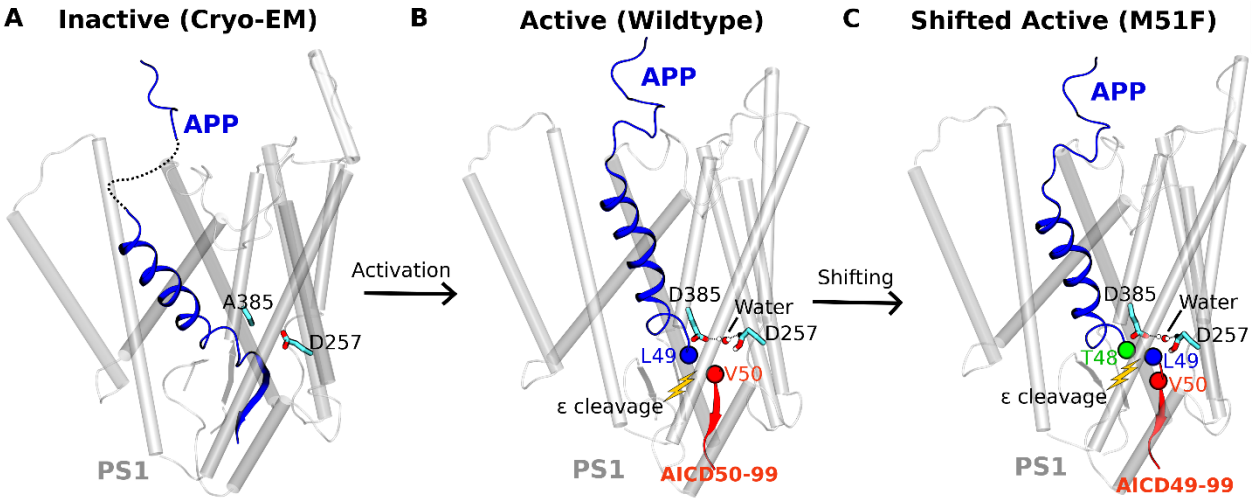

Figure 6: Summary of the (A) inactive cryo-EM, (B) active (wildtype) and (C) shifted active (M51F) conformational states of the APP substrate-bound γ-secretase. Distinct AICD products were generated from the wildtype and M51F mutant APP. The complementary simulations and experiments have revealed mechanisms of the γ-secretase activation and its ε cleavage of the APP substrate.

Appendix

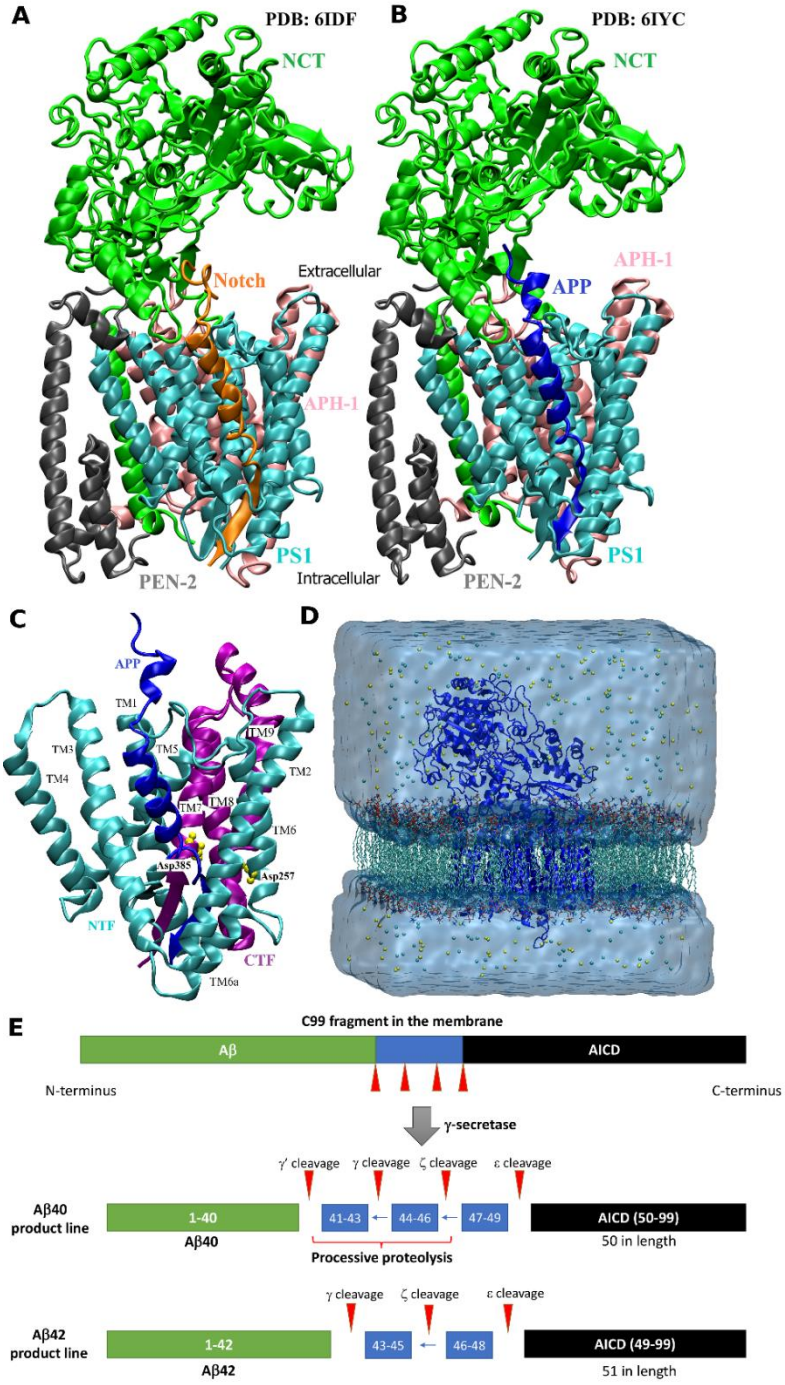

Figure S1: Ribbon representations of the (A) Notch- and (B) APP-bound γ-secretase complexes that include the presenilin (PS1), presenilin enhancer 2 (PEN2), anterior pharynx-defective 1 (APH1) and Nicastrin (NCT) subunits. (C) Representation of the catalytic PS1 domain of APP-bound γ-secretase. The transmembrane (TM) helices and active-site Asp385 and Asp257 residues are labelled. The N-terminal fragment (NTF) is colored in cyan and C-terminal

fragment (CTF) is colored in purple (D) Computational model of γ-secretase complex in GaMD simulations. The protein was embedded into a POPC lipid bilayer and solvated in an aqueous medium of 0.15 M NaCl. (E) Schematic representation of APP substrate processing by γ-secretase.

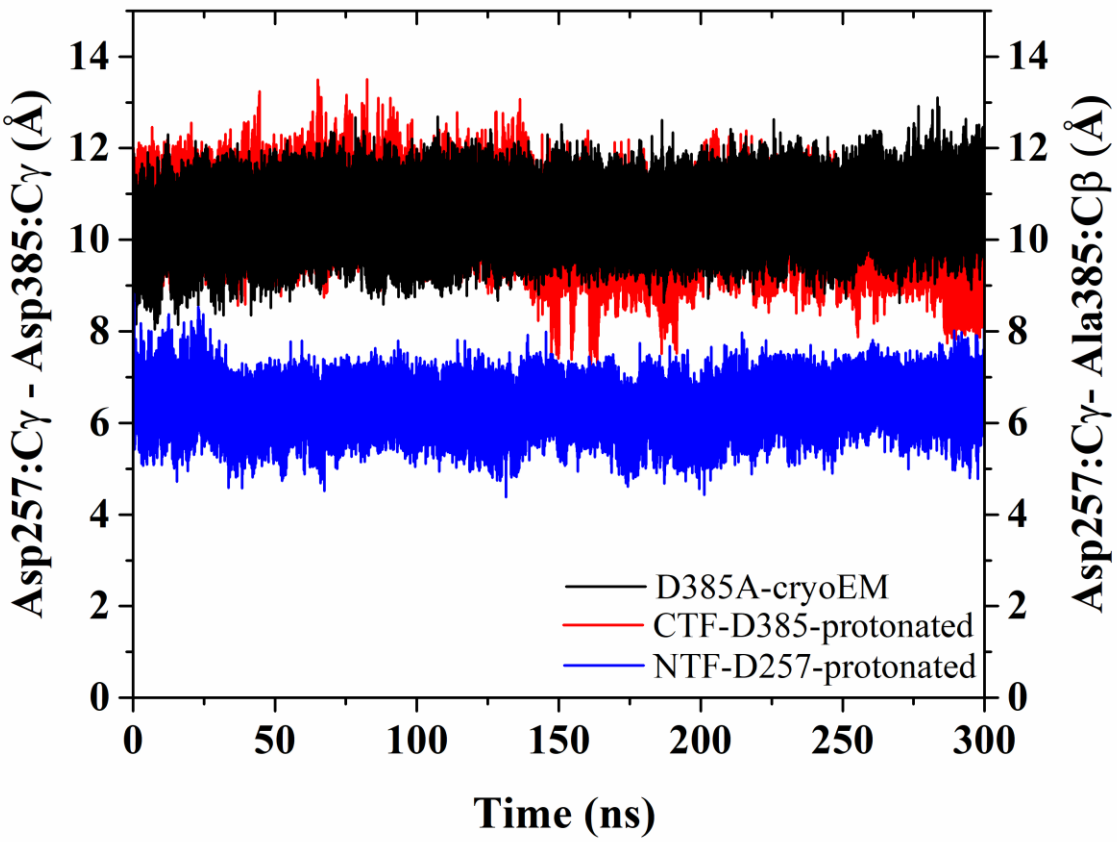

Figure S2: Time course of the Asp257:Cγ - Ala385:Cβ distance calculated from GaMD simulations of the cryo-EM D385A system (black) and Asp257:Cγ - Asp385:Cγ distance calculated from GaMD simulations of D385-protonated (red) and D257-protonated (blue) systems of Notch-bound γ-secretase complex. The disulfide bond between the N-terminus of Notch and PS1 HL1 loop was kept in these simulations.

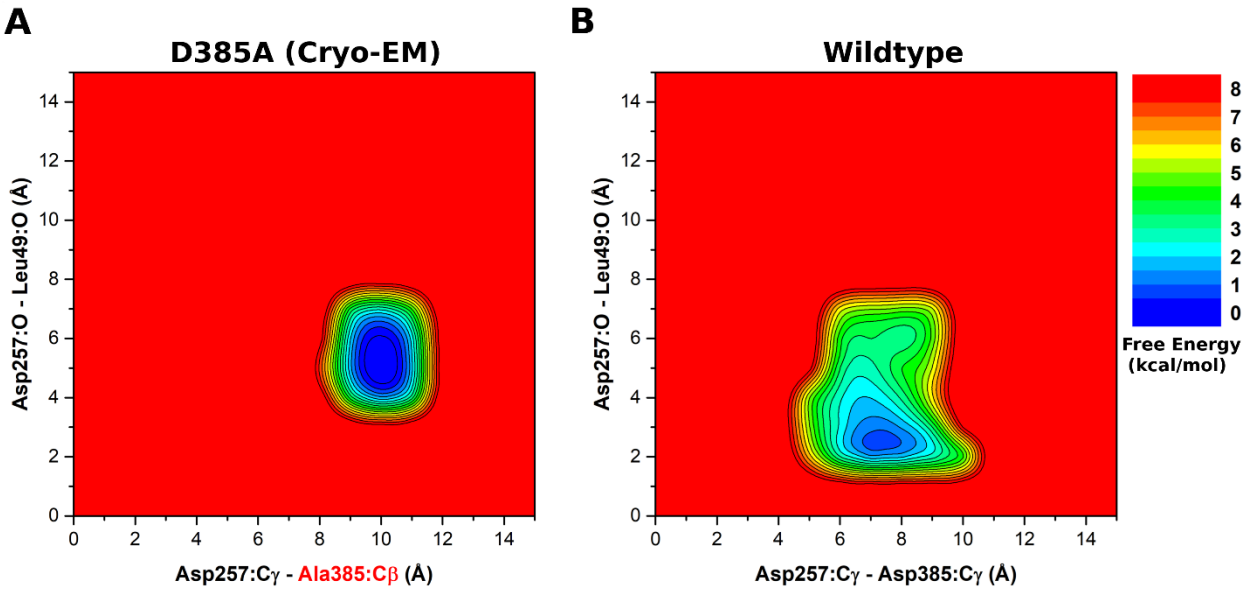

Figure S3: (A) 2D free energy profile of the Asp257:Cγ - Ala385:Cβ and Asp257:sidechain O - Leu49:O distances calculated from GaMD simulations of the D385A cryo-EM structure of γ-secretase bound by APP. (B) 2D free energy profile of the Asp257:Cγ - Asp385:Cγ and Asp257:O - Leu49:protonated O distances calculated from GaMD simulations of the wildtype γ-secretase. The disulfide bond between the N-terminus of APP and PS1 HL1 loop was kept in these simulations.

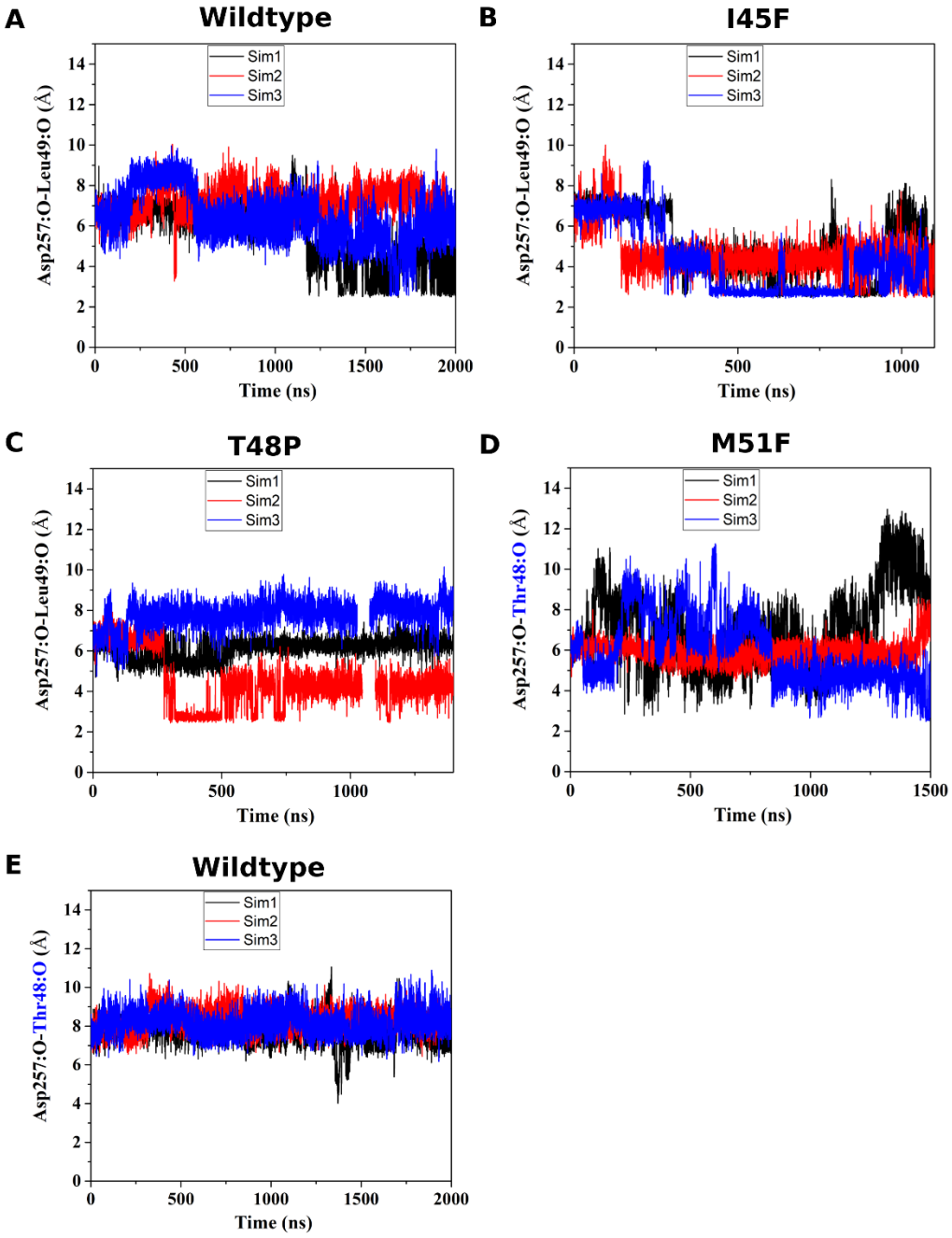

Figure S4: Time courses of the Asp257:protonated O - Leu49:O distance calculated from GaMD simulations of (A) wildtype, (B) I45F, and (C) T48P APP bound γ-secretase, and the Asp257:protonated O – Thr48:O distance calculated from GaMD simulations of the (D) M51F and (E) wildtype APP bound γ-secretase.

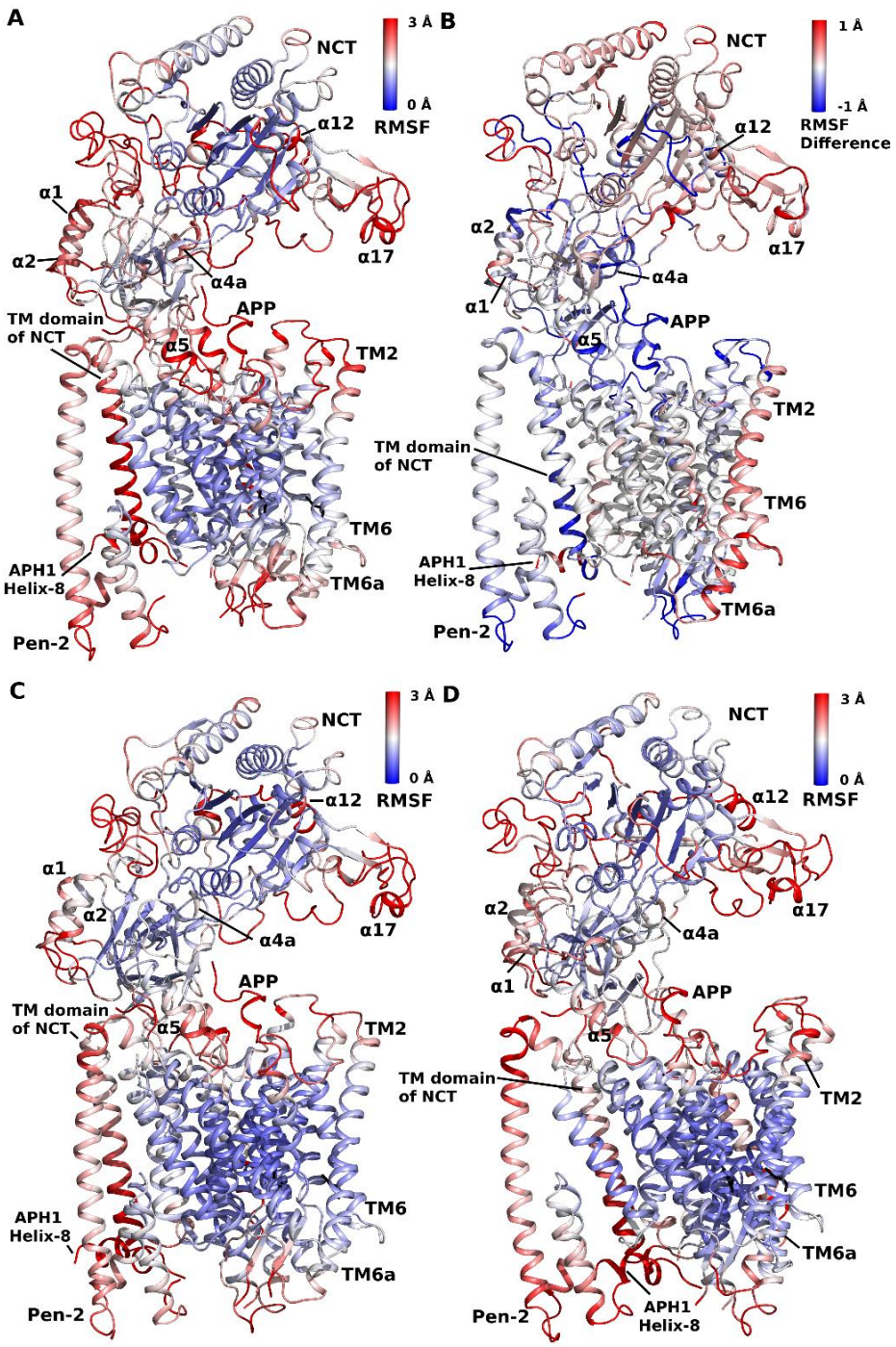

Figure S5: Comparison of structural flexibility of wildtype and mutant APP-bound γ-secretase calculated from GaMD simulations. (A) Root-mean-square fluctuations (RMSFs) of the wildtype APP-bound γ-secretase, (B) RMSF difference between the M51F and wildtype APP bound γ-secretase, (C) RMSFs of the I45F APP bound γ-secretase and (D) RMSFs of the T48P mutant APP-bound γ-secretase.

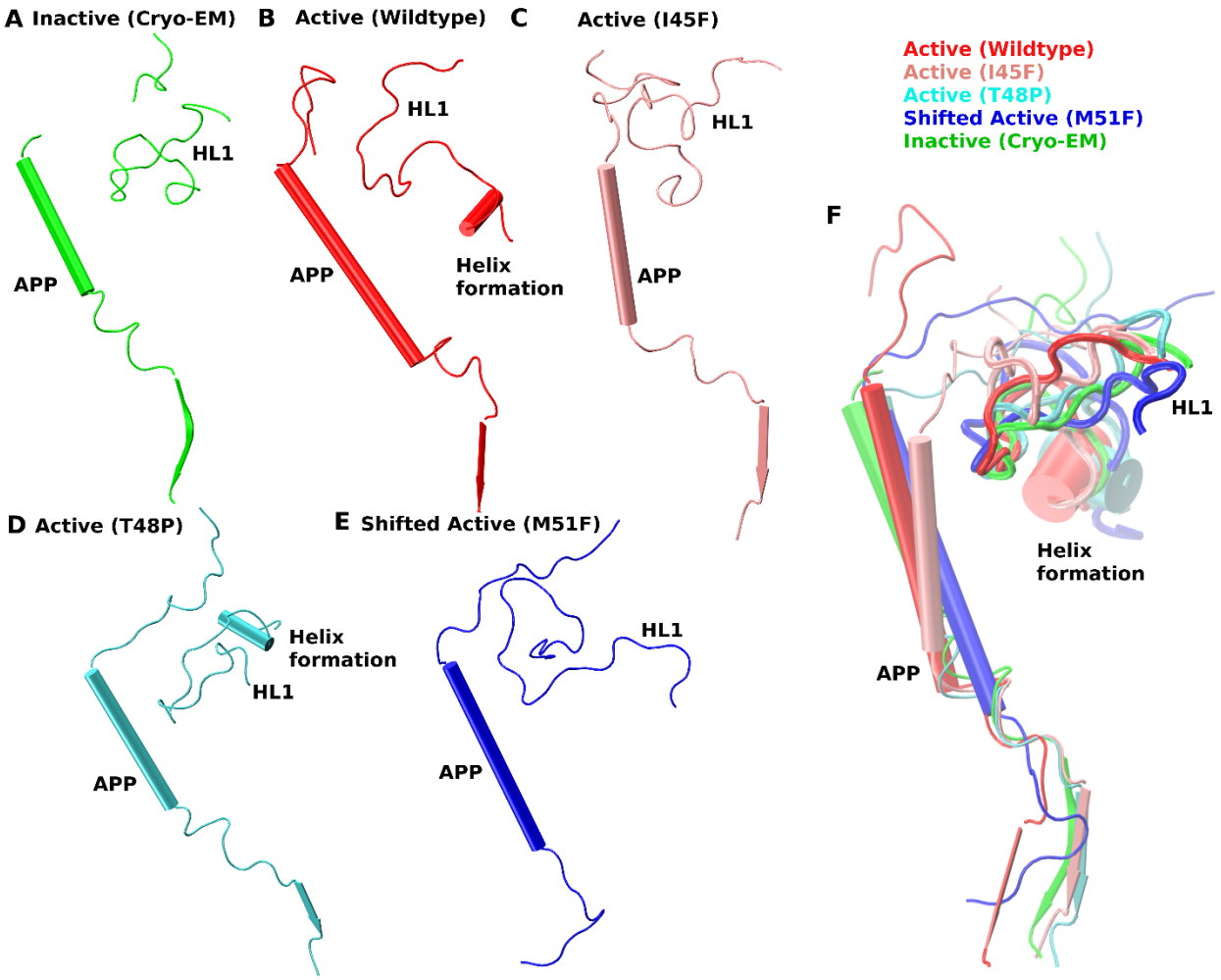

Figure S6: Side view of APP substrate and PS1 HL1 loop in the (A) inactive, representative active conformations of the (B) wildtype, (C) I45F and (D) T48P mutant APP-bound γ-secretase and the (E) Shifted conformation of M51F mutant APP-bound γ-secretase complex obtained from the GaMD simulations. (F) Comparison of APP substrate and PS1 HL1 loop in the different systems of APP-bound γ-secretase complex.

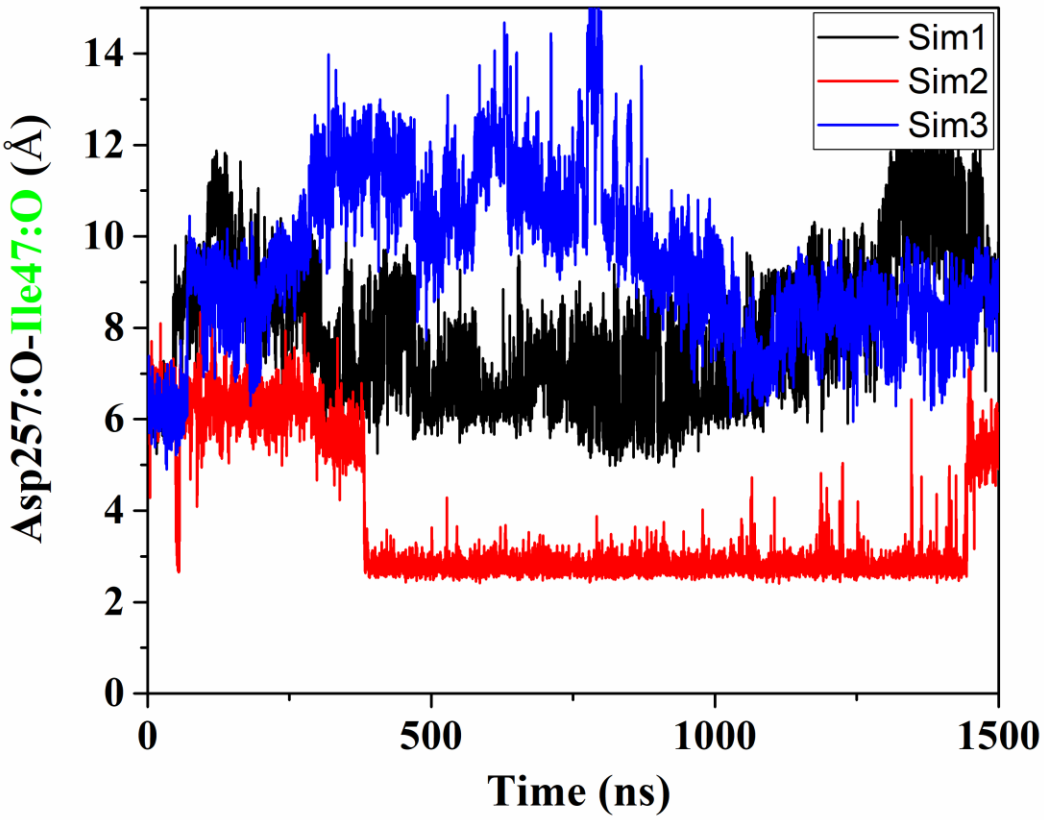

Figure S7: Time course of the Asp257:protonated O - Ile47:O distance calculated from GaMD simulations of the M51F APP bound γ-secretase.

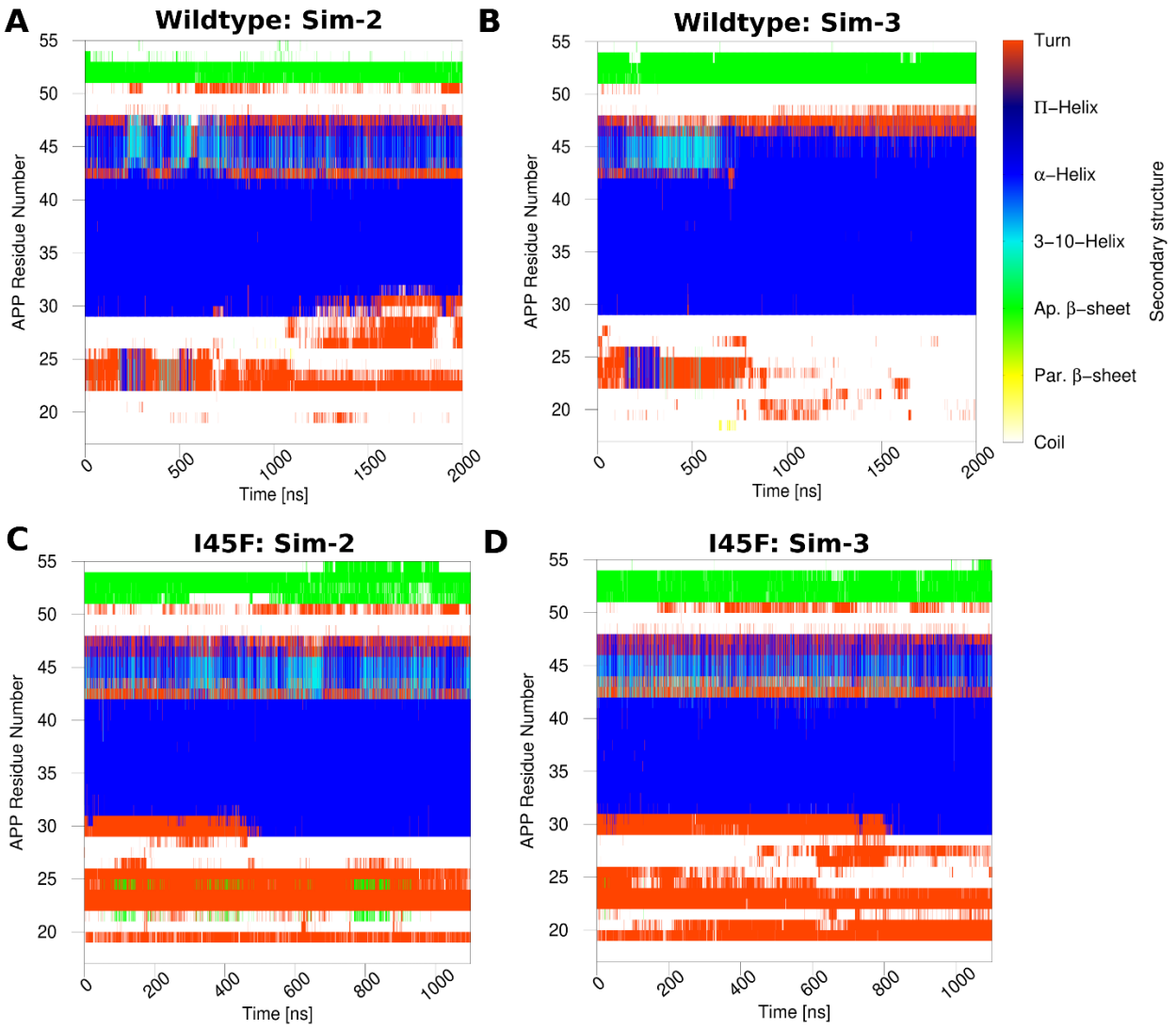

Figure S8: Time courses of the APP secondary structures in the wildtype and I45F forms as bound to γ-secretase calculated from GaMD simulations: (A) Sim-2 and (B) and Sim-3 for the wildtype (Sim-1 in Figure 4A), and (C) Sim-2 and (D) Sim-3 for the I45F mutant (Sim-1 in Figure 4B).

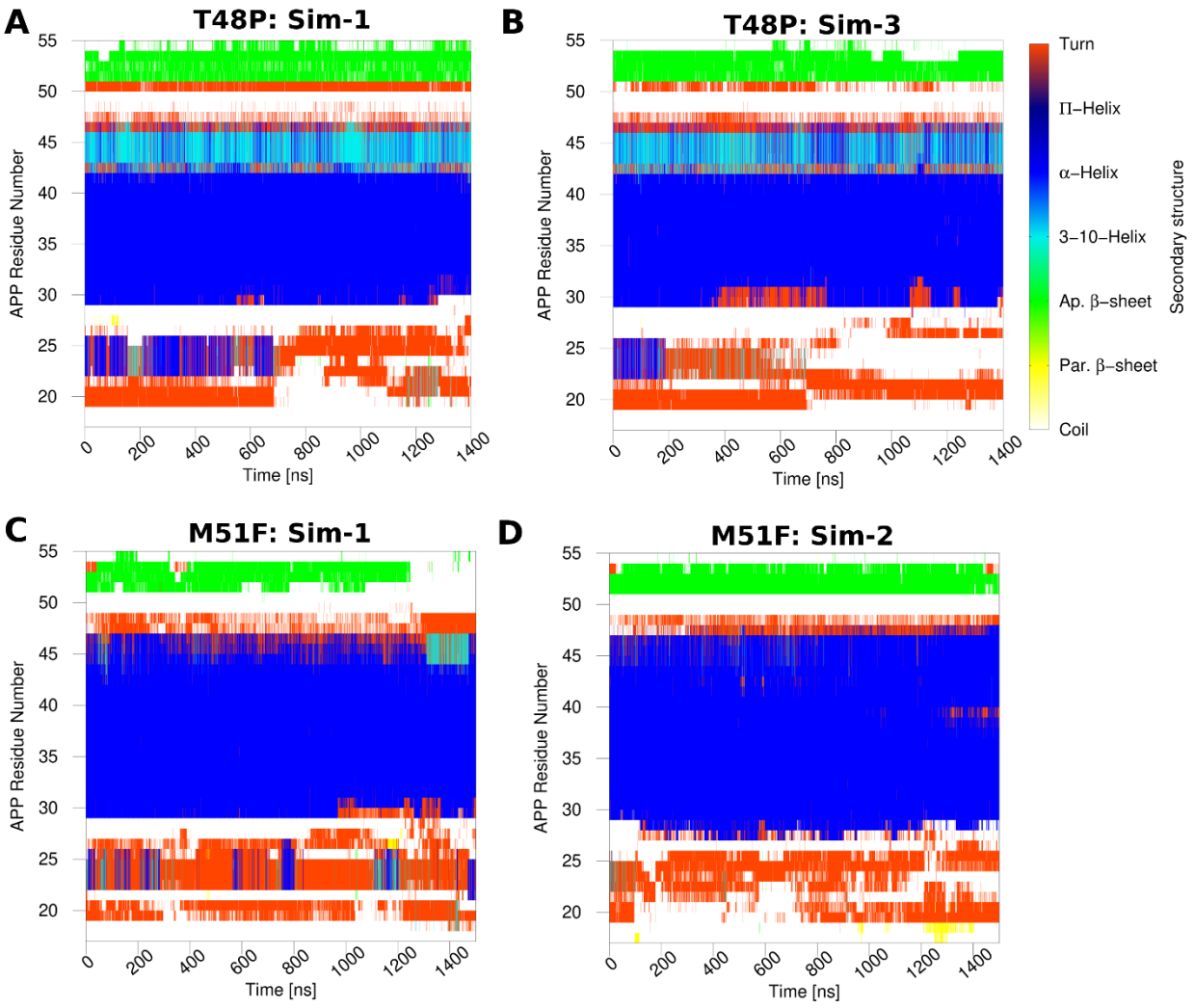

Figure S9: Time courses of the APP secondary structures in the T48P and M51F forms as bound to γ-secretase calculated from GaMD simulations: (A) Sim-1 and (B) and Sim-3 for the T48P mutant (Sim-2 in Figure 4C), and (C) Sim-1 and (D) Sim-2 for the M51F mutant (Sim-3 in 4D).

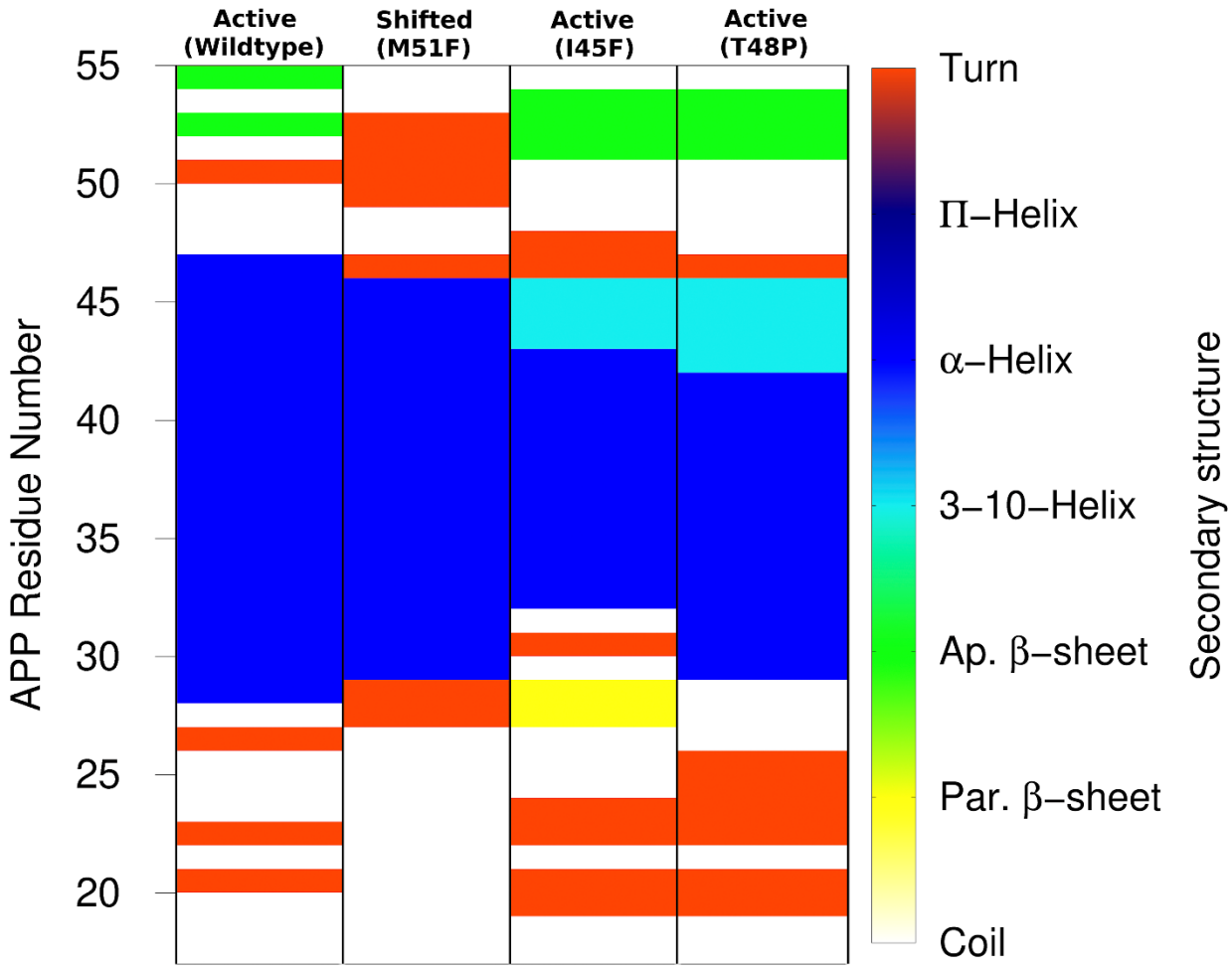

Figure S10: Secondary structures of the APP substrate in the representative active conformations of the wildtype, I45F and T48P mutant APP-bound γ-secretase and the shifted active conformation of M51F mutant APP-bound γ-secretase using the top ranked PS1-APP structural clusters obtained from the corresponding GaMD simulations.

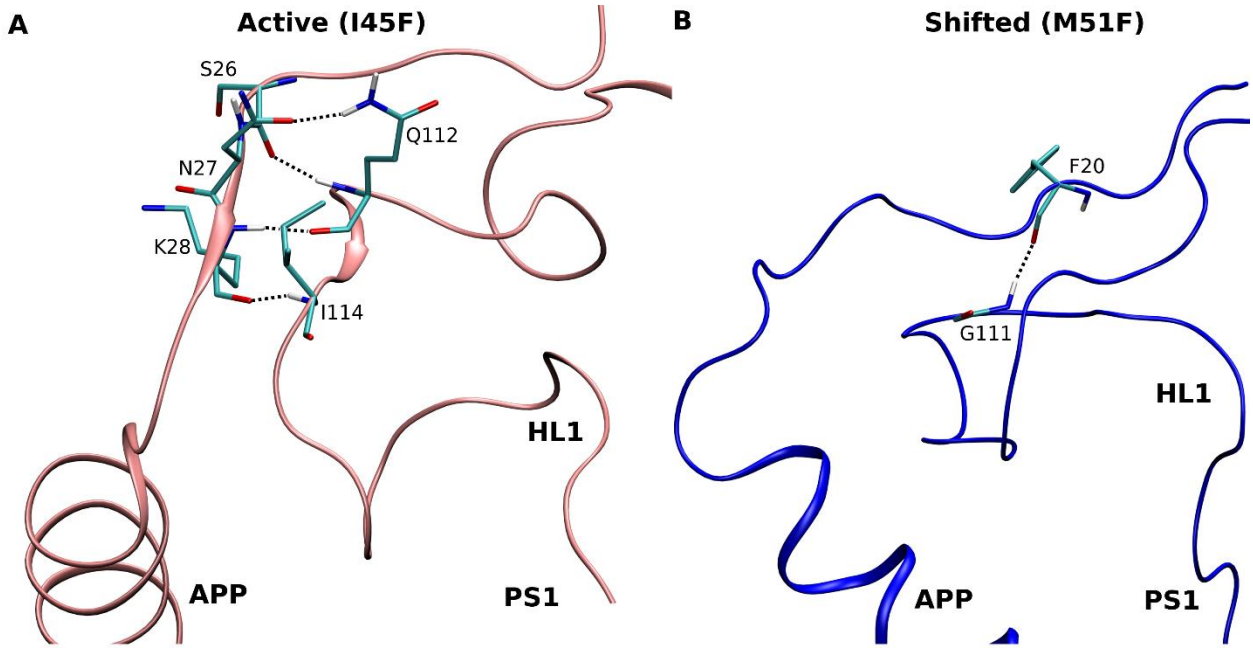

Figure S11: Hydrogen bonds formed between residues G111, Q112 and I114 of the PS1 HL1 loop and the N-terminus of APP substrate in the (A) Active and (B) Shifted conformations of γ-secretase obtained from the GaMD simulations of the I45F and M51F mutant APP systems.

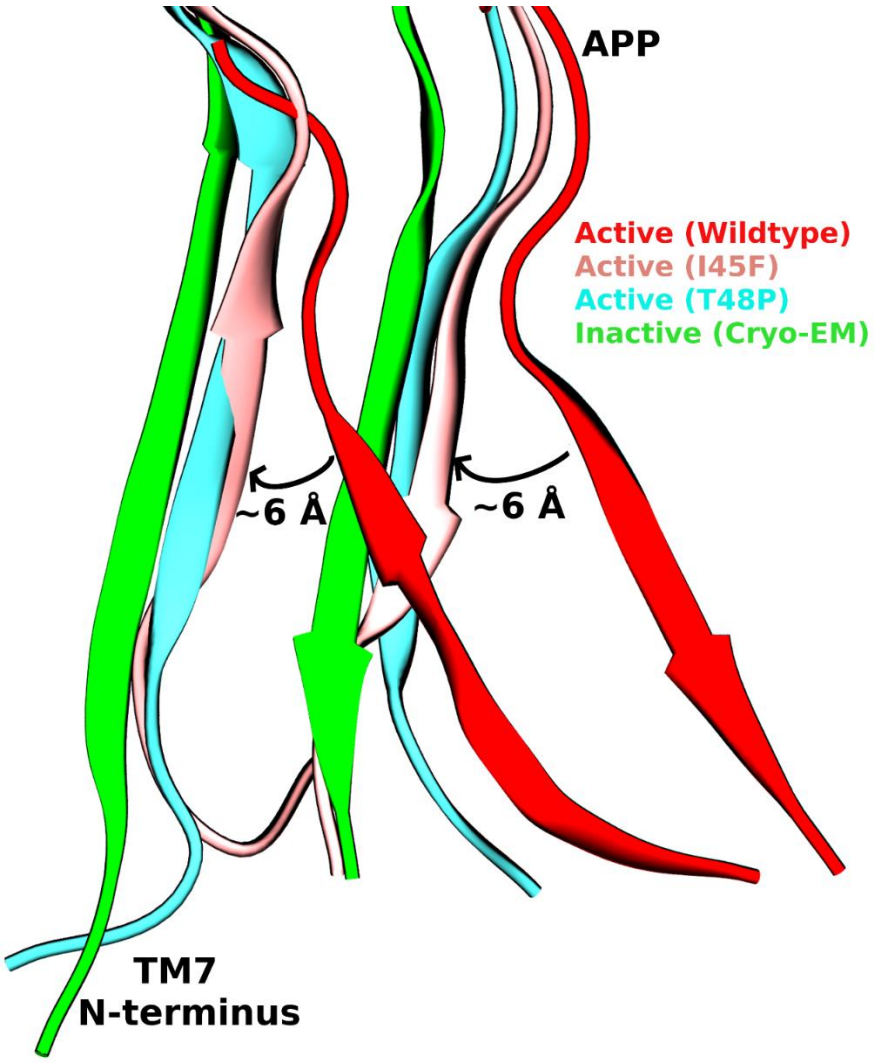

Figure S12: Comparison of the β-sheet conformational changes in the N-terminus of PS1 TM7 and the C-terminus of APP substrate in the inactive (cryo-EM) and active conformations of the wildtype, I45F and T48P mutant APP-bound γ-secretase observed in the GaMD simulations. Relative to the wildtype active conformation, the N-terminus of PS1 TM7 and the C-terminus of APP substrate moved towards the PS1 TM6a by ~6 Å to maintain the β-sheet structure in the I45F and T48P active conformations.

Table S1: List of amino acid residues constituting the S1', S2' and S3' subpockets in the representative wildtype active, I45F active, T48P active and M51F shifted active conformations of γ-secretase obtained from the corresponding GaMD simulations. The residues that are within 5 Å of APP substrate residues P1', P2' and P3' are listed in the table.

System	S1'	S2'	S3'
Active (Wildtype)	V261 L268 R269 L271 V272 L381 G382 P433	V272 I287 V379 K380 L381 G382 L425 K430 A431 L432	Y154 L271 V272 A275 T281 L282 F283 I287 G378 V379 K380 L381 G382
Shifted (M51F)	I143 T147 L150 Y256 D257 A260 V261 L271 V272 A275 E276 P433	D257 V261 R269 V272 E273 Q276 E277	V261 V379 K380 L381 Y389 L421 T422 L425 A431 L432 P433 A434 L435
Active (I45F/T48P)	V261 V272 F283 I287 K380 L381 G382 D385 A431 L432 P433 A434 L435	V82 L85 V261 V379 K380 L381 G382 D385 L418 T421 L422 L425 A431 L432 P433 A434 L435 P436	V272 A275 L282 I287 V379 K380 L381 L425 K430 A431 L432

Chapter 2

Mechanism of Tripeptide Trimming of Amyloid β-Peptide 49 by γ-Secretase

Apurba Bhattarai, Sujan Devkota, Hung Nguyen Do, Jinan Wang, Sanjay Bhattarai, Michael S. Wolfe and Yinglong Miao

[1]Center for Computational Biology and Department of Molecular Biosciences, [2]Department of Medical Chemistry, School of Pharmacy, University of Kansas, Lawrence, KS 66047, USA

Reprinted (Adapted) with permission from Bhattarai, A., Devkota, S., Do, H. N., Wang, J., Bhattarai, S., Wolfe, M. S., & Miao, Y., Mechanism of Tripeptide Trimming of Amyloid β-Peptide 49 by γ-Secretase, *Journal of the American Chemical Society* (2022), DOI: 10.1021/jacs.1c10533[70].

Abstract

The membrane-embedded γ-secretase complex processively cleaves within the transmembrane domain of amyloid precursor protein (APP) to produce 37-to-43-residue amyloid β-peptides (Aβ) of Alzheimer's disease (AD). Despite its importance in pathogenesis, the mechanism of processive proteolysis by γ-secretase remains poorly understood. Here, mass spectrometry and western blotting were used to quantify the efficiency of the tripeptide trimming (Aβ49→Aβ46) of wildtype (WT) and familial AD (FAD) mutant APP substrate. In comparison to WT APP, the efficiency of this first trimming step was similar for the I45F, A42T and V46F APP FAD mutants, but substantially diminished for the I45T and T48P mutants. In parallel with biochemical experiments, all-atom simulations using a novel Peptide Gaussian accelerated molecular dynamics (Pep-GaMD) method were applied to investigate tripeptide trimming of Aβ49 by γ-secretase. The starting structure was active γ-secretase bound to Aβ49 and APP intracellular domain (AICD), as generated from our previous study that captured activation of γ-secretase for the initial endoproteolytic cleavage of APP (Bhattarai *et al., ACS Cent Sci*, 2020, 6:969-983). Pep-GaMD simulations captured remarkable structural rearrangements of both the enzyme and substrate, in which hydrogen-bonded catalytic aspartates and water became poised for tripeptide trimming of Aβ49 to Aβ46. These structural changes required a positively charged N-terminus of endoproteolytic coproduct AICD, which could dissociate during conformational rearrangements of the protease and Aβ49. The simulation findings were highly consistent with biochemical experimental data. Taken together, our complementary biochemical experiments and Pep-GaMD simulations have enabled elucidation of the mechanism of tripeptide trimming of Aβ49 by γ-secretase.

Keywords: amyloid precursor protein (APP), proteolysis, Alzheimer's disease (AD), familial AD (FAD), Peptide Gaussian accelerated molecular dynamics (Pep-GaMD), mass spectrometry (MS).

Introduction

Alzheimer's disease (AD) contributes to more than 80% of all dementia cases [71]. Deaths related to AD in the United States increased by 89% between 2000 and 2014, and more than 6.2 million Americans are affected with AD in 2021 (www.alz.org). AD is characterized by deposition of longer amyloid β-peptides (Aβ) in the form of cerebral plaques. The amyloid β-protein precursor (APP) is successively processed by β-secretase and γ-secretase to produce Aβ peptides. γ-Secretase first sheds the APP extracellular domain to produce transmembrane peptide C99, followed by processive proteolysis by γ-secretase to produce Aβ peptides of varying lengths [72]. Membrane-embedded γ-secretase is a multi-domain aspartyl protease with presenilin as the catalytic subunit. γ-Secretase is considered "the proteasome of the membrane", with more than 100 known substrates, including APP and the Notch family of cell-surface receptors [73,74]. The location of the proteolysis and the number of cleavages within the APP transmembrane domain by γ-secretase determines the length of final Aβ products and the likelihood of forming plaques.

Of the many transmembrane substrates, processive proteolysis of APP by γ-secretase is the most studied. γ-Secretase first carries out endoproteolytic (ε) cleavage of C99 peptide near the cytosolic end of the transmembrane domain, producing Aβ49 and Aβ48 peptides and their respective AICD co-products (AICD50-99 and AICD49-99, respectively)[75]. These initially formed long Aβ peptides are then cut generally every three residues from their C-termini to release tripeptide (and one tetrapeptide) co-products. The two general pathways of γ-secretase processive proteolysis are Aβ48→Aβ45→Aβ42→Aβ38 and Aβ49→Aβ46→Aβ43→Aβ40 [16,17], producing Aβ42 and Aβ40 as their dominant products, respectively. Among these two, the longer Aβ42 peptide is more prone to aggregate and forms plaques [76]. Moreover, early-onset

familial AD (FAD) APP mutants can bias the enzyme to produce longer Aβ peptides that are pathological and cause AD [77]. The trimming of APP substrate by γ-secretase enzyme is dictated by active site S1', S2' and S3'subpockets that respectively bind to P1', P2' and P3' substrate residues [78].

Critical gaps remain in understanding the mechanism of intramembrane processive proteolysis by γ-secretase. Recently reported cryo-EM structures of γ-secretase bound to Notch and APP substrates provided valuable insights into the structural basis of substrate recognition of the enzyme [58,79]. However, artificial structural constraints were included that could affect the enzyme-substrate interactions. Molecular dynamics (MD) simulations have proven useful in understanding the structural dynamics of γ-secretase, notably the enzyme-substrate interactions, including many previous studies[23-33,80-84]. Recently, we computationally restored the wildtype (WT) enzyme-substrate co-structure and applied all-atom simulations using the Gaussian accelerated molecular dynamics (GaMD) method to build the first dynamic model of γ-secretase activation [8]. GaMD is an enhanced sampling technique that works by adding a harmonic boost potential to smooth the potential energy surface and reduce system energy barriers [4]. Our GaMD simulations captured the extremely slow motions underlying enzyme activation, with the two catalytic aspartates and a coordinated water molecule poised for proteolysis of APP at the ε cleavage site. We showed that the I45F and T48P FAD mutations in APP enhanced the ε cleavage of the amide bond between Leu49-Val50 compared with the WT APP. In contrast, the M51F mutation in APP shifted the ε cleavage to the adjacent Thr48-Leu49 amide bond, changing the proteolysis from the Aβ49 to the Aβ48 pathway. Despite these advances, the detailed atomistic mechanism of processive proteolysis by γ-secretase remains elusive. This is consistent with γ-secretase being a well-known slow-acting enzyme (k_{cat} for APP ε proteolysis

~ 2-6 per hour) [45,85], making it difficult to capture the dynamic transitions comprising large energy barriers in MD simulations. Hence, despite its importance in the pathogenesis of AD, the mechanism of processive proteolysis (tripeptide trimming) by γ-secretase remains poorly understood.

Here, we report the first dynamic model of tripeptide trimming of Aβ49 to Aβ46 (ζ cleavage) by γ-secretase. Extensive all-atom simulations using a novel Peptide GaMD (Pep-GaMD) method [86] captured the slow dynamic molecular transition from the ε to ζ proteolytic cleavage step. In Pep-GaMD, a boost potential is applied selectively to the essential potential energy of the peptide to effectively model its high flexibility and accelerate its dynamic motions[86]. In addition, another boost potential is applied on the protein and solvent to enhance conformational sampling of the protein and facilitate peptide binding. Pep-GaMD has been demonstrated on binding of model peptides to the SH3 protein domains. Independent 1-μs dual-boost Pep-GaMD simulations have captured repetitive peptide dissociation and binding events, which enable calculation of peptide binding thermodynamics and kinetics. The calculated binding free energies and kinetic rate constants agreed very well with the available experimental data [87].

In this study, we have combined biochemical experiments, including matrix-assisted laser desorption/ionization—time-of-flight) mass spectrometry (MALDI-TOF MS), liquid chromatography—tandem mass spectrometry (LC-MS/MS) and western blotting, with Pep-GaMD enhanced sampling simulations to elucidate the mechanism of tripeptide trimming of Aβ49 by γ-secretase. Our findings from Pep-GaMD simulations of WT and five FAD mutants (I45F, A42T, V46F, I45T and T48P) of Aβ49 bound to γ-secretase were highly consistent with

quantitative biochemical analysis of their specific proteolytic products, providing important mechanistic insights into tripeptide trimming by the enzyme.

Results

Probing ζ cleavage of WT and FAD-mutant Aβ49 by γ-secretase in biochemical experiments.

To compare the ζ cleavage of the WT and FAD mutants of APP by γ-secretase, we performed *in vitro* cleavage assay experiments using purified γ-secretase and recombinant APP-based substrate C100-FLAG, which contained the C99 APP C-terminal fragment with an N-terminal start methionine and a C-terminal FLAG epitope tag [88]. Efficiency of the cleavage of substrate Aβ49 to products Aβ46 and tripeptide was calculated by measuring Aβ49 production and Aβ49 degradation. To quantify Aβ49 production by ε cleavage of APP substrate, levels of co-products AICD 50-99 were determined using a combination of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and quantitative western blotting.

First, AICD produced in the assay was immunoprecipitated with anti-FLAG antibodies and detected by MALDI-TOF MS (**Fig. 1A**). For the WT, A42T, V46F and I45T APP substrate, the signal intensities corresponding to AICD 49-99 and AICD 50-99 show higher level of AICD 49-99 than AICD 50-99. However, for mutants I45F and T48P APP substrate, signal intensities show higher level of AICD 50-99 than AICD 49-99. This suggests I45F and T48P favor production of Aβ49 rather than production of Aβ48 while A42T, V46F and I45T favor production of Aβ48 rather than production of Aβ49.

The same reaction mixtures were subjected to quantitative western blotting with anti-FLAG antibodies (**Fig. 1B**), where standards of known concentrations of C100-FLAG were also run to make a standard curve, plotting band intensity against concentration of FLAG-tagged C100. From this standard curve, the concentration of total AICD-FLAG product obtained in the enzyme reaction was quantified (**Fig. 1C**). Quantification of the total AICD revealed increased total AICD production for V46F mutant substrate and decreased total AICD production for A42T, I45F, I45T, and T48P mutant substrates compared to AICD production for the WT. The concentration of AICD 50-99 was calculated using the total AICD level determined by quantitative western blot and the ratio of AICD 49-99 to AICD 50-99 determined from MALDI-TOF MS. (**Fig. 1D**). The calculated concentrations of AICD 50-99 thus provided the level of production of co-product Aβ49. Aβ49 production was slightly increased for I45F mutant, while for all other mutants A42T, V46F, I45T and T48P decreased Aβ49 production was observed compared to Aβ49 production of the WT.

To determine the degradation of Aβ49, we calculated and quantified trimming product tripeptide ITL. The mixtures from the cleavage assay were subjected to LC-MS/MS analysis to detect tripeptides. All substrate constructs studied produced ITL except for T48P mutant which produced IPL due to the replacement of T with P. For quantification of these tripeptides production, standard curves of each peptide were generated by plotting the concentration of synthetic peptide against the integrated areas of the three most abundant ion fragments from MS/MS (**Fig. S1**). The ITL and IPL peptide generated in the γ-secretase cleavage was monitored and quantified. (**Fig. 1E**). The quantification of the trimming product (ITL or IPL) or the Aβ49 degradation reveal decrease in Aβ49 degradation for A42T, V46F, I45T and T48P. For I45F, Aβ49 degradation is similar to that of the WT. Concentration of both Aβ49 production as well

as Aβ49 degradation was used to calculate the percent efficiency (**Fig. 1F**). For all constructs, cleavage efficiency was close to 100% except that for two mutants I45T and T48P, the cleavage efficiencies decreased substantially to 35% and 34%, respectively.

We selected these particular FAD mutations in APP substrate based on their different effects on the Aβ49→Aβ46 trimming step in our recently reported study[85]. In that study, we examined the effects of 14 different FAD mutations in APP substrate on all proteolytic steps carried out by γ-secretase. Moreover, we determined the Aβ42/40 ratios for these and other FAD mutations in APP substrate and found the relative effects of these mutations on this ratio compared to that seen with WT enzyme to be generally consistent with those reported from other groups[89-91]. To the best of our knowledge, the effects of the I45F, A42T, V46F, I45T and T48P FAD mutations of the substrate on the Aβ49→Aβ46 trimming step have not been reported by any other groups.

Activation of γ-secretase for tripeptide trimming of Aβ49 was captured in Pep-GaMD simulations.

In parallel with the biochemical experiments, Pep-GaMD simulations were carried out on the γ-secretase bound by the WT and the I45F, A42T, V46F, I45T and T48P mutants of Aβ49 **(Table 1)**. The active WT APP-bound γ-secretase was obtained from our previous study [8], and the amide bond between Aβ49 and AICD50-99 was cleaved as the new simulation starting structure **(Fig. S2,** see details in **Methods)**. We initially performed dual-boost GaMD simulations on the γ-secretase bound to Aβ49 with AICD50-99 removed. However, even after running ~6 μs GaMD simulations, we could not effectively sample conformational transitions

of the system for ζ cleavage of Aβ49 to Aβ46 **(Fig. S3)**. The distance between the enzyme Asp257 catalytic residue and substrate Val46-Ile47 amide bond presented a computational challenge for conformational sampling, with apparently high energy barriers to overcome. To address the challenge, we applied our recently developed Pep-GaMD [86] method, which selectively boosts the essential potential energy of the peptide to effectively model the peptide flexibility and further improve sampling. We built four Pep-GaMD simulation systems with γ-secretase bound to Aβ49 in the presence of AICD50-99 and a system in the absence of AICD50-99 **(Fig. S4)**. The C-terminus of Aβ49 and the N-terminus of AICD50-99 was either charged or neutral combined to form four different Pep-GaMD enzyme systems. Spontaneous activation of γ-secretase for ζ cleavage of Aβ49 was observed during 600 ns Pep-GaMD simulations with "charged C-terminal Aβ49 and charged N-terminal AICD50-99" **(Figs. S5** and **S6** and **Movie S1 and Table S2)**. The enzyme activation for ζ cleavage was characterized by coordinated hydrogen bonding between the enzyme Asp257 and carbonyl oxygen of substrate Val46. The catalytic aspartates were at a distance of ~7-8 Å between their Cγ atoms, which could accommodate a water molecule for nucleophilic attack of the carbonyl carbon of the scissile amide bond **(Fig S5)**. The water molecule formed hydrogen bonds with both catalytic aspartates and was at ~4 Å distance away from the carbonyl carbon of substrate Val46 residue. The activated γ-secretase conformation was well poised for cleavage of amide bond between Val46 and Ile47 for ζ cleavage of the Aβ49. In the γ-secretase bound to WT charged C-terminal Aβ49 and charged N-terminal AICD50-99system, we observed AICD50-99 dissociation in addition to enzyme activation for ζ cleavage **(Fig. S7 and Movie S2)**. The AICD50-99, initially located near the Aβ49, slowly moved downwards to the intracellular PS1 pocket and then dissociated completely from the enzyme. Meanwhile, the AICD50-99 transitioned from β-sheet to a

loop/un-structured conformation during the Pep-GaMD simulations **(Fig. S7)**. Similarly, γ-secretase bound to "neutral C-terminal Aβ49 and charged N-terminal AICD50-99" was also observed to become activated for ζ cleavage of Aβ49. In comparison, the γ-secretase systems bound to "charged C-terminal Aβ49" (in the absence of AICD50-99), "charged C-terminal Aβ49 and neutral N-terminal AICD50-99", and "neutral C-terminal Aβ49 and neutral N-terminal AICD50-99" could not sample enzyme activation for ζ cleavage of Aβ49 (**Table S2**, **Fig. S6A-B, D** and **S8A-C**). This showed that the presence of charged N-terminal AICD50-99 was crucial for the enzyme activation for ζ cleavage of Aβ49. Therefore, systems for γ-secretase bound by charged C-terminal Aβ49 and charged N-terminal AICD50-99 were set up for running Pep-GaMD simulations of the FAD mutants of Aβ49."

Free energy profiles were calculated from Pep-GaMD simulations to characterize the activation of γ-secretase for ζ cleavage of the Aβ49 substrate, for which the distance between the enzyme catalytic aspartates and the distance between the enzyme protonated Asp257 and substrate residue Val46 were selected as reaction coordinates **(Fig. 1G-1L, S6 and S9 and Table S2)**. In the WT Aβ49, three low-energy conformational states were identified from the free energy profile, including "Final", "Intermediate" and "Initial" **(Figs. 1G, S6C and S10 and Table S2)**. In the "Final" conformational state, the aspartates were ~7-8 Å apart to accommodate the water molecule in between. The substrate Val46 maintained a distance of ~3 Å from the active site Asp257 to form hydrogen bond in the Final active state. In the "Initial" state, the substrate Val46 was distant (~8-9 Å) from the active site Asp257, while the inter-aspartate distance was ~6-7 Å. The "Initial" state represented the active state for the ζ cleavage of APP. In the "Intermediate" state, the aspartates remained ~6-7 Å apart, while the Aβ49

peptide (carbonyl oxygen of Val46) was at a distance of ~6 Å from the protonated Asp257 **(Fig. 1G)**.

In the I45F mutant system, two low-energy conformational states, "Initial" and "Final", were identified from the free energy profile of Pep-GaMD simulations **(Figs. 1H, S11A and S12A and Movie S3)**. Two out of three Pep-GaMD simulations could capture the activation process, as the Asp257 could form stable hydrogen bond with Val46 as reflected in the distance time course plot (**Fig. S11A**). The "Final" state in the free energy profile represented the active conformation of the enzyme for ζ cleavage of the scissile amide bond between Val46 and Ile47 APP residues. In the "Final" conformational state, the two catalytic aspartates were ~7-8 Å apart, and APP Val46 was ~3 Å distance away from the protonated aspartate. In the "Initial" state, the substrate Val46 was further away from the catalytic aspartate (~6-7 Å), and the aspartates were ~7-8 Å distance away from each other **(Fig. 1H)**.

In the A42T mutant APP system, four low-energy conformational states were identified from the free energy profile **(Figs. 1I, S11B and S12B and Movie S4)**. Mutation of Ala42 to Thr42 caused the enzyme-substrate complex to sample a larger conformational space. In addition to the "Initial" and "Final" states, two new "Inhibited-1" and "Inactive" conformational states were identified for the A42T mutant system. The catalytic aspartates were ~4-5 Å (too close) apart in the "Inhibited-1" state and 13 Å away (too far) in the "Inactive" state. In the "Inhibited" state, the catalytic aspartates could not accommodate a water molecule between them and hence was inhibited from proteolytic activation. APP Val46 was ~4-5 Å from the protonated Asp257 in this "Inhibited-1" state. In the "Inactive" state, the aspartates were ~13 Å apart and thus too far to form the dual hydrogen bonds with the water in between them, even

though the Asp257 could form a hydrogen bond with the Val46 carbonyl oxygen. This hindered activation required for ζ cleavage.

In the V46F mutant system, two low-energy conformational states were identified, including "Inhibited-2" and "Final" **(Figs. 1J, S11C and S12C and Movie S5)**. Like other γ-secretase systems, the "Final" state corresponded to the active conformation of the enzyme poised for ζ cleavage of Aβ49. Moreover, the "Inhibited-2" state had the two aspartates at proximity (~4-5 Å) between the Cγ atoms and unable to accommodate a water molecule in between for enzyme activation. APP substrate was ~10 Å away from active site Asp257 in the "Inhibited-2" state.

Furthermore, Pep-GaMD simulations were carried out on I45T and T48P mutant Aβ49-bound γ-secretase **(Figs. 1K-1L, S11D-S11E and S12D-S12E)**. Both of these mutant systems were not able to activate the enzyme for ζ cleavage, being consistent with the experimental results where the ζ cleavage efficiency dropped to about one third compared to that of the WT. In the Pep-GaMD free energy profile of the I45T mutant system, only one "Intermediate" low-energy conformational state was identified. This "Intermediate" state was the same low-energy conformation as the one in the WT system. For the T48P system, the hydrogen bond between APP Val46 and the protonated Asp257 was formed for a certain time in one of the three Pep-GaMD production simulations **(Fig. S11E)**. However, in the free energy profile, we could identify two low-energy conformational states, including "Initial" and "Inhibited-1", but not the "Final" active state **(Fig. 1L)**. The "Inhibited-1" state resembled the one identified in the A42T mutant system. The "Initial" conformational state was the same as the one identified in

the WT, I45F and A42T systems. These Pep-GaMD simulation findings were consistent with the biochemical experiments, verifying the I45T and T48P systems as negative controls.

Conformational changes in activation of γ-secretase for tripeptide trimming of Aβ49

We calculated root mean square fluctuations (RMSFs) of γ-secretase bound by the WT and FAD-mutant APP from Pep-GaMD simulations **(Fig. S13 and Movie S6)**. In the WT Aβ49-bound γ-secretase, the TM2, TM6, TM6a and C-terminus of TM9 helix were flexible in the catalytic PS1 subunit. The Pen-2 subunit exhibited high fluctuations with ~ 3 Å RMSF. Helices α1, α2, α5, α12, α17, and TM domain of nicastrin were also flexible during the Pep-GaMD simulations. Structural clustering was performed on Pep-GaMD snapshots of the system using hierarchical agglomerative algorithm in CPPTRAJ [69] (see **Methods**). The top-ranked cluster was selected as the representative "Final" active conformation for the ζ cleavage of Aβ49. The starting structure from ζ cleavage of APP was obtained as the "Initial" active conformation. The catalytic PS1 of the "Final" conformation was compared to that of the "Initial" conformation in **Fig. 2A**. Relative to the "Initial" conformation, the substrate helical domain tilted by ~50 degrees in the "Final" conformation **(Figs. 2A** and **2B)**. Residue Leu49 in the substrate C-terminus moved downwards by ~5 Å **(Figs. 2B**, **4A** and **S14)**. The last residue in a helical conformation in the "Final" state of Aβ49 was Thr43 whereas it was Ile45 in the "Initial" state. In transition from the "Initial to the "Final" conformational state, two substrate residues, Val44 and Ile45 unwound their helical conformation and changed to a turn/loop conformation. Residues Thr43 and Ile45 were in similar positions in the "Initial" and "Final" active conformations relative to the membrane perpendicular axis (**Fig. S14**). In comparison, the

substrate C-terminal Leu49 moved downwards by ~5 Å while straightening the C-terminal loop **(Figs. 2B and S14B)**.

At the enzyme active site, the catalytic Asp385 did not have significant movement during the adjustments for substrate peptide trimming (**Fig. 2C**). In comparison, the protonated catalytic Asp257 moved by ~3 Å towards the substrate. Asp257 moved forward to form a hydrogen bond with the carbonyl oxygen of the scissile amide bond between the substrate residues Val46 and Ile47. Similarly, TM3 moved outwards by ~2 Å **(Fig. 2D)**, and TM6a moved downwards by ~2 Å **(Fig. 2E)**. Flexibility in these helices involved important FAD mutation sites including Tyr154, His163, Ala164, Leu166, Trp165, Ser169, Ile168, Tyr256, Ala260, Leu262, Cys263, Pro264, Pro267, Arg269, Val272 and Leu271 (www.alzforum.org). Trp165 and His163 from TM3 and Arg269 from TM6a showed significant movements in their side chains. With a major part of C-terminus of APP absent (as AICD dissociates, see next section) the β2 loop at N-terminus of TM7 moved away from the APP by ~5 Å in the Final state as compared to the Initial state **(Fig. 2F)**. FAD mutation residues in the β2-TM7 region including Arg377, G378, L383 and G384 showed flexibility in the simulations. In particular, residue Arg377 reoriented its side chain in the "Final" conformational state.

Changes in secondary structures of the WT and FAD-mutant Aβ49 during tripeptide trimming

Secondary structures of the WT and FAD-mutant Aβ49 bound to γ-secretase were recorded during the Pep-GaMD simulations and plotted in **Figs. 3 and S15**. Changes in secondary structures of Aβ49 during ζ cleavage were compared to that of APP substrate ("Initial" active

conformation) during ζ cleavage from our previous study [8] **(Fig. S16)**. Unwinding of the helix C-terminus in Aβ49 during ζ cleavage was observed in the secondary structure plot. During the ζ cleavage, the C-terminus of the WT APP substrate could maintain helical conformation up to Ile45/Val46 **(Fig. S16)**. In comparison, WT Aβ49 was helical up to Thr43 in the C-terminal region **(Figs. 3A and 2B)**. About 2-3 residues unwound near the ζ cleavage site to expose the scissile amide bond between Val46 and Ile47 to the catalytic aspartates and the coordinated water for activation. A new helix was formed for residues Ser26 to Ala30 in the Aβ49 during the transition from ε to ζ cleavage in the WT system **(Figs. 3A and S17)**. With the 50° tilt of Aβ49 peptide in the space between TM2 and TM3, the N-terminus is exposed to the hydrophobic lipid bilayer **(Fig. S17)**. This helped the N-terminal loop to transition to a α-helical conformation. The effects of the mutations on the new helical conformation is mentioned and explained in the next paragraph. A turn/unstructured conformation at residues Ala30-Ile31 separated these two helices. In addition, the N-terminus of Aβ49 lost its interactions with the hydrophobic loop 1 (HL1) because of the tilting away from this loop.

Similarly, secondary structural changes were recorded for the I45F, A42T and V46F Aβ49 mutant systems **(Figs. 3B-3D and S15)**. Like the WT, the I45F and A42T Aβ49 mutants maintained a helical conformation up to Thr43 at the N-terminus during the Pep-GaMD simulations. C-terminal residues after the Thr43, which included the ζ cleavage site bond between Val46 and Ile47, were observed mostly in a turn/unstructured conformation. This allowed the catalytic aspartates and water to approach the scissile amide bond for forming coordinated hydrogen bonds required for this cleavage. Likewise, bands of new helix formation were observed in the secondary structure plots from Asn27 to Ile31 and from Asp23 to Lys28 for I45F- and A42T-mutant Aβ49 systems, respectively **(Fig. 3B and 3C)**. The new helix formed

was due to its exposure to the hydrophobic lipid membrane. V46F Aβ49 was observed to be the most dynamic in terms of secondary structure changes **(Fig. 3D)**. A band of helix was observed between Gly29 to Thr43, with a turn conformation formed between Leu34 – Val36. Thr43 to Ile47 transitioned between helix and turn conformations during the Pep-GaMD simulations of the V46F Aβ49. Like the WT and other mutant systems, new N-terminal helix formation was observed at residues Phe20 to Gly25 in the V46F mutant APP **(Fig. 3D)**. The hydrophobic lipid environment helped these residues transition from turn to a helical conformation in the V46F mutant APP.

Active-site subpockets formed in γ-secretase for tripeptide trimming.

The "Final" active conformational state of Aβ49-bound γ-secretase was further analyzed for the P1', P2' and P3' substrate residues at the ζ cleavage active site and the respective S1', S2' and S3' subpockets in which they reside [78]. The S1' subpocket accommodating the P1' residue in the WT Aβ49 was formed by residues from PS1 TM6a helix, β1 loop, TM3 helix and TM7 N-terminal region **(Figs. 2 and 4A)**. The residues that formed the subpockets are listed in **Table S1**. The S2' subpocket occupied by the P2' substrate residue consisted of residues from PS1 TM6 helix, TM6a helix, PAL motif of TM9 helix, β1 and β2 loop region. Moreover, the S3' subpocket accommodating the P3' residue was formed by residues from PS1 TM6 helix, TM6a helix and β1 loop. In reference to Aβ49, S1' and S3' pockets were located on the same side (TM6a and TM3 helices) whereas the S2' pocket was located on the opposite side (TM6 and TM9 helices).

Similarly, the "Final" active conformational states of the I45F and A42T mutant Aβ49-bound γ-secretase systems had the same subpockets formed at the active site during the ζ cleavage as that of the WT system **(Fig. 4B-4C and Table S1)**. In the I45F and A42T "Final" active conformation, the S1' and S3' subpockets occupied by the respective P1' and P3' substrate residues consisted of residues from PS1 TM6 helix, TM6a helix, TM7 helix, β1 and β2 loop region. In comparison, the S2' subpocket was located on the opposite side of Aβ49 and consisted of residues from PS1 TM6 helix, TM6a helix, PAL motif of TM9 helix, β1 and β2 loop. Furthermore, in the V46F "Final" active conformation, the locations of the S1' and S2' subpockets accommodating P1' and P2' Aβ49 substrate residues, respectively, were different as compared to that of the WT system **(Fig. 4C-D)**. The S1' pocket occupied by the P1' residue of Aβ49 consisted of residues from TM6 helix, TM6a helix and TM2 helix **(Table S1)**. The S2' subpocket occupied by the P2' residue of the V46F mutant in the "Final" active state was the same as the S1' subpocket in the "Final" active state of the WT, I45F and A42T systems **(Fig. 4C-D)**. Moreover, the S3' subpocket accommodating the P3' substrate residue in the V46F mutant was the same as the one of the WT, I45F and A42T systems (**Fig. 4E**).

Discussion

Current AD treatments ease symptoms, but none has been clearly demonstrated to slow or halt disease progression. While the molecular cause of AD remains poorly understood, the hallmark pathological criteria for AD diagnosis is the deposition of amyloid-β (Aβ) plaques in the brain [92]. Aβ peptides are products of processive proteolysis by γ-secretase. Dominant missense mutations in the substrate (APP) and the enzyme (presenilin component of γ-secretase) cause early-onset FAD, and these mutations result in deficient carboxypeptidase trimming of initially

formed long Aβ peptides to shorter secreted forms [85,93,94]. Yet the mechanism of processive proteolysis of APP by γ-secretase is unknown. Recent reports of cryo-EM structures of γ-secretase bound to APP and Notch substrates as well as to γ-secretase inhibitors and modulators revealed details of the structural basis of substrate recognition as well as enzyme inhibition and modulation [58,79,95]. Regardless, static conformations of the enzyme cannot explain the underlying mechanism of enzyme activation and substrate processing. Essentially nothing is known about the dynamic mechanism of processive proteolysis by γ-secretase.

It would require quantum mechanics/molecular mechanics (QM/MM) calculations to fully understand the catalytic mechanism of proteolysis by γ-secretase. The catalytic step is likely the rate-limiting step of the enzyme proteolysis, being slower than the substrate-enzyme interaction dynamics. Nevertheless, the latter (dynamic motions of the substrate-enzyme interaction) has been suggested to take place over minutes[45]. This is still considered as slow dynamics and extremely long timescales that is way beyond the reach of *state-of-the-art* conventional MD simulations, but amenable to enhanced sampling simulations. We initially performed ~6 μs regular dual-boost GaMD simulations but could not sample stable enzyme-substrate hydrogen bond that characterizes system conformation for ζ cleavage of Aβ49 (Figure S3). Then we turned to our recently developed Pep-GaMD method, which selectively boosts the essential potential energy of the peptides. Pep-GaMD has been demonstrated to greatly accelerate protein-peptide binding simulations by orders of magnitude[86]. Compared with previous GaMD, Pep-GaMD is a more powerful method that can be applied for further improved enhanced sampling of protein-peptide interactions. The new Pep-GaMD simulations allowed us to capture the ζ cleavage activation in 600 ns. In this context, novel Pep-GaMD simulations have, *for the first time*, captured slow dynamic conformational transitions in both

the enzyme and substrate for tripeptide trimming of the wildtype and FAD mutants of Aβ49, being consistent with MS and western blotting biochemical experiments.

Here, we have applied the combination of novel Pep-GaMD enhanced sampling simulations and biochemical experiments to address the issue. Different systems of γ-secretase bound by the WT and FAD-mutant Aβ49 substrates were investigated to understand tripeptide trimmin, ζ cleavage **(Fig. 5)**. Five γ-secretase systems—bound to the WT, I45F, A42T and V46F charged C-terminal Aβ49 in presence charged N-terminal AICD50-99 and bound to WT neutral C-terminal Aβ49 and charged N-terminal AICD50-99— underwent activation for ζ cleavage during 600 ns Pep-GaMD simulations **(Fig. 5B)**. This was consistent with biochemical experiments, as these mutant systems showed similar efficiencies for the Aβ49 to Aβ46 proteolytic step (ζ cleavage). In comparison, γ-secretase bound by I45T and T48P Aβ49 showed little or no sample activation **(Fig. 5C)**. Furthermore, Aβ49-bound γ-secretase in the absence of AICD50-99 was not able to sample the "Final" active state for ζ cleavage of the substrate **(Fig. S8A)**, similarly for γ-secretase bound to WT charged C-terminal Aβ49 and neutral N-terminal AICD50-99 and γ-secretase bound to WT neutral C-terminal Aβ49 and neutral N-terminal AICD50-99 **(Fig. S8B-S8C)**. This highlighted the importance of AICD50-99 and its N-terminal charge in facilitating processive proteolysis by γ-secretase. Following ζ cleavage, both the C-terminus of Aβ49 and N-terminus of AICD50-99 at the active site could be exposed to water molecules and thus charged at physiological pH 7 (as carboxylate and ammonium, respectively). The charged state likely aided movement toward the polar aqueous environment and away from the hydrophobic transmembrane interior of the PS1 active site. Indeed, the AICD50-99 with charged N-terminus could dissociate from PS1 in the Pep-GaMD simulations that helped prepare for the next cleavage during processive proteolysis by γ-secretase.

During ζ cleavage activation, two residues unwound from the C-terminus of the Aβ49 helix, changing to a turn conformation **(Fig. 5B)**. This was observed in the time courses of the substrate secondary structures as well. Unlike the helical conformation, the loop/turn conformation facilitated exposure of the scissile amide bond to the catalytic aspartates and the coordinated water molecule. In parallel, positions of the Thr43 and Ile45 residues in the "Initial" and "Final" states relative to the membrane were similar, whereas the C-terminal residue Leu49 moved downwards by ~5 Å. Moreover, the helical domain of Aβ49 tilted by ~50 degrees **(Fig. 5B)**. Thus, tilting of the helical domain and unwinding of C-terminal helix in the substrate apparently facilitated the proteolytic progression from ε to ζ cleavage by γ-secretase. Helix unwinding was accompanied by straightening of the C-terminal loop/turn and downward movement of the terminal residue Leu49. Similarly, the β-sheet conformation between the APP C-terminus and the β1 loop was broken as ε cleavage product AICD50-99 dissociates. This caused the β1 loop to move away from the APP C-terminus by ~5 Å. This region has been suggested to be important for substrate recognition and proteolytic processing [79]. Similarly, γ-secretase inhibitors (GSIs) and transition state analogs (TSAs) bind to this region [95]. The present study also shows an important role of this region in activation of γ-secretase for ζ cleavage of Aβ49.

Relevant to this study, Hitzenberger et. al.[23] performed restraint MD simulations to produce γ-secretase complex structure bound to the Ab49, Ab46 and Ab43 peptides. Simulations on these complexes showed that both helix unwinding and sliding of active site aspartates towards the scissile amide bond are responsible for peptide repositioning during substrate processing by γ-secretase. During repositioning of the Ab peptides, the N-terminus was anchored to maintain its interaction with PS1 subunit. However, these enzyme-substrate

model complexes were generated by combining apo γ-secretase missing the Nicastrin subunit and C99 peptide using restrained MD. AICD50-99 peptide was not included in the γ-secretase study and charges on the terminal ends of the Ab49 and AICD50-99 peptide were not considered. In comparison, our model was based on the holo enzyme activated for ε cleavage from our previous study[8] generated using the APP bound γ-secretase cryo-EM structure. To the best of our knowledge, we are unaware of previous studies on the molecular dynamics of tripeptide trimming, in particular the first trimming step of Aβ49 to Aβ46, by the γ-secretase complex.

Pep-GaMD captured the enzyme activation for ζ cleavage for γ-secretase systems bound to WT and three FAD-mutant (I45F, A42T and V46F) APP substrates. The low-energy "Final" active conformation was identified in the Pep-GaMD free energy profiles of all these systems. However, the PMF profiles representing each enzyme system was different in terms of distinct low-energy states and the conformational space sampled by the enzyme during ζ cleavage. The I45F and V46F mutant systems sampled two low-energy conformations with the I45F system being the least conformationally dynamic **(Fig. 1H, J)**. Three and four low-energy states were identified from free energy profiles of the WT and the A42T mutant systems, respectively, with the A42T mutant system being more dynamic **(Fig. 1G, I)**. Each system had its own set of conformations and a distinct activation pathway. This suggested that the enzyme is remarkably dynamic, consistent with its ability to cleave over 100 different substrates[96].

In the "Final" active state of γ-secretase poised for the ζ cleavage, subpockets were formed in the active site that were different from that formed for the ε cleavage **(Fig. 4 and S20)**. This finding was consistent with the observation that the C-terminus of Aβ49 during ζ

cleavage did not form a β-sheet conformation with the PS1 TM6a β2 region, instead adopting a loop conformation **(Fig. 2F)**. The locations of the active site subpockets formed for ζ cleavage were compared to those formed for ε cleavage **(Fig. S20)**. Moreover, the locations of S1', S2' and S3' subpockets formed for ζ cleavage were the same for different γ-secretase systems bound to the WT and mutant Aβ49 except for the V46F mutant system. The S2' subpocket accommodating the T48 P2' residue formed for ζ cleavage was the same as the S1' subpocket accommodating the V50 P1' subpocket formed for ε cleavage. In contrast, the ζ cleavage S3' subpocket for the L49 P3' residue was the same as the ε cleavage S3' subpocket for the L52 P3' residue. The S1' subpocket for the I47 P1' residue for ζ cleavage and the S2' subpocket for the M51 P2' residue for ε cleavage had their own unique location in their respective Final active states. Regardless, S1'/S3' and S2' subpockets were located on opposite sides of the substrate in both of the "Initial" and "Final" active states.

During 600 ns of Pep-GaMD simulations, we did not observe the enzyme activation at other cleavage sites except for ζ cleavage at the amide bond between Val46-Ile47 of Aβ49 peptide. This can be observed from the time course plots of the distance between the protonated Asp257 and the carbonyl oxygen atoms of residues Ile47 (for cleavage at the second position) and Ile45 (for cleavage at the fourth position) (Figure S21). Moreover, the pathway of tripeptide trimming for ζ activation is energetically more favorable compared to that of the second or fourth amino acid residue cleavage. In this context, even though Pep-GaMD was able to capture the slow dynamic transitions of the enzyme activation for ζ cleavage during 600 ns simulation time, the simulations appeared to still suffer from insufficient sampling of the entire system conformational space and the calculated free energy profiles remained un-converged. Hence,

the free energy profiles reflect semi-quantitative picture of the tripeptide trimming process rather than the exact correctness of the free energy values.

Here, we have investigated both the wildtype and 5 FAD mutants (including I45F, A42T, V46F, I45T and T48P) of the substrate. In a recent report[85], we determined the Aβ42/40 ratios for these and other FAD mutations in APP substrate and found the relative effects of these mutations on this ratio compared to that seen with WT enzyme to be generally consistent with those reported from other groups[89-91,97]. To the best of our knowledge, there are no other studies comprehensively exploring APP FAD mutations and how they are processed by γ-secretase beyond our own[85]. Even with respect to Aβ42/40 ratios, a standard measure in the field, the only other comprehensive study of the 14 APP TMD FAD mutations was by our group[78], which gave closely similar relative changes in Aβ42/40 ratios with all these mutations, even though the systems were different (cellular transfection of APP[78] vs. C100Flag and purified proteins[85]).

Other reports on effects of APP mutations on Aβ42/40 have studied only very selected mutations. Among the 5 mutations studied in the current work: (1) Aβ42 levels cannot be determined by ELISA for A42T, as this changes the ELISA epitope; (2) No reported data can be found for T48P; Similar Aβ42/40 changes are seen with I45F and I45T[89-91]; The only discrepancy is with V46F: Devkota et al.[85] showed no change vs. WT, while Bolduc et. al.[78], Lichtenthaler et. al.[90], and Tamaoka et. al.[97] report ~4-fold increases in Aβ42/40. The reason for this discrepancy is unclear, however, only Devkota et. al. [85] used purified proteins, while the other reports measured secreted peptides in transfected cells.

In summary, we have presented here the first dynamic model of tripeptide trimming—of Aβ49 to Aβ46—by γ-secretase, which was highly consistent with mass spectrometry (MS)

and western blotting biochemical experiments. Specifically, MS and western blotting were used to quantify the efficiency of tripeptide trimming (Aβ49→Aβ46) of wildtype (WT) and familial Alzheimer's disease (FAD) mutant APP substrates. In comparison to WT APP, the efficiency of this first trimming step was similar for the I45F, A42T and V46F APP FAD mutants, but substantially diminished for the I45T and T48P mutants. All-atom simulations performed in parallel with the biochemical experiments captured remarkable structural rearrangements of both the enzyme and substrate, in which hydrogen-bonded catalytic aspartates and water became poised for tripeptide trimming of Aβ49 to Aβ46. Our complementary biochemical experiments and all-atom simulations have enabled elucidation of the mechanism of tripeptide trimming of γ-secretase. It will guide our future studies on subsequent cleavage steps of the APP substrate and processive cleavage of the other substrates of γ-secretase. Detailed mechanistic understanding of these processes is expected to greatly facilitate rational drug design of this critical enzyme.

Table 1: Summary of Pep-GaMD simulations performed on different systems of γ-secretase bound by Aβ49 and AICD50-99 peptides. [a]N_{atoms} is the total number of atoms in the simulation systems. [b]ΔV_{avg} and [c]$\sigma_{\Delta V}$ are the average and standard deviation of the Pep-GaMD boost potential, respectively.

System	[a]N_{atoms}	Dimension (Å^3)	Simulation (ns)	[b]ΔV_{avg} (kcal/mol)	[c]$\sigma_{\Delta V}$ (kcal/mol)
WT Aβ49	254,233	152 x 123 x 146	600 x 3	106.9	14.0
WT neutral Aβ49 - neutral AICD	254,337	152 x 123 x 146	600 x 3	155.9	11.8
WT neutral Aβ49 - charged AICD	254,340	152 x 123 x 146	600 x 3	134.2	11.3
WT charged Aβ49 - neutral AICD	254,334	152 x 123 x 146	600 x 3	133.9	11.1
WT charged Aβ49 - charged AICD	254,377	152 x 123 x 146	600 x 3	134.0	10.9
I45F charged Aβ49 - charged AICD	254,335	152 x 123 x 146	600 x 3	137.5	12.1
A42T charged Aβ49 - charged AICD	254,341	152 x 123 x 146	600 x 3	148.9	11.4
V46F charged Aβ49 - charged AICD	254,329	152 x 123 x 146	600 x 3	177.9	11.9
I45T charged Aβ49 - charged AICD	254,323	152 x 123 x 146	600 x 3	149.2	11.4
T48P charged Aβ49 - charged AICD	254,328	152 x 123 x 146	600 x 3	137.9	11.2

Figure 1

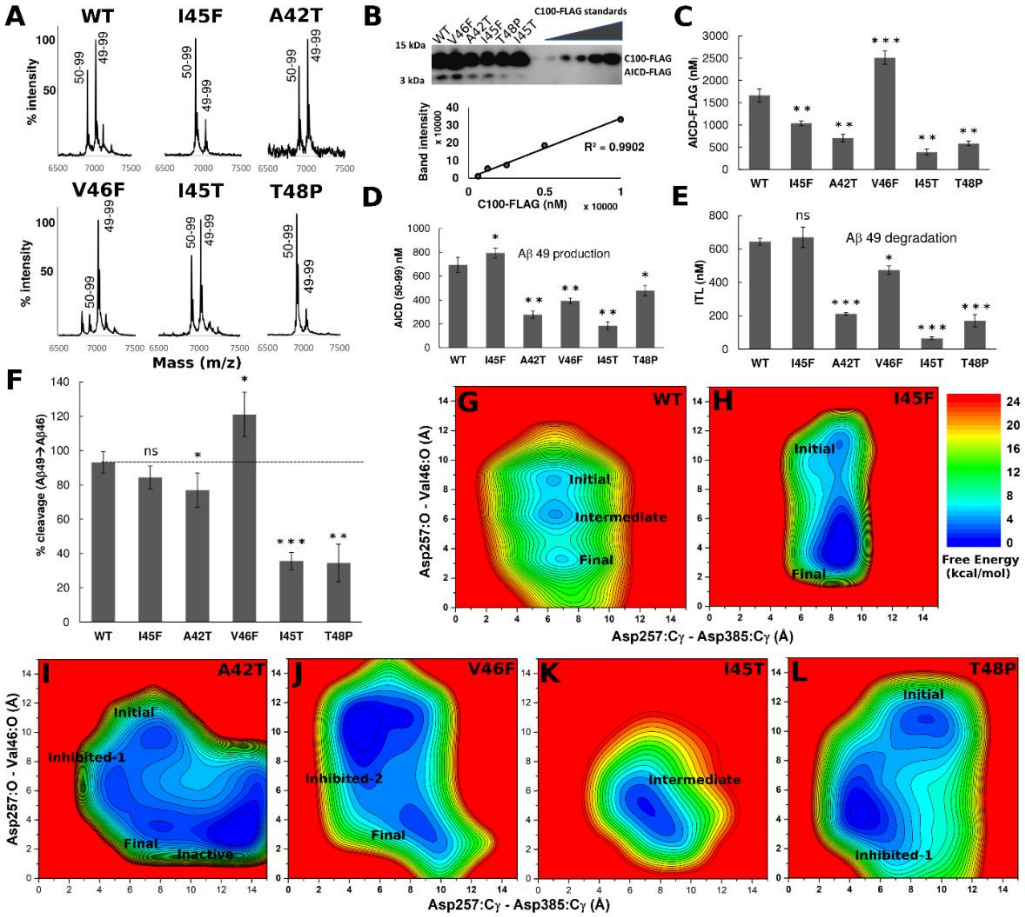

Figure 1: Tripeptide trimming of the wildtype (WT) and FAD mutants of Aβ49 γ-secretase characterized by MS, western blotting and Pep-GaMD simulations. (A) MALDI-TOF MS detection of AICD 50-99 and AICD 49-99 products, (B) Anti-FLAG immunoblot of total AICD-FLAG levels. Purified C100-FLAG at a range of known concentrations was used to generate a standard curve, (C) Quantification of total AICD-FLAG levels from immunoblot by densitometry, (D) Quantification of AICD 50-99 using total AICD levels determined from immunoblot and intensity ratios determined from MALDI-TOF MS, (E) Quantification of ITL tripeptides generated from the trimming of WT and FAD mutants of Aβ49, (F) Cleavage efficiency of the first trimming (ζ) step. Grey dotted line denotes cleavage efficiency from WT APP substrate. (G-L) 2D free energy profiles calculated from the Pep-GaMD simulations of (G) WT, (H) I45F, (I) A42T, (J) V46F, (K) I45T and (L) T48P Aβ49 bound to γ-secretase. The distances between the Cγ atoms of Asp257 and Asp385 in PS1 and between the hydroxyl oxygen of PS1 Asp257 and the carbonyl oxygen of Aβ49 Leu49 were selected as the reaction coordinates.

Figure 2

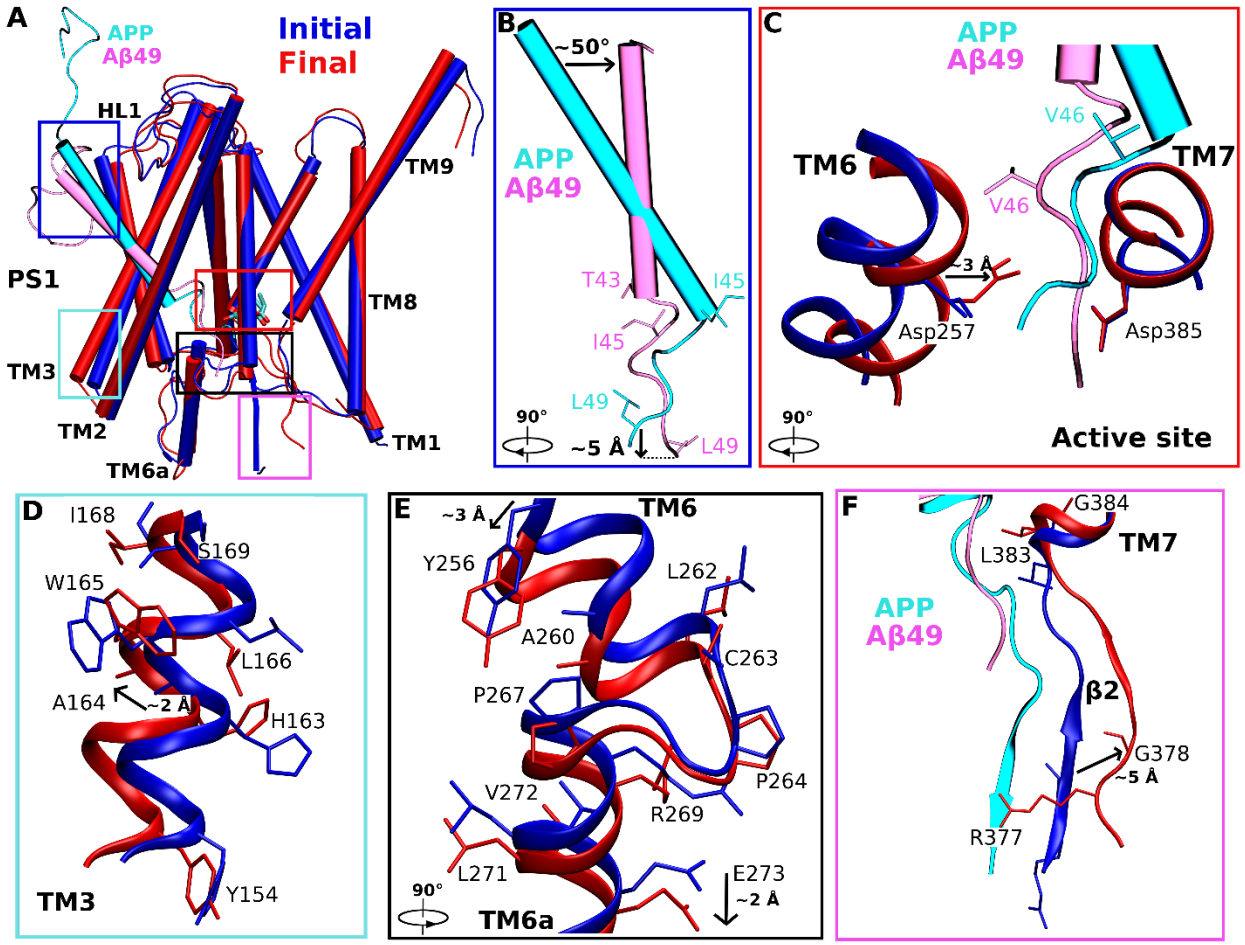

Figure 2: Conformational changes of the PS1 catalytic subunit and substrate during activation of γ-secretase for tripeptide trimming of Aβ49 in Pep-GaMD simulations. (A) Comparison of the Initial (active for ζ cleavage, blue) and Final (active for ζ cleavage, red) conformations of the Aβ49-bound PS1. The enzyme activation for tripeptide trimming was characterized by coordinated hydrogen bonding between the enzyme Asp257, carbonyl oxygen of Aβ49 Val46 and a water molecule accommodated between the two aspartates poised for cleavage of the amide bond between Val46 and Ile47 residues. (B-F) Conformational changes of (B) Aβ49 substrate, (C) catalytic aspartates, (D) TM3, (E) TM6 and TM6a, and (F) β2 strand from the Initial to the Final conformational state. The helical domain of Aβ49 tilted by ~50º and residue Leu49 at the C-terminus of Aβ49 moved downwards by ~5 Å. Protonated catalytic Asp257 moved ~3 Å towards the Aβ49 substrate. The enzyme TM3 moved outwards by ~2 Å and TM6a moved downwards by ~2 Å. The enzyme β2 strand (N-terminus of TM7) moved away from APP and closer towards the β1 strand (C-terminus of TM6a) by ~5 Å.

Figure 3

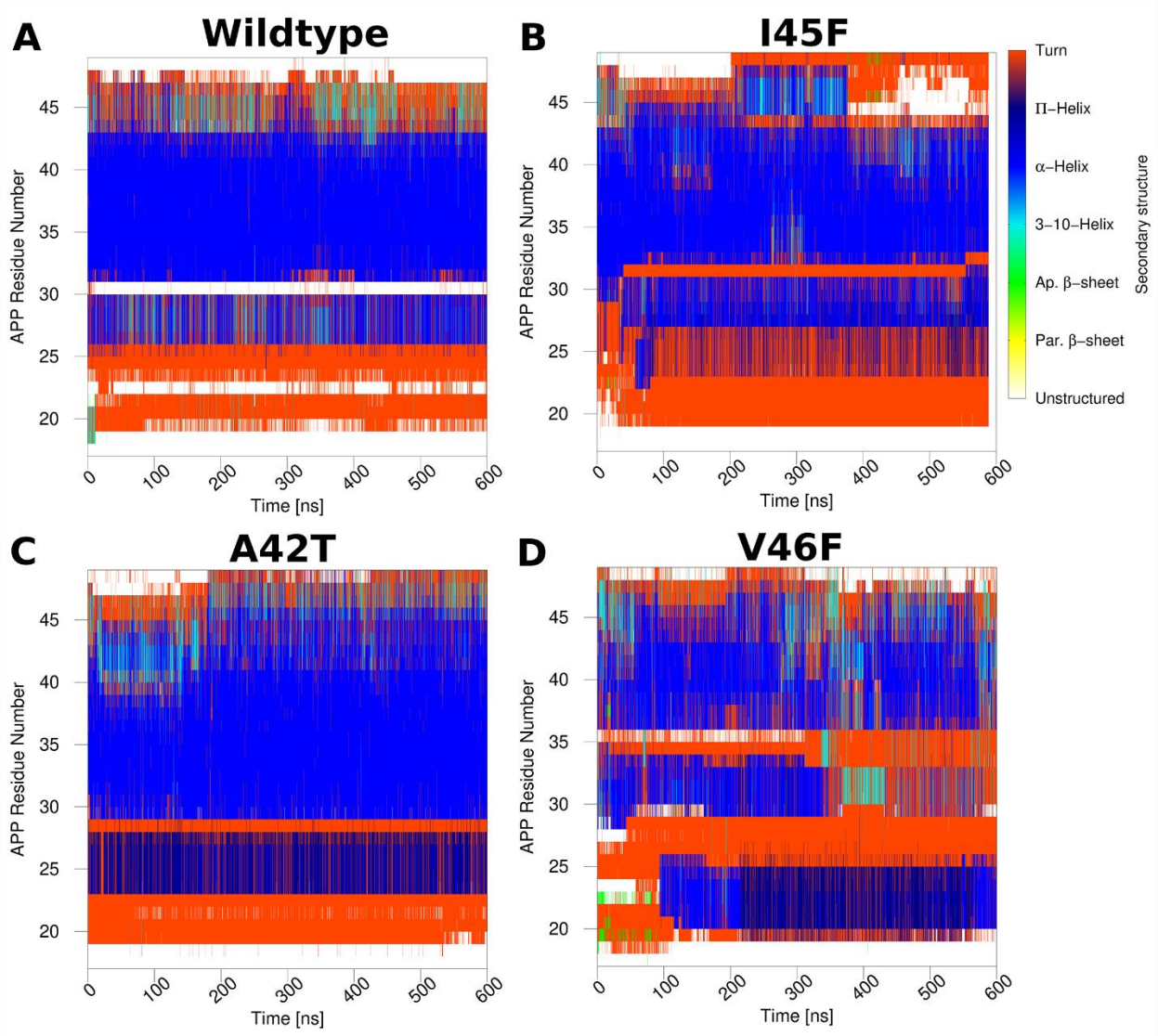

Figure 3: Time-dependent secondary structures of Aβ49 bound to γ-secretase calculated from the Pep-GaMD simulations. (A) WT, (B) I45F, (C) A42T and (D) V46F systems of Aβ49. Results from other simulations are plotted in Figs. S12 and S15.

Figure 4

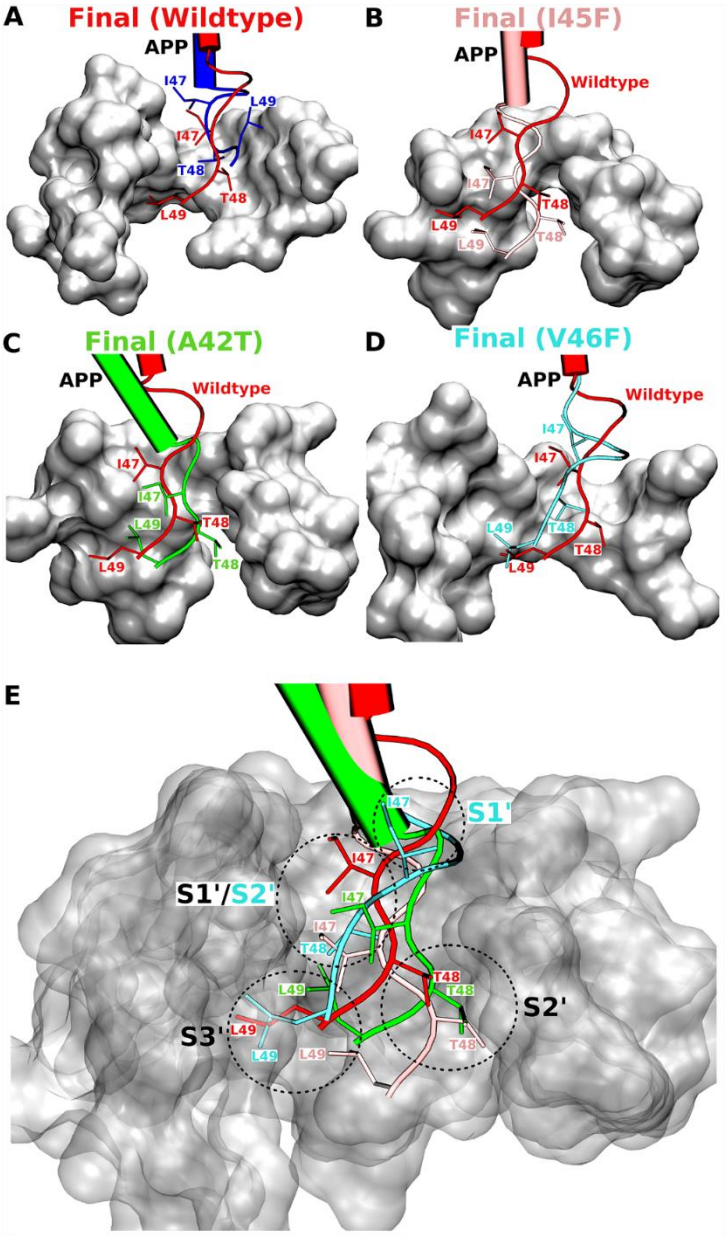

Figure 4: Active-site conformations of γ-secretase for tripeptide trimming of Aβ49 observed in the Pep-GaMD simulations. (A-D) Conformations of the substrate P1', P2' and P3' residues in the Final active conformations of the (A) WT (red), (B) I45F (pink), (C) A42T (green) and (D) V46F (cyan) Aβ49 systems. (E) Comparison of the PS1 active-site S1', S2' and S3' pockets that accommodate the WT and mutants of Aβ49.

Figure 5

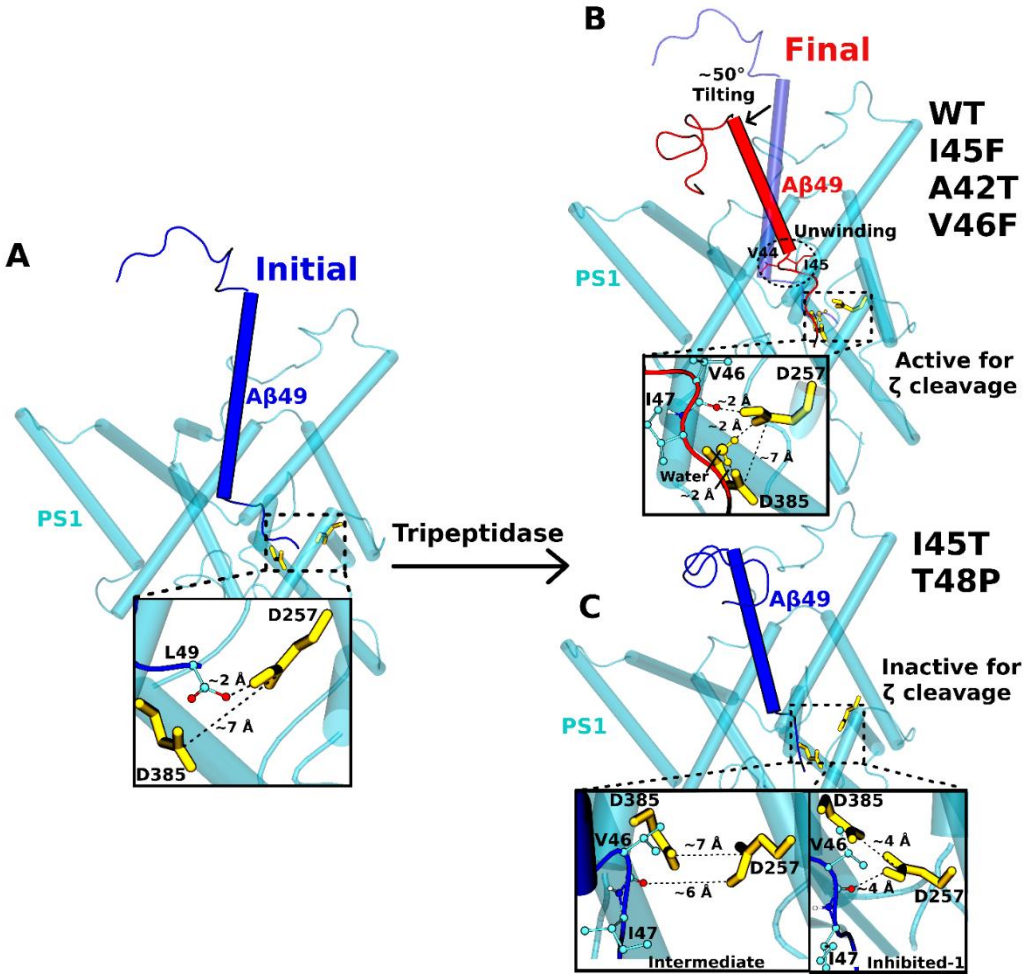

Figure 5: Dynamic model of tripeptide trimming of Aβ49 by γ-secretase. (A) The "Initial" conformational state of Aβ49 bound γ-secretase. (B) The WT Aβ49 and its I45F, A42T and V46F mutants were able to transition to the "Final" state with ~50° tilting of the helical domain and unwinding of the helix C-terminus (residues V44-I45) and became poised for ζ cleavage of the V46-I47 amide bond by γ-secretase. (C) In contrast, the I45T and T48P mutant Aβ49-bound γ-secretase were trapped in the "Intermediate" or "Inhibited-1" state, being inactive for ζ cleavage of the substrate.

Materials and Methods

C100-FLAG substrates expression and purification

C100-FLAG constructs [88] were transformed into *E. coli* BL21 cells. *E. coli* BL21 cells were grown in LB media at 37^0C in the incubator shaker until OD_{600} reached 0.6. Cells were induced with 0.5 mM IPTG and grown for 4 hours shaking at 37 ^{o}C. The cells were then pelleted and resuspended in lysis buffer composed of 50 mM HEPES pH 8 with 1% Triton X-100. The cells were lysed by French press three times and lysate was centrifuged to remove cell debris. The clear lysate was mixed with anti-FLAG M2-agarose beads (Sigma-Aldrich) for 16 h at 4 ^{o}C. Substrates were eluted from the beads with 100 mM glycine pH 2.5 with 0.25% NP-40, following by washing of the beads 3 times with lysis buffer. The elute was neutralized with Tris HCl and stored at -80^0C.

γ-Secretase assays

γ-Secretase purification and assays were carried out as described previously [8]. Briefly, 30 nM γ-secretase was incubated for 30 min at 37 °C in assay buffer composed of 50 mM Hepes pH 7.0, 150 mM NaCl, and 0.25% 3-[(3-cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate (CHAPSO) detergent supplemented with 0.1% phosphatidylcholine (DOPC) and 0.025% phosphatidylethanolamine (DOPE). Reactions were initiated by addition purified C100-FLAG substrate to a final concentration of 5 μM and performed by incubating at 37 °C for 16 h.

The enzymes used for all reactions were purified wild-type γ-secretase from the same enzyme preparation. Essentially, all of the purified enzymes were active, as determined using a stoichiometric γ-secretase inhibitor. There was no loss of specific activity, and any reduction in endoproteolytic cleavage of APP substrate or subsequent tripeptide trimming was due to the mutation in the substrate. Each substrate was purified to homogeneity and analyzed for its integrity and identity as we have recently reported[85]. Moreover, all enzyme reactions were conducted under substrate saturation and at a time point within the linear range of product formation (i.e., product levels are proportional to reaction rates).

Detection of AICD species

After 16 h, AICD-FLAG produced from the enzymatic assay was isolated by immunoprecipitation. The assay mixture was incubated with anti-FLAG M2 beads (SIGMA) in 10 mM MES pH 6.5, 10 mM NaCl, 0.05% DDM detergent for 16 h at 4 ^{o}C. AICD products were eluted from the anti-FLAG beads with acetonitrile:water (1:1) with 0.1% trifluoroacetic acid. The elutes were run on a Bruker autoflex MALDI-TOF mass spectrometer in linear mode.

Western blotting

Samples from γ-secretase assays and C100-FLAG standards were run on 4-12% Bis-Tris gel and transferred to PVDF membrane. The membrane was treated with 5% dry milk in PBS Tween-20 for 1 h at ambient temperature. The membrane was then incubated with anti-FLAG M2 antibodies at 4 ^{o}C overnight. The membrane was washed 3 times with PBS Tween-20 and

incubated with anti-mouse secondary antibodies for 1 h. The membrane was washed and imaged for chemiluminescence and band signal intensity was measured by densitometry.

LC-MS/MS tandem mass spectrometry

Small peptides were analyzed using an ESI Quadrupole Time-of-Flight (Q-TOF) mass spectrometer (Q-TOF Premier, Waters) by LC-MS/MS experiment, as previously described [85]. Briefly, assay samples and standard peptides were loaded onto a C18 column and eluted with a step gradient of 0.08% aqueous formic acid (A), acetonitrile (B), isopropanol (C), and a 1:1 acetone/dioxane mixture (D). The gradient well separated the lipids and detergent present in the buffer from the small peptides. The three most abundant collision-induced dissociation (CID) fragments were identified from the MS/MS for each small peptide. The peptide chromatographic area was obtained from the summed signals from three most abundant ions.

Simulation system setup

All-atom simulations using the Pep-GaMD method[86] were performed on the γ-secretase activation for ζ cleavage of Aβ49. Active APP-bound γ-secretase was taken from the previous study[8] and the amide bond between Aβ49 and AICD50-99 was cleaved as the starting structure. The enzyme was based on previously published cryo-EM structure[79] **(Fig. S1)** with Asp385 computationally restored, artificial enzyme-substrate disulfide bond removed and missing residues on APP N-terminus added. The Ala385 residue in the cryo-EM structure was computationally mutated back to Asp385. Two artificial disulfide bonds between Cys112 of

PS1-Q112C and Cys4 of PS1-V24C were removed as the wildtype residues were restored. SWISS-MODEL[65] homology modeling was used to restore 5 N-terminal APP residues that were missing in the cryo-EM structure. The simulation systems of γ-secretase bound by wildtype and mutant Aβ49 **(Figure S1)** were then prepared similarly as in the previous study for APP-bound γ-secretase and summarized in **Table 1**. For APP-mutant simulations systems, isoleucine, alanine, valine, isoleucine and threonine residues were mutated to phenylalanine, threonine, phenylalanine, threonine and proline computationally at the 45^{th}, 42^{nd}, 46^{th}, 45^{th} and 48^{th} residue of APP substrate, respectively. These corresponded to I45F, A42T, V46F, I45T and T48P mutations as per the numbering based on C99, although the actual substrate in the model was based on C83.

In the Pep-GaMD simulations, boost potential was applied selectively to the essential potential energy of the peptide (Aβ49 and AICD50-99) to effectively model its high flexibility and accelerate its dynamic motions. In addition to the γ-secretase systems bound by charged C-terminal Aβ49 and charged N-terminal AICD50-99, and neutral C-terminal Aβ49 and charged N-terminal AICD50-99, we tested Pep-GaMD simulations on enzyme systems bound by Aβ49 in the absence and presence of charged C-terminal Aβ49 and neutral N-terminal AICD50-99, and neutral C-terminal Aβ49 and neutral N-terminal AICD50-99 **(Fig. S4)**. The neutral and charged N-terminus of the AICD50-99 was characterized by the presence of $-NH_2$ and NH_3^+ functional groups at the N-terminal end, respectively. Similarly, the neutral and charged C-terminus of Aβ49 was characterized by the presence of COOH and COO- functional groups at the C-terminal end, respectively. Unlike the charged N-terminal AICD50-99 systems, activation was not observed during the 600 ns of Pep-GaMD of either of the enzyme systems bound by

Aβ49 in the absence and presence of neutral N-terminal AICD50-99 **(Figs. S6 and S8 and Table S2)**. Free energy profiles were plotted for the Pep-GaMD simulations of all these enzyme systems. Two low energy conformational states were identified in the system without the AICD bound including "Inhibited-2" and "Intermediate" **(Fig. S8A)**. Similarly, "Initial" and "Intermediate" low energy conformational states were identified in the free energy profile of the enzyme system bound to charged C-terminal Aβ49 in the presence of neutral N-terminal AICD50-99 **(Fig. S8B)**. The free energy profile of γ-secretase system bound by neutral C-terminal Aβ49 in presence of neutral N-terminal AICD50-99 identified two low energy states including "Initial" and "Intermediate" **(Fig. S8C)**. "Final" and "Intermediate" low energy states were identified in the free energy profiles of the enzyme system bound by neutral C-terminal Aβ49 and charged N-terminal AICD50-99 **(Fig. S8D)**. The "Intermediate" and the "Inhibited-2" conformational states here were same as the one identified in the wildtype and the V46F mutant γ-secretase systems, respectively **(Fig. 2A and 2D)**. The "Initial" conformational state resembled the one identified in the wildtype, I45F and A42T mutant systems **(Fig. 2A-2C)**. In comparison, "Final" active conformational state was identified in the wildtype system bound to Aβ49 and charged N-terminal AICD50-99 **(Fig. 2A)**. Therefore, systems for γ-secretase bound by Aβ49 and charged N-terminal AICD50-99 were used for final Pep-GaMD simulations.

Simulation Protocol

The CHARMM36m[98] parameter set was used for the protein and POPC lipids. Initial energy minimization and thermalization of the γ-secretase complex followed the same protocol as used in the previous study[8]. Then dual-boost Pep-GaMD simulations were performed to investigate

the γ-secretase enzyme activation for ζ cleavage (**Table 1**). The threshold energy E for adding boost potential was set to the upper bound, i.e., $E = V_{min} + (1/k)^{4,6}$. The simulations included 50 ns equilibration after adding the boost potential and then multiple independent production runs lasting 600 ns with randomized initial atomic velocities. Pep-GaMD production simulation frames were saved every 0.2 ps for analysis.

Simulation analysis

VMD[66] and CPPTRAJ [69] were used to analyze the Pep-GaMD trajectories. The distance between the catalytic aspartates was calculated between the Cγ atoms. Hydrogen bond distance was calculated between the donor protonated oxygen atom of PS1 Asp257 and the acceptor carbonyl oxygen atom of APP substrate residue Val46. Root-mean-square fluctuations (RMSFs) were calculated for the protein residues, averaged over three independent Pep-GaMD simulations and color coded for schematic representation of each complex system. The CPPTRAJ was used to calculate the protein secondary structure plots. The *PyReweighting* toolkit[62] was applied to reweight Pep-GaMD simulations for free energy calculations by combining all simulation trajectories for each system. Bin size of 1-3 Å was used for the PMF calculation of distances. The cutoff was set to 500-1000 frames in each bin for calculating the 2D PMF profiles. Protein snapshots were taken every 1 ps for structural clustering. Clustering was performed on the Pep-GaMD simulations of wildtype, I45F, A42T and V46F mutant Aβ49 bound γ-secretase based on the RMSD of PS1 using hierarchical agglomerative algorithm in CPPTRAJ [69] generating ~10 representative structural clusters for each system. The top structural cluster was identified as the representative Final active conformational states for each γ-secretase system.

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 project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231, and the Research Computing Cluster at the University of Kansas. This work was supported in part by the startup funding in the College of Liberal Arts and Sciences at the University of Kansas (Y.M.) and GM122894 from the National Institutes of Health (M.S.W.).

Appendix

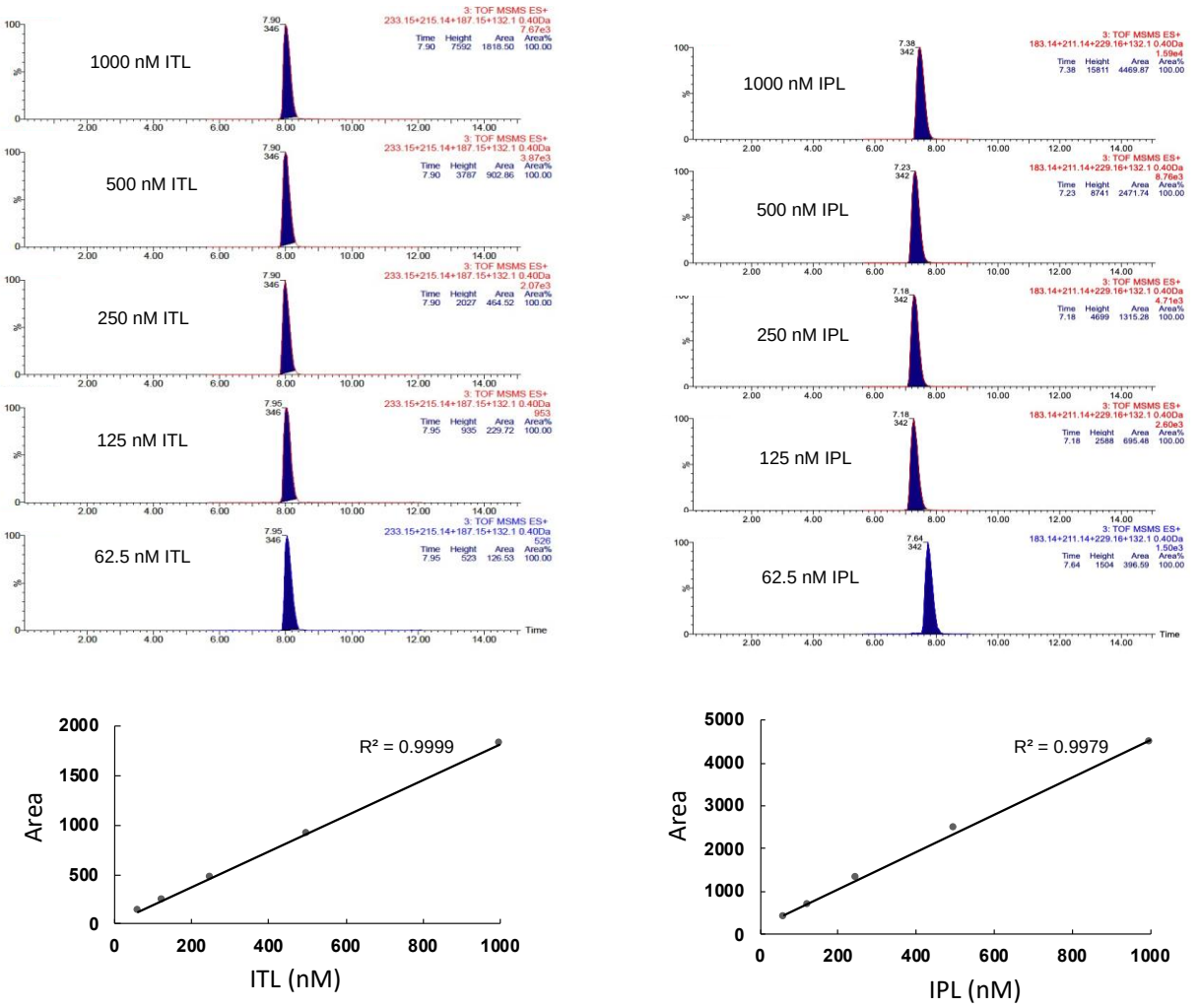

Figure S1: LC-MS/MS of all small-peptide standards (ITL and IPL) predicted to be generated for C100 substrates tested after γ-secretase digestion of substrates. Chromatograms are selected ion plots of the three most abundant sequence-specific product ions, selected with a 0.03 unit window. Standard curves for all small peptides were generated by plotting of the resulting peak areas of ion plots against the small-peptide concentration.

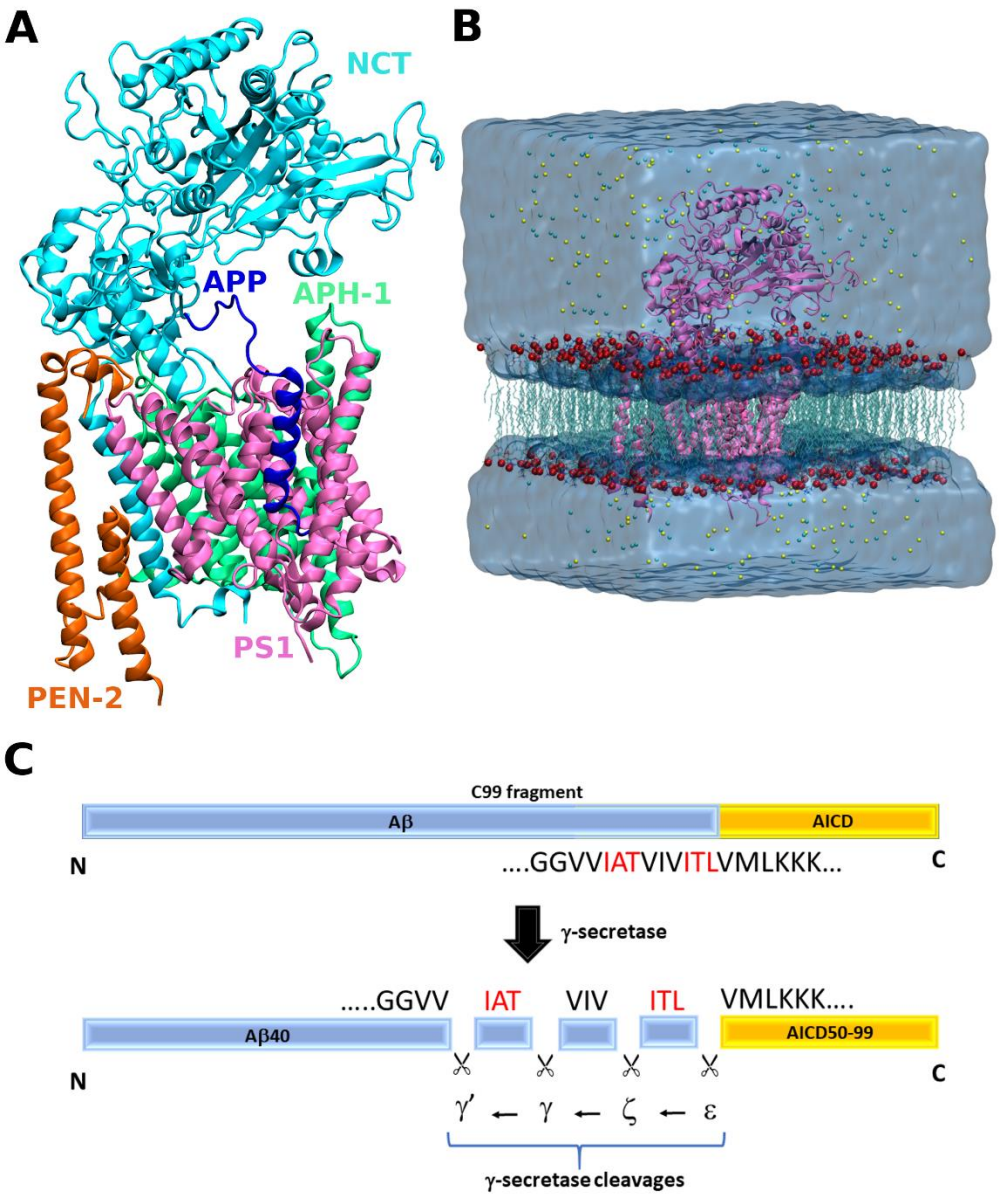

Figure S2: (A) γ-secretase structure bound to Aβ49 substrate (blue) with Niscastrin (NCT, cyan), Presenilin-1 (PS1, pink), Anterior Pharynx-Defective 1 (APH-1, green) and Presenilin Enhancer-2 (PEN-2, orange) subunits. The enzyme-substrate complex is represented in ribbons. (B) Pep-GaMD computational model of γ-secretase complex. The protein was embedded into a POPC lipid bilayer and solvated in an aqueous medium of 0.15 M NaCl. (C) Schematic representation of ε cleavage and processive proteolysis of APP substrate by γ-secretase.

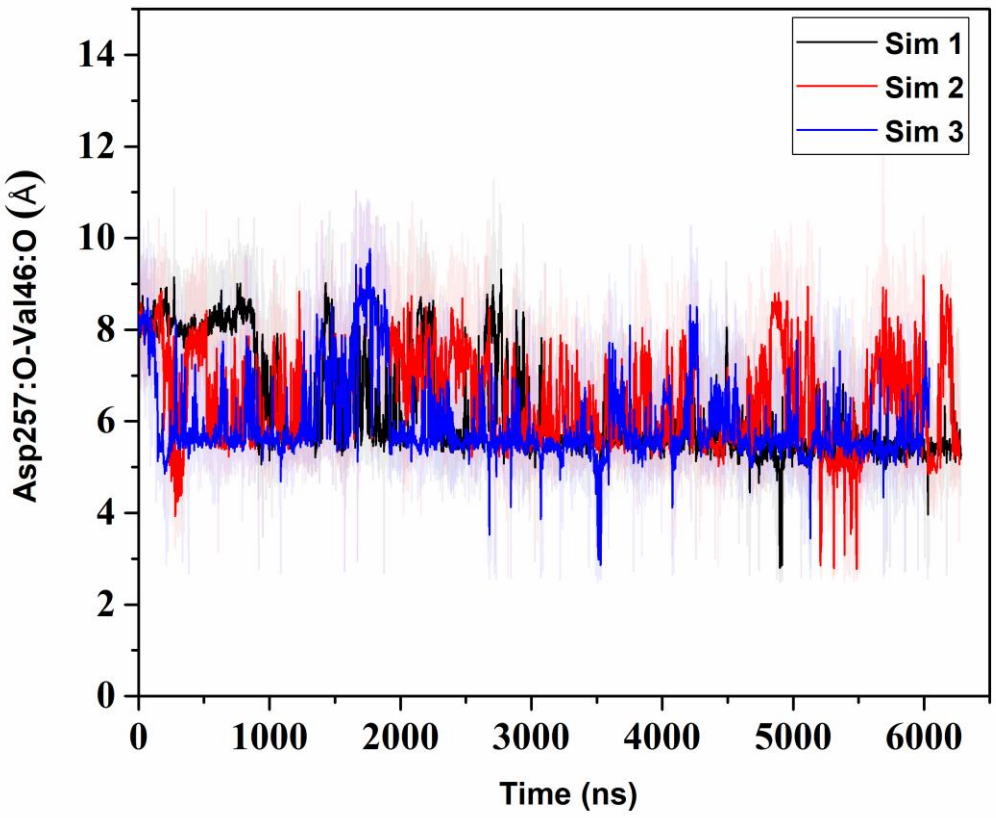

Figure S3: Timecourse of the Asp257:protonated O - Val46:O distance calculated from GaMD simulations of WT γ-secretase system. More than 6 μs long GaMD simulations of the enzyme could not capture stable activation for ζ cleavage.

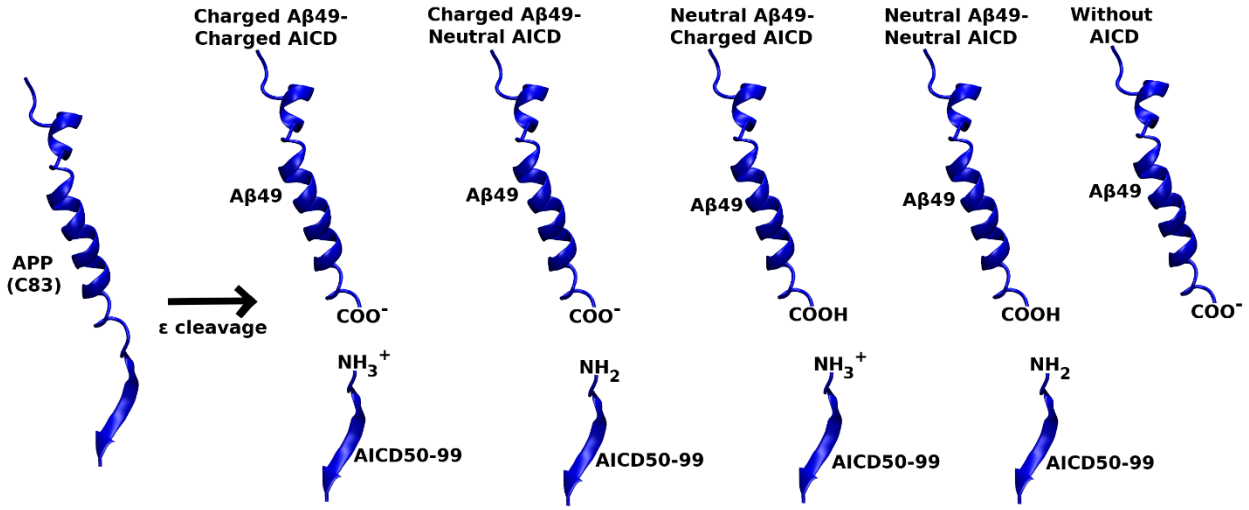

Figure S4: Ribbon representation of Aβ49 peptide and differently charged N-terminal AICD50-99 after ε cleavage of APP substrate by γ-secretase. Five different systems of γ-secretase bound to neutral and charged C-terminal Aβ49 in the absence and presence of neutral and charged N-terminal AICD were used for Pep-GaMD simulations.

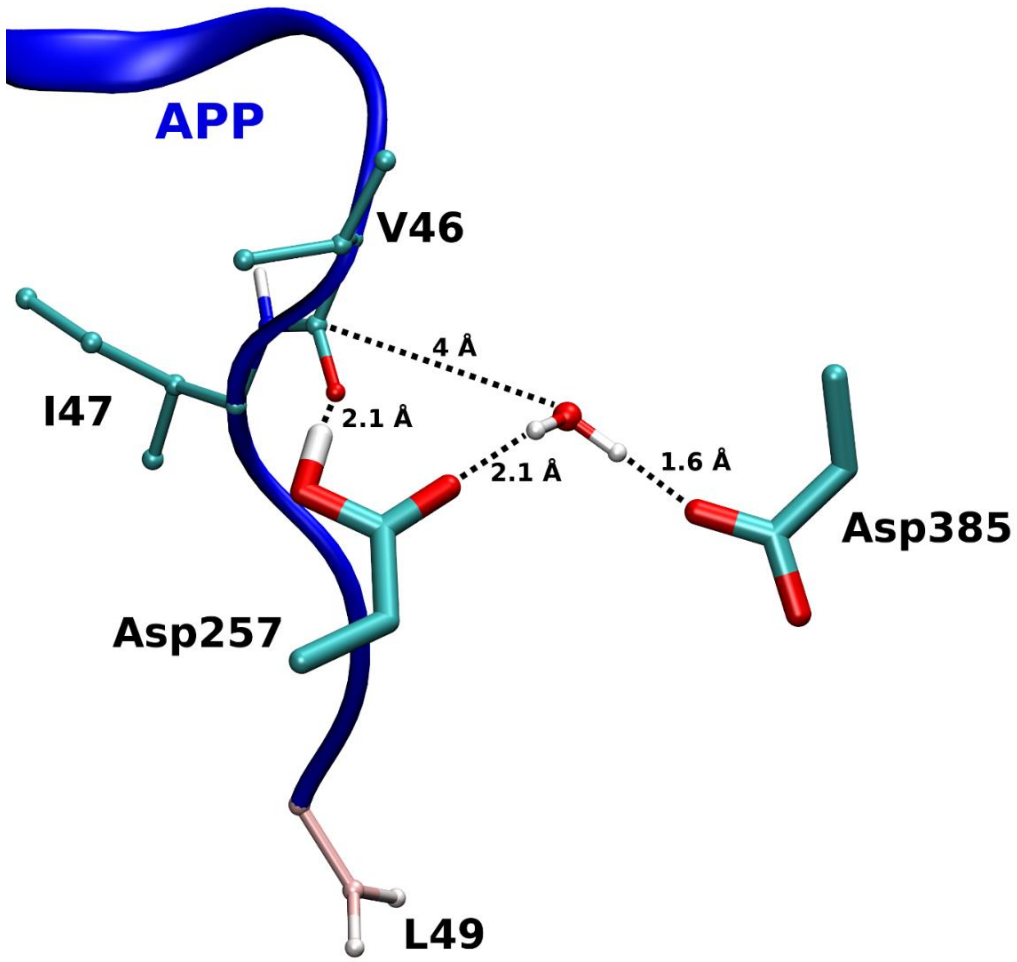

Figure S5: The active site poised for ζ cleavage proteolysis. The enzyme activation for ζ cleavage was characterized by coordinated hydrogen bonding between the enzyme Asp257 and carbonyl oxygen of C99 Val46. The water molecule could form hydrogen bond interactions with both catalytic aspartates and is at ~4 Å distance away from the carbonyl carbon of Val46 residue. The APP substrate (blue), aspartates and APP residues are shown as ribbon, stick and, balls and sticks, respectively.

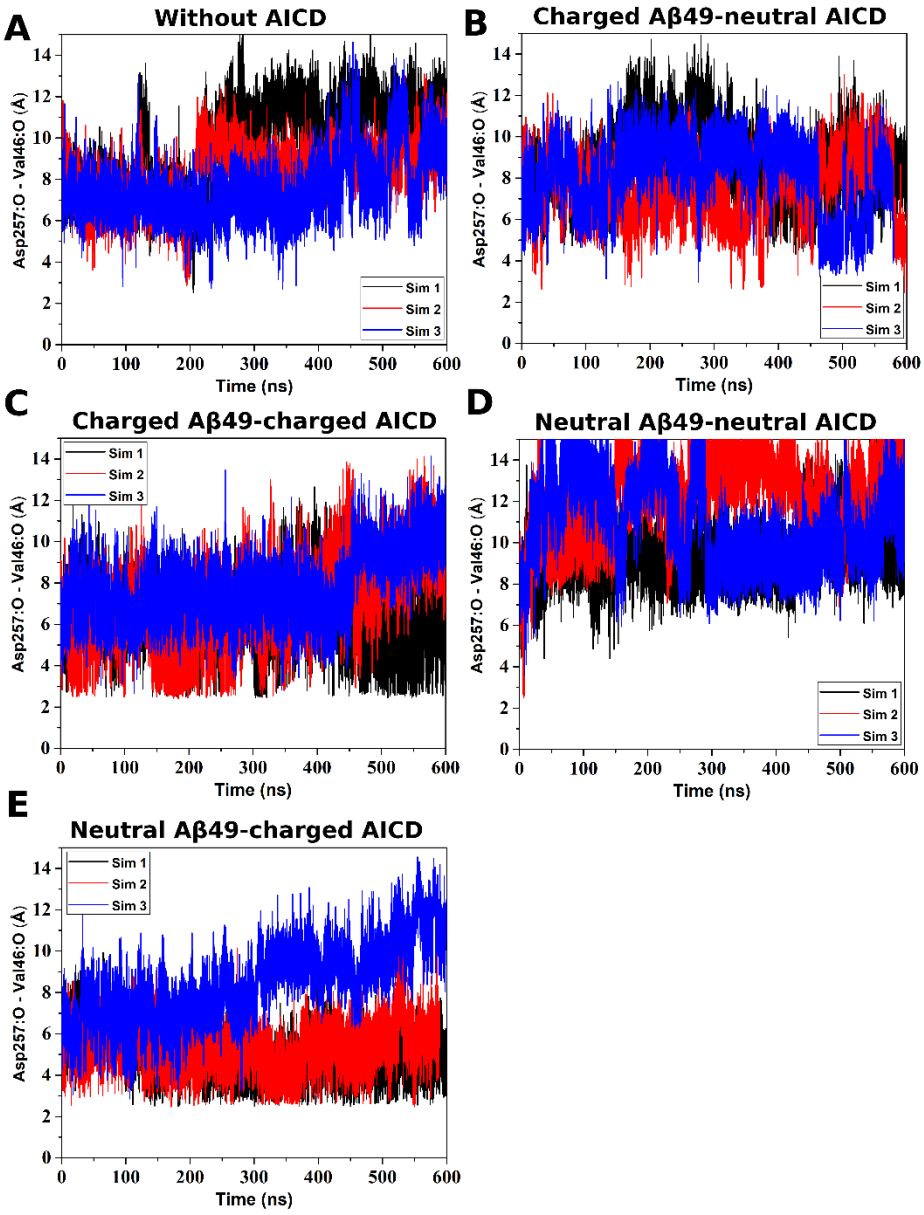

Figure S6: Time courses of the Asp257:protonated O – Val46:O distances calculated from Pep-GaMD simulations of (A) WT without AICD , (B) WT with charged Aβ49 and neutral AICD, (C) WT with charged Aβ49 and charged AICD, (D) WT with neutral Aβ49 and neutral AICD, and (E) with neutral Aβ49 and charged AICD bound γ-secretase systems.

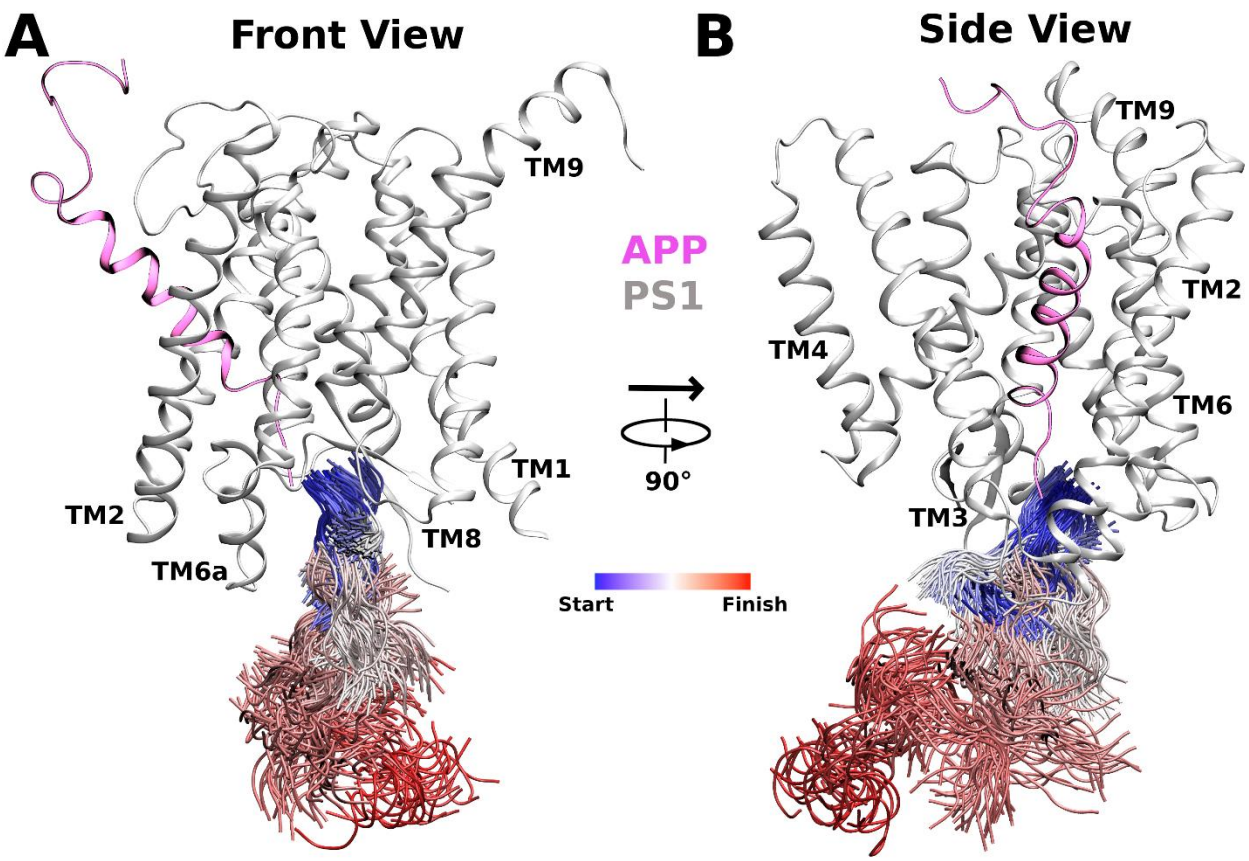

Figure S7: AICD50-99 dissociation pathway observed in Pep-GaMD simulations of γ-secretase system bound to wildtype APP colored by simulations time in a blue-white-red (BWR) color scheme.

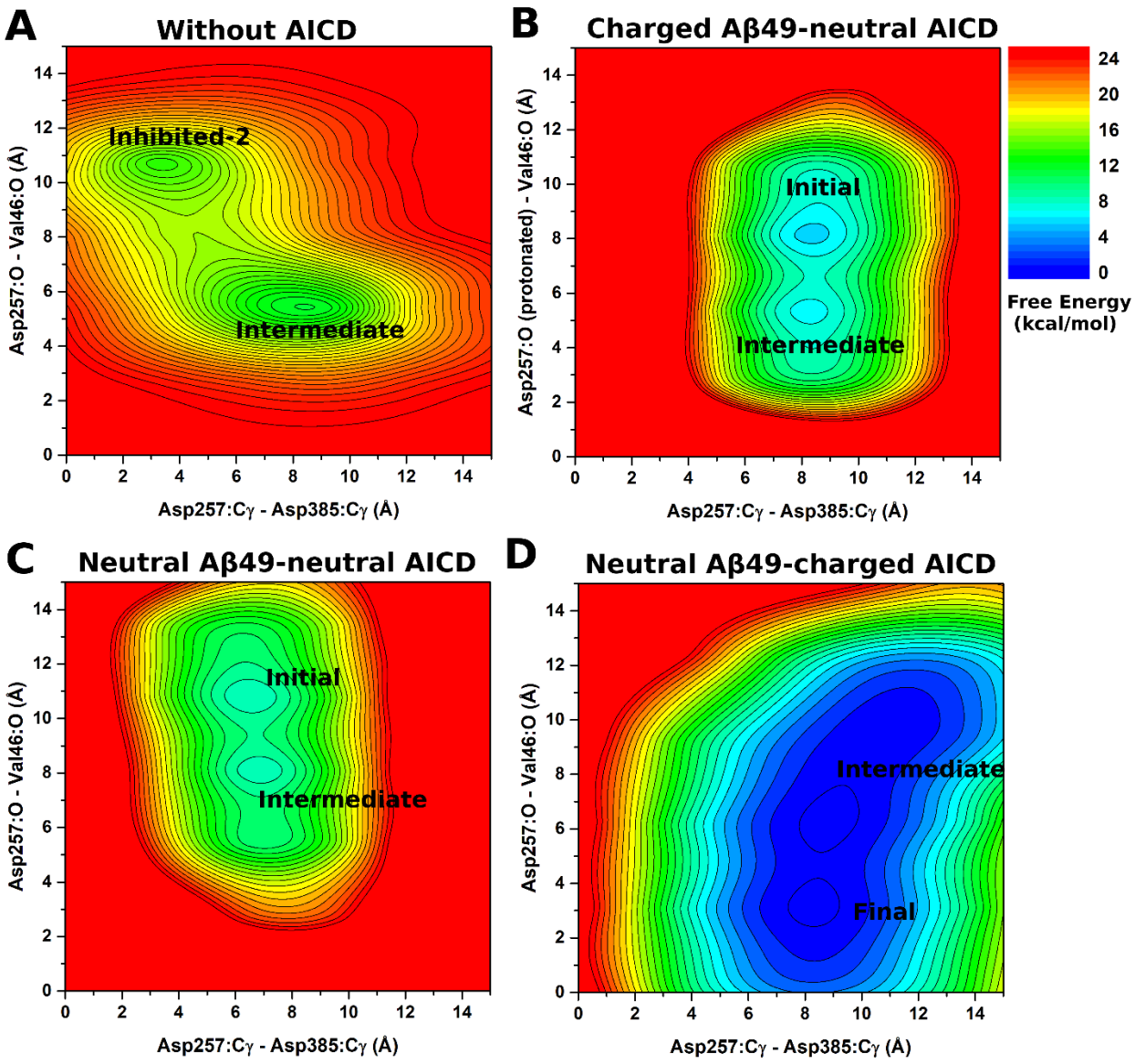

Figure S8: 2D free energy profiles of the Asp257:Cγ - Asp385:Cγ and Asp257:protonated O - Leu49:O distances calculated from Pep-GaMD simulations of (A) wildtype without AICD, (B) wildtype with charged C-terminal Aβ49 and neutral N-terminal AICD50-99, (C) wildtype with neutral C-terminal Aβ49 and neutral N-terminal AICD50-99, and (D) wildtype with neutral C-terminal Aβ49 and charged N-terminal AICD50-99 γ-secretase systems.

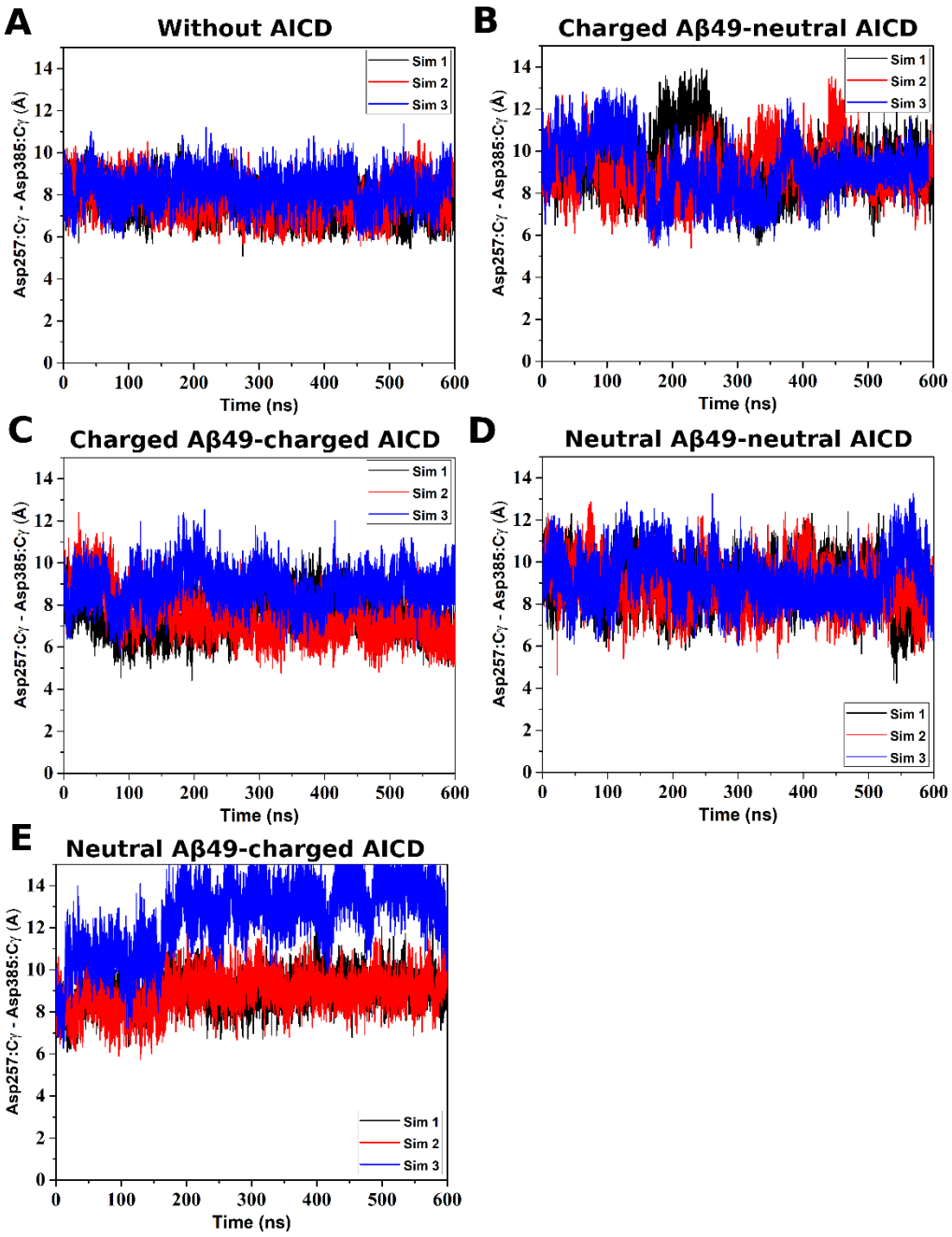

Figure S9: Time courses of the Asp257:Cγ - Asp385:Cγ distances calculated from Pep-GaMD simulations of (A) WT without AICD, (B) WT with charged Aβ49 and neutral AICD, (C) WT with charged Aβ49 and charged AICD, (D) WT with neutral Aβ49 and neutral AICD, and (E) with neutral Aβ49 and charged AICD bound γ-secretase systems.

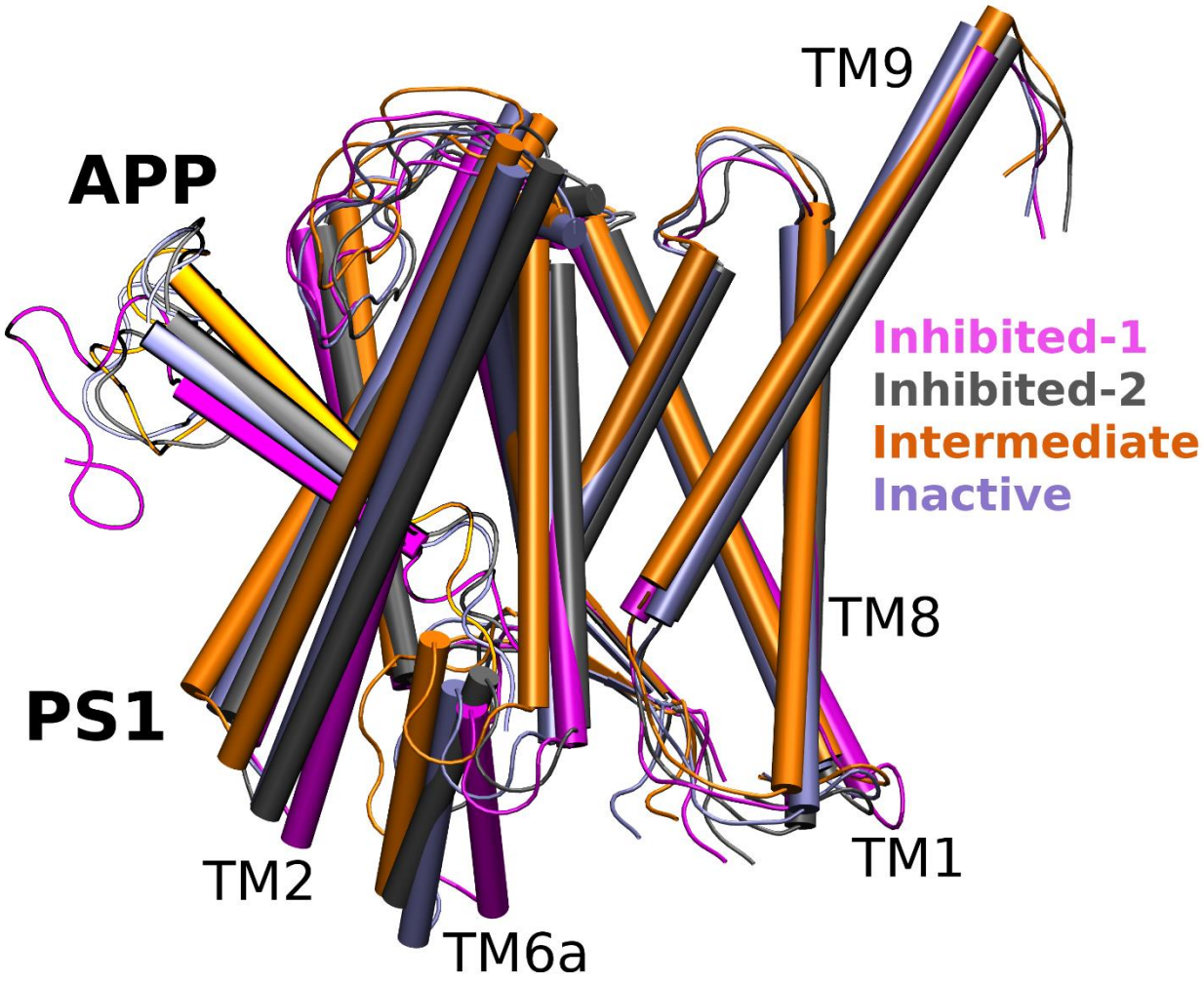

Figure S10: Comparison of different low energy state conformations as identified from Pep-GaMD free energy profiles of wildtype and mutant APP bound γ-secretase systems including Inhibited-1 (magenta), Inhibited-2 (gray), Intermediate (orange) and Inactive (ice blue) states.

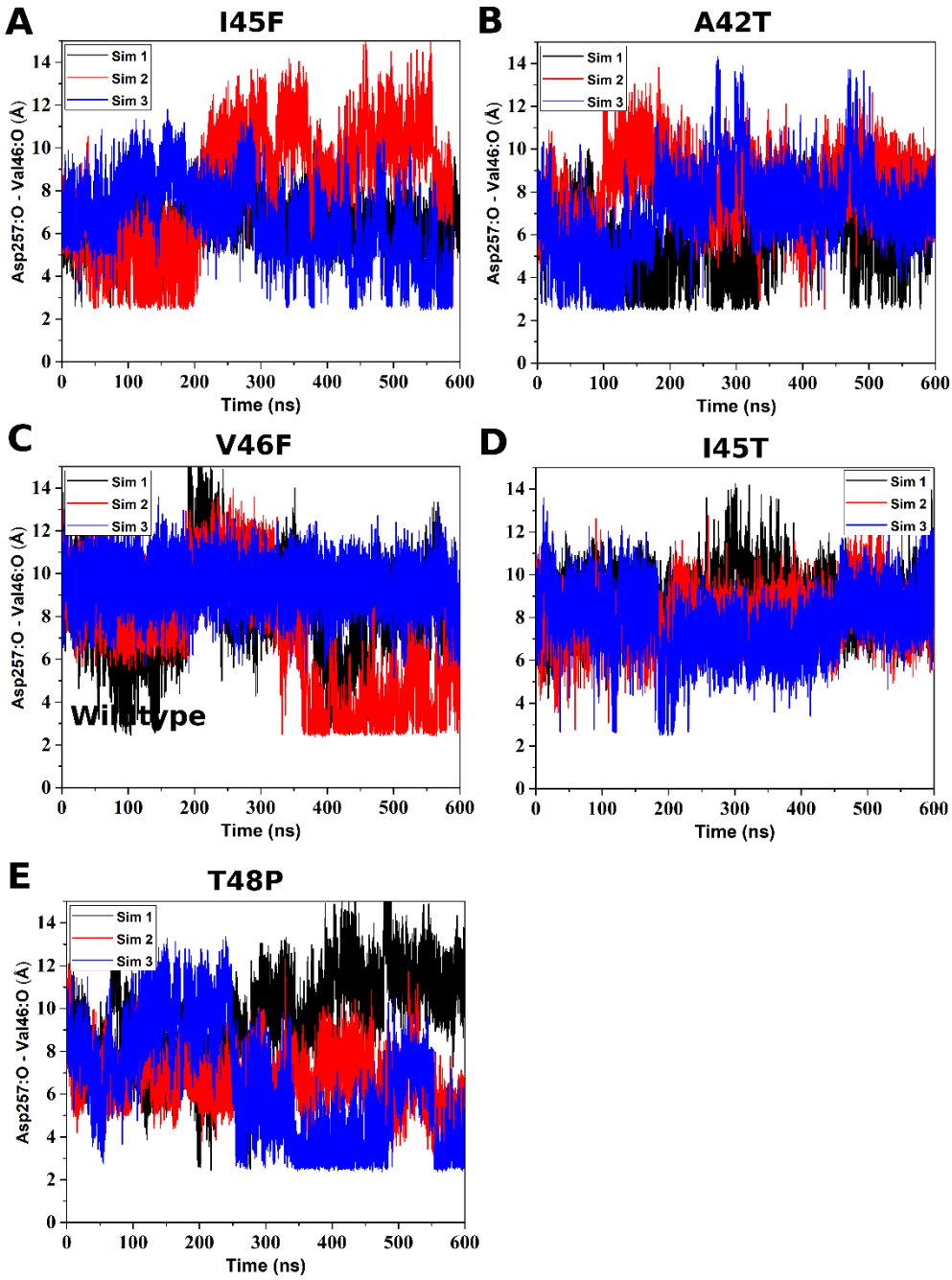

Figure S11: Time courses of the Asp257:protonated O – Val46:O distances calculated from Pep-GaMD simulations of (A) I45F, (B) A42T, (C) V46F, (D) I45T and (E) T48P APP, all with C-terminally charged Aβ49 in presence of N-terminally charged AICD, bound γ-secretase systems.

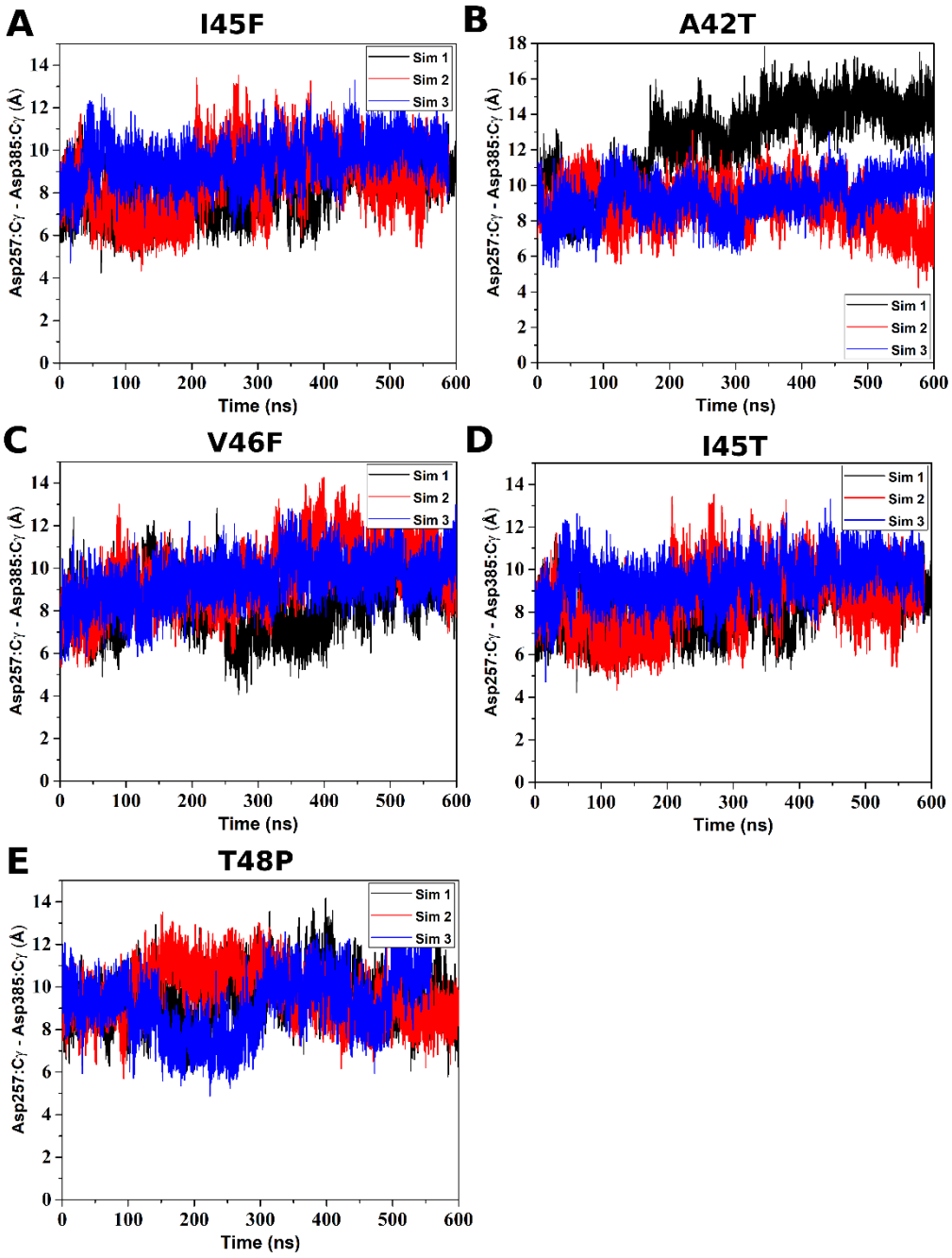

Figure S12: Time courses of the Asp257:Cγ - Asp385:Cγ distances calculated from Pep-GaMD simulations of (A) I45F, (B) A42T, (C) V46F, (D) I45T and (E) T48P APP, all with C-terminally charged Aβ49 in presence of N-terminally charged AICD, bound γ-secretase systems.

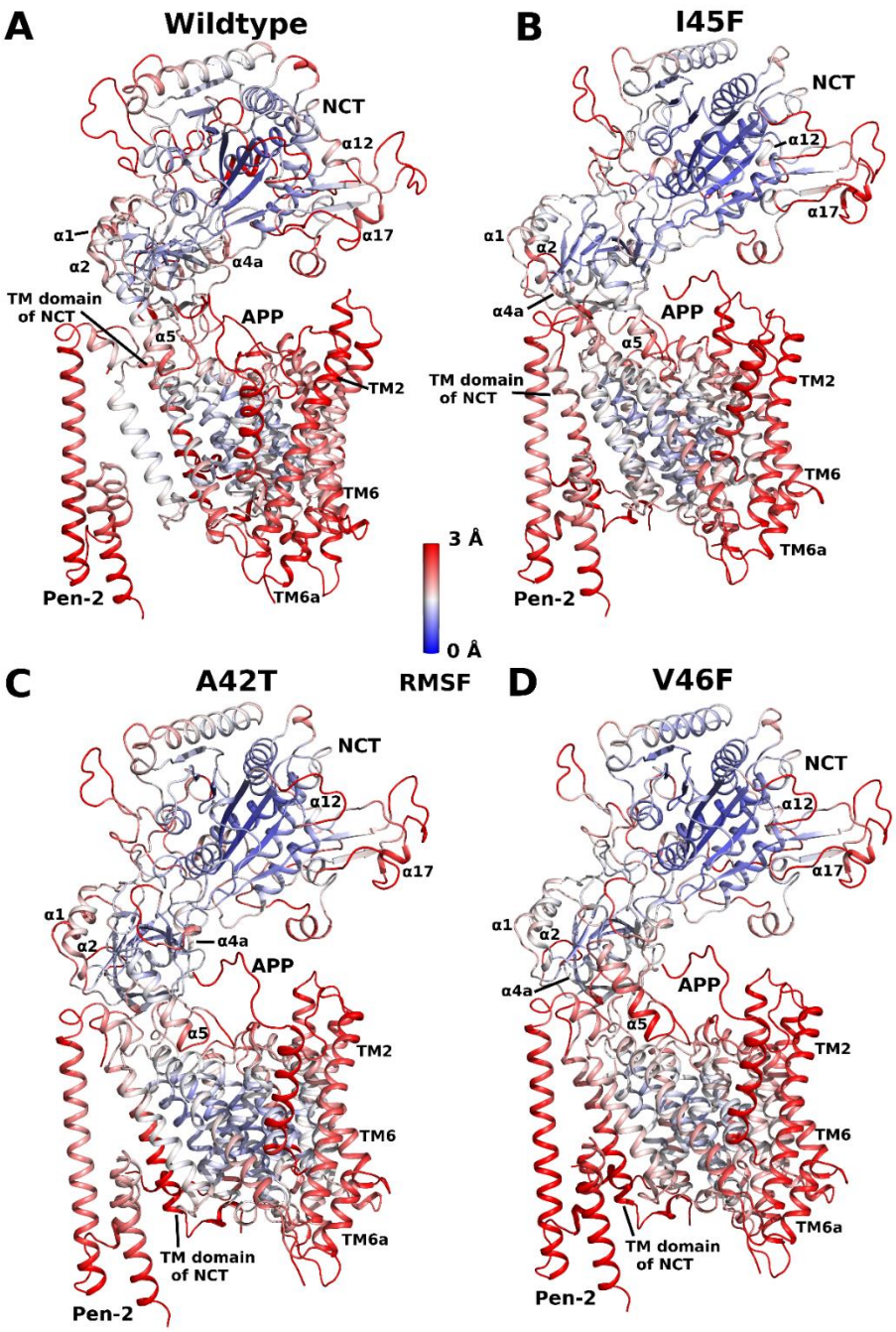

Figure S13: Root mean square fluctuation (RMSF) plots of different γ-secretase systems bound to (A) wildtype, (B) I45F mutant, (C) A42T mutant, and (D) V46F mutant APP. The RMSF is shown in blue-white-red color scheme for 0-3 Å of fluctuations in the enzyme-substrate complex.

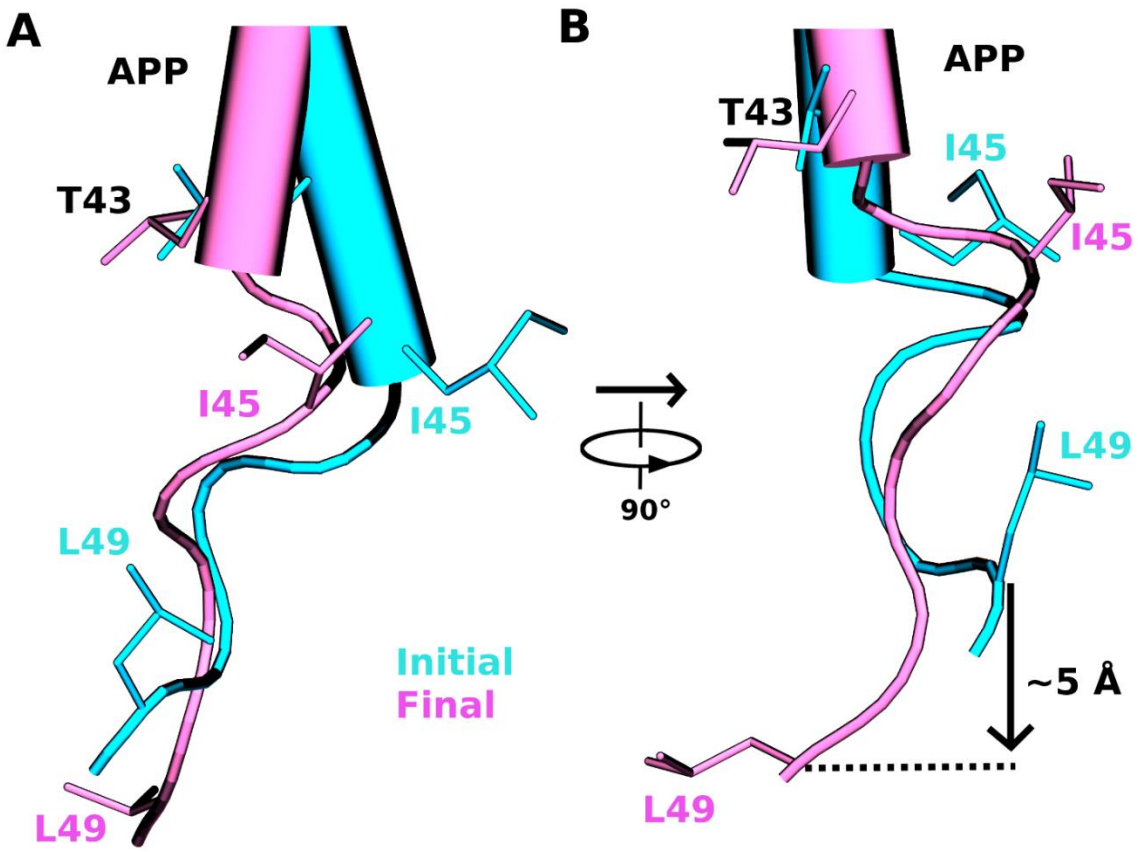

Figure S14: (A) Front and **(B)** side view comparison of relative positions of APP residues T43, I45 and L49 in the Initial and Final active states of the γ-secretase.

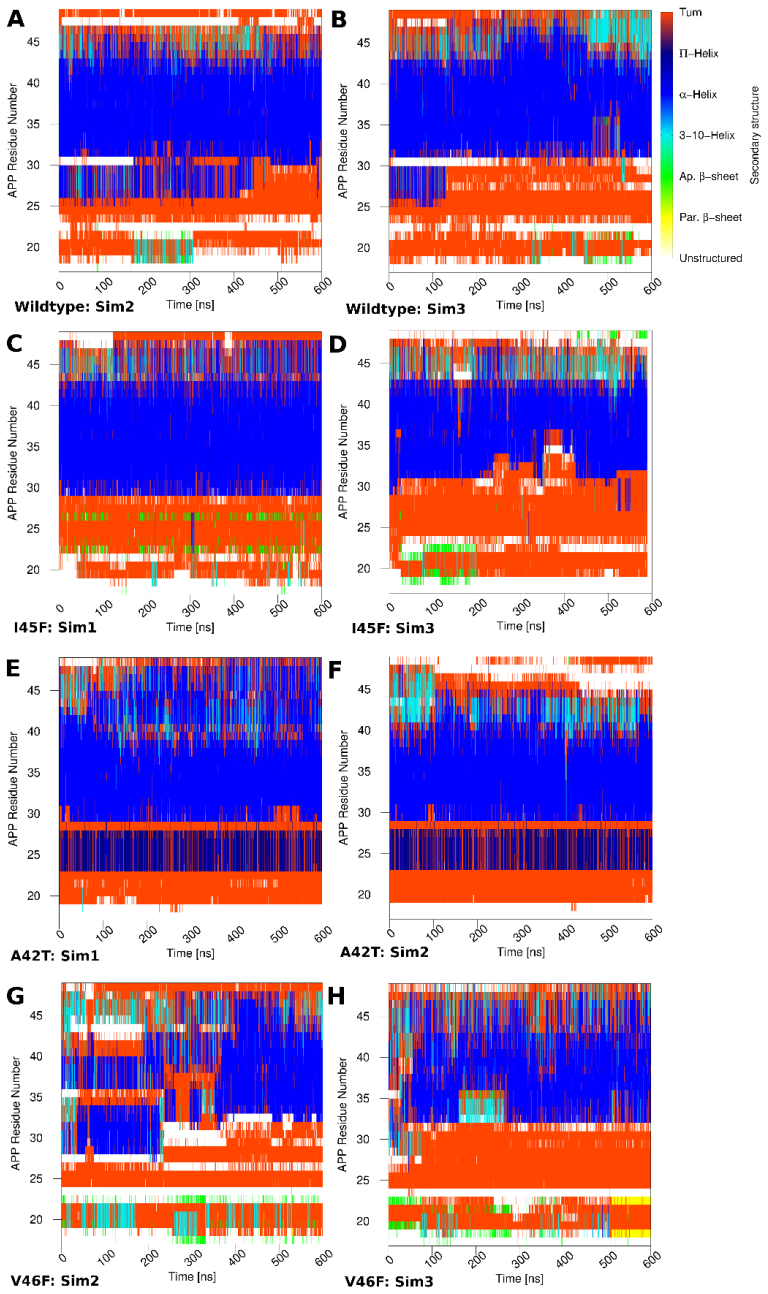

Figure S15: Time courses of secondary structures changes in the (A) Sim 2 and (B) Sim3 Pep-GaMD simulations of WT, (C) Sim 1 and (D) Sim 3 Pep-GaMD simulations of I45F, (E) Sim1 and (F) Sim2 Pep-GaMD simulations of A42T, (G) Sim2 and (H) Sim3 Pep-GaMD simulations of V46F mutant APP bound to γ-secretase.

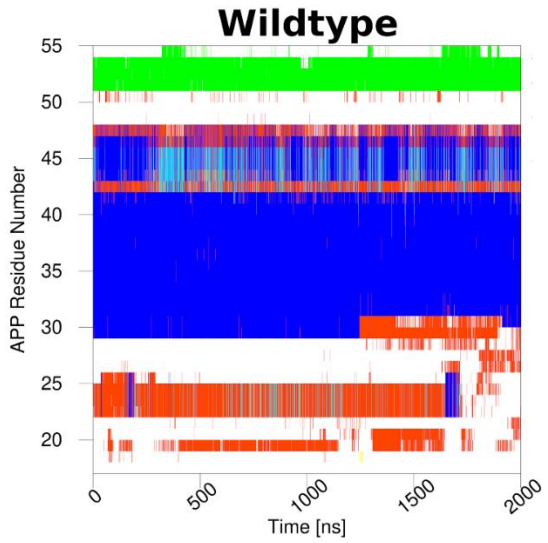

Figure S16: Time course of secondary structure changes in the GaMD simulations of WT APP substrate-bound γ-secretase recorded during enzyme activation for e cleavage. This plot is extracted from our previous study (Bhattarai, Apurba, et al. *ACS central science* 6.6 (2020): 969-983.).

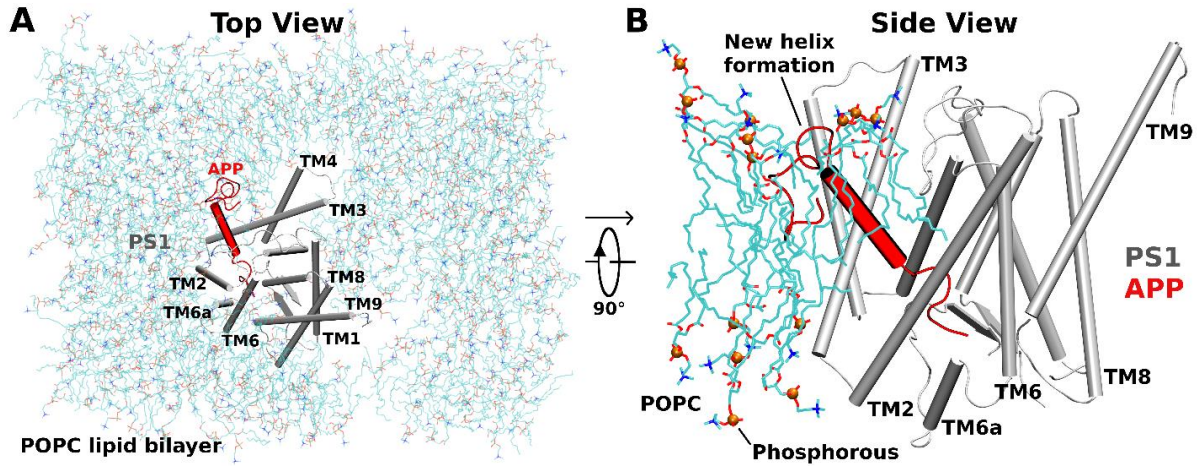

Figure S17: (A) Top and (B) Side view of APP (Aβ49) bound γ-secretase PS1 interacting with the POPC lipid bilayer membrane. The N-terminus of APP substrate during the ζ cleavage activation bends and interacts with the hydrophobic lipid bilayer to form α-helix conformation.

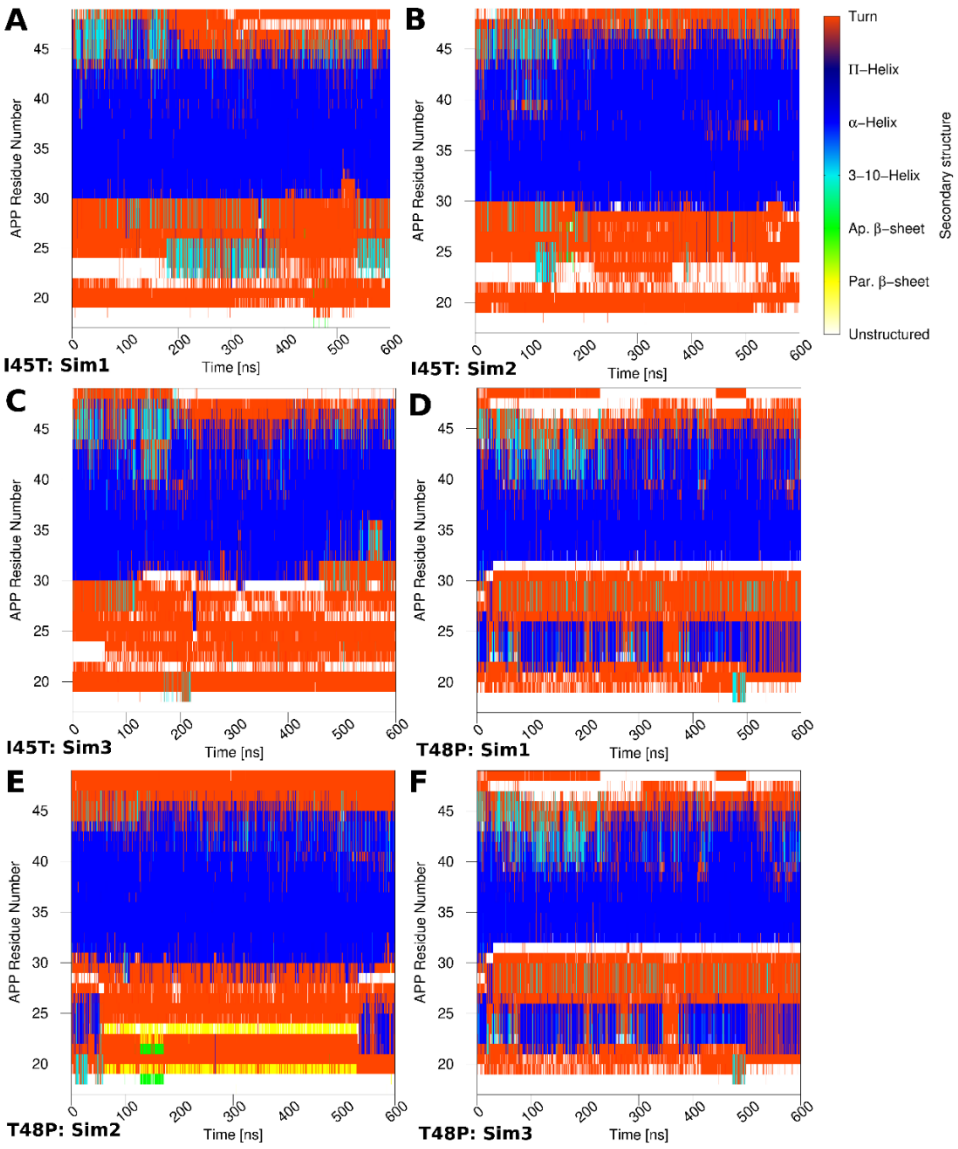

Figure S18: Time courses of secondary structure changes in the (A) Sim 1, (B) Sim2 and (C) Sim3 Pep-GaMD simulations of I45T mutant and (D) Sim 1, (E) Sim2 and (F) Sim3 Pep-GaMD simulations of T48P mutant APP bound to γ-secretase.

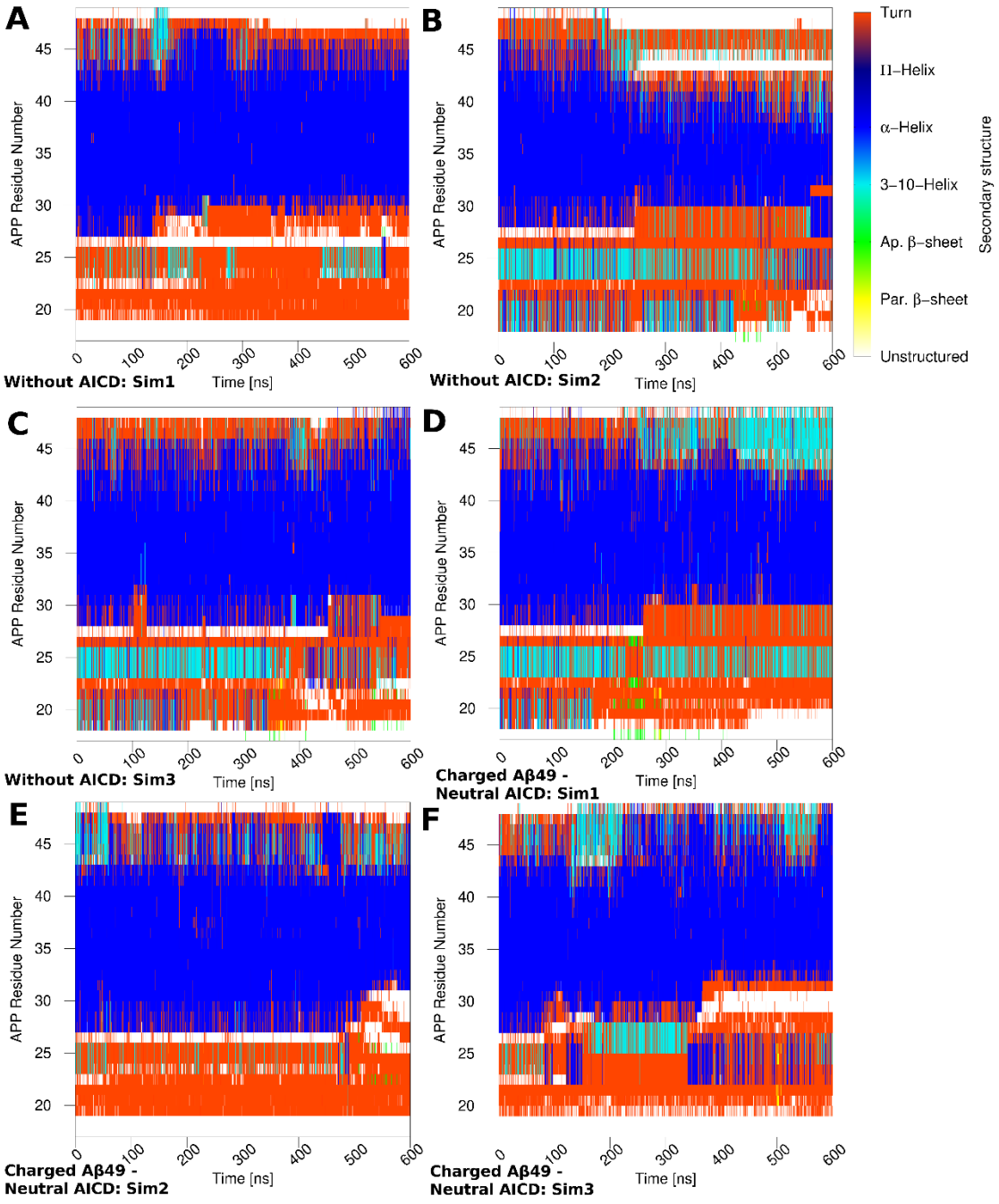

Figure S19: Time courses of secondary structure changes in the (A) Sim 1, (B) Sim2 and (C) Sim3 Pep-GaMD simulations of wildtype without AICD and (D) Sim 1, (E) Sim2 and (F) Sim3 Pep-GaMD simulations of wildtype with neutral AICD bound to γ-secretase.

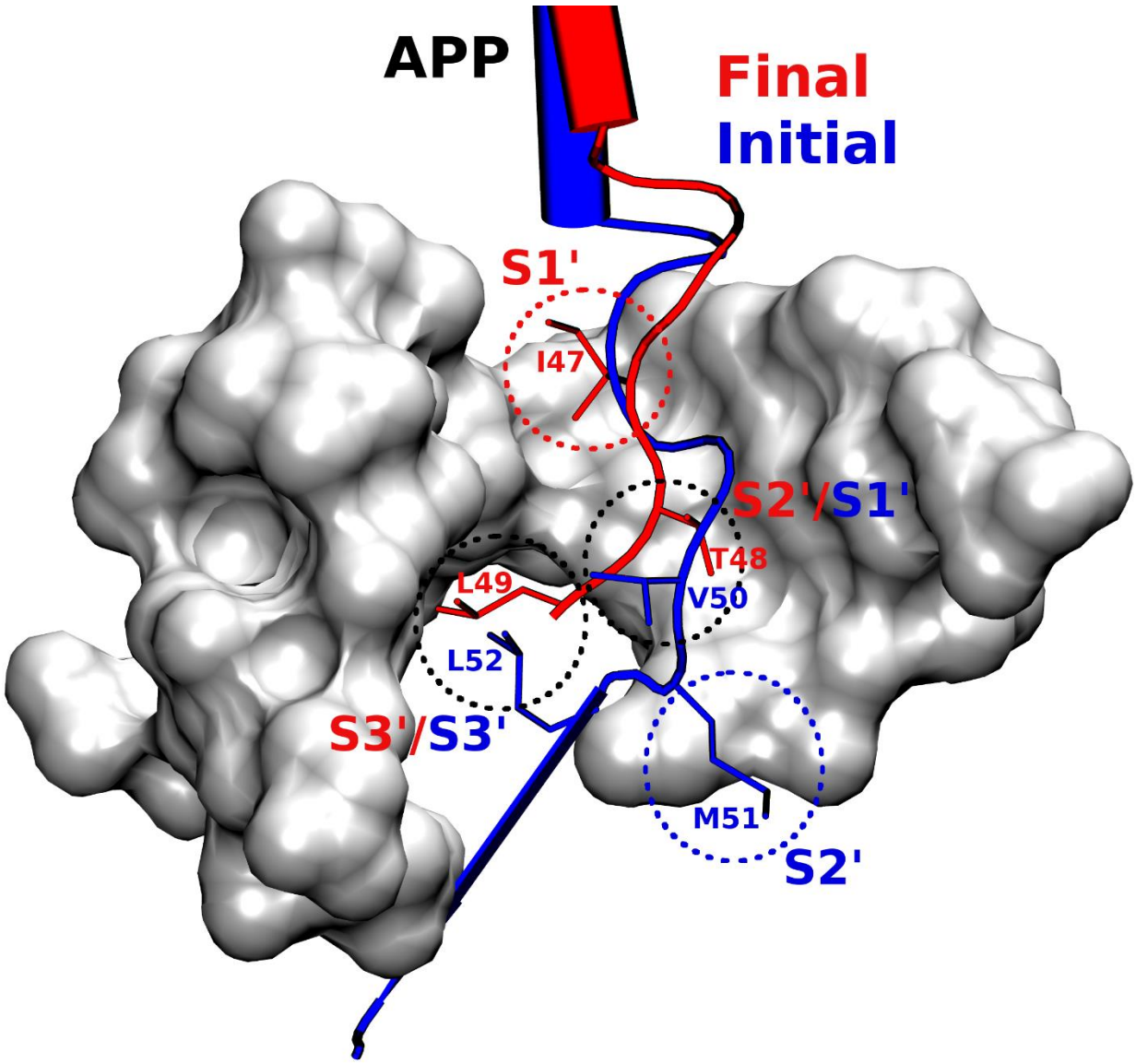

Figure S20: Comparison of active site subpockets of the Final active state during ζ cleavage to that of the Initial active state during ε cleavage.

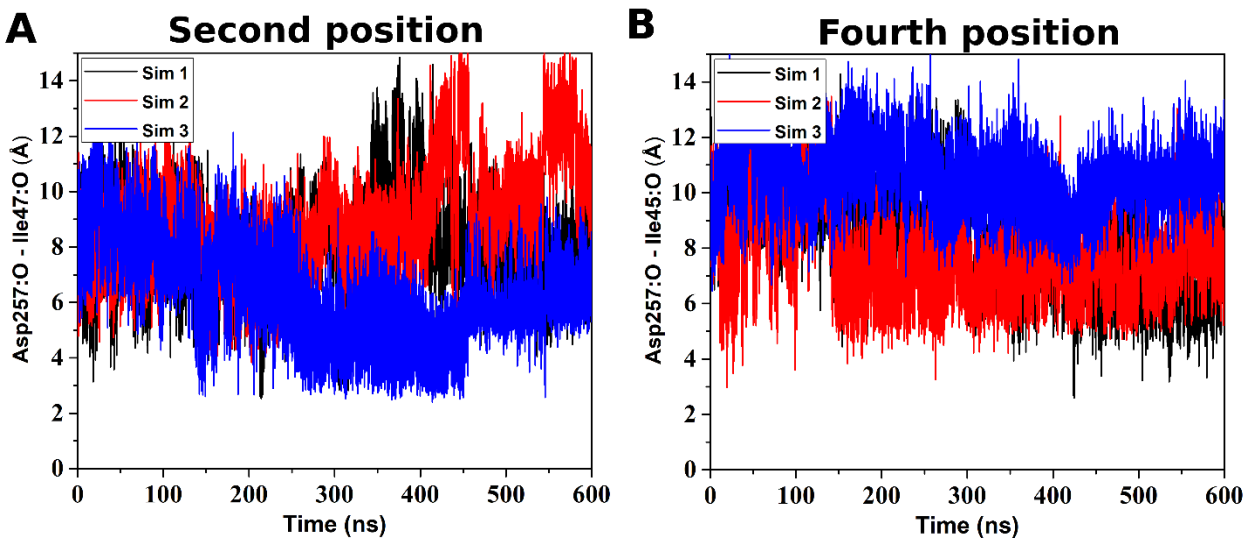

Figure S21: Time courses of the (A) Asp257:protonated O – Ile47:O and (B) Asp257:protonated O – Ile45:O distances calculated from Pep-GaMD simulations of WT with charged Aβ49 and charged AICD bound γ-secretase systems representing two and four amino acid residues shift for cleavage starting from activated enzyme for ε cleavage, respectively.

Table S1: Residues constituting the active site subpockets S1', S2' and S3' occupied by P1', P2' and P3' Aβ49 residues for different γ-secretase systems bound to wildtype and I45F, A42T and V46F FAD mutant Aβ49.

System	S1'	S2'	S3'
Wildtype/I45F/A42T	L249 Y256 L268 I287 L286 L271 L272 L150 T147 L282 F283 W165 V261 G382 L383 G384	P433 L435 L258 V261 L268 K380 L381 G382	L268 L271 V272 A275 L282 F283 I287 L381 G382 L383 I287
V46F	I253 T147 Y256 L268 L271 M146	L249 Y256 L268 I287 L286 L271 L272 L150 T147 L282 F283 W165 V261 G382 L383 G384	L268 L271 V272 A275 L282 F283 I287 L381 G382 L383 I287

Table S2: Summary of Pep-GaMD simulations on five different γ-secretase systems bound to neutral/charged C-terminal Aβ49 in the absence or presence of neutral/charged N-terminal AICD50-99 peptide. "Activation" denotes that the Pep-GaMD simulations could capture activation of γ-secretase for the ζ cleavage of Aβ49, and otherwise denoted "No activation".

Aβ49 C-terminus / **AICD N-terminus**	**Neutral**	**Charged**
Neutral	No activation	No activation
Charged	Activation	Activation
Absent	-	No activation

Chapter 3

G-Protein-Coupled Receptor-Membrane Interactions Depend on the Receptor Activation state

Apurba Bhattarai [1], Jinan Wang[1], and Yinglong Miao[1*]

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

Reprinted (Adapted) with permission from Bhattarai, Apurba, Jinan Wang, and Yinglong Miao. "G-protein-coupled receptor–membrane interactions depend on the receptor activation state." *Journal of computational chemistry* 41.5 (2020): 460-471[42].

Abstract

G-protein-coupled receptors (GPCRs) are the largest family of human membrane proteins and serve as primary targets of ~1/3 of currently marketed drugs. In particular, adenosine A_1 receptor (A_1AR) is an important therapeutic target for treating cardiac ischemia-reperfusion injuries, neuropathic pain and renal diseases. As a prototypical GPCR, the A_1AR is located within a phospholipid membrane bilayer and transmits cellular signals by changing between different conformational states. It is important to elucidate the lipid-protein interactions in order to understand the functional mechanism of GPCRs. Here, all-atom simulations using a robust Gaussian accelerated molecular dynamics (GaMD) method were performed on both the inactive (antagonist bound) and active (agonist and G protein bound) A_1AR, which was embedded in a 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) lipid bilayer. In the GaMD simulations, the membrane lipids played a key role in stabilizing different conformational states of the A_1AR. Our simulations further identified important regions of the receptor that interacted distinctly with the lipids in highly correlated manner. Activation of the A_1AR led to differential dynamics in the upper and lower leaflets of the lipid bilayer. In summary, GaMD enhanced simulations have revealed strongly coupled dynamics of the GPCR and lipids that depend on the receptor activation state.

Keywords: G-protein-coupled receptors, Adenosine A_1 receptor, Enhanced sampling, Gaussian accelerated molecular dynamics, Protein-lipid interactions.

Introduction

G-protein-coupled receptors (GPCRs) are primary cell surface receptors that account for vital physiological and pathological functions in the human body. About 1/3 of currently marketed drugs approved by the Food and Drug Administration (FDA) target GPCRs. Four subtypes of adenosine receptors (ARs), the A_1AR, A_{2A}AR, A_{2B}AR and A_3AR, mediate a broad range of physiological functions. Particularly, the Adenosine A_1 Receptor (A_1AR) has emerged as an important therapeutic target for treating cardiac ischemia-reperfusion injuries, neuropathic pain and renal diseases.[99] Being a GPCR, the A_1AR is embedded in cell membrane, maintaining close contact with lipid molecules. Lipids have been suggested to affect the receptor conformation and dynamics, which play an important role in transmitting cellular signals from extracellular environment to the cytoplasm. Similarly, lipid metabolites are also known to bind proteins and act as messengers[100]. These include lysolipids, sphingo-1-phosphate (S1P), diacylglycerol and fatty acyl derivates. In addition, lipids help in the partitioning of membrane and receptors. Membrane proteins are affected by lipid compositions and function differently in healthy and diseased individuals.[101] Therefore, it is important to study GPCR-membrane interactions in order to elucidate functional mechanism of the membrane proteins.

Experimental techniques including fluorescence resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS), fluorescence recovery after photobleaching (FRAP) and fluorescence-based monitoring of solvent relaxations rates have been utilized to study protein-membrane biology[102-105]. Experiments showed that cholesterol could affect the stability, oligomerization, and ligand binding properties of membrane proteins[106-119]. X-ray crystal structures identified allosteric sites for cholesterol binding to GPCRs[111,113,120]. Phospholipids were found to modulate dynamic processes of GPCRs such as G protein

association and ligand binding[121,122]. Recently, Dawaliby et. al. showed experimentally that lipids with different head groups favor different activation states of the β_2-adrenergic receptors (β_2AR)[123]. Lipids with phosphatidylglycerol (PG) headgroups preferred agonist binding and receptor activation, whereas lipids with phosphatidylethanolamine (PE) headgroups preferred antagonist binding and inactive state of the β_2AR. Despite these advances, there remains a knowledge gap in the understanding of protein-membrane interactions. From atomic motions of lipid molecules to curvature change across the cell membrane, protein-membrane interactions span a wide range of time scales[124]. It is often difficult to directly examine protein-membrane interactions in experiments due to limited time resolution.

Molecular dynamics (MD) simulation has emerged as a powerful computational technique to bridge the gap of knowledge for studying membrane-protein interactions. Both atomistic and coarse-grained MD simulations have been applied to study the effects of lipids in protein dynamics and cellular signaling.[101] Bruzzese et al. confirmed the above mentioned experimental results obtained by Dawaliby et. al. that different charges of PG and PE lipid headgroups affected the GPCR activation and deactivation in MD simulations[125]. The net negative charge in PG molecules favored interaction with positively-charged residues in the intracellular loop 3 (ICL3) and intracellular end of transmembrane helix 6 (TM6). This stabilized the outward movement of TM6 and hence the active state of the β2AR. Neale et. al.[126] showed that the PG lipid blocked formation of the R1313.50-E2686.30 ionic lock by interacting with R1313.50 in the β2AR. Residue superscripts denote Ballesteros and Weinstein (BW) numbering of GPCRs[127]. The R3.50 and E6.30 are highly conserved residues in GPCRs and often form an ionic lock in inactive receptors. In contrast, net-neutral zwitterionic lipids such as PE with strongly favored the inactive structure of β2AR and destabilized the active

structure[125]. Salas-Estrada et. al[128] showed that activation in rhodopsin induced changes in the membrane structure, including increase in the local order and effective length of lipid acyl chains in the vicinity of the protein. Dror et. al. showed that gradual inactivation of the β2AR occurred in the neutral lipid 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC)[129,130]. This was consistent with another study favored partial deactivation of the β2AR was found in the 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) lipids[125]. In coarse-grained MD simulations, Song et. al. showed that PIP2 (Phosphatidylinositol 4,5-bisphosphate) stabilized the outward movement of TM6 by binding in the crevice between TM6 and TM7 of adenosine A2A receptor (A2AAR)[131].

In addition, enhanced sampling techniques have been applied to investigate protein membrane interaction. Using steered MD[132] and umbrella sampling[133] techniques, Song et. al. suggested that the PIP2 facilitated recruitment of the G protein by forming bridging interactions with basic residues of the Gα subunit and hence stabilizing the active A2AAR[131]. However, these enhanced sampling techniques require predefined collective variables, which are often difficult to identify in the context of protein-membrane interactions. In this context, Gaussian accelerated MD (GaMD) is a robust technique that provides unconstrained enhanced sampling without the need to set predefined collective variables4,5. GaMD simulations have been successfully applied to investigate GPCR activation[4,134,135], protein folding[4,135], ligand binding and unbinding[4,134,135], protein-protein interactions[136-138] and protein-nucleic acid interactions[139,140].

Here, we have applied GaMD to investigate lipid interactions with the A1AR in two different conformational states, the cryo-EM structure of the active adenosine ADO-bound A1AR coupled with the G_i protein (referred to as ADO-A_1AR-Gi)[141] and the X-ray

structure[142,143] of the inactive antagonist PSB36-bound A_1AR (referred to as PSB36-A_1AR). Our simulations showed that the protein-membrane interactions depended on different conformational states of the A_1AR. The membrane lipids played an important role in stabilizing different conformations of the A_1AR. GaMD simulations further identified important regions of the receptor that interacted distinctly with the lipids. Activation of the A_1AR led to differential dynamics in the upper and lower leaflets of the lipid bilayer.

Materials and Methods

Gaussian Accelerated Molecular Dynamics (GaMD)

GaMD is an enhanced sampling technique, in which a harmonic boost potential is added to reduce the system energy barriers[4]. GaMD is able to accelerate biomolecular simulations by orders of magnitude[59,135]. GaMD does not need predefined collective variables. Moreover, because GaMD boost potential follows a gaussian distribution, biomolecular free energy profiles can be properly recovered through cumulant expansion to the second order[4]. GaMD has successfully overcome the energetic reweighting problem in free energy calculations that was encountered in the previous aMD (accelerated molecular dynamics) method[37,39] for free energy calculations. GaMD has been implemented in widely used software packages including AMBER [4,60] and NAMD[6]. A brief summary of GaMD is provided here.

Consider a system with *N* atoms at positions $\vec{r} = \{\vec{r}_1, \cdots, \vec{r}_N\}$. When the system potential $V(\vec{r})$ is lower than a reference energy *E*, the modified potential $V^*(\vec{r})$ of the system is calculated as:

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}),$$

$$\Delta V(\vec{r}) = \begin{cases} \frac{1}{2}k\big(E - V(\vec{r})\big)^2, & V(\vec{r}) < E \\ 0, & V(\vec{r}) \geq E \end{cases} \qquad (1)$$

where *k* is the harmonic force constant. The two adjustable parameters *E* and *k* are automatically determined based on three enhanced sampling principles[4]. The reference energy needs to be set in the following range:

$$V_{max} \leq E \leq V_{min} + \frac{1}{k}, \qquad (2)$$

where V_{max} and V_{min} are the system minimum and maximum potential energies. To ensure that Eqn. (2) is valid, *k* has to satisfy: $k \leq \frac{1}{V_{max} - V_{min}}$ Let us define $\equiv k_0 \frac{1}{V_{max} - V_{min}}$, then $0 < k_0 \leq 1$. The standard deviation of ΔV needs to be small enough (i.e., narrow distribution) to ensure proper energetic reweighting[62]: $\sigma_{\Delta V} = k\big(E - V_{avg}\big)\sigma_V \leq \sigma_0$ where V_{avg} and σ_V are the average and standard deviation of the system potential energies, $\sigma_{\Delta V}$ is the standard deviation of ΔV with σ_0 as a user-specified upper limit (e.g., $10k_B$T) for proper reweighting. When *E* is set to the lower bound $E=V_{max}$, k_0 can be calculated as:

$$k_0 = \min(1.0, k_0') = \min\,(1.0, \frac{\sigma_0}{\sigma_V}\frac{V_{max} - V_{min}}{V_{max} - V_{avg}}). \qquad (3)$$

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

$$k_0 = k_0'' \equiv (1 - \frac{\sigma_0}{\sigma_V})\frac{V_{max} - V_{min}}{V_{max} - V_{avg}}, \qquad (4)$$

if k_0'' is found to be between *0* and *1*. Otherwise, k_0 is calculated using Eqn. (3).

Similar to aMD, GaMD provides schemes to add only the total potential boost ΔV_P, only dihedral potential boost ΔV_D, or the dual potential boost (both ΔV_P and ΔV_D). The dual-boost

simulation generally provides higher acceleration than the other two types of simulations[63]. The simulation parameters comprise of the threshold energy E for applying boost potential and the effective harmonic force constants, k_{0P} and k_{0D} for the total and dihedral potential boost, respectively.

Energetic Reweighting of GaMD Simulations

To calculate potential of mean force (PMF)[64] from GaMD 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:

$$p(A_j) = p^*(A_j)\frac{\langle e^{\beta\Delta V(\vec{r})}\rangle_j}{\sum_{i=1}^{M}\langle p^*(A_i)e^{\beta\Delta V(\vec{r})}\rangle_i},\ j = 1, \dots, M, \qquad (5)$$

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:

$$\langle e^{\beta\Delta V(\vec{r})}\rangle = exp\left\{\sum_{k=1}^{\infty}\frac{\beta^k}{k!}C_k\right\}, \qquad (6)$$

where the first two cumulants are given by

$$\begin{aligned} C_1 &= \langle\Delta V\rangle, \\ C_2 &= \langle\Delta V^2\rangle - \langle\Delta V\rangle^2 = \sigma_v^2. \end{aligned} \qquad (7)$$

The boost potential obtained from GaMD simulations usually follows near-Gaussian distribution. Cumulant expansion to the second order thus provides a good approximation for

computing the reweighting factor[4,62]. The reweighted free energy $F(A) = -k_B T \ln p(A)$ is calculated as:

$$F(A) = F^*(A) - \sum_{k=1}^{2} \frac{\beta^k}{k!} C_k + F_c, \quad (8)$$

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

Lipid -S_{CD} Order Parameter

The -S_{CD} order parameter measures orientational anisotropy of the C-H bond in sn-2 acyl chains of lipids that are usually obtained from NMR experiments. It is a function of the angle between the C-H bond and lipid bilayer normal. It is defined by following equation:

$$S_{CD} = \frac{1}{2}\langle 3cos^2\theta - 1\rangle, \quad (9)$$

where θ is the angle between the bilayer normal and C-H bond and $\langle \dots \rangle$ denotes an ensemble average. Here, the -S_{CD} order parameter is averaged over all the lipids in the system and all the frames in the simulation trajectory. The -S_{CD} order parameter calculated from GaMD simulations is not reweighted due to complexity of the function. However, since GaMD maintains the overall shape of the original potential energy surface[4], the resulting order parameter is found to be close to the experimental values[144] (see **Results**). The -S_{CD} value usually ranges from -0.25 to 0.5, with 0.5 for the C-H bond being fully ordered along the bilayer normal and -0.25 being parallel to the bilayer plane. The -S_{CD} approximates the mobility of each C-H bond and hence estimates the membrane fluidity.

System Setup

The cryo-EM structure of the ADO-A_1AR-Gi complex (PDB: 6D9H[141]) and X-ray structure of PSB36-A_1AR complex (PDB: 5N2S[142]) were used to prepare the simulation systems. As helix 8 region was missing in the crystal structure of PSB36-A_1AR, atomic coordinates were added using another X-ray structure of the inactive A_1AR (PDB: 5UEN[143]) after aligning the receptor TM domain. All chain termini were capped with neutral groups, i.e. the acetyl group (ACE) for the N-terminus and methyl amide group (CT3) for C terminus. Protein residues were set to the standard CHARMM protonation states at neutral pH with the *psfgen* plugin in VMD[66]. Then the receptor was inserted into a POPC bilayer with all overlapping lipid molecules removed using the *Membrane* plugin in VMD[66]. The system charges were then neutralized at 0.15 M NaCl using the *Solvate* plugin in VMD[66]. Periodic boundary conditions were applied on the simulation systems. The simulation systems of the active and inactive A_1AR systems are summarized in **Table 1**.

Simulation Protocol

The CHARMM36 parameter set[67] was used for the protein and POPC lipids. For agonist ADO and antagonist PSB36, the force field parameters were obtained from the CHARMM ParamChem web server[145,146]. Initial energy minimization and thermalization of the A_1AR system follow the same protocol as used in the previous GPCR simulations[68]. The simulation proceeded with equilibration of lipid tails. With all the other atom fixed, the lipid tails were energy minimized for 1000 steps using the conjugate gradient algorithm and melted with constant number, volume, and temperature (NVT) run for 0.5ns at 310 K. Each system was further equilibrated using constant number, pressure, and temperature (NPT) run at 1 atm and

310 K for 10 ns with 5 kcal (mol Å^2)$^{-1}$ harmonic position restraints applied to the protein. Further equilibration of the systems was performed using an NPT run at 1 atm and 310 K for 0.5ns with all atoms unrestrained. Conventional MD simulation was performed on each system for 10 ns at 1atm pressure and 310 K with a constant ratio constraint applied on the lipid bilayer in the X-Y plane. The GaMD simulations were carried out using NAMD2.13[6,147]. Both dihedral and dual-boost GaMD simulations were then performed to study the protein-membrane interactions in the inactive and active A_1AR systems (**Table 1**). In the GaMD simulations, the threshold energy E for adding boost potential is set to the lower bound, i.e. $E = V_{max}$[4,6]. The simulations included 50ns equilibration after adding the boost potential and then multiple independent production runs lasting 150 – 300 *ns* with randomized initial atomic velocities. GaMD production simulation frames were saved every 0.2ps for analysis.

Simulation analysis

The VMD[66] and CPPTRAJ[69] tools were used for trajectory analysis. In particular, distance was calculated between the Cα atoms residues Arg$^{3.50}$ and Glu$^{6.30}$. Root-mean-square fluctuations (RMSFs) were calculated for the protein residues and ligands, averaged over two independent GaMD simulations and color coded for schematic representation of each complex system. MEMBPLUGIN, a plugin for the VMD package was used to calculate the -S_{CD} order parameter for POPC lipid tails[148]. The -S_{CD} order parameters were averaged over all lipids and frames of the two independent GaMD simulations for each system. The CPPTRAJ tool was used to calculate the correlation matrices. The Cα atoms of the receptor and phosphorous atoms in the POPC lipid head groups were used for the calculations. In addition to the phosphorous atom,

the C_8 and C_{18} atoms representing different regions of the lipids were also used to calculate dynamic correlations with the receptor. The *PyReweighting* toolkit[62] was applied to reweight GaMD simulations for free energy calculations by combining independent trajectories for each system. A bin size of 1 Å was used for the $Arg^{3.50}$-$Glu^{6.30}$ distance and 1 for the number of lipids. The cutoff was set to 500 for calculating the 2D PMF profiles.

Results

Structural flexibility of the A_1AR depended on the receptor conformational state

All-atom GaMD simulations were performed on two different conformational states of the A_1AR, active (ADO-A_1AR-Gi) and inactive (PSB36-A_1AR) states (**Table 1**). For the inactive A_1AR system, the boost potential was 4.47±1.81 kcal/mol and 8.45±3.33 kcal/mol in dihedral and dual-boost GaMD simulations, respectively. For the active A_1AR system, the boost potential was 5.04±2.22 kcal/mol and 9.94±2.57 kcal/mol in dihedral and dual-boost GaMD simulations, respectively (**Table 1**). Thus, dual-boost GaMD provided higher acceleration in the simulations with greater boost potential. In the dihedral GaMD simulations of the inactive A_1AR, the TM helices of the receptor were rather rigid. Only the intracellular end of TM6, the terminus of helix 8 (H8), extracellular end of TM1 and extracellular loop 2 (ECL2) regions were flexible (**Figure 1A**). Similar results were obtained for the active A_1AR in the dihedral GaMD simulations (**Figure 1B**). However, the intracellular ends of TM6 and TM5 of the A_1AR exhibited more fluctuations in the active state compared to the inactive state. The ECL2 region was relatively more flexible in the active A_1AR than in the inactive A_1AR. In both systems, the ligands remained stably bound at the orthosteric site throughout the simulations. In comparison,

the G protein coupled to the active A_1AR exhibited higher fluctuations. In particular, C terminus of the α5 helix, α4-β5 loop and α4-β6 loop of the Gα subunit and terminal ends of the $G_{\beta\gamma}$ subunits exhibited high fluctuations up to 3 Å. Similar results were also found in the dual-boost GaMD simulations of the inactive and active A_1AR systems (**Figure S1**).

Lipids in the lower leaflet of the active A_1AR system showed higher fluidity than in the inactive-A_1AR system

The lipid -S_{CD} order parameters were calculated for the upper (extracellular) and lower (cytoplasmic) leaflets from GaMD simulations of the inactive and active A_1AR systems (**Figure 2**). In both inactive and active A_1AR systems, lower leaflet was more fluid than the upper leaflet with smaller -S_{CD} order parameters. The lower leaflet exhibited significant differences between the inactive and active A_1AR systems. In particular, the -S_{CD} order parameter of the fifth carbon atom in POPC was ~0.20 in the lower leaflet in the inactive A_1AR system (**Figure 2A, 2C**), but decreased to ~0.17 in the active A_1AR system (**Figure 2B, 2D**). This indicated higher inclination of C-H bonds being ordered along the bilayer normal in the lower leaflet of the active A_1AR system. This appeared to correlate with the outward movement of TM6 as the A_1AR changed from the inactive to active state. In the inactive A_1AR, the $R131^{3.50}$-$E268^{6.30}$ distance at the free energy minimum was ~7 Å (**Figure 3A, 3C**). In comparison, this distance at the free energy minimum increased to ~17 Å in the active A_1AR (**Figure 3B, 3D**). The lateral movement of the TM6 could push the surrounding lipids. Higher flexibility of the receptor TM6 intracellular end was accompanied by increased fluidity of lipids in the lower leaflet of membrane. Similar results were found from the dihedral and dual-boost GaMD simulations

(**Figure 3**). Therefore, the structural conformation and flexibility of the GPCR are strongly coupled with the surrounding membrane lipids.

The inactive A_1AR attracted more lipids in the upper leaflet to the TM6 than the active A_1AR

Considering significant conformational changes (especially in the TM6) during receptor activation, we hypothesized that the number of lipids interacting with the active and inactive receptor are different. In order to identify the low-energy states of the membrane-receptor interactions, potential of mean force (PMF) profiles were calculated by reweighting the GaMD simulations (**Figure 3**). The $R131^{3.50}$-$E268^{6.30}$ distance was chosen as one reaction coordinate to characterize activation of the GPCR. The number of POPC phosphate head groups within 5 Å of TM6 was calculated as the other reaction coordinate (**Figures S2** and **S3**). In the upper leaflet, approximately one lipid molecule was found interacting with TM6 in the inactive A_1AR (**Figure 3A, 3C**). But no lipid in the upper leaflet was found within 5 Å of TM6 in the active A_1AR (**Figure 3B, 3D**). Further analysis revealed that one positively-charged residue was located in the receptor ECL3 ($K265^{ECL3}$), being close to the extracellular end of the TM6 (**Figure 5A**). In the inactive A_1AR, this lysine pointed towards the lipid membrane and thus attracted the negatively-charged phosphate head group of a POPC molecule. Instead, the positively-charged side chain of $K265^{ECL3}$ formed a stable salt-bridge with negatively-charged glutamate ($E172^{ECL2}$) of ECL2 in the active A_1AR (**Figure S4**). Residue $K265^{ECL3}$ did not interact with the lipid in the active A_1AR. Therefore, the inactive A_1AR interacted with more phospholipids in the upper leaflet compared to the active A_1AR.

The active A_1AR attracted more lipids in the lower leaflet to the TM6 than the inactive A_1AR

In contrast to the upper leaflet, the lower leaflet had more lipids within 5 Å of TM6 in the active A_1AR than in the inactive system (**Figure 4**). In the lowest energy state, the inactive A_1AR interacted with approximately two lipids within 5 Å of TM6 (**Figure 4A, 4C**). In comparison, the active A_1AR exhibited a relatively broader energy well. The TM6 intracellular domain interacted with ~2-4 lipid molecules (**Figure 4B, 4D**). Upon activation of the A_1AR, the TM6 moved outwards by ~10 Å and exposed its positively-charged residues to the membrane. Therefore, the lipids that diffused in the lower leaflet of the membrane interacted more frequently with the receptor. Fewer lipids interacted with the inactive A_1AR because the receptor has a narrower curvature in the TM6 intracellular region. Moreover, the lower leaflet had a larger number of lipids interacting with receptor TM6 than the upper leaflet. Out of five positively-charged residues in TM6, four were located in the intracellular region ($K224^{6.25}$, $K228^{6.29}$, $K231^{6.32}$ and $K234^{6.35}$) (**Figure 5B**). The negatively-charged phosphate head groups of POPC tended to interact with these positively-charged lysine residues to stabilize the active receptor conformation. Therefore, the lipid-GPCR interaction should play an important role in the conformational changes during activation of the A_1AR.

GaMD simulations revealed strongly coupled dynamics between the GPCR and membrane lipids

Dynamic correlations were identified between residues of the A_1AR and lipids in both the upper and lower leaflets. The Cα atoms in the receptor residues and the phosphorous atoms in the

lipid head groups were used to calculate the correlation matrices (see details in **Methods**). Similar results were obtained using the C_8 and C_{18} atoms in the lipid hydrophobic tails to calculate the dynamic correlation matrices (**Figure S5**). In all the simulation systems, motions of the receptor N-terminus, ECL1, ECL2 and ECL3 regions were positively correlated to those of lipids in upper leaflet (**Figure 6**). Similarly, motions of the receptor ICL1, ICL2 and ICL3 were positively correlated to those of lipids in the lower leaflet (**Figure 7**). For the inactive A_1AR, most TM helix residues exhibited negatively correlated motions with the lipids. In this regard, the TM helices appeared to move in the opposite direction relative to the lipids in the inactive A_1AR simulation system (**Figure 6A, 6C, 7A and 7A**). For the active A_1AR, correlations between TM helix residues and the lipids in both the upper and lower leaflets were very weak, being close to zero (**Figure 6B, 6D, 7B and 7D**). However, marked positive correlations were identified between the intracellular region of the receptor TM6 and lipids in the lower leaflet (**Figures 7B and 7D**). Therefore, the TM6 intracellular region of the active A_1AR appeared to move in the same direction with the surrounding lipids. This was highly consistent with the simulation finding that significantly more lipids were found within 5 Å of the TM6 intracellular domain in the active A_1AR system (**Figure 4**) and they formed remarkably stronger electrostatic interactions (**Figure 5**) compared with the inactive A_1AR system. In summary, the GaMD simulations revealed strongly coupled dynamics between the GPCR and membrane lipids.

Discussion

In this study, we have applied all-atom GaMD simulations to investigate GPCR-membrane interactions, using the A_1AR as a model receptor. In the GaMD simulations, the inactive and active A_1AR showed different structural flexibility profiles. The ECL2 region, intracellular

ends of TM6 and TM5 exhibited higher fluctuations in the active A_1AR compared to the inactive A_1AR. The receptor TM domain was rigid and the ligands remained tightly bound at the orthosteric site. However, the G protein coupled to the active A_1AR exhibited high flexibility during the simulations, especially in the α5 helix, α4-β5 loop and α4-β6 loop of the Gα subunit and terminal ends of the $G_{\beta\gamma}$ subunits. These results were consistent to our earlier simulation findings of AR-G protein complexes[149].

The $-S_{CD}$ order parameter values obtained from GaMD simulations were consistent with experimental data. In NMR experiments, the $-S_{CD}$ order parameter for the fifth carbon C-H bond of POPC was observed to be at ~0.18-0.20[144]. The $-S_{CD}$ order parameter of POPC's fifth carbon atom was ~0.20±0.02 in the lower leaflet in the inactive A_1AR system. It decreased to ~0.17±0.02 in the active A_1AR system. The $-S_{CD}$ order parameter of the ninth carbon C-H bond in POPC was measured as ~0.10 in NMR experiments[144], for which the same value was obtained from GaMD simulations. Furthermore, the GaMD simulations showed that POPC lipids in the lower leaflet of the active A_1AR system were more fluid than in the inactive A_1AR system. The $-S_{CD}$ order parameters for the lower leaflet in the active A_1AR system were smaller than those in the inactive A_1AR system. This finding correlated with the outward movement of TM6 in the active A_1AR, which caused higher inclination of the C-H bonds to be aligned along the bilayer normal. The smaller $-S_{CD}$ order parameters suggested higher membrane fluidity in the lower leaflet of the active A_1AR system.

In the GaMD simulations, the inactive A_1AR attracted more lipids in the upper leaflet than the active A_1AR. The membrane facing positively-charged lysine residue ($K265^{ECL3}$) interacted with the negatively-charged phosphate head group of POPC. In contrast, this lysine pointed towards ECL2 in the active A_1AR. This was consistent with our previous study[150], in

which the positive allosteric modulator (PAM) enhanced the agonist binding at the orthosteric site by forming a salt-bridge between $E172^{ECL2}$-$K265^{ECL3}$. Moreover, the active A_1AR attracted more lipids in the lower leaflet compared with the inactive A_1AR. When exposed to the membrane, the positively-charged residues in intracellular region of TM6 of the A_1AR interacted with negatively-charged head groups. This was further verified by the correlation matrix. The only positive correlation between the transmembrane helices and the lipids was observed between the intracellular region of TM6 and lipids in the lower leaflet of active A_1AR. Four lysine residues ($K263^{6.25}$, $K267^{6.29}$, $K270^{6.32}$ and $K273^{6.35}$) present in the intracellular end in the β_2AR of TM6 were also known to interact with negatively-charged headgroups of 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) lipid molecules and thus stabilize the GPCR active state[125].

In summary, all-atom GaMD simulations have revealed strongly coupled dynamics between a GPCR and the membrane lipids that depend on the receptor activation state. The GaMD method has greatly enhanced sampling of the lipid-protein interactions, which would take significantly longer simulation time using cMD. Nonetheless, the activation or deactivation conformational transitions of the GPCR were not observed in the presented GaMD simulations. Longer GaMD simulations (e.g., microseconds) are expected to capture such conformational transitions [151] and the related effects of lipid-receptor interactions will be investigated in the future. Furthermore, the effects different lipid types (e.g., cholesterol, PIP_2, etc.) on GPCR-membrane interactions are subject to future studies. It is important to study specific lipid interactions with GPCRs during the receptor activation. Developments of enhanced sampling methodologies and computing power would aid to further address these challenges.

Acknowledgements

We dedicate this manuscript to Prof. Benoit Roux's 60th birthday for his contributions in particularly membrane protein simulations and free energy calculations. 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 project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. This work was supported in part by the American Heart Association (Award 17SDG33370094), the National Institutes of Health (R01GM132572) and the startup funding in the College of Liberal Arts and Sciences at the University of Kansas.

Table 1: Summary of GaMD simulations performed on the adenosine A_1 receptor (A_1AR).

System	[a]N_{atoms}	Dimension (Å^3)	[b]Method	Simulation	[c]ΔV_{avg} (kcal/mol)	[d]$\sigma_{\Delta v}$ (kcal/mol)
PSB36-A_1AR	77,809	93x100x101	GaMD_Dih	300 ns x 2	4.47	1.81
			GaMD_Dual	300 ns x 2	8.45	3.33
ADO-A_1AR-G_i	180,394	93x111x167	GaMD_Dih	300 ns x 2	5.04	2.22
			GaMD_Dual	150 ns x 2	9.94	2.57

[a]N_{atoms} is number of atoms in the simulation systems.
[b]GaMD_Dih and GaMD_Dual represent the dihedral and dual boost GaMD simulations respectively.
[c]ΔV_{avg} and [d]$\sigma_{\Delta V}$ are the average and standard deviation of the GaMD boost potential.

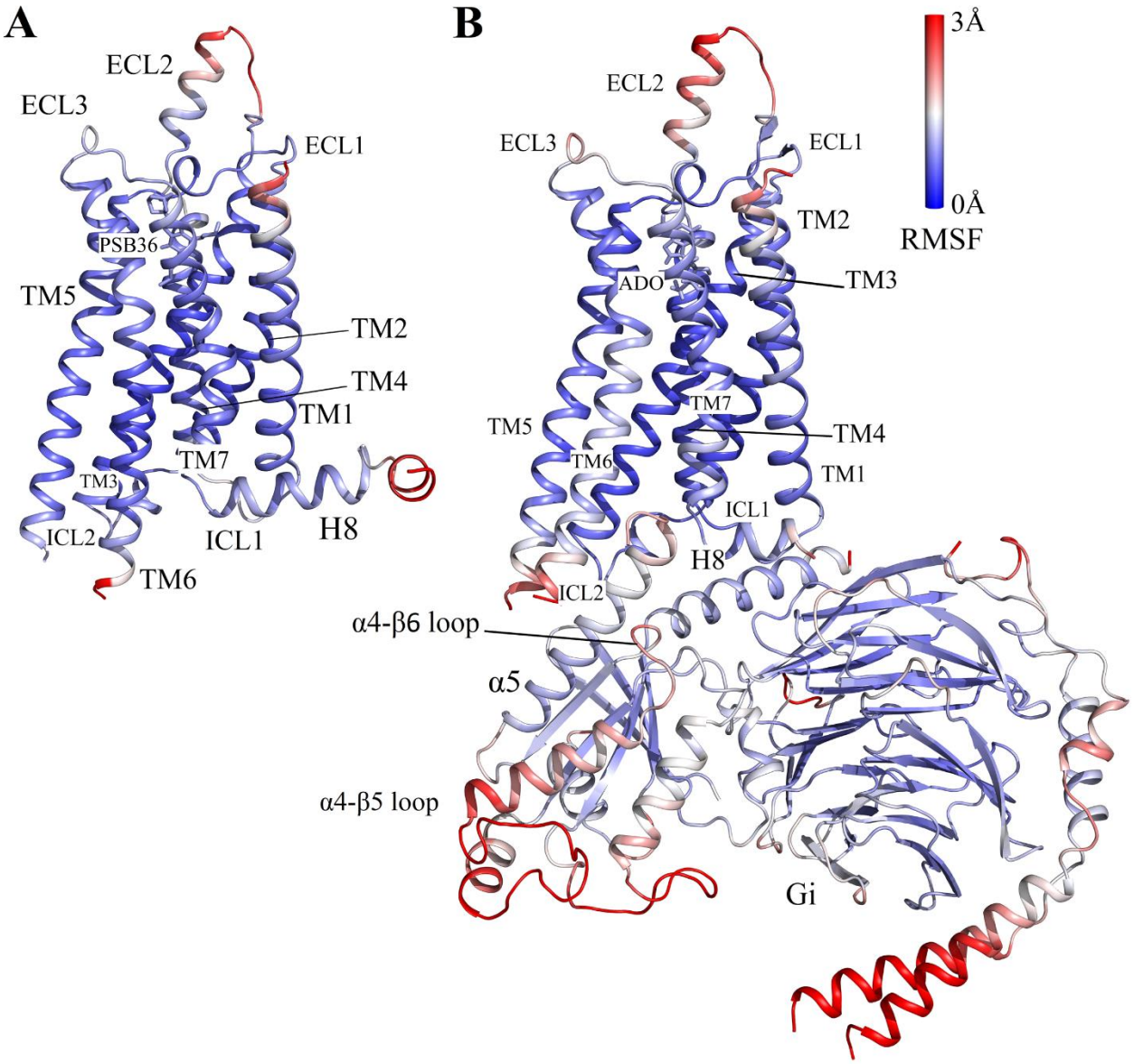

Figure 1: Comparison of structural flexibility of the inactive and active A1AR systems obtained from dihedral GaMD simulations: (A) Root-mean-square fluctuations (RMSFs) of the inactive PSB36-A1AR complex. (B) RMSFs of the active ADO-A1AR-Gi protein complex. A color scale of 0 Å (blue) to 3 Å (red) was used.

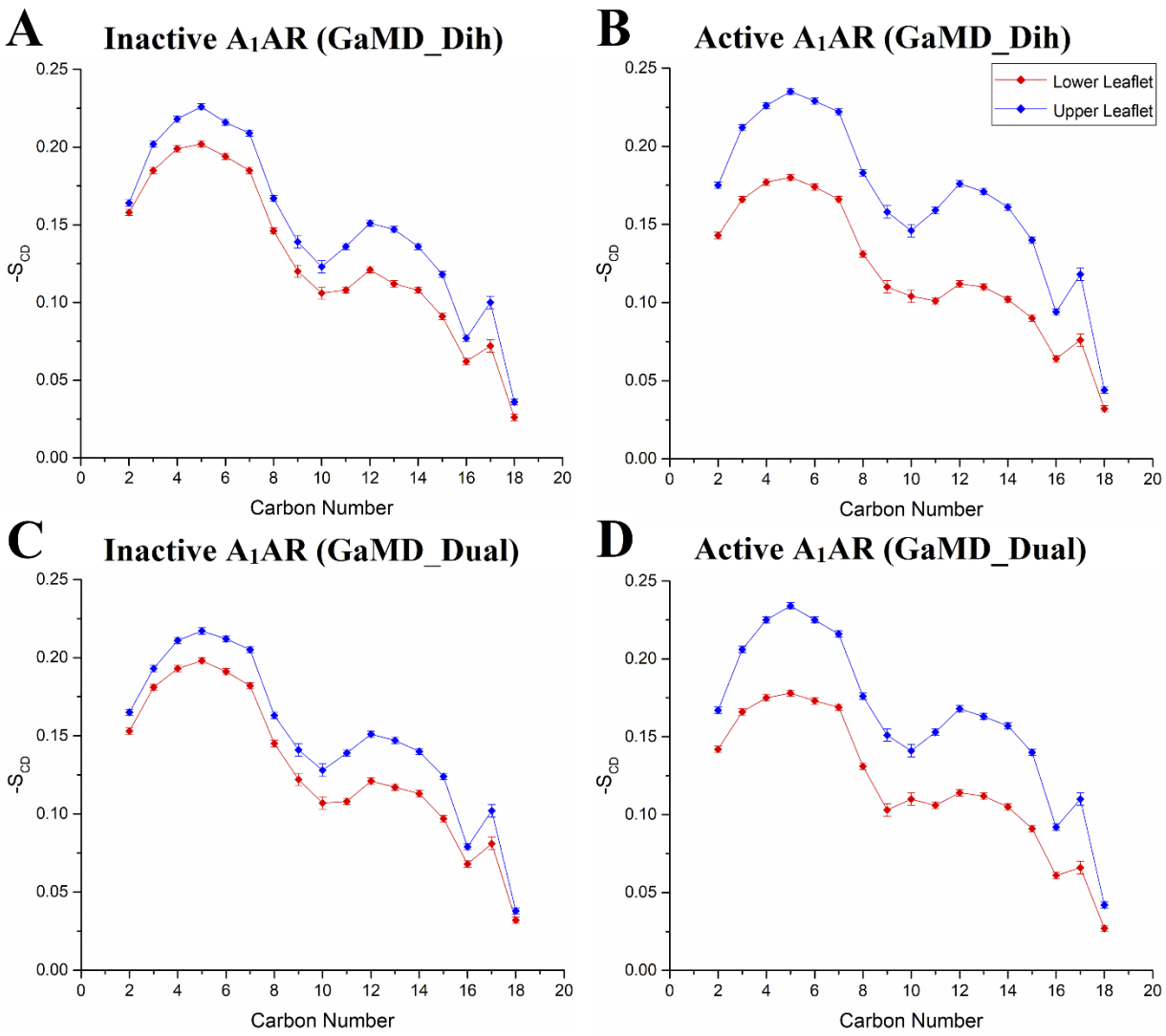

Figure 2: The $-S_{CD}$ order parameters calculated for sn-2 acyl chains of POPC lipids in different simulation systems: (A) Inactive A1AR using dihedral-boost GaMD, (B) Active A1AR using dihedral-boost GaMD, (C) Inactive A1AR using dual-boost GaMD and (D) Active A1AR using dual-boost GaMD. Red diamond lines represent the average $-S_{CD}$ order parameters for the cytoplasmic lower leaflet and blue diamond lines for the extracellular upper leaflet.

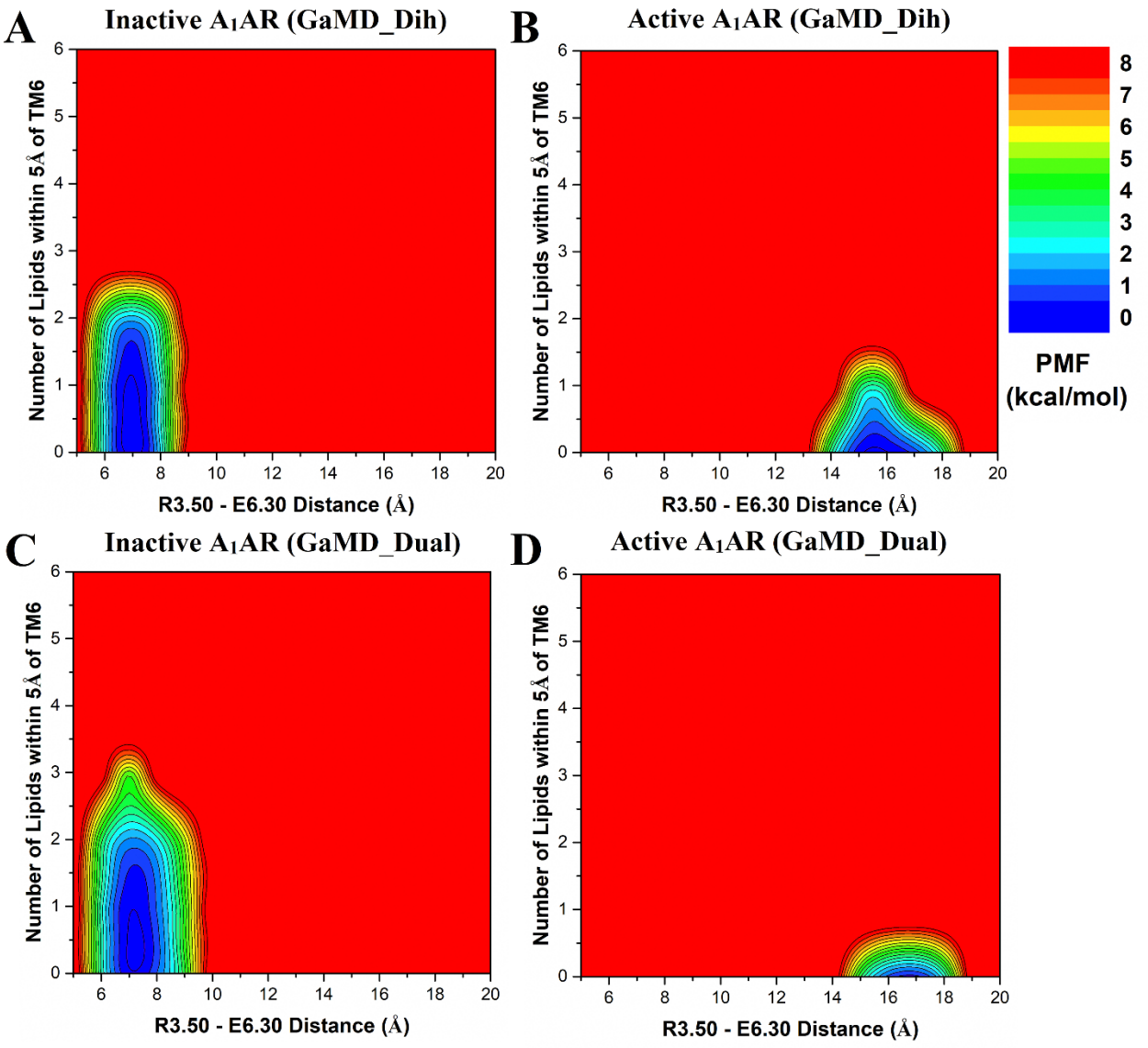

Figure 3: Free energy profiles of the extracellular upper leaflet of membrane in different simulation systems regarding the number of lipids within 5 Å of the receptor TM6 and the receptor R3.50 – E6.30 distance: (A) Inactive A_1AR using dihedral-boost GaMD, (B) Active A_1AR using dihedral-boost GaMD, (C) Inactive A_1AR using dual-boost GaMD and (D) Active A_1AR using dual-boost GaMD. The R3.50 – E6.30 distance is ~7 Å in the inactive A1AR and increases to ~17 Å in the active A1AR due to outward movement of TM6.

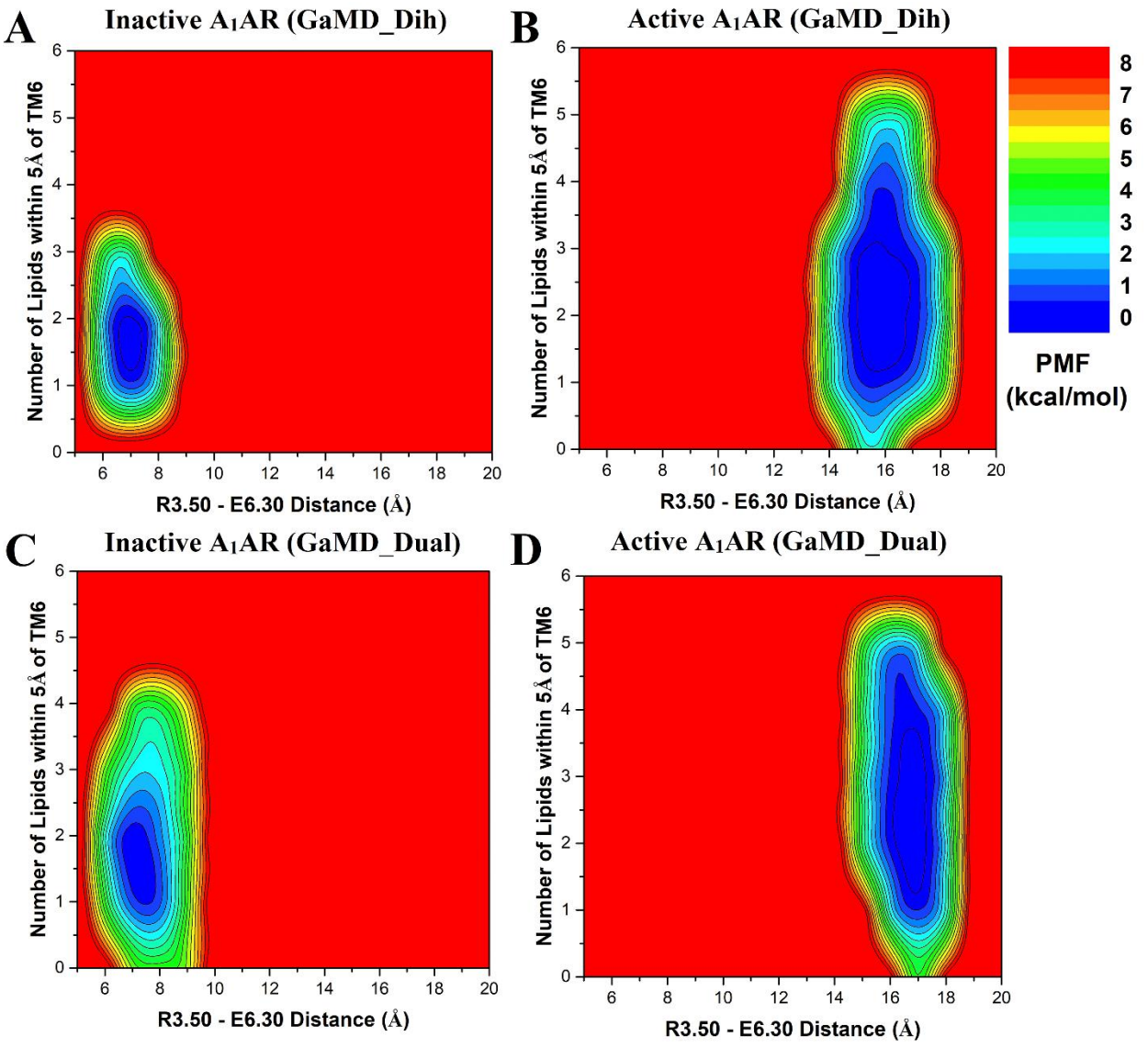

Figure 4: Free energy profiles of the cytoplasmic lower leaflet of membrane in different simulation systems regarding the number of lipids within 5 Å of the receptor TM6 and the receptor R3.50 – E6.30 distance: (A) Inactive A_1AR using dihedral-boost GaMD, (B) Active A_1AR using dihedral-boost GaMD, (C) Inactive A_1AR using dual-boost GaMD and (D) Active A_1AR using dual-boost GaMD. The R3.50 – E6.30 distance is ~7 Å in the inactive A1AR and increases to ~17 Å in the active A1AR due to outward movement of TM6.

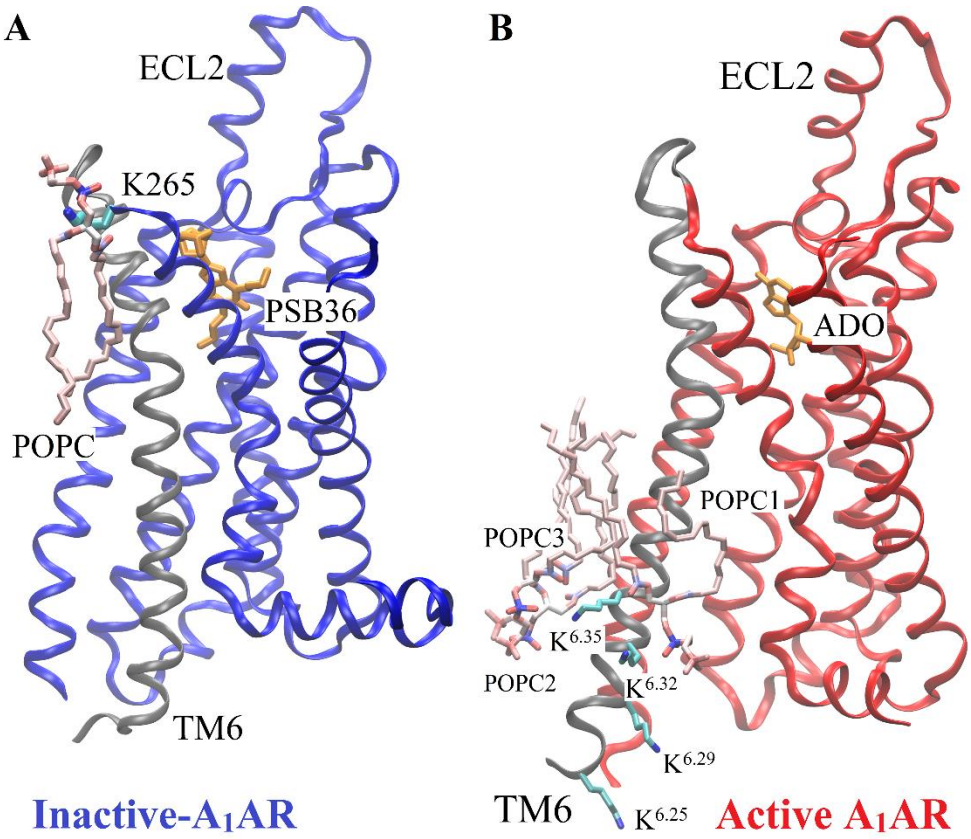

Figure 5: Minimum energy states of POPC lipid interacting with the positively-charged lysine residues in TM6 of the receptor obtained from dihedral GaMD simulations. (A) One POPC molecule in the upper leaflet interacts with one lysine residue ($K265^{ECL3}$) of the inactive A_1AR. (B) Three POPC molecules (POPC1, POPC2, POPC3) in the lower leaflet interact with four Lysine residues ($K263^{6.25}$, $K267^{6.29}$, $K270^{6.32}$ and $K273^{6.35}$) of the active A_1AR. The receptor TM6 is colored in gray.

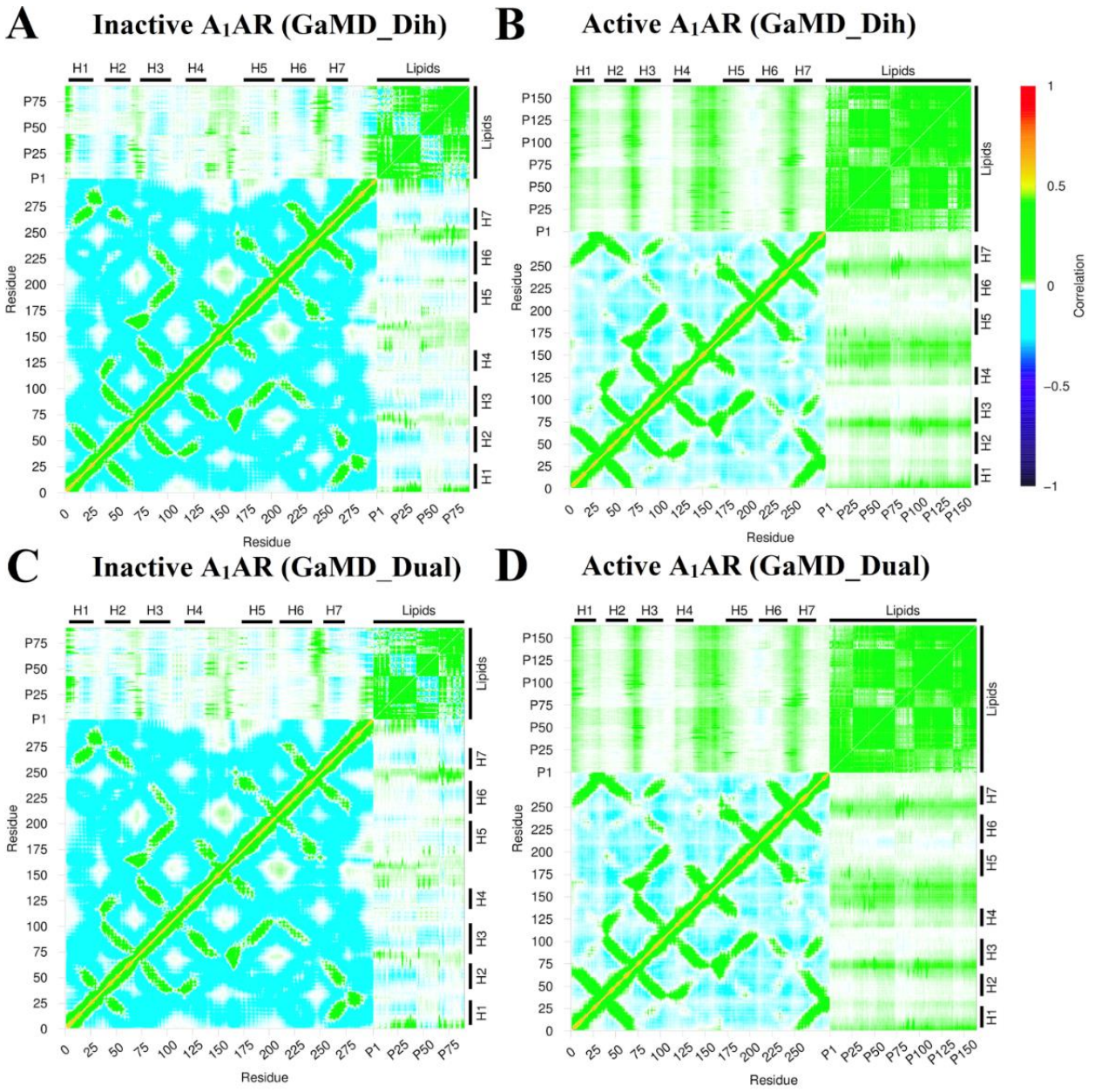

Figure 6: Dynamic correlation matrices calculated for lipids in the extracellular upper leaflet with residues in the A1AR in different simulation systems: (A) Inactive A_1AR using dihedral-boost GaMD, (B) Active A_1AR using dihedral-boost GaMD, (C) Inactive A_1AR using dual-boost GaMD and (D) Active A_1AR using dual-boost GaMD. The Cα atoms of the receptor and phosphorous atoms in the lipid head groups were used for calculating the correlation matrices here. Similar results were obtained using the C_8 and C_{18} atoms in the lipid hydrophobic tails as shown in **Figure S5**. The receptor ICL1, ICL2 and ICL3 represent intracellular loops between TM helices 1-2, 3-4, and 5-6 respectively. Similarly, the receptor ECL1, ECL2 and ECL3 represent extracellular loops between TM helices 2-3, 4-5, and 6-7 respectively.

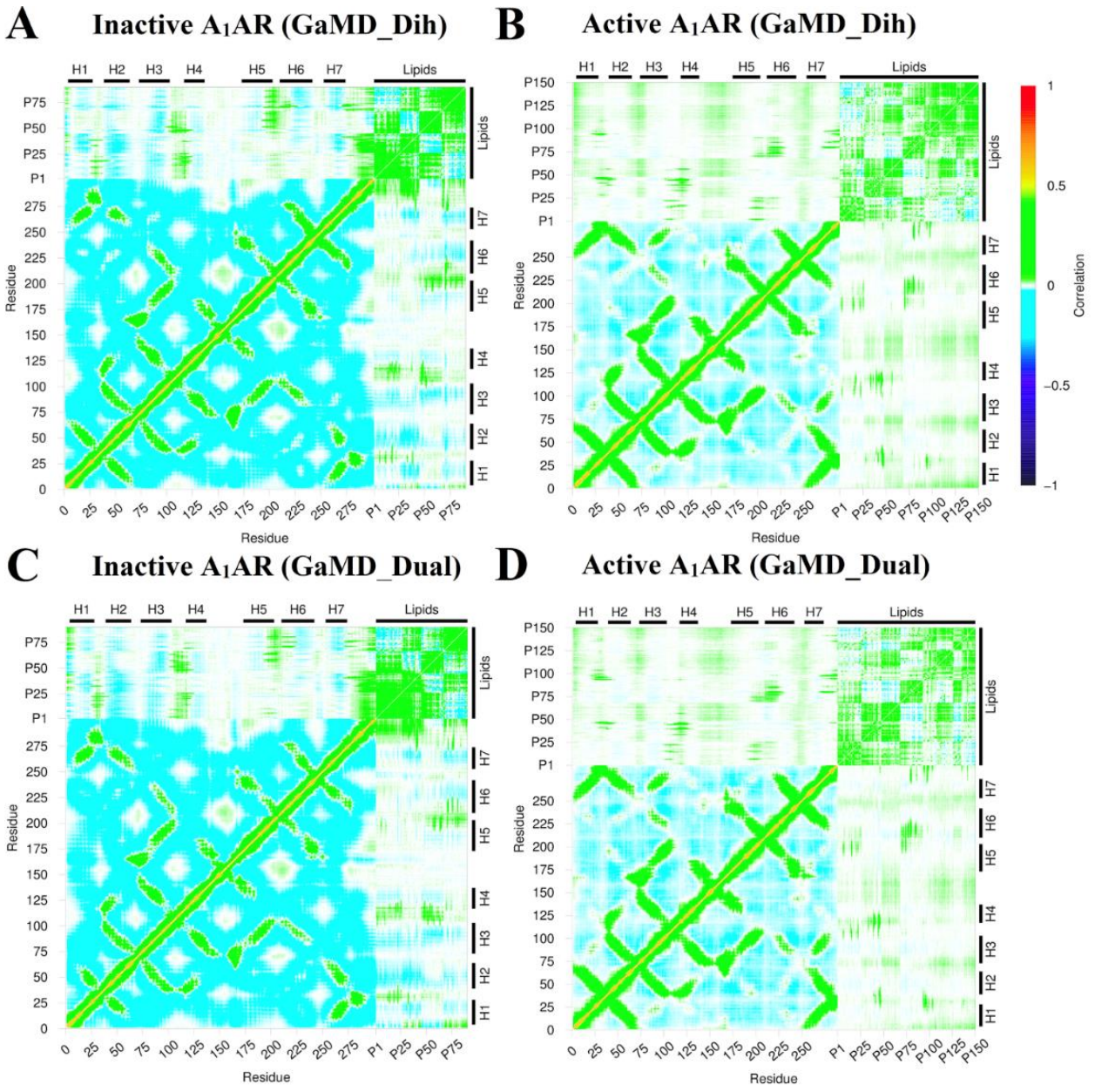

Figure 7: Dynamic correlation matrices calculated for lipids in the intracellular lower leaflet with residues in the A1AR in different simulation systems: (A) Inactive A_1AR using dihedral-boost GaMD, (B) Active A_1AR using dihedral-boost GaMD, (C) Inactive A_1AR using dual-boost GaMD and (D) Active A_1AR using dual-boost GaMD. The Cα atoms of the receptor and phosphorous atoms in the lipid head groups were used for calculating the correlation matrices here. Similar results were obtained using the C_8 and C_{18} atoms in the lipid hydrophobic tails as shown in **Figure S5**. The receptor ICL1, ICL2 and ICL3 represent intracellular loops between TM helices 1-2, 3-4, and 5-6 respectively. Similarly, the receptor ECL1, ECL2 and ECL3 represent extracellular loops between TM helices 2-3, 4-5, and 6-7 respectively.

Appendix

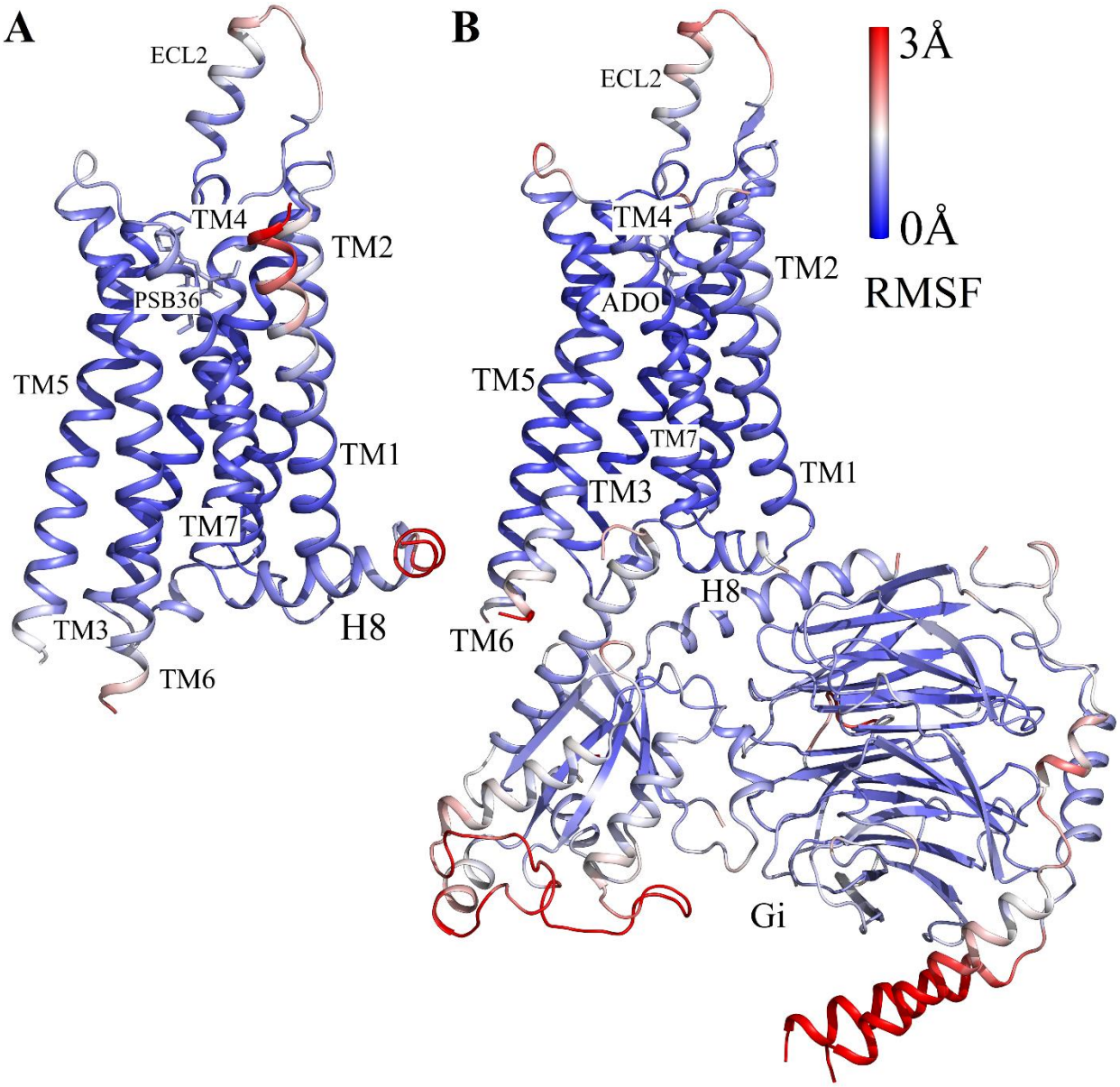

Figure S1: Comparison of structural flexibility of the inactive and active A_1AR systems obtained from dual-boost GaMD simulations: (A) Root-mean-square fluctuations (RMSFs) of the inactive PSB36-A_1AR complex. (B) RMSFs of the active ADO-A_1AR-Gi complex. A color scale of 0 Å (blue) to 3 Å (red) is used.

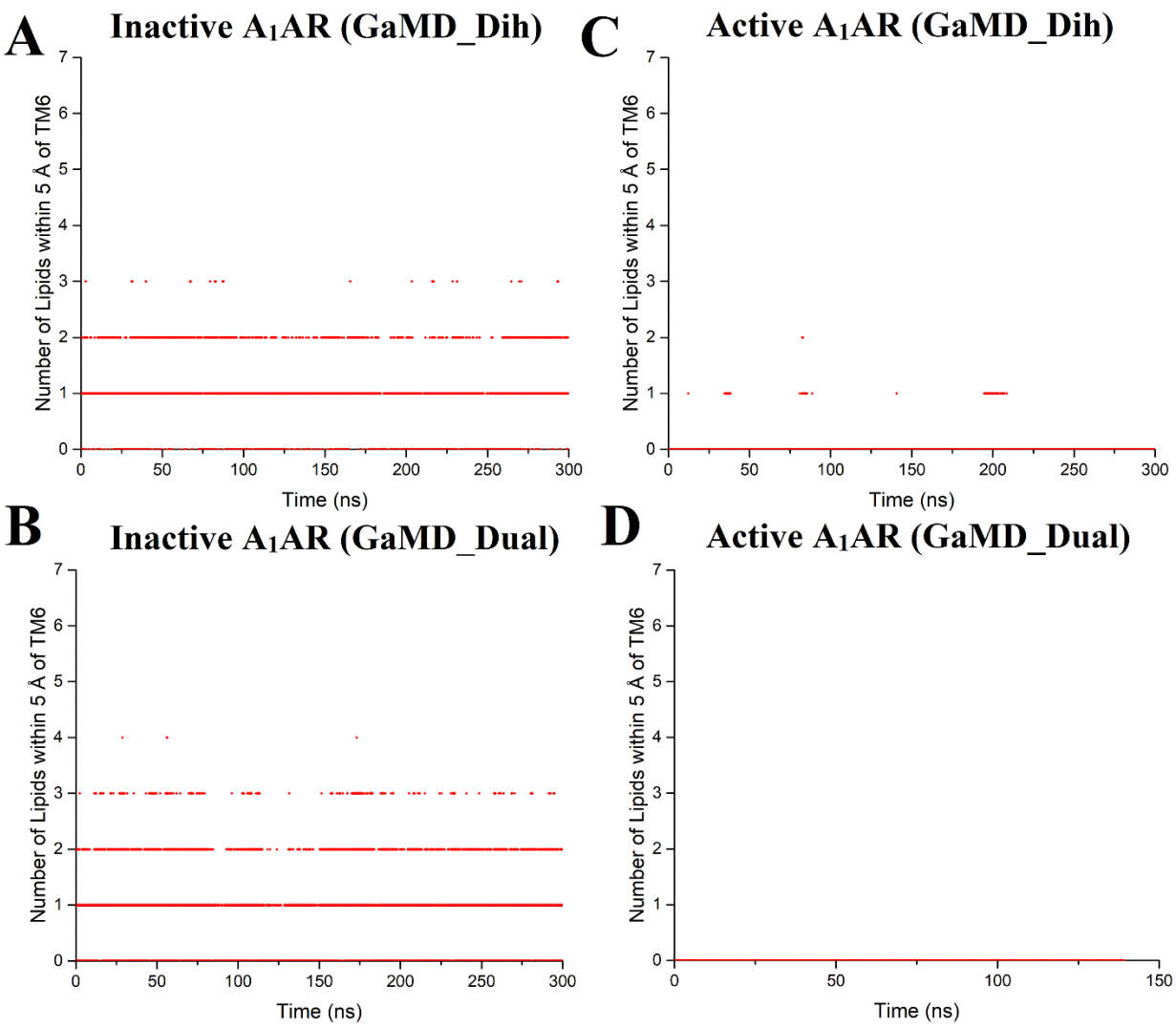

Figure S2: Time courses for number of POPC molecules within 5 Å of TM6 in the upper leaflet of different simulation systems. (A) Inactive A_1AR using dihedral-boost GaMD, (B) Active A_1AR using dihedral-boost GaMD, (C) Inactive A_1AR using dual-boost GaMD and (D) Active A_1AR with using dual-boost GaMD.

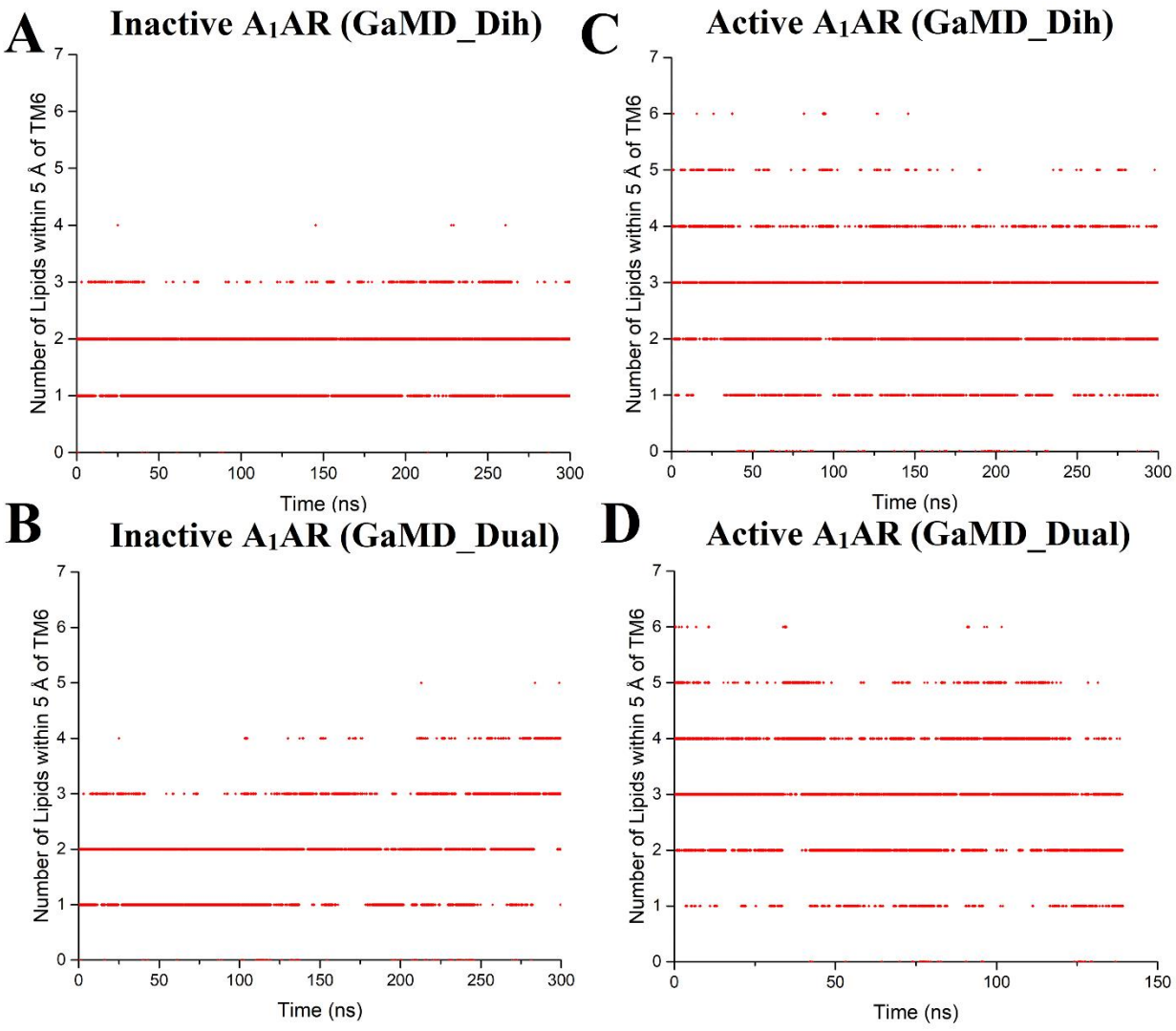

Figure S3: Time courses for number of POPC molecules within 5 Å of TM6 in the lower leaflet of different simulation systems. (A) Inactive A_1AR using dihedral-boost GaMD, (B) Active A_1AR using dihedral-boost GaMD, (C) Inactive A_1AR using dual-boost GaMD and (D) Active A_1AR with using dual-boost GaMD.

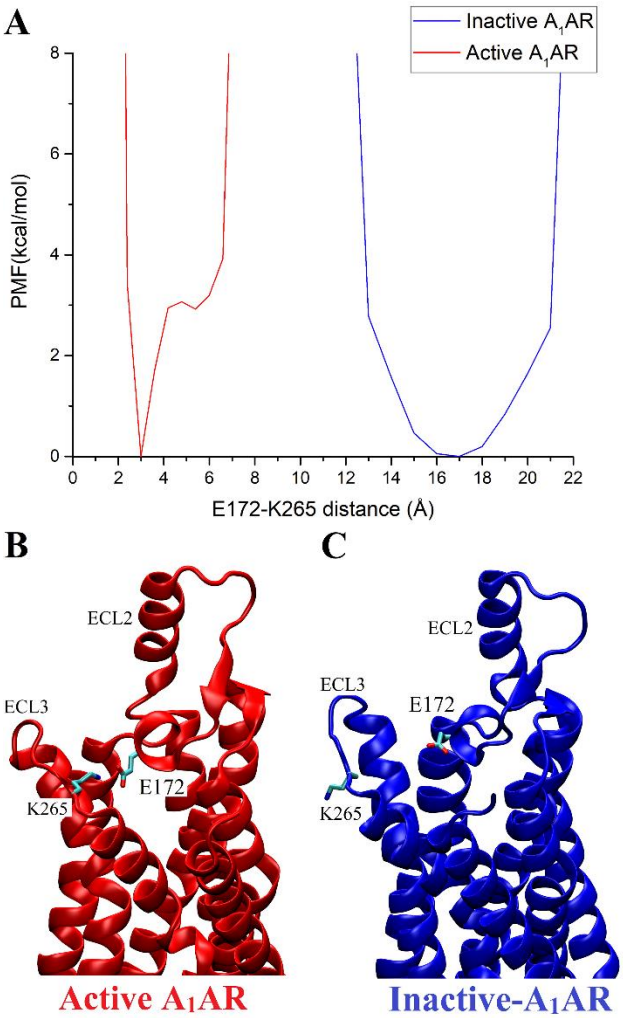

Figure S4: Free energy profile for inactive and active A_1AR systems regarding $E172^{ECL2}$-$K265^{ECL3}$ distance (A). Minimum energy state snapshots of inactive (C) and active (B) A_1AR systems showing the residues $E172^{ECL2}$ and $K265^{ECL3}$.

Chapter 4

Retrospective Ensemble Docking of Allosteric Modulators in an Adenosine G-Protein-Coupled Receptor

Apurba Bhattarai, Jinan Wang, and Yinglong Miao *

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

Reprinted (Adapted) with permission from Bhattarai, Apurba, Jinan Wang, and Yinglong Miao. "Retrospective ensemble docking of allosteric modulators in an adenosine G-protein-coupled receptor." *Biochimica et Biophysica Acta (BBA)-General Subjects* 1864.8 (2020): 129615[152].

Abstract

Ensemble docking has proven useful in drug discovery and development. It increases the hit rate by incorporating receptor flexibility into molecular docking as demonstrated on important drug targets including G-protein-coupled receptors (GPCRs). Adenosine A_1 receptor (A_1AR) is a key GPCR that has been targeted for treating cardiac ischemia-reperfusion injuries, neuropathic pain and renal diseases. Development of allosteric modulators, compounds binding to distinct and less conserved GPCR target sites compared with agonists and antagonists, has attracted increasing interest for designing selective drugs of the A_1AR. Despite significant advances, more effective approaches are needed to discover potent and selective allosteric modulators of the A_1AR. Ensemble docking that integrates Gaussian accelerated molecular dynamic (GaMD) simulations and molecular docking using *Autodock* has been implemented for retrospective docking of known positive allosteric modulators (PAMs) in the A_1AR. Ensemble docking outperforms docking of the receptor cryo-EM structure. The calculated docking enrichment factors (EFs) and the area under the receiver operating characteristic curves (AUC) are significantly increased. Receptor ensembles generated from GaMD simulations are able to increase the success rate of discovering PAMs of A_1AR. It is important to account for receptor flexibility through GaMD simulations and flexible docking.

Keywords: Adenosine A_1 Receptor, Allosteric Modulators, Gaussian accelerated Molecular Dynamics, Ensemble Docking, G-protein-coupled receptors.

Introduction

G-protein-coupled receptors (GPCRs) are the largest family of human membrane proteins. They mediate cellular responses to hormones, neurotransmitters, chemokines and the senses of sight, olfaction and taste. GPCRs have served as the primary targets for about one third of currently marketed drugs [153-155]. Particularly, the adenosine A_1 receptor (A_1AR) has emerged as an important therapeutic target for treating cardiac ischemia-reperfusion injuries, neuropathic pain and renal diseases[156]. However, development of the A_1AR agonists as effective drugs has been greatly hindered. Several candidates could not progress into the clinic due to low efficacy and/or safety issues related to off-target effects. Four receptor subtypes, the A_1, A_{2A}, A_{2B} and A_3, share a highly conserved endogenous agonist binding ("orthostatic") site. It is challenging to design effective agonists with high selectivity.

It is appealing to design positive allosteric modulators (PAMs) that bind a less conserved, topographically distinct site and increase the responsiveness of the A_1AR to endogenous adenosine in local regions of its production. The first selective PAM of the A_1AR, PD81,723, was introduced in 1990[157,158]. Since then, many research groups have performed extensive structure-activity relationship studies [158-167]. Several refined compounds were identified. Notably, T62 evaluated by King Pharmaceuticals progressed to Phase IIB clinical trial but failed due to lack of potency [168,169]. Overall, these compounds still suffer from major limitations such as low solubility and potency for pharmaceutical use. It remains difficult to discover PAMs of higher potency for the A_1AR.

Virtual screening has become increasingly important in the development of therapeutic drugs and discovery of novel GPCR ligands, including the antagonists, agonists, and allosteric modulators [170,171]. Molecular docking is a widely used virtual screening technique. Early

docking studies were performed using static crystal structures of target receptors and flexible ligands, which were successful in certain cases [172]. However, proteins have been often considered rigid, largely limiting the successful rate of docking. In fact, proteins are usually flexible and involve conformational changes upon ligand binding [173]. Therefore, computational approaches are needed to deal with the flexibility of proteins.

Ensemble docking of small probe molecules for flexible pharmacophore modeling was first introduced in 1999 [173]. Since then, ensemble-based [174] methods has been implemented to identify novel ligands of different proteins [175], including the immunophilin FKBP [176], HIV-1 integrase [177], RNA-editing ligase [178], membrane fusion protein [179], prothrombinase enzyme [180] and fibroblast growth factor 23 [181]. Additionally, ensemble docking has been applied to applied interactions between the drug and off-target proteins [182] and identify protein targets of natural products [183,184].

Furthermore, ensemble docking has been applied to discover novel allosteric modulators of GPCRs. Huang et al.[185] discovered a unique PAM ogerin of GPR68 and allosteric modulators of GPR65 by docking to receptor ensembles generated from homology modeling. Long-timescale accelerated molecular dynamics (aMD) simulations were incorporated into ensemble docking to design allosteric modulators of the M2 muscarinic GPCR [186]. A number of 12 compounds with affinities $\leq$30 μM was identified, four of which were confirmed as new negative allosteric modulators (NAMs) and one as a PAM of the M2 receptor.

For the A_1AR, X-ray structures have been determined in the inactive antagonist-bound form [143] and a cryo-EM structure in the active agonist-Gi-bound complex [141]. However, there is still no published structure of the A_1AR bound by allosteric modulators. This has greatly

hindered structure-based design of potent and selective PAMs of the A_1AR [187]. Nevertheless, mutagenesis and molecular modeling studies have suggested that the A_1AR allosteric site may reside within the second extracellular loop (ECL2) [188,189]. We previously applied Gaussian accelerated molecular dynamics (GaMD) simulations to determine binding modes of two prototypical A_1AR PAMs, PD81723 and VCP171[190]. The GaMD simulations allowed us to identify low-energy binding modes of the PAMs at an allosteric site formed by the receptor ECL2, being highly consistent with the experimental data [188,189]. Additionally, we performed GaMD simulations on both the inactive antagonist-bound and active agonist-Gi-bound A_1AR [191,192]. The GPCR-membrane interactions were found to highly depend on the receptor activation state and motions in the GPCR and membrane lipids are strongly coupled [191]. These studies provided an excellent starting point for ensemble docking and rational drug design of the A_1AR. In this study, structural ensembles were generated from GaMD simulations of the A_1AR and used for retrospective docking of allosteric modulators (**Fig. 1 and 2A**). Receptor snapshots were taken every 0.2 ps from the GaMD simulations and clustered for the ECL2 target site. Ten representative structural clusters were obtained for molecular docking. We performed retrospective docking of 25 known PAMs of the A_1AR [167,193-196] and 2475 drug-like decoys generated from the Directory of Useful Decoys, Enhanced (DUD-E) [197]. *AutoDock* was applied for both rigid-body and flexible docking [198-200]. Enrichment factors (EFs) [201] and area under the receiver operating characteristic curve (AUC) were calculated to evaluate the docking performances. Retrospective docking of the PAMs allowed us to validate structural ensembles of the A_1AR and optimize our molecular docking protocol for future virtual screening.

Materials and Methods

Gaussian accelerated Molecular Dynamics (GaMD)

GaMD is an enhanced sampling technique, in which a harmonic boost potential is added to reduce the system energy barriers[202]. GaMD is able to accelerate biomolecular simulations by orders of magnitude[203,204]. GaMD does not need predefined collective variables. Moreover, because GaMD boost potential follows a Gaussian distribution, biomolecular free energy profiles can be properly recovered through cumulant expansion to the second order[202]. GaMD has successfully overcome the energetic reweighting problem in free energy calculations that was encountered in the previous accelerated molecular dynamics (aMD) method[205,206] for free energy calculations. GaMD has been implemented in widely used software packages including AMBER [202,207] and NAMD[208]. A brief summary of GaMD is provided here.

Consider a system with *N* atoms at positions $\vec{r} = \{\vec{r}_1, \cdots, \vec{r}_N\}$. When the system potential $V(\vec{r})$ is lower than a reference energy *E*, the modified potential $V^*(\vec{r})$ of the system is calculated as:

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}),$$

$$\Delta V(\vec{r}) = \begin{cases} \frac{1}{2}k\left(E - V(\vec{r})\right)^2, & V(\vec{r}) < E \\ 0, & V(\vec{r}) \geq E \end{cases} \qquad (1)$$

where *k* is the harmonic force constant. The two adjustable parameters *E* and *k* are automatically determined based on three enhanced sampling principles[202]. The reference energy needs to be set in the following range:

$$V_{max} \leq E \leq V_{min} + \frac{1}{k}, \qquad (2)$$

Where V_{max} and V_{min} are system minimum and maximum potential energies. To ensure that Eqn. (2) is valid, k has to satisfy: $k \leq \frac{1}{V_{max}-V_{min}}$ Let us define $k \equiv k_0 \frac{1}{V_{max}-V_{min}}$, then $0 < k_0 \leq 1$. The standard deviation of ΔV needs to be small enough (i.e., narrow distribution) to ensure proper energetic reweighting[209]: $\sigma_{\Delta V} = k(E - V_{avg})\sigma_V \leq \sigma_0$ where V_{avg} and σ_V are the average and standard deviation of the system potential energies, $\sigma_{\Delta V}$ is the standard deviation of ΔV with σ_0 as a user-specified upper limit (e.g., $10k_B$T) for proper reweighting. When E is set to the lower bound $E=V_{max}$, k_0 can be calculated as:

$$k_0 = \min(1.0, k_0') = \min\ (1.0, \frac{\sigma_0}{\sigma_V}\frac{V_{max}-V_{min}}{V_{max}-V_{avg}}). \qquad (3)$$

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

$$k_0 = k_0'' \equiv (1 - \frac{\sigma_0}{\sigma_V})\frac{V_{max}-V_{min}}{V_{avg}-V_{min}}, \qquad (4)$$

if k_0'' is found to be between *0* and *1*. Otherwise, k_0 is calculated using Eqn. (3).

Similar to aMD, GaMD provides schemes to add only the total potential boost ΔV_P, only dihedral potential boost ΔV_D, or the dual potential boost (both ΔV_P and ΔV_D). The dual-boost simulation generally provides higher acceleration than the other two types of simulations[210]. The simulation parameters comprise of the threshold energy E for applying boost potential and the effective harmonic force constants, k_{0P} and k_{0D} for the total and dihedral potential boost, respectively.

Energetic reweighting of GaMD simulations

To calculate potential of mean force (PMF)[211] from GaMD 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:

$$p(A_j) = p^*(A_j)\frac{\langle e^{\beta\Delta V(\vec{r})}\rangle_j}{\sum_{i=1}^{M}\langle p^*(A_i)e^{\beta\Delta V(\vec{r})}\rangle_i},\ j = 1, \dots, M, \quad (5)$$

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:

$$\langle e^{\beta\Delta V(\vec{r})}\rangle = exp\left\{\sum_{k=1}^{\infty}\frac{\beta^k}{k!}C_k\right\}, \quad (6)$$

where the first two cumulants are given by

$$\begin{aligned} C_1 &= \langle\Delta V\rangle, \\ C_2 &= \langle\Delta V^2\rangle - \langle\Delta V\rangle^2 = \sigma_v^2. \end{aligned} \quad (7)$$

The boost potential obtained from GaMD simulations usually follows near-Gaussian distribution. Cumulant expansion to the second order thus provides a good approximation for computing the reweighting factor[202,209]. The reweighted free energy $F(A) = -k_B T \ln p(A)$ is calculated as:

$$F(A) = F^*(A) - \sum_{k=1}^{2}\frac{\beta^k}{k!}C_k + F_c, \quad (8)$$

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

System setup

The cryo-EM structure of the ADO-A_1AR-Gi complex (PDB: 6D9H[141]) were used to prepare the simulation systems. In order to prepare PAM-bound structures of the active A_1AR for docking, the VCP171 was docked with *Autodock* to the ECL2 allosteric site of the A_1AR [188-190]. Then the binding conformation of VCP171 with the highest docking score was chosen as initial structure of the PAM-bound A_1AR. For both PAM-free (*apo*) and PAM-bound (*holo*) structures of the A_1AR, all chain termini were capped with neutral groups, i.e. the acetyl group (ACE) for the N-terminus and methyl amide group (CT3) for C terminus. Protein residues were set to the standard CHARMM protonation states at neutral pH with the *psfgen* plugin in VMD[212]. Then the receptor was inserted into a POPC bilayer with all overlapping lipid molecules removed using the *Membrane* plugin in VMD[212]. The system charges were then neutralized at 0.15 M NaCl using the *Solvate* plugin in VMD[212]. Periodic boundary conditions were applied on the simulation systems.

The CHARMM36 parameter set [213] was used for the protein and POPC lipids. For agonist ADO and PAM VCP171, the force field parameters were obtained from the CHARMM ParamChem web server[188,214]. Initial energy minimization and thermalization of the A_1AR system follow the same protocol as used in the previous GPCR simulations[215]. The simulation proceeded with equilibration of lipid tails. With all the other atom fixed, the lipid tails were energy minimized for 1000 steps using the conjugate gradient algorithm and melted with constant number, volume, and temperature (NVT) run for 0.5ns at 310 K. Each system was further equilibrated using constant number, pressure, and temperature (NPT) run at 1 atm and

310 K for 10 ns with 5 kcal (mol Å^2)$^{-1}$ harmonic position restraints applied to the protein. Further equilibration of the systems was performed using an NPT run at 1 atm and 310 K for 0.5ns with all atoms unrestrained. Conventional MD simulation was performed on each system for 10 ns at 1atm pressure and 310 K with a constant ratio constraint applied on the lipid bilayer in the X-Y plane. Both dihedral and dual-boost GaMD simulations (GaMD_Dih and GaMD_dual respectively) were then performed using NAMD2.13[208,216] and AMBER 18 [202] to generate receptor ensembles for docking. Simulations of the A_1AR were summarized in **Table 1.** In the GaMD simulations, the threshold energy *E* for adding boost potential is set to the lower bound, i.e. $E = V_{max}$[202,208]. The simulations included 50ns equilibration after adding the boost potential and then three independent production runs lasting 300 *ns* with randomized initial atomic velocities. GaMD production simulation frames were saved every 0.2ps. Snapshots of all three GaMD production simulations were combined for clustering to identify representative structures for docking using the hierarchical agglomerative algorithm in CPPTRAJ[217]. The *PyReweighting* toolkit[209] was applied to reweight GaMD simulations by combining independent trajectories for each system. Free energy values were calculated for the top-ranked structural clusters of the receptor.

Retrospective docking of known allosteric ligands to the A_1AR

A number of 25 known PAMs of the A_1AR were collected from literature [167,193-196] (**Fig. 2B**) and 2475 decoys were obtained from the DUD-E database [197]. This compound library served as a dataset to validate the receptor ensembles constructed from GaMD simulations of the A_1AR. Docking was performed on these compounds to the receptor ECL2 site using *Autodock*[198].

Default parameters (number of individuals in population (ga_pop_size), 1,500; maximum number of generations (ga_num_generations), 27,000 and number of genetic algorithms run (ga_run), 10) were used for the *Autodock* docking calculations unless described otherwise. A set of docking algorithms were extensively tested, including the flexible docking and rigid-body docking at different levels (the long level with ga_num_evals of 25,000,000, medium level with ga_num_evals of 2,500,000 and short level with ga_num_evals of 250,000). For the flexible docking, residues within 5 Å of VCP171 in the docking pose were selected as flexible residues. The representative structures obtained from GaMD simulations and the cryo-EM structure of the A_1AR (PDB ID: 6D9H) were used for the receptor. The ADO agonist was kept in all the docking calculations. Firstly, all polar hydrogens were added and Gasteiger charges were assigned to atoms in the receptor. Secondly, a 3D search grid was created with the *AutoGrid* algorithm[218] to calculate binding energies of the ligands and decoys in the A_1AR. The center of mass of the receptor ECL2 was chosen as the grid center and a box size of 60*60*60 $Å^3$ was applied for docking. The Lamarckian genetic algorithm[218] was applied to model protein-ligand interactions.

For docking of GaMD simulation structural clusters, the predicted binding energy for each conformation was calculated by its raw docking score and reweighted according to the following:

$$BE_i = E_{dock} + PMF_i \tag{9}$$

where PMF_i is the reweighted free energy of the *i*th receptor structural cluster and E_{dock} is the corresponding raw docking score. Ranking of the docked compounds was examined in terms of both the minimum predicted binding energy obtained for each compound against any of the

given receptor structures ("BE_{min}") and the average of predicted binding energies ("BE_{avg}") [186]. The average binding energy is calculated as:

$$BE_{avg} = \frac{1}{N_c}\sum_{i=1}^{N_c} BE_i \quad (10)$$

where *Nc* is the total number of receptor structural clusters, BE_i is the binding energy of a compound to the *i*th cluster. The minimum predicted binding energy is given by:

$$BE_{min} = Min\left(BE_1, BE_1, \cdots BE_{N_C}\right) \quad (11)$$

Enrichment factor (EF) is calculated by:

$$EF = \frac{Hits\_sampled/N_{sampled}}{Hits\_total\ /N_{total}} = \frac{N_{total}}{N_{sampled}} \cdot \frac{Hits\_sampled}{Hits\_total} \quad (12)$$

where the N_{total}, $N_{sampled}$, *Hits_total* and *Hits_sampled* are the number of total compounds in the database, number of compounds for ranking, number of total hits in the database and number of hits found among the ranked compounds, respectively. The "early" enrichment factor (EF') that accounts for the rank of each of the *Hits_sampled* known actives is calculated by:

$$EF' = \frac{50\%}{APR_{sampled}} \cdot \frac{Hits\_sampled}{Hits\_total} \quad (13)$$

Where the $APR_{sampled}$ is the average percentile rank of the *Hits_sampled* known actives.

Results

Construction of receptor ensembles through structural clustering of GaMD simulations

All-atom GaMD simulations were performed on different systems of the active A_1AR with varied levels of acceleration, i.e. the dihedral and dual boost (see **Methods** and **Table 1**). Two

different software packages, NAMD2.13[208,216] and AMBER 18[202,207] were used for the simulations. Six sets of GaMD simulations, including the ADO-A_1AR-Gi (NAMD, GaMD_Dih), ADO-A_1AR-Gi (NAMD, GaMD_Dual), ADO-A_1AR-Gi (AMBER, GaMD_Dih), ADO-A_1AR-Gi (AMBER, GaMD_Dual), ADO-A_1AR-Gi-VCP171 (AMBER, GaMD_Dih) and ADO-A_1AR-Gi-VCP171 (AMBER, GaMD_Dual) were obtained to generate receptor ensembles. Overall, GaMD_Dual provided higher boost potential than the GaMD_Dih. AMBER provided higher boost potential than NAMD in the GaMD simulations due to different algorithms used to calculate the potential average and standard deviation[190]. The GaMD_Dual simulations of ADO-A_1AR-Gi with AMBER recorded an average boost potential of 21.43 kcal/mol and 6.50 kcal/mol standard deviation compared with 10.14 kcal/mol average and 2.59 kcal/mol standard deviation in the GaMD_Dual simulations with NAMD. The GaMD_Dih simulations of ADO-A_1AR-Gi using AMBER also exhibited higher average boost potential of 6.89 kcal/mol and 4.06 kcal/mol standard deviation compared with 5.04 kcal/mol average and 2.22 kcal/mol standard deviation in the GaMD_Dih simulations using NAMD. For the VCP171 bound A_1AR system, the GaMD_Dual simulations with AMBER recorded a 24.80 kcal/mol average boost potential and 6.61 kcal/mol standard deviation compared with the average of 6.80 kcal/mol and of 2.47 kcal/mol standard deviation in the GaMD_Dih simulations with AMBER.

Ten representative structures of the receptor ECL2 allosteric site were obtained by root-mean-square deviation (RMSD)-based structural clustering of the receptor snapshots for each set of the GaMD simulations. The representative structural clusters were characterized by the fraction of simulation frames in the cluster, the GaMD reweighted PMF free energy values of each cluster, and their sampled conformational space (**Table 2**). Overall, the top clusters in each receptor ensemble contributed to significantly higher fractions of simulation frames than the

lower ranked clusters. However, the free energy values were ordered differently for most of the structural clusters with GaMD reweighting. It was thus important to take GaMD reweighted free energies into account in the ensemble docking. Notably, structural clusters obtained from AMBER GaMD_Dual simulations of ADO-A_1AR-Gi-VCP171 system sampled the largest conformational space, as characterized by ~4.5-5.9 Å average distance between the cluster centroids (AvgCDist) and 2.31 Å average distance between simulation frames (AvgDist) of the lowest free energy/top-ranked cluster **(Table 2)**. In comparison, clusters from AMBER GaMD_Dih simulations of the VCP171-bound A_1AR sampled the smallest conformational space with ~2.1-2.4 Å average cluster centroid distance and 1.78 Å average distance between simulation fames in the top cluster. Clusters obtained from GaMD simulations of the ADO-A_1AR-Gi system sampled conformational space in the medium range **(Table 2)**. These receptor structural ensembles were subsequently used for retrospective docking of PAMs in the A_1AR.

Correction with GaMD reweighting improved retrospective docking performances

The GaMD receptor ensembles were used for docking of the compound library consisting of 25 known PAMs and 2475 decoys of the A_1AR. The EF and EF' enrichment factors and AUC values were calculated to evaluate the ensemble docking performances [201]. The EF and EF' were calculated for top 5% and 10% of total sampled compounds (**Table 3**). For the ten representative conformations obtained from each set of GaMD simulations, both the minimum binding energy (BE_{min}) and the average binding energy (BE_{avg}) were used to assess the docking performances. To evaluate the effects of GaMD reweighting, we compared the binding energies using both the raw scores obtained from docking calculations ("**raw"** in **Table 3**) and the

reweighted scores corrected with cluster free energy values calculated by GaMD reweighting **("reweighted"** in **Table 3**). For a total of 60 calculated docking metrics regarding the AUC, EF (5%), EF (10%), EF' (5%) and EF' (10%) as listed in **Table 3**, 32 (53.3%) of them showed better performance using the GaMD-reweighted scores than using the raw scores. While only 10 (16.7%) of them showed decreased performance and 18 (30%) with the same performance (**Table 3**).

With GaMD reweighting, the calculated AUC using ranking by the minimum binding energy (BE_{min}) increased by ~10-14% for receptor ensembles obtained from NAMD GaMD_Dih, NAMD GaMD_Dual and AMBER GaMD_Dih simulations of the ADO-A_1AR-Gi system. The AUC using ranking by the average binding energy (BE_{avg}) increased from 0.44 to 0.61 for receptor ensemble from AMBER GaMD_Dih of the ADO-A_1AR-Gi system. More importantly, while the EF and EF' values for 5% of sampled compounds were generally small for both raw and reweighted docking scores, the EF and EF' for 10% of sampled compound increased significantly with reweighted scores for receptor ensembles of all the simulations (**Table 3**). This would be more relevant for drug design projects since only top ranked compounds could often be tested in experiments. Therefore, binding energies corrected by GaMD reweighting improved the docking performances.

Furthermore, we compared the calculated AUC, EF and EF' in terms of ranking using the minimum binding energy (BE_{min}) and average binding energies (BE_{avg}). Although with exceptions, the usage of BE_{avg} mostly outperformed that of BE_{min}, especially in the calculated EF' (5%) and EF' (10%) (**Table 3**). Notably, the EF'(5%) increased by ~70% to 7 times using BE_{avg} in ensemble docking of the ADO-A_1AR-Gi system from the NAMD GaMD_Dih, NAMD

GaMD_Dual and AMBER GaMD_Dih simulations than using the BE_{min}, and ~2-5 times for the EF'(10%) (**Table 3**). The average predicted binding energy (BE_{avg}) was thus applied for further parameter testing of the docking calculations.

Flexible docking performed better than different levels of rigid-body docking

Next, we compared the performances of flexible docking and different levels of rigid-body docking with *Autodock* (**Table 4**). In a total of 30 calculated docking metrics regarding the AUC, EF(5%), EF(10%), EF'(5%) and EF'(10%), 22 (73.3%) of them showed improved performance with flexible docking compared with rigid-body docking at the short, medium and long levels, while only 2 (6.7%) showed the same performance and 6 (20%) with decreased performance. Among the 22 metrics that showed improved performance with flexible docking, 17 of these values increased by more than 2 times compared with rigid-body docking (**Table 4**).

The AUC increased with flexible docking of all the receptor ensembles except the ADO-A_1AR-Gi ensemble from the AMBER GaMD_Dih and GaMD_Dual simulations (**Table 4**). The EF (5%) increased from 0.0 with different levels of rigid-body docking to 0.8 with flexible docking for all the receptor ensembles except the ADO-A_1AR-Gi ensemble from the AMBER GaMD_Dih simulations. Similar significant increase was observed for the EF' (5%) values with flexible docking of these receptor ensembles. Notably, AMBER GaMD_Dual simulations of the PAM VCP171-bound ADO-A_1AR-Gi and PAM-free ADO-A_1AR-Gi systems combined with flexible docking provided the highest EF' (5%) for, i.e., 8.33 and 6.58, respectively. For the EF (10%) and EF'(10%), flexible docking led to significantly higher performance values

for most of the simulations receptor ensembles than the rigid-body docking (**Table 4**).In summary, flexible docking provided significantly improved performances than rigid-body docking, suggesting that the flexibility of protein side chains played an important role even in the ensemble docking calculations.

Optimized ensemble docking protocol and comparison with docking using cryo-EM structure

Comparing flexible docking of all six receptor ensembles with ranking by the GaMD reweighted the average binding energy (BE_{avg}), we found that the AMBER GaMD_Dual simulations of the PAM VCP171 bound ADO-A_1AR-Gi system provided the best docking performance (**Table 4**). The GaMD simulations of VCP171-bound A_1AR sampled more relevant receptor conformations that favored binding of the PAMs, although the ECL2 allosteric site was highly flexible [191,192] and the VCP171 PAM exhibited high fluctuations. Such conformations were not well sampled in GaMD simulations of the PAM-free (*apo*) A_1AR started from the cryo-EM structure of ADO-A_1AR-Gi.

For comparison, we performed flexible docking using *Autodock* with only the cryo-EM structure of the active ADO-A_1AR-Gi (PDB: 6D9H). While the EF(5%) and EF(10%) of flexible docking of the cryo-EM structure exhibited the same values as ensemble docking of the PAM-bound A_1AR, the AUC, EF'(5%) and EF'(10%) values decreased significantly to 0.53, 2.88 and 3.47, respectively, compared to the latter (**Table 5**). It is worth noting that the performance of docking the PAMs in the current study was relatively lower than that of the orthosteric ligands that bind inside class A GPCRs with generally higher affinities [186]. Nonetheless, with GaMD simulations and flexible docking, we have optimized our ensemble

docking protocol to effectively account for the receptor flexibility, which will facilitate virtual screening of PAMs for the A_1AR.

Discussion

In this study, we have carried out extensive retrospective docking calculations of PAM binding to the A_1AR using GaMD enhanced simulations and *AutoDock*. The GaMD simulations have been performed using the AMBER and NAMD simulation packages at different acceleration levels (dihedral and dual boost). The rigid-body docking at different short, medium and long levels and flexible docking have been all evaluated in our ensemble docking protocol.

Overall, flexible docking performed significantly better than the rigid-body docking at different levels with *AutoDock*. This suggested that the flexibility of protein side chains in ensemble docking is also important. The side chains of representative receptor structures obtained from GaMD simulations might be still in unfavored conformations for PAM binding. Flexible docking of protein side chains could then alleviate this problem to achieve better performance. The Glide induced fit docking was also found to outperform rigid-body docking in a previous study to identify allosteric modulators of the M2 muscarinic GPCR [186]. Generally speaking, flexible docking greatly helps accounting for protein flexibility in the side chains and improves docking performances[219].

To fully account for the protein flexibility, especially the backbone, we further incorporated GaMD enhanced sampling simulations in an ensemble docking protocol (**Fig. 1**). In the GaMD simulations, the boost potential obtained for one system with AMBER was greater than with NAMD due to different algorithms used to calculate the system potential statistics[190].

Accordingly, larger conformation space of the receptor was sampled in the AMBER simulations, which resulted in improved docking performance. Correction of binding energy by the GaMD reweighted free energy of the receptor structural cluster improved the docking performances. With GaMD reweighted scores, ranking by the average binding energy (BE_{avg}) performed better than by the minimum binding energy (BE_{min}) in terms of the AUC, EF and EF' docking metrics.

Receptor ensemble obtained from AMBER GaMD_Dual simulations of the VCP171 PAM-bound ADO-A_1AR-Gi outperformed other receptor ensembles for docking. This ensemble consists of snapshots of the *holo* A_1AR with PAM bound at the ECL2 allosteric site. Interactions between the PAM and receptor ECL2 could induce more suitable conformations for PAM binding, which were otherwise difficult to sample in the simulations of PAM-free (*apo*) A_1AR. Dual-boost GaMD was observed to perform better than the dihedral-boost GaMD for ensemble docking, suggesting that GaMD at the higher dual-boost acceleration level was needed to sufficiently sample conformational space of the GPCR PAM binding site. This finding is in contrast to the previous docking study of allosteric modulators to the M_2 receptor that showed dihedral-boost aMD outperformed dual-boost aMD for ensemble docking [186]. Such discrepancy likely resulted from the relatively lower boost potential with a new formula applied in the GaMD simulations compared with the previous aMD method [202,208]. The aMD simulations especially with dual boost appeared to provide too high acceleration and sample receptor conformations that did not facilitate compound docking. In comparison, dual-boost GaMD generated more appropriate receptor ensembles for docking. GaMD also solved the energetic noise problem of aMD in the protein simulations. The reweighted free energies of receptor

structural clusters obtained from GaMD simulations have been shown to further improve ranking of the compounds and ensemble docking performance as demonstrated here.

In summary, we have optimized the protocol for ensemble docking of allosteric modulators to the A_1AR, a prototypical GPCR. The ensemble docking integrates all-atom dual-boost GaMD simulations of the PAM-bound (*holo*) agonist-A_1AR-Gi complex, flexible docking with *AutoDock* and compound scoring with the GaMD reweighted average binding energy. Enhanced sampling simulations and flexible docking have greatly improved the docking performance by effectively accounting for the protein flexibility in both the backbone and side chains. Such an ensemble docking protocol will greatly facilitate future allosteric drug design of the A_1AR and other GPCRs.

Acknowledgements

We would like to dedicate this manuscript to the 60th birthday of Prof. Jeremy Smith. We appreciate the preliminary docking work of Shulammite Lim and thank Prof. Arthur Christopoulos, Dr. Lauren May, Dr. Anh Nguyen and Prof. Jens Carlsson for valuable discussions. 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 project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231, and the Research Computing Cluster at the University of Kansas. This work was supported by the American Heart Association (Award 17SDG33370094), the National Institutes

of Health (R01GM132572) and the startup funding in the College of Liberal Arts and Sciences at the University of Kansas.

Table 1: Summary of GaMD simulations performed on the adenosine A_1 receptor (A_1AR).

System	[a]N_{atoms}	Dimension(Å^3)	Program	[b]Method	Simulation	[c]ΔV_{avg} (kcal/mol)	[d]$\sigma_{\Delta v}$ (kcal/mol)
ADO-A_1AR-Gi	180,394	93x111x167	NAMD	GaMD_Dih	300 ns x 2	5.04	2.22
				GaMD_Dual	300 ns x 2	10.14	2.59
ADO-A_1AR-Gi	180,394	93x111x167	AMBER	GaMD_Dih	300 ns x 3	6.89	4.06
				GaMD_Dual	300 ns x 3	21.43	6.50
ADO-A_1AR-Gi-VCP171	180,403	93x111x167	AMBER	GaMD_Dih	300 ns x 3	6.80	2.74
				GaMD_Dual	300 ns x 3	24.80	6.61

[a]N_{atoms} is number of atoms in the simulation systems.
[b]GaMD_Dih and GaMD_Dual represent the dihedral and dual boost GaMD simulations respectively.
[c]ΔV_{avg} and [d]$\sigma_{\Delta V}$ are the average and standard deviation of the GaMD boost potential.

Table 2: Summary of receptor structural clusters obtained from different GaMD simulations of the A_1AR. Ten representative structural clusters are listed with their fraction of simulation frames, the GaMD reweighted PMF free energy values (kcal/mol), the average distance of each cluster centroid to every other cluster centroid (AvgCDist) and the average distance between frames in the cluster (AvgDist).

Cluster ID	Cluster Properties	ADO-A_1AR-Gi (NAMD, GaMD_Dih)	ADO-A_1AR-Gi (NAMD, GaMD_Dual)	ADO-A_1AR-Gi (AMBER, GaMD_Dih)	ADO-A_1AR-Gi (AMBER, GaMD_Dual)	ADO-A_1AR-Gi-VCP171 (AMBER, GaMD_Dih)	ADO-A_1AR-Gi-VCP171 (AMBER, GaMD_Dual)
Cluster 1	Fraction	0.22	0.18	0.50	0.78	0.73	0.56
	Reweighted PMF (kcal/mol)	0.19	1.78	0.00	1.64	0.00	0.00
	AvgCDist (Å)	3.01	3.50	4.18	2.76	2.10	4.62
	AvgDist (Å)	2.40	2.29	2.28	1.82	1.78	2.31
Cluster 2	Fraction	0.16	0.16	0.16	0.09	0.15	0.16
	Reweighted PMF (kcal/mol)	0.40	1.27	0.84	2.83	0.97	1.40
	AvgCDist (Å)	3.26	3.04	4.51	3.02	2.37	4.51
	AvgDist (Å)	2.30	2.40	1.99	1.69	1.64	2.62
Cluster 3	Fraction	0.15	0.14	0.14	0.04	0.07	0.14
	Reweighted PMF (kcal/mol)	1.11	0.00	0.61	3.52	1.45	0.35
	AvgCDist (Å)	3.05	3.12	3.75	2.97	2.24	5.41
	AvgDist (Å)	2.35	2.31	2.60	1.88	1.62	2.73
	Fraction	0.14	0.13	0.11	0.04	0.02	0.05

Cluster 4	Reweighted PMF (kcal/mol)	1.14	1.41	1.19	3.70	1.60	2.93
	AvgCDist (Å)	3.03	3.57	3.94	3.42	2.33	4.80
	AvgDist (Å)	2.38	2.20	2.25	1.74	1.57	2.65
Cluster 5	Fraction	0.10	0.12	0.04	0.03	0.02	0.03
	Reweighted PMF (kcal/mol)	1.38	1.25	1.99	3.41	1.83	2.09
	AvgCDist (Å)	2.98	3.10	3.77	3.11	2.34	4.60
	AvgDist (Å)	2.23	2.25	2.00	1.66	1.35	2.78
Cluster 6	Fraction	0.07	0.06	0.03	0.01	0.004	0.02
	Reweighted PMF (kcal/mol)	2.75	0.85	1.54	4.15	2.93	2.79
	AvgCDist (Å)	2.95	3.31	4.62	3.06	2.22	4.76
	AvgDist (Å)	2.14	2.19	2.47	1.69	1.49	2.30
Cluster 7	Fraction	0.06	0.06	0.01	0.001	0.004	0.02
	Reweighted PMF (kcal/mol)	2.23	0.71	1.82	4.86	3.41	1.17
	AvgCDist (Å)	3.12	3.51	4.34	3.07	2.22	5.42
	AvgDist (Å)	2.33	2.35	2.23	1.74	1.71	2.45
Cluster 8	Fraction	0.06	0.06	0.004	0.00	0.00	0.01
	Reweighted PMF (kcal/mol)	2.18	2.32	2.38	5.69	3.36	1.86
	AvgCDist (Å)	3.13	3.40	4.01	2.73	2.22	5.33
	AvgDist (Å)	2.28	2.16	1.63	1.41	1.42	1.88
Cluster 9	Fraction	0.03	0.05	0.003	0.001	0.001	0.01
	Reweighted PMF (kcal/mol)	2.28	1.02	3.43	0.73	4.01	4.10
	AvgCDist (Å)	3.03	3.38	3.83	3.36	2.24	4.78
	AvgDist (Å)	2.11	2.10	1.60	0.00	1.19	1.73
Cluster 10	Fraction	0.02	0.03	0.002	0.001	0.001	0.00
	Reweighted PMF (kcal/mol)	0.00	1.07	4.77	0.00	2.86	5.78
	AvgCDist (Å)	3.90	3.09	3.77	3.57	2.22	5.87
	AvgDist (Å)	2.22	2.14	1.65	0.00	1.35	0.00

Table 3: Summary of docking results for the AUC and the EF and EF' values with 5% and 10% of sampled compounds calculated using different GaMD simulation ensembles of the A_1AR. Among the listed 60 docking performances metrics, 32 of them (highlighted in bold) showed better performance using the GaMD-reweighted scores than using the raw scores, while only 18 and 10 of them showed the same and decreased performance values, respectively.

System	Ranking	AUC		EF (5%)		EF' (5%)		EF (10%)		EF' (10%)	
		raw	reweighted	raw	reweighted	raw	reweighted	raw	reweighted	raw	reweighted
ADO-A_1AR-Gi (NAMD, GaMD_Dih)	BE_{avg}	0.48	0.45	0.00	0.00	0.51	0.78	0.00	0.92	0.00	0.96
	BE_{min}	0.46	0.51	0.00	0.00	0.77	0.46	0.00	1.09	0.40	0.48
ADO-A_1AR-Gi (NAMD, GaMD_Dual)	BE_{avg}	0.45	0.45	0.00	0.00	0.78	0.78	0.00	0.96	0.78	0.96
	BE_{min}	0.46	0.51	0.00	0.00	1.47	0.46	0.80	2.47	0.46	0.48
ADO-A_1AR-Gi (AMBER, GaMD_Dih)	BE_{avg}	0.44	0.61	0.00	0.00	0.36	3.73	0.00	0.71	1.20	5.03
	BE_{min}	0.43	0.49	0.00	0.00	0.42	0.53	0.00	0.76	0.00	0.97
ADO-A_1AR-Gi	BE_{avg}	0.43	0.43	0.00	0.00	0.48	0.47	0.00	0.71	0.00	0.71

(AMBER, GaMD_Dih)	BE_{min}	0.42	0.35	0.00	0.00	0.91	0.42	0.00	1.10	0.00	0.67
ADO-A_1AR-Gi-VCP171 (AMBER, GaMD_Dih)	BE_{avg}	0.47	0.47	0.00	0.00	0.42	0.42	0.00	0.77	0.00	0.77
	BE_{min}	0.49	0.50	0.00	0.00	1.48	0.60	0.40	1.53	0.00	1.10
ADO-A_1AR-Gi-VCP171 (AMBER, GaMD_Dual)	BE_{avg}	0.47	0.44	0.00	0.00	0.37	0.37	0.00	0.67	0.00	0.66
	BE_{min}	0.49	0.46	0.00	0.00	0.73	0.38	0.00	1.10	0.00	0.73

Table 4: Summary of docking results for the flexible and rigid-body docking at the "short", "medium" and "long" levels using *Autodock* regarding the AUC, EF and EF' values with 5% and 10% of sampled compounds calculated using different GaMD simulation ensembles of the A_1AR. Among the listed 30 docking performance metrics, 22 of them (highlighted in bold) showed better performance using flexible docking than using rigid docking, while only 2 and 6 of them showed the same and decreased performance values, respectively.

System	Levels of Docking	ADO-A_1AR-Gi (NAMD, GaMD_Dih)	ADO-A_1AR-Gi (NAMD, GaMD_Dual)	ADO-A_1AR-Gi (AMBER, GaMD_Dih)	ADO-A_1AR-Gi (AMBER, GaMD_Dual)	ADO-A_1AR-Gi-VCP171 (AMBER, GaMD_Dih)	ADO-A_1AR-Gi-VCP171 (AMBER, GaMD_Dual)
AUC	Short	0.45	0.45	0.61	0.43	0.47	0.44
	Medium	0.49	0.49	0.67	0.50	0.50	0.48
	Long	0.50	0.50	0.23	0.54	0.50	0.52
	Flexible	0.53	0.53	0.53	0.51	0.54	0.60
EF (5%)	Short	0.00	0.00	0.80	0.00	0.00	0.00
	Medium	0.00	0.00	3.20	0.00	0.00	0.00
	Long	0.00	0.00	0.80	0.00	0.00	0.00
	Flexible	0.80	0.80	0.00	0.80	0.80	0.80
EF' (5%)	Short	0.78	0.78	3.73	0.47	0.42	0.37
	Medium	0.95	0.95	4.81	0.89	1.11	1.14
	Long	0.10	0.10	3.52	1.36	0.87	0.88
	Flexible	5.00	5.00	1.41	6.58	3.97	8.33
EF (10%)	Short	0.00	0.00	1.20	0.00	0.00	0.00
	Medium	0.00	0.00	0.40	0.00	0.40	0.40
	Long	0.00	0.00	2.00	0.40	0.00	0.00
	Flexible	0.40	0.40	0.40	0.40	0.40	0.80
EF' (10%)	Short	0.96	0.96	5.03	0.71	0.77	0.66
	Medium	1.12	1.12	7.41	1.06	1.16	1.13
	Long	0.20	0.20	2.72	1.69	1.11	1.03
	Flexible	3.05	3.05	1.42	2.46	2.32	4.98

Table 5: Summary of docking results for the AUC and the EF and EF' values in 5% and 10% of sampled compounds calculated using the finally selected ensemble docking model and cryo-EM structure of the A_1AR.

System	AUC	EF (5%)	EF' (5%)	EF (10%)	EF' (10%)
ADO-A_1AR-Gi-VCP171 (AMBER, GaMD_Dual)	0.60	0.80	8.33	0.80	4.98
ADO-A_1AR-Gi (Cryo-EM, PDB:6D9H)	0.53	0.80	2.88	0.80	3.47

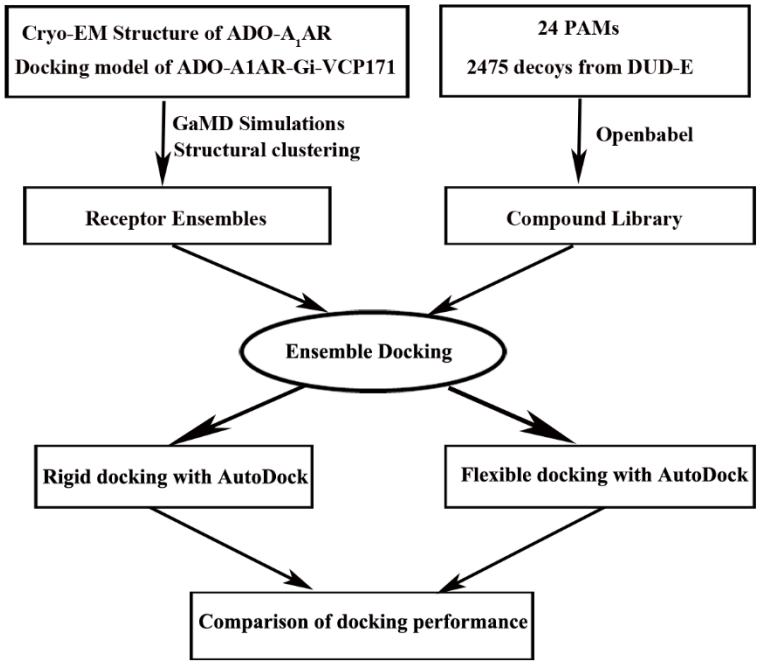

Fig. 1: Overview flowchart for retrospective docking of positive allosteric modulators (PAMs) in the A_1AR. Starting from the cryo-EM structure of the active ADO-Gi-bound A_1AR (6D9H) and docking model of PAM VCP171-bound A_1AR (ADO-A_1AR-Gi-VCP171), GaMD simulations were carried out to construct structural ensembles to account for the receptor flexibility. Meanwhile, a compound library was prepared for 25 known PAMs of the A_1AR and 2450 decoys obtained from the DUD-E with *openbabel* 2.4.1. Ensemble docking was then performed to identify the PAMs, for which the AUC and enrichment factors were calculated to evaluate docking performance. Both rigid-body and flexible docking were tested using *Autodock*.

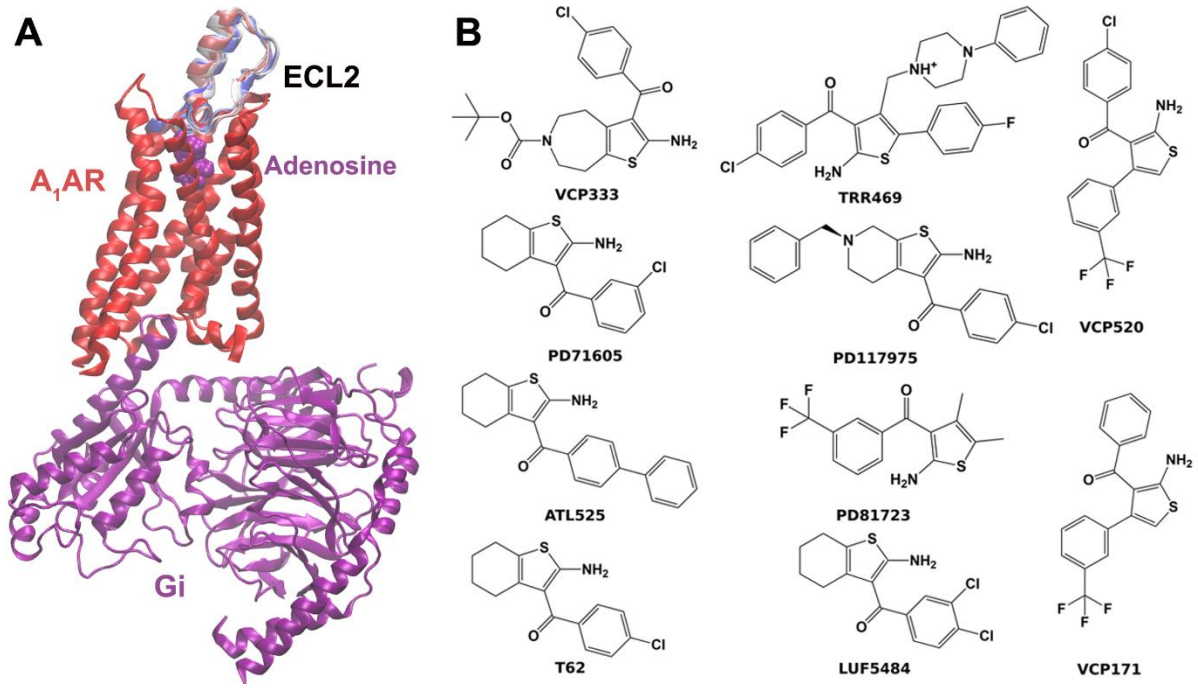

Fig. 2: (A) Structural ensembles of the target allosteric site located in extracellular loop 2 (ECL2) generated for the A_1AR. (B) Example PAMs of the A_1AR used for retrospective docking of the receptor ensembles.

Chapter 5

Mechanism of Ligand Recognition by Human ACE2 Receptor

Apurba Bhattarai[1], Shristi Pawnikar[1] and Yinglong Miao[1]*

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

Reprinted (Adapted) with permission from Bhattarai, Apurba, Shristi Pawnikar, and Yinglong Miao. "Mechanism of ligand recognition by human ACE2 receptor." *The journal of physical chemistry letters* 12.20 (2021): 4814-4822[220].

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 have addressed this challenge through all-atom simulations using a novel Ligand Gaussian accelerated molecular dynamics (LiGaMD) method. Microsecond-timescale LiGaMD simulations have unprecedentedly captured multiple times of spontaneous binding and unbinding of a potent inhibitor MLN-4760 in the ACE2 receptor. With ligand far away in the unbound state, the ACE2 receptor samples distinct Open, Partially Open, Closed and Fully Closed conformations. Upon ligand binding to the active site, conformational ensemble of the ACE2 receptor is biased towards the Closed state as observed in the X-ray experimental structure. The LiGaMD simulations thus suggest a conformational selection mechanism for ligand recognition by the highly flexible ACE2 receptor, which is expected to facilitate rational drug design targeting human 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[221,222]. 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[221]. 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[223,224]. ACE2 also stabilizes amino acid transporter B°AT1 to regulate the gut microbiome and amino acid homeostasis[225].

Human ACE2 has been identified as the functional receptor for severe acute respiratory syndrome coronaviruses including SARS-CoV and SARS-CoV-2[226]. SARS-CoV-2 is responsible for 2019 coronavirus pandemic (COVID-19). By April 2021, ~154 million people have been infected by COVID-19 with ~3.22 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[227], 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 S1A)**. 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., ~16° bending upon inhibitor binding[227]. 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. 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[221] **(Figure S1A)**. 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)[227]. Despite tremendous efforts to determine these experimental structures[223,227-233], the dynamics and functional mechanism of the ACE receptor are still poorly understood[234].

MLN-4760 is a highly selective and potent (IC_{50}: 0.44 nM) small-molecule inhibitor of the ACE2 receptor[235]. The inhibitor has two carboxylate groups contributing to -2 net charge of the molecule (**Figure S1B**). 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.

Materials and Methods

Ligand Gaussian accelerated molecular dynamics (LiGaMD)[236] is an enhanced sampling computational technique for efficient simulations of both dissociation and binding of ligand molecules. It is developed based on Gaussian accelerated molecular dynamics (GaMD), which works by adding a harmonic boost potential to smooth the biomolecular potential energy surface[4]. GaMD greatly reduces energy barriers and accelerates biomolecular simulations by orders of magnitude[35]. 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[4]. In LiGaMD[236], 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[236].

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. The MLN-4760 inhibitor-bound ACE2 receptor structure (PDB: 1R4L)[227] was used to set up the simulation system (**Figure S1C**, see detail in the Supporting Information). 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. This resulted in ~14 mM ligand concentration, being close to the ligand solubility of >5mg/ml or ~12 mM in aqueous solution. The simulation system was prepared as such since the simulation time needed to observe ligand binding would be inversely proportional to the ligand concentration in the solvent. During the LiGaMD equlibration, 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. After the equilibration simulation, ten independent 700 – 2000 ns LiGaMD production simulations ("Sim1" – "Sim10") were further performed with randomized intitial atomic velocities (**Table S1**).

Results

Both binding and unbinding of MLN-4760 inhibitor to the active site of the ACE2 receptor was observed in three of the ten LiGaMD simulations ("Sim1", "Sim2" and "Sim3" in **Table S1**), during which RMSD of the ligand relative to the 1R4L X-ray structure reached a minimum of ~0.99 Å. During "Sim1" LiGaMD simulation, the MLN-4760 inhibitor bound to the active site of the ACE2 receptor during ~500-1400 ns and then dissociated into the bulk solvent (**Figure 1A**). The inhibitor then bound to the receptor active site and dissociated quickly in two different

events at ~1500ns and ~1800ns time, respectively. During "Sim2" LiGaMD simulation, the MLN-4760 inhibitor bound to the active site of the ACE2 receptor during ~100-500 ns (**Figure 1B**) and then dissociated at ~500 ns into the bulk solvent. The inhibitor then bound to the receptor active site and disscoiated quickly at ~700ns. It bound to the receptor active site again at ~1000ns for ~80ns and dissociated at ~1080 ns. Similarly, during "Sim3" LiGaMD simulation, the MLN-4760 inhibitor bound to the active site of the ACE2 receptor during ~1780-1800 ns and then dissociated into bulk solvent (**Figure S3A**). Meanwhile, the receptor underwent large-scale conformational changes with fluctuations in the interdomain distance (**Figure S4**) and sampled different conformations inlcuding the "Open", "Partially Open", "Closed" and "Fully Closed", out of which three states ("Open", "Partially Open" and "Closed") were consistent with the receptor experimental structures and "Fully Closed" was a new low-energy conformational state discovered in the LiGaMD simulations. In the other seven LiGaMD production simulations ("Sim4" to "Sim10" in **Table S1**), no complete, stable ligand binding was observed **(Figure S3)**. The three LiGaMD simulations that successfully captured ligand binding and dissociation ("Sim1" to "Sim3") were used for further analysis of ligand binding pathways.

During the "Sim2" 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 α5 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 **(Figures 1B and 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 α2 and α4 helices as the subdomain I transitioned from "Closed" to "Open" conformation (**Figure 1D**). In contrast, the ligand bound through the space just above the 3_{10} H4 and α5 helices **(Figure 1C)**. However, during multiple binding events in "Sim1" and "Sim3", the inhibitor bound and dissociated from the active site through the space between the receptor α2 and α4 helices. The inhibitor dissociated through the same pathway in the LiGaMD equilibration as well **(Figure S2A)**. This showed that the inhibitor can take either of these pathways (between the receptor α2 and α4 helices or through the interface between the receptor 3_{10} H4 and α5 helices) for binding to the active site and just one pathway (between the receptor α2 and α4 helices) for dissociation. During the ligand binding and dissociation in "Sim1" - "Sim3" LiGaMD trajectories, subdomain I of the receptor sampled different conformations. However, such conformational changes were also observed in other seven simulations ("Sim4" - "Sim10") regardless of ligand binding/dissociation.

A 2D potential of mean force (PMF) free energy profile was calculated with the ligand root-mean-square deviation (RMSD) relative to X-ray conformation and the interdomain distance by combining the 10 independent LiGaMD production trajectories **(Figures 2A, S4 and S5)**. The protein interdomain distance was calculated between the Cα atoms of residues Glu56 and Ser128, which were located at the tip of the α2 and α4 helices, respectively. Nine 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)", "Intermediate-4 (I-4)", "Unbound-1 (U-1)", "Unbound-2 (U-2)", "Unbound-3 (U-3)" and "Ubound-4 (U-4)". Particularly, the system adopted the "Bound" state with ligand RMSD < 5 Å, the "Unbound" state with ligand RMSD > 35 Å and intermediate states with 5 – 35 Å ligand RMSD relative to the 1R4L X-ray structure. The PMF free energy profiles were calculated

through reweighting of the LiGaMD simulations with trajectories of all ten ligands considered. The PMF minima indeed highlight the lowest energy states of ligand binding to the ACE2 receptor.

In the "Bound" state, the ligand bound at the protein active site and the protein interdomain distance was ~12-14 Å. The ligand exhibited a minimum RMSD of 0.99 Å compared with the X-ray structure **(Figures 2A and 2B)**. The system sampled four different intermediate states during ligand binding, *i.e.*, "Intermediate-1 (I-1)", "Intermediate-2 (I-2)", "Intermediate-3 (I-3)" and "Intermediate-4 (I-4)". In the "I-1" state, the ligand RMSD was ~9.8 Å and the interdomain distance was ~18-22 Å **(Figures 2A and 2C)**. The ligand was located near the active site making interactions with residues of the α5 helix, α11 helix, α14 helix, α14 helix and 3_{10}H4 helix in the two protein subdomains. In the "I-2" state, the ligand RMSD was ~25.6 Å and the interdomain distance was ~5-7 Å **(Figures 2A and 2C)**. The ligand interacted with the α17 helix, α18 helix and α19 helix of the subdomain II in the receptor. In the "I-3" state, the ligand RMSD was ~30.4 Å and the interdomain distance was ~13-20 Å **(Figures 2A** and **2C)**. The ligand interacted with the α2 helix of subdomain I and the α4 helix of subdomain II in the receptor. In the "I-4" state, the ligand RMSD was ~ 31.7 Å and the interdomain distance was ~25-26 Å **(Figures 2A** and **2C)**. The ligand interacted with the α8 helix, α14 helix, and 3_{10}H4 in the protein subdomain II (**Figure 2C)**. The system sampled four "Unbound-1 (U-1)", "Unbound-2 (U-2)", "Unbound-3 (U-3)" and "Unbound-4 (U-4)" states, where the ligand RMSD was ~80 Å and the interdomain distances were ~5-7 Å, 10-12 Å, ~20-21 Å and ~25 Å, respectively. In these states, the ligand was found far away from the receptor in the bulk solvent and the receptor could change among the "Fully Closed", "Closed", "Partially Open" and "Open" conformations **(Figures 2A** and **2D)**.

Low-energy intermediate conformational states “I-1”, “I-2”, “I-3” and “I-4” 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, 3C** and **3D**, 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 ~18-22 Å interdomain distance. The ligand molecule was located near the receptor active site with ~9.8 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 Arg273 **(Figure 3A)**. Another carboxylate group formed ionic interaction with Asn149 postively charged nitrogen group. Similary, the ligand’s central ring formed π-π interaction with the aromatic protein residue Phe274. The ligand chloride group formed polar interactions with protein residues Thr371, S409 and Thr445. In the intermediate “I-2” state, the receptor adopted a “Fully Closed” conformation and the ligand formed polar and hydropobic interactions with the residues of receptor subdomin II. One of the chloride atoms formed polar interactions with residue Glu552. One of the ligand’s negatively charged carboxylate groups formed ionic interactions with the protein positively charged groups of residue Arg559 and Asn572. The ligand also formed hydrophobic interactions with protein residue Leu568.

In the intermediate “I-3” state, the receptor adopted the “Partially Open” conformation and the ligand interacted with the α2 helix of subdomain I and the α4 helix of subdomain II in the receptor (**Figure 3C**). One of the ligand chloride atoms formed polar interactions with

protein residues Ser70 and Asn117, while the other chloride atom formed polar interactions with protein residue Ser113. One of the ligand's negatively charged carboxylate groups formed ionic interactions with the protein postively charged groups of residues Lys114 and Asn64, while the other carboxylate group formed ionic interactions with positively charged nitrogen of Gln60. The ligand central ring's nitrogen atom formed ionic interactions with negatively charged oxygen group of protein residue Thr118 (**Figure 3C**). In the intermediate "I-4" state, one of the ligand chloride atoms formed polar interactions with protein residues Ser280 of subdomain II (**Figure 3D**). One of the ligand charged carobxylate groups formed polar interactions with positively charged protein residue Lys247. One of the nitrogen atoms in the ligand's central ring formed ionic interactions with negatively charged oxygen of Pro284 and Gln287 in the receptor. In addition to free energy profiles calculated by combining all the 10 LiGaMD simulations ("Sim1" to "Sim10"), we calculated PMF profiles for each of the ten independent LiGaMD production simulations as shown in **Figure S5**. These free energy profiles showed clear differences, in terms of positions and free energy values of the PMF minima and energy barrier heights, suggesting that the LiGaMD simulations were still not converged.

During the LiGaMD simulations, Zn^{2+} was stabilized by ionic interactions with the Glu375 and Glu402 residues near the active site of human ACE2 receptor. The distances between the positively charged zinc and the C_δ atoms of Glu375 and Glu402 were maintained at ~2-3 Å. This was observed consistently in all the 10 LiGaMD independent simulations **(Figure S6)**.

In the LiGaMD simulations, while the protein subdomain II was stable maintaining the 1R4L X-ray conformation with ~2-4 Å RMSD **(Figure 4B)**, subdomain I in the human ACE2 receptor exhibited high flexibility and underwent large conformational changes with ~3-10 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. Four low-energy conformational states were identified in the PMF profile, including the "Open", "Partially Open", "Closed" and "Fully Closed" **(Figure 4C)**.

In the "Closed" conformation, subdomain I moved near subdomain II closing the active site. The receptor interdomain distance was ~14-15 Å and RMSD of subdomain I was ~5 Å compared with the 1R4L X-ray structure **(Figures 4C and 4D)**. In the "Fully Closed" conformation, the subdomain I can further move towards the subdomain II with interdomain distance ~12-13 Å. The subdomain I RMSD was ~4.5 Å. In the "Partially Open" conformation, the receptor interdomain distance increased to ~17-18 Å and RMSD of subdomain I RMSD was ~6 Å compared with the 1R4L X-ray structure **(Figures 4C and 4D)**. Finally, the receptor interdomain distance could increase further to ~19-20 Å and the subdomain I RMSD relative to the 1R4L X-ray structure increased to ~7 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 **(Figures 4C and 4D)**. 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[237], 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. When the ligand binds to the active site of the receptor in the Bound state with the ligand RMSD <5 Å, the interdomain distance was confined to be small at ~14 Å, being closed. This suggested that the 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 has two carboxylate groups contributing to net -2 charge of the molecule. Hence, ligands repelled each other with no significant ligand-ligand interactions observed in the simulations. 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[236,238].

Despite our encouraging simulation findings, it is important to note that we sampled ligand dissociation and binding events in 3 out of 10 LiGaMD simulations. The free energy profiles calculated for each of the ten individual simulations (**Fig. S5**) and all ten simulations combined (**Fig. 5A and 4C**) showed clear differences, in terms of the positions and free energy values of the PMF minima and energy barrier heights. This suggested that the LiGaMD simulations were still not converged and the calculated free energy profiles were not accurate for direct comparison with experimental ligand binding free energy. It remained challenging to sample enough events of repetitive ligand binding and unbinding along with the large protein conformational changes. More sufficient sampling would be needed in order to obtain converged simulations and calculate accurate ligand binding free energies and kinetic rates. 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[239,240] and Markov state models[241]**.**

In this context, the MLN-4760 inhibitor binds to the human ACE2 receptor with high affinity (IC_{50}: 0.44 nM)[227]. It is extremely difficult to simulate the ligand dissociation and binding with long-timescale conventional MD (cMD) and even the enhanced sampling methods. A recent study[234] 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.

We further highlighted the dynamic nature of the ACE2 receptor in terms of the large-scale movement of subdomain I upon ligand binding. Furthermore, since the human ACE2

receptor shows conformational selection for ligand binding as revealed from the LiGaMD simulations, virtual screening using ensemble docking[242-244] 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, being consistent with previous experimental structures. Ligand binding could bias the receptor conformational ensemble towards the Closed state, suggesting a conformational selection mechanism. Therefore, the LiGaMD simulations have allowed us to understand the mechanism of ligand recognition by the ACE2 receptor, which is expected to facilitate rational drug design targeting ACE2 for the therapeutic treatments of COVID-19 and other related human diseases.

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.

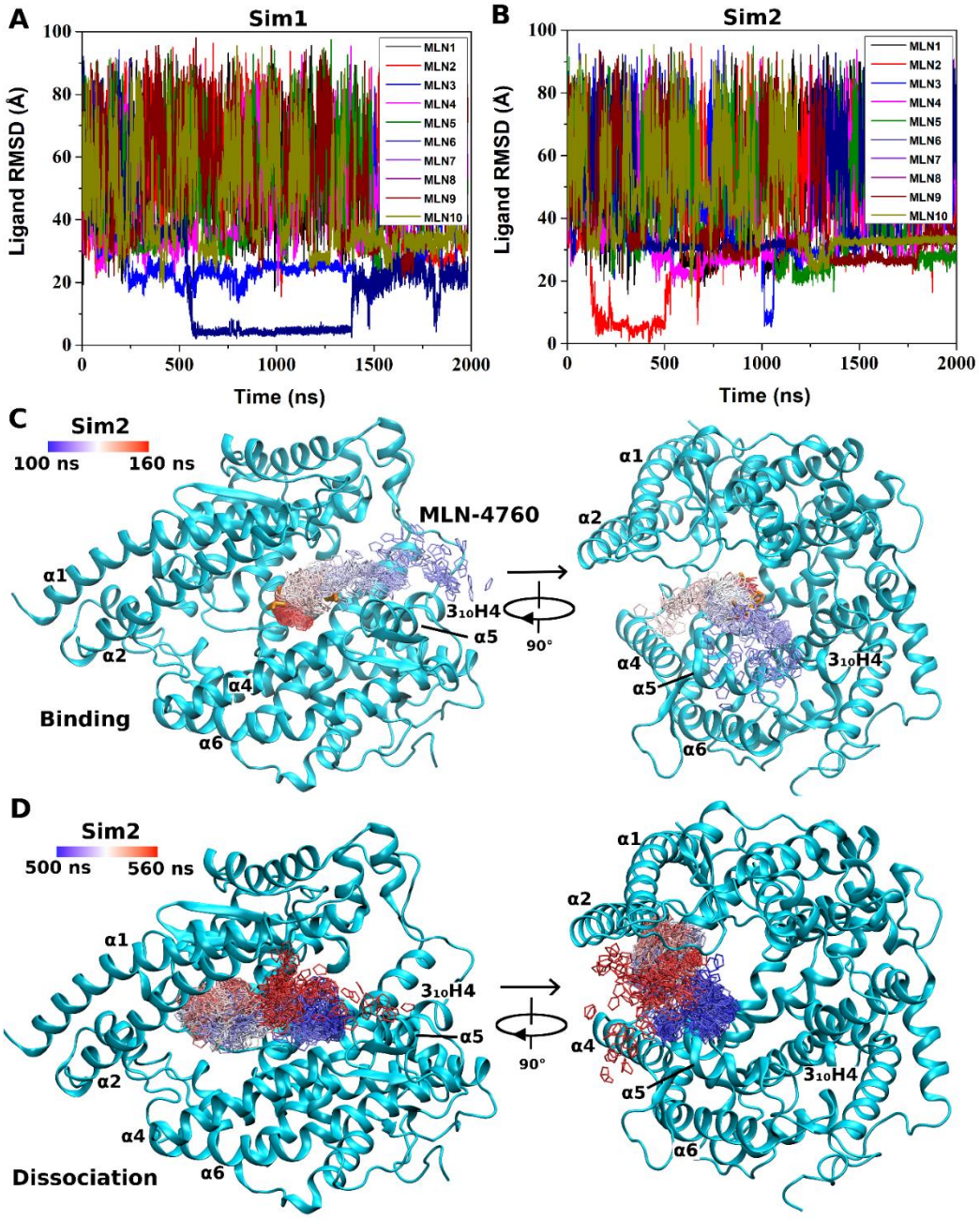

Figure 1. Root-mean-square deviations (RMSDs) of ten MLN-4760 inhibitor molecules relative to the bound X-ray conformation (PDB: 1R4L) are calculated from the 2000 ns (A) "Sim1" and (B) "Sim2" LiGaMD trajectories, in which ligand RMSD reached a minimum of ~0.99 Å. (C) Two views of the ligand binding pathway observed in "Sim2", 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 ligand dissociation pathway observed in "Sim2", 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.

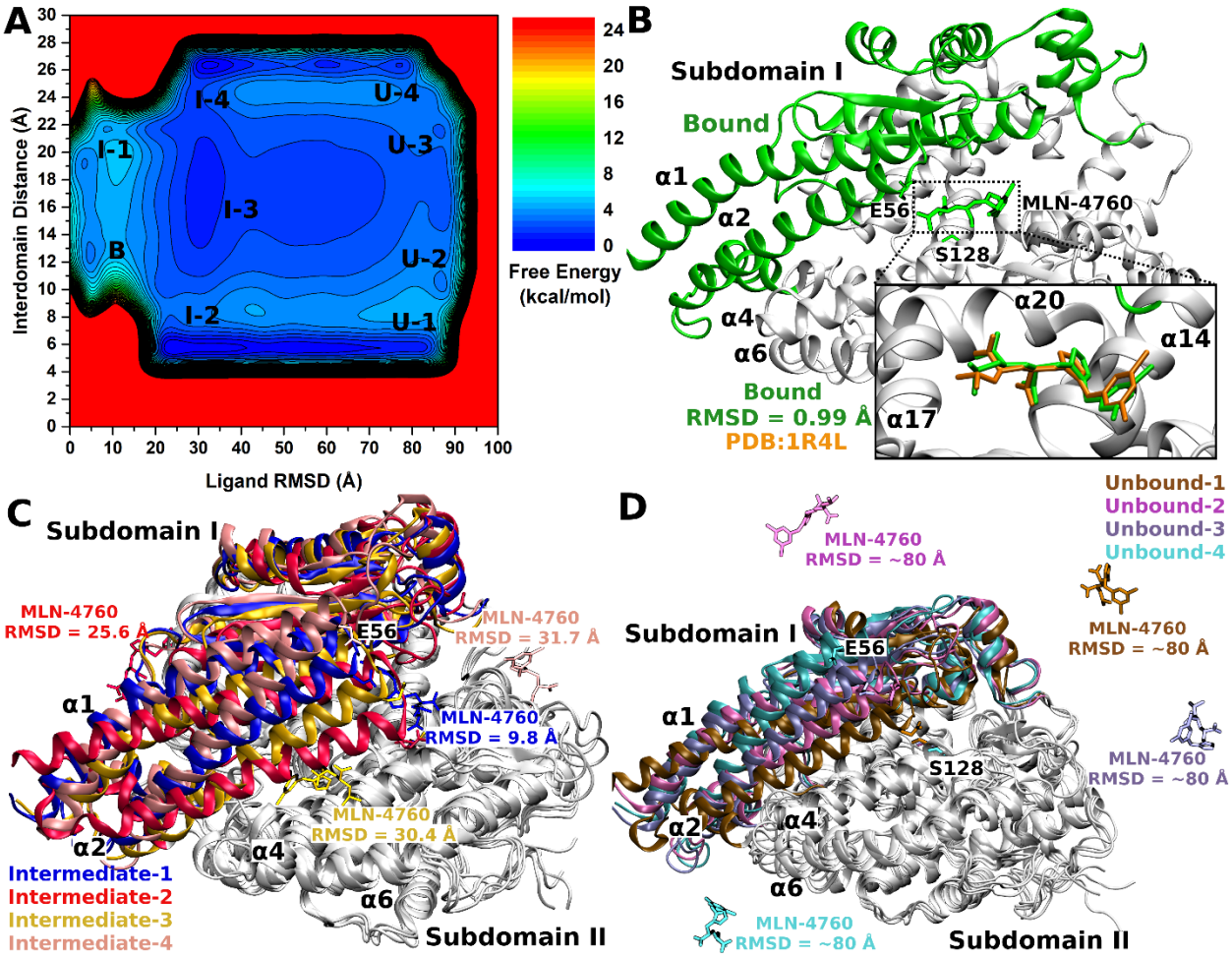

Figure 2. (A) The 2D potential of mean force (PMF) free energy profile of the ligand RMSD and interdomain distance calculated by combining the 10 independent LiGaMD production simulations of human ACE2 receptor. Nine low energy conformational states were identified, including the "Bound (B)", "Intermediate-1 (I-1)", "Intermediate-2 (I-2)", "Intermediate-3 (I-3)", "Intermediate-4 (I-4)", "Unbound-1 (U-1)", "Unbound-2 (U-2)", "Unbound-3 (U-3)" and "Unbound-4 (U-4)". (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), "Intermediate-2 (I-2)" (red), "Intermediate-3 (I-3)" (yellow) and "Intermediate-4 (I-4)" (pink) states. (D) Conformations of the ACE2 receptor in the "Ubound-1 (U-1)" (brown), "Unbound-2 (U-2)" (rose), "Ubound-3 (U-3)" (voilet), and "Unbound-4 (U-4)" (cyan) states. The protein is shown as ribbons and ligand as sticks.

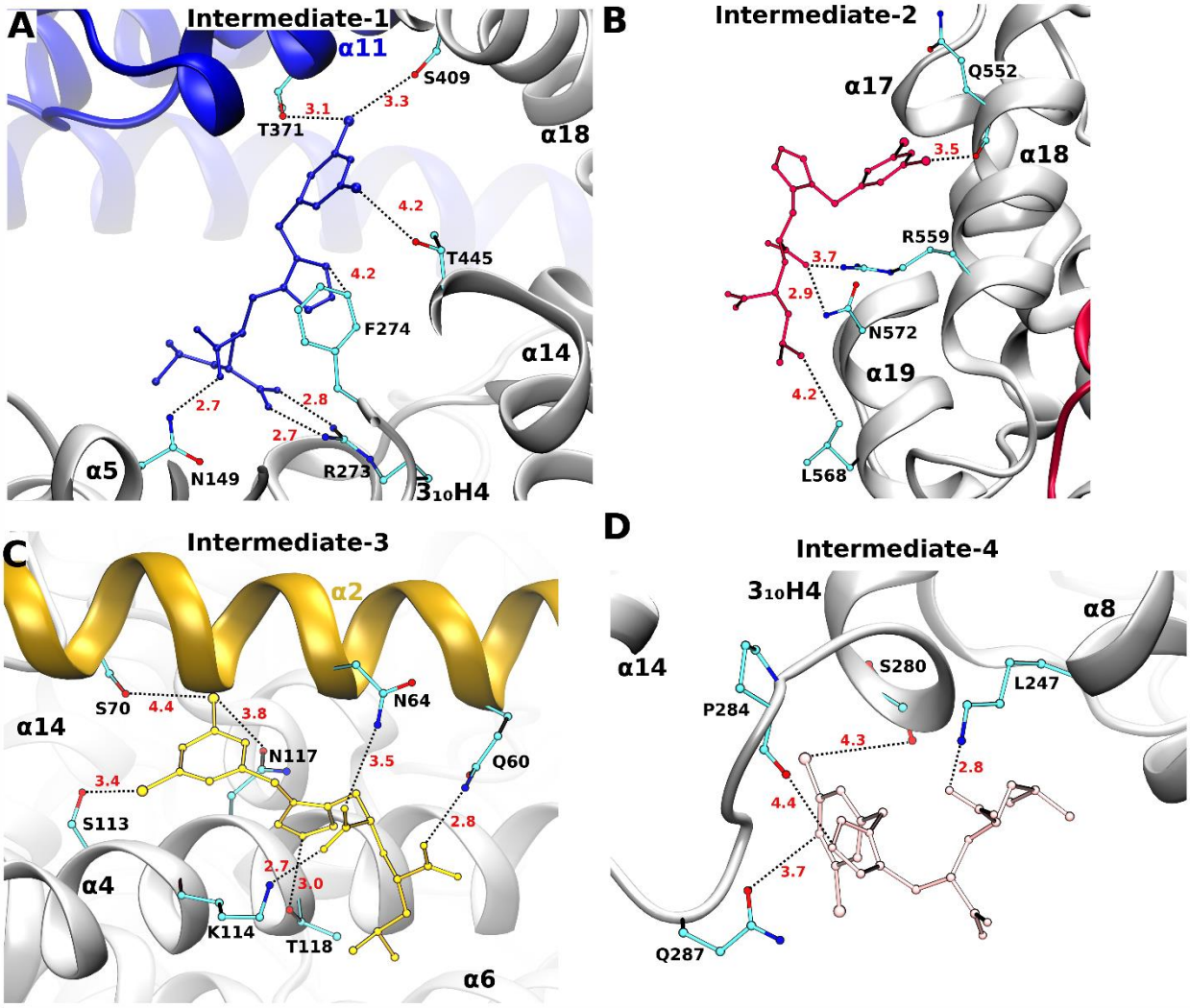

Figure 3. (A) The "Intermediate-1 (I-1)" conformational state of the MLN-4760 ligand (blue balls and sticks) bound to the ACE2 receptor (ribbons). Residues of subdomain I (blue) and subdomain II (white) including N149, R273, F274, T445, T371 and S409 formed interactions with the ligand. (B) The "Intermediate-2 (I-2)" conformational state of MLN-4760 (red) bound to the ACE2 receptor (ribbons). Residues of subdomain II (white) including Q552, R559, N572 and L568 formed interactions with the ligand. (C) "The "Intermediate-3 (I-3)" conformational state of MLN-4760 (yellow) bound to the ACE2 receptor (ribbons). Residues of subdomain I (yellow) and subdomain II (white) including K114, T118, Q60, N64, N117, S113 and S70 formed polar interactions with the ligand. (C) The "Intermediate-4 (I-4)" conformational state of MLN-4760 (pink) bound to the ACE2 receptor (ribbons). Residues of subdomain II (gray) including L247, S280, P284 and Q287 formed polar interactions with the ligand.

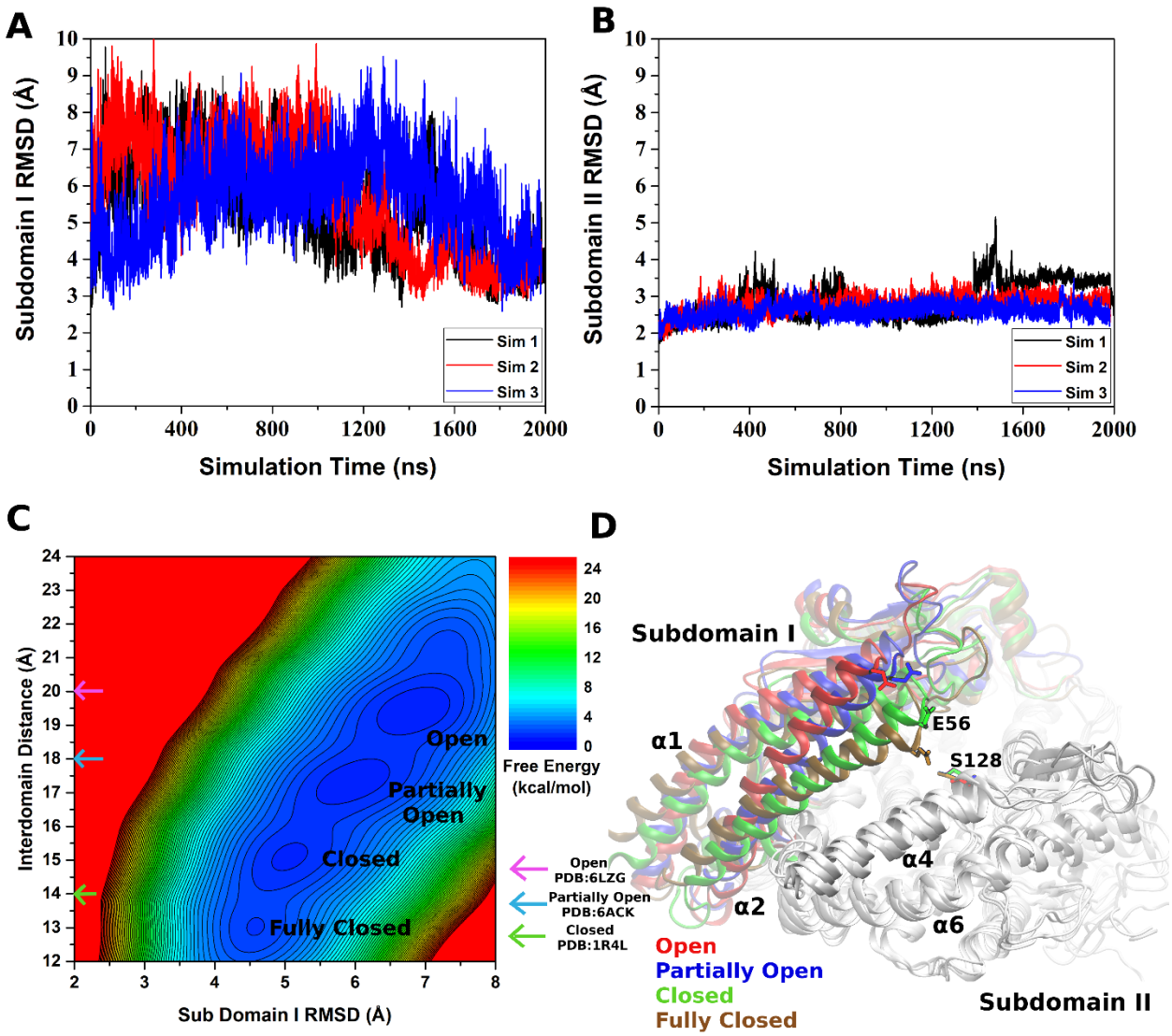

Figure 4. RMSDs of (A) subdomain I and (B) subdomain II of the ACE2 receptor relative to the closed X-ray conformation (PDB:1R4L) are calculated from three independent LiGaMD production simulations. (C) 2D potential of mean force (PMF) of the subdomain I RMSD and interdomain distance calculated by combining the ten LiGaMD simulations. Four low energy conformational states of the receptor are identified in the PMF profile, including the "Fully Closed", "Closed", "Partially Open" and "Open". "Closed", "Partially Open" and "Open" low energy conformational states 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) "Closed" (green), and "Fully Closed" (brown) states in the LiGaMD simulations. Subdomain II is stable and colored in white.

Appendix

Table S1: Summary of LiGaMD simulations performed on the ACE2 receptor in the presence of the MLN-4760 inhibitor.

Simulation	[a]N_{atoms}	Dimension (Å^3)	Simulation (ns)	[b]ΔV_{avg} (kcal/mol)	[c]$\sigma_{\Delta v}$ (kcal/mol)
Sim1	100,449	129 x 93 x 93	2000	126.8	9.76
Sim2	100,449	129 x 93 x 93	2000	127.6	9.62
Sim3	100,449	129 x 93 x 93	2000	127.2	9.14
Sim4	100,449	129 x 93 x 93	1600	129.2	10.3
Sim5	100,449	129 x 93 x 93	1600	129.5	10.5
Sim6	100,449	129 x 93 x 93	1000	129.3	10.4
Sim7	100,449	129 x 93 x 93	1000	129.5	10.5
Sim8	100,449	129 x 93 x 93	700	129.4	10.4
Sim9	100,449	129 x 93 x 93	700	129.4	10.4
Sim10	100,449	129 x 93 x 93	700	129.4	10.4

[a]N_{atoms} is the number of atoms in the simulation system.
[b]ΔV_{avg} and [c]$\sigma_{\Delta V}$ are the average and standard deviation of the LiGaMD boost potential, respectively.

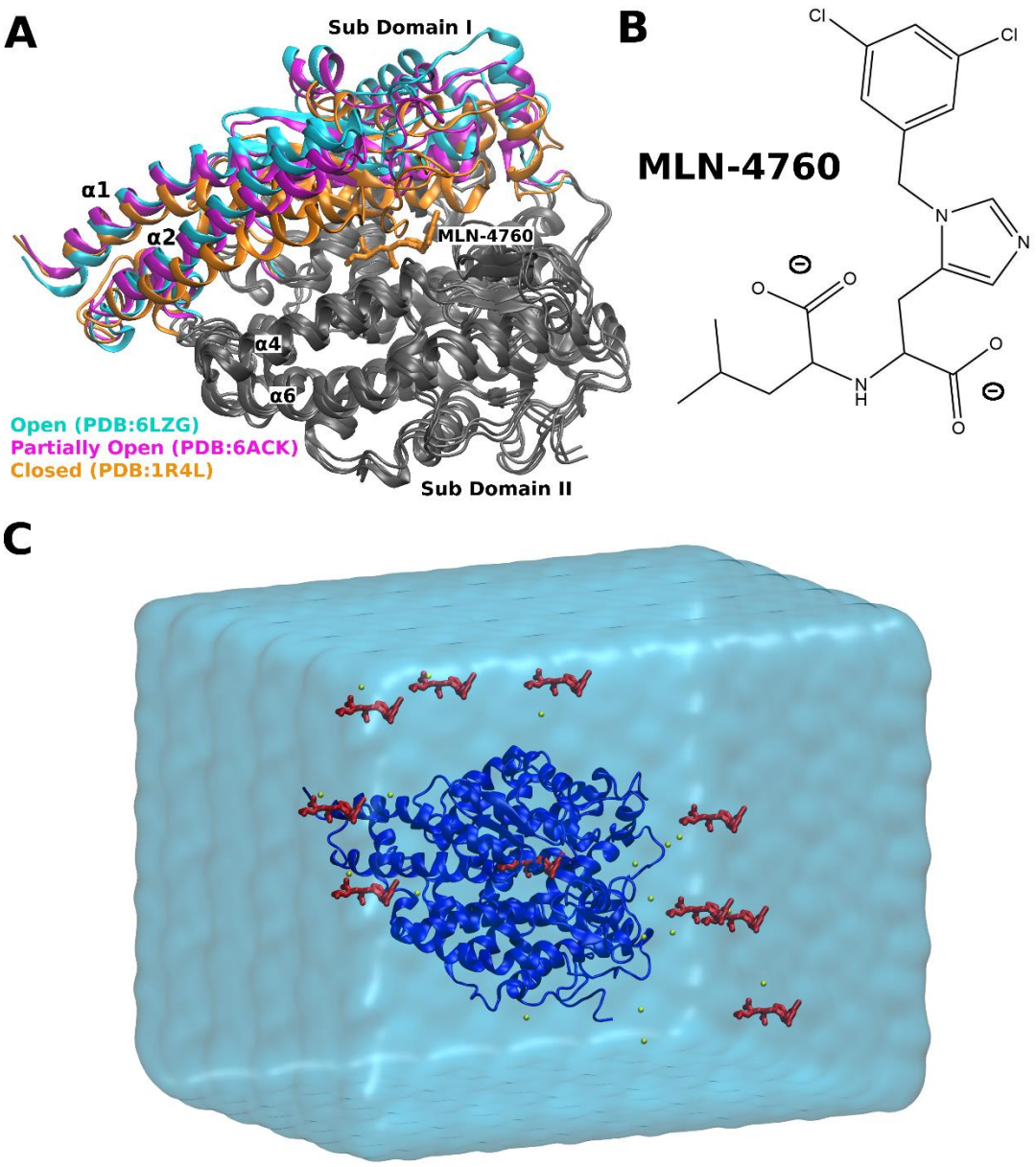

Figure S1: (A) X-ray and cryo-EM strucutres of the ACE2 receptor with subdomain I in the "Open" (cyan, PDB: 6LZG), "Partially Open" (magenta, PDB: 6ACK) and "Closed" (orange, PDB: 1R4L) conformations. Subdomain II is stable and colored in gray. In the "Closed" conformation, the receptor is bound by the MLN-4760 inhibitor. The protein is shown as ribbons and the ligand as sticks (orange). (B) Structure of the MLN-4760 inhibitor molecule. (C) Computational model of the ACE2 receptor (blue ribbons) with 10 MLN-4760 ligand molecules (red sticks) (one in the X-ray bound conformation and another nine placed randomly in the solvent) used in LiGaMD simulations. The system was neutralized by adding counter ions and immersed in a cubic TIP3P water box, which was extended 10 Å from the receptor surface.

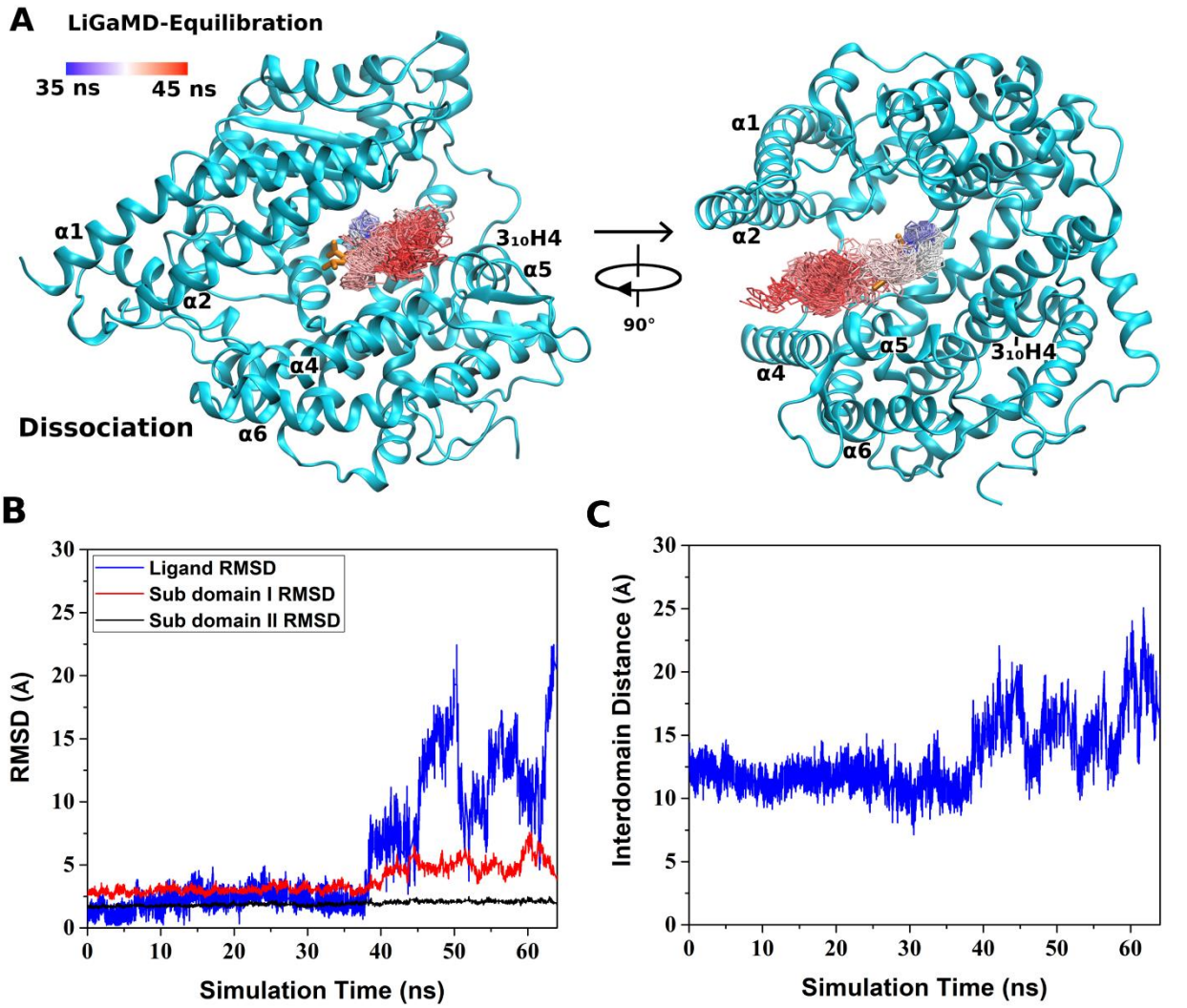

Figure S2: (A) Two views of the ligand dissociation pathway observed in LiGaMD equilibration, for which the center ring of MLN-4760 is represented by lines and colored by simulation time in a blue-white-red (BWR) color scale. (B) Time course of the MLN-4760 ligand RMSD (blue), sub domain I RMSD (red) and sub domain II RMSD (black) calculated from LiGaMD equilibration trajectory, in which the MLN-4760 ligand dissociated from the active site of the receptor. (C) Time course of the interdomain distance (Glu56:CA – Ser128:CA) calculated from LiGaMD equilibration trajectory.

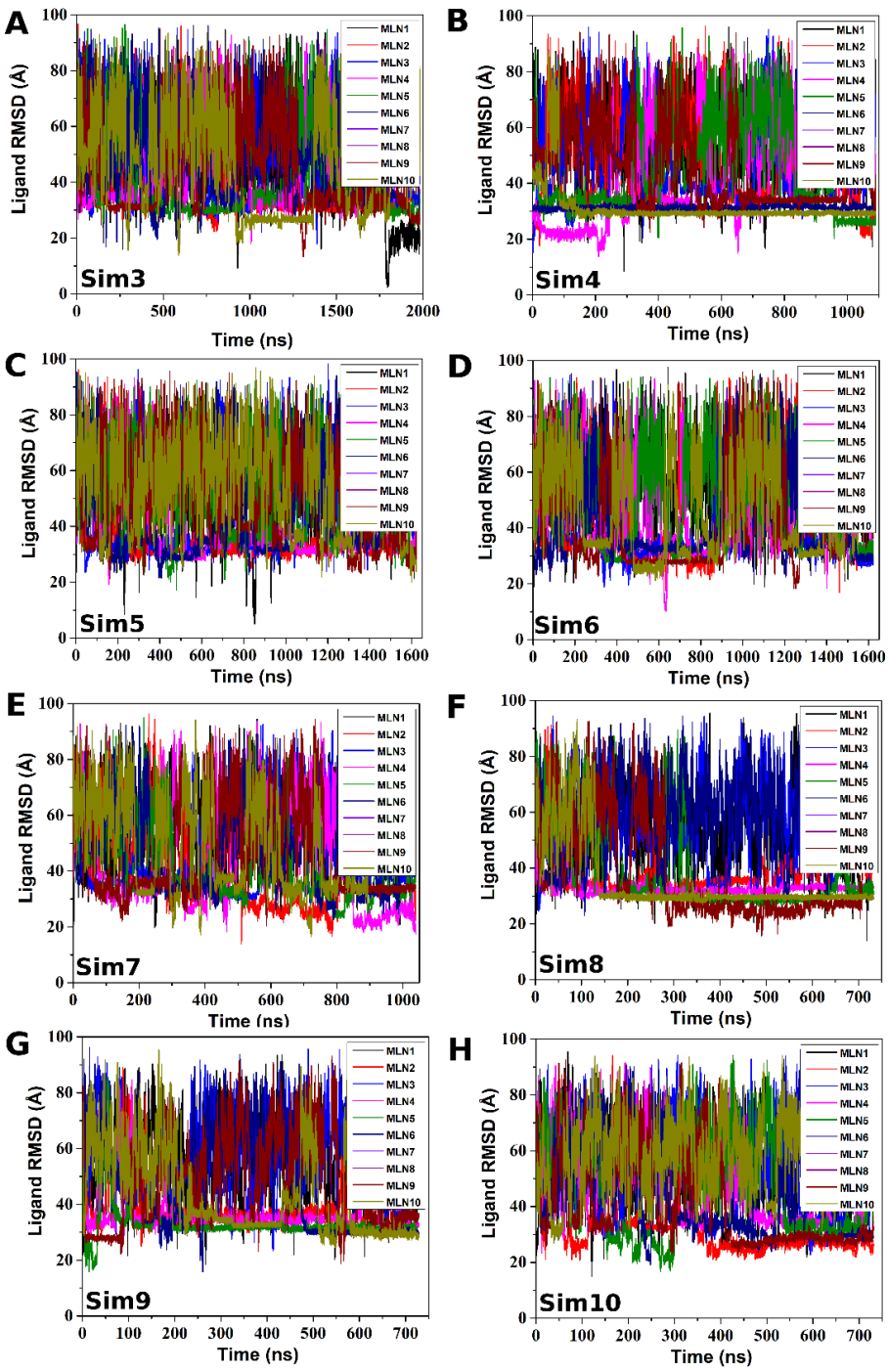

Figure S3: Time courses of the MLN-4760 ligand RMSD relative to the bound X-ray conformation (PDB: 1R4L) calculated from (A) "Sim3", (B) "Sim4", (C) "Sim5", (D) "Sim6", (E) "Sim7", (F) "Sim8", (G) "Sim9" and (H) "Sim10" LiGaMD trajectories of varying lengths ranging ~700-1600 ns, in which the MLN-4760 ligand did not bind to the active site of the receptor.

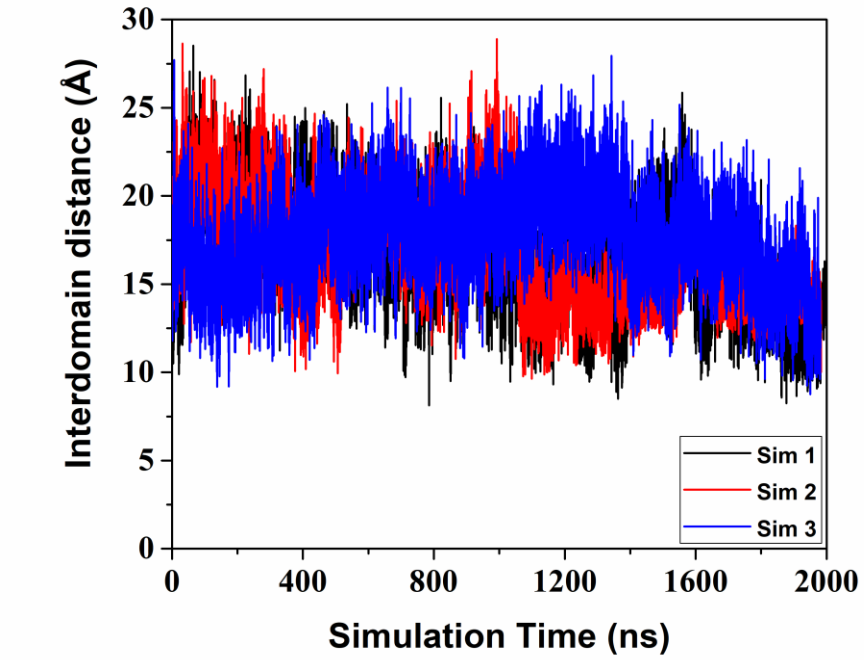

Figure S4: Time course of the interdomain distance (Glu56:CA – Ser128:CA) calculated from three independent LiGaMD trajectories including “Sim 1” (black), “Sim 2” (red) and “Sim 3” (blue).

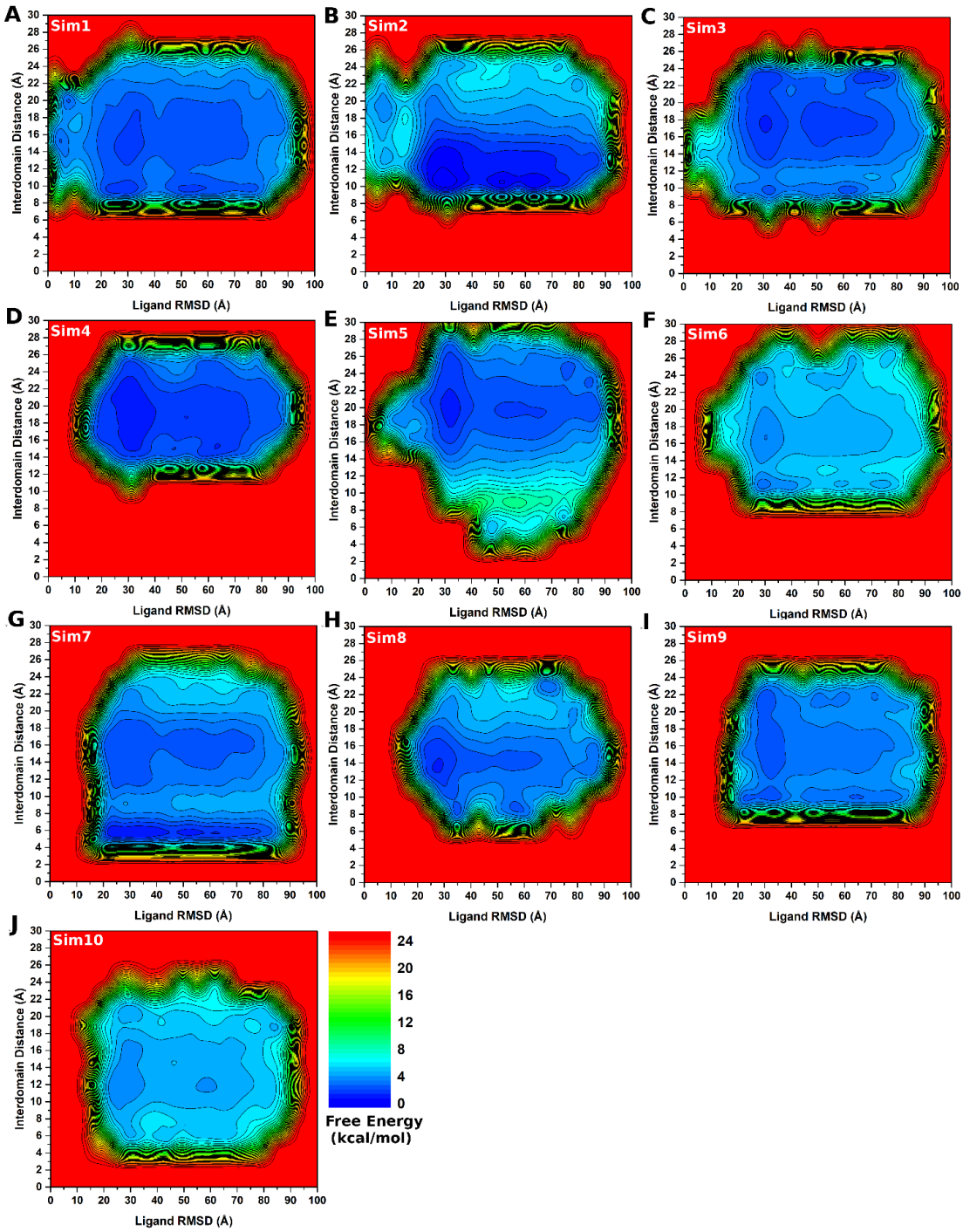

Figure S5: The 2D potential of mean force (PMF) free energy profile of the ligand RMSD and interdomain distance calculated for individual (A) "Sim1", (B) "Sim2", (C) "Sim3" (D) "Sim4", (E) "Sim5", (F) "Sim6", (G) "Sim7", (H) "Sim8", (I) "Sim9" and (J) "Sim10" LiGaMD productio simulations of human ACE2 receptor.

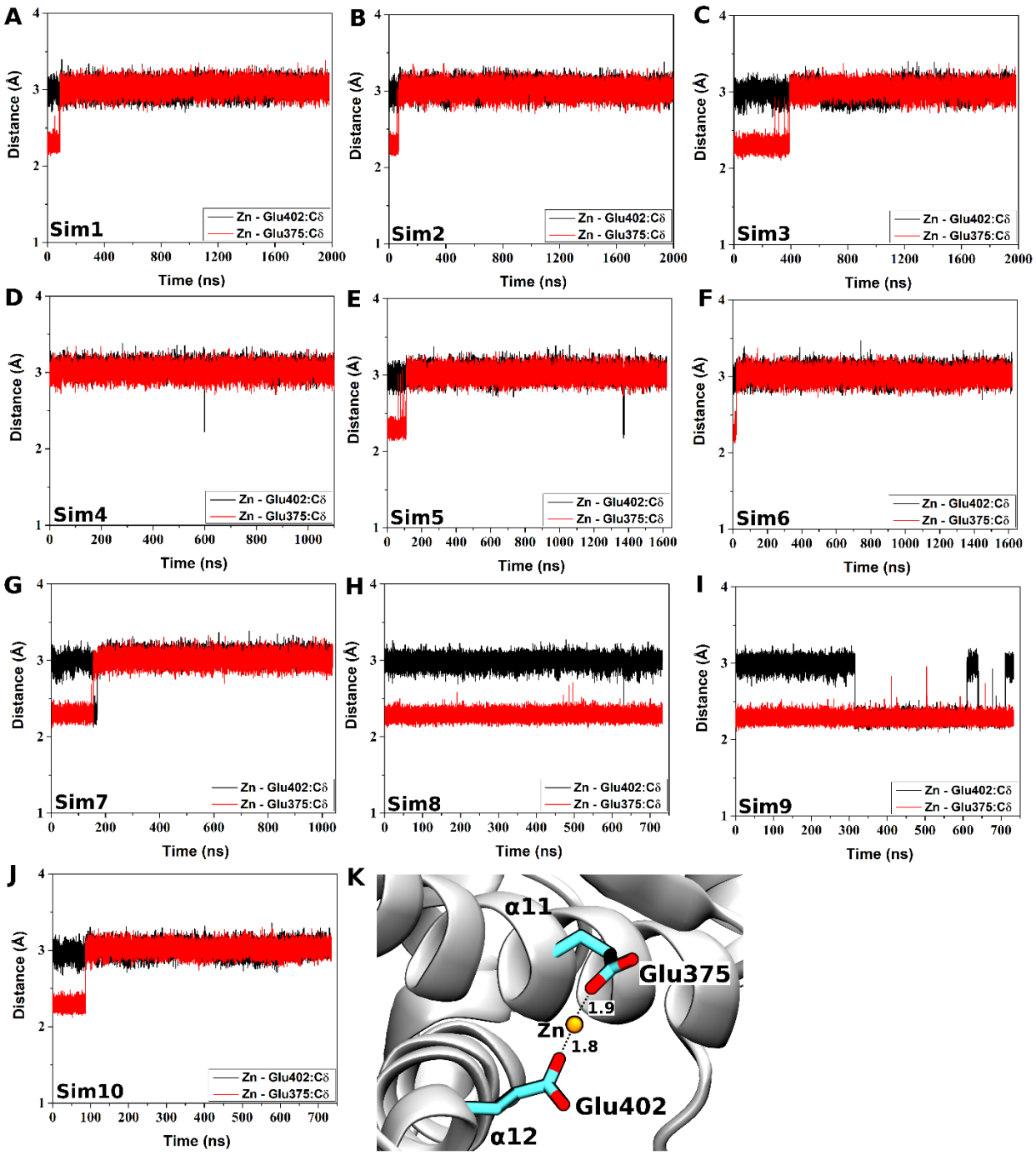

Figure S6: Time course of the distance between Zn^{2+} and Glu402:Cδ (black) and Glu375:Cδ (red) calculated from (A) "Sim1", (B) "Sim2", (C) "Sim3", (D) "Sim4", (E) "Sim5", (F) "Sim6", (G) "Sim7", (H) "Sim8", (I) "Sim9", (J) "Sim10" LiGaMD trajectories of varying lengths ranging ~700-2000 ns. (K) Zinc (orange ball) near the active site of the ACE2 receptor stabilized by ionic interactions with the Glu375 and Glu402 residues (cyan sticks). The receptor is shown in white ribbons.

Conclusions

Membrane proteins are essential to human cell physiology and are medically important. However, little is often known about the structural dynamic mechanisms of membrane proteins, which has impeded drug design against many deadly human diseases including Alzheimer's disease, heart failure, neuropathic pain, and SARS-COV-1/2. Here, in my dissertation, I have implemented GaMD, LiGaMD and Pep-GaMD enhanced sampling methods in combination with complementary advanced computational techniques including molecular docking, homology modeling and free energy calculations for accelerated simulations and drug design of different membrane proteins including γ-secretase, GPCR and human ACE2 receptor.

Using GaMD simulations, as mentioned in Chapter 1, I was able to capture spontaneous activation of the γ-secretase enzyme for the ε cleavage of the wildtype and FAD mutant APP substrates (*ACS Central Science*, 2020)[8]. Detailed analysis of the GaMD simulations allowed us to identify distinct low-energy conformational states of γ-secretase, different secondary structures of the wildtype and mutant APP substrate, and important active-site sub-pockets for catalytic function of the enzyme. In Chapter 2, Pep-GaMD simulations captured remarkable structural rearrangements of both the γ-secretase and substrate, in which hydrogen-bonded catalytic aspartates and water became poised for tripeptide trimming of Aβ49 to Aβ46 (*Journal of American Chemical Society, 2022*)[70]. These structural changes required a positively charged N-terminus of endoproteolytic coproduct AICD, which could dissociate during conformational rearrangements of the protease and Aβ49. The simulation findings were highly consistent with experimental analyses of APP proteolytic products using mass spectrometry and western blotting. Detailed mechanistic understanding of the proteolytic processes of γ-secretase is expected to greatly facilitate rational drug design of this critical enzyme.

In Chapter 3, I was able to discover that interactions between GPCR and lipid membrane depend on the receptor activation state (*Journal of Computational Chemistry*, 2020)[42]. GPCRs are the largest family of human membrane proteins and represent the primary targets for ~1/3 of currently marketed drugs. I performed all-atom GaMD simulations on a membrane-embedded A_1AR in the inactive antagonist bound and active agonist-G protein-bound states. The GaMD simulations have revealed distinct patterns of GPCR-membrane interactions of the inactive and active GPCR. The dynamics of GPCR and lipids are strongly coupled. This study provides important mechanistic insights into the effects of membrane in GPCR function. In Chapter 4, I optimized a computational protocol termed "ensemble docking" for allosteric modulator design of GPCRs, in which GaMD simulations and flexible molecular docking are combined to effectively account for the receptor flexibility (*BBA-General Subjects*, 2020)[152]. I demonstrated improved performance of this protocol in retrospective docking of known allosteric modulators in the A_1AR. This optimized protocol will facilitate future drug design of other GPCRs and receptors.

In Chapter 5, I performed LiGaMD simulations on human ACE2, the receptor hijacked by SARS-CoV-2 (COVID-19 corona virus) for viral infection (*Journal of Physical Chemistry Letters, 2021*)[220]. The simulations, for the first time, revealed both dissociation and binding of an inhibitor drug in human ACE2 receptor. The receptor with remarkable flexibility changes its structure into Closed conformation upon inhibitor binding. Overall, the receptor could sample distinct Closed, Partially Open and Open conformational states, being consistent with previous experimental structures. Ligand binding could bias the receptor conformational ensemble towards the Closed state, suggesting a conformational selection mechanism. The LiGaMD simulations thus helped us to understand the functional mechanism of human ACE2, with

implications in developing treatments against the SARS-CoV-2 and other related human diseases.

Outlook

Membrane proteins are common but critical biological components. Membrane proteins remain one of the most studied biomolecules in the pharmaceutical and biomedical industry as they represent more than 60% of drug targets. Both experimentally and computationally, studying the structure and functional mechanism of membrane proteins remains a challenge. During my Ph.D., I tried to bridge gaps of knowledge in understanding of structural dynamics and functional mechanisms of a few critical membrane proteins. In the future, there are several other interesting and important questions that need to be addressed in follow-up studies.

With the knowledge of molecular mechanisms of γ-secretase activation for the ε and ζ cleavage of the APP substrate and the more powerful Pep-GaMD sampling method, the next obvious question in the field could be "What is the mechanism of the next γ cleavage (Aβ46→Aβ43) and subsequent cleavage steps?". Likewise, new cryo-EM structures bound to small molecule inhibitors and γ-secretase modulators (GSMs) were released recently. They provided new insights into structural recognition of the enzyme by small molecules. However, the dynamic mechanisms of inhibition and modulation by these small molecules remain elusive. Current AD treatments ease symptoms, but none has been clearly demonstrated to slow or halt disease progression. Therefore, designing new GSM drugs is imperative to the field. One good strategy is using virtual screening and retrospective docking protocol that combines MD simulations and molecular docking calculations.

GPCR drug design is a very active research area. A new paradigm that is relevant and exciting is the use of allosteric drug leads in combination with endogenous ligands rather than just using the orthosteric ligand. This provides potential selectivity and eliminates drug side effects. That being said, using the optimized ensemble docking protocol followed by virtual

screening to design new allosteric drugs of the A_1AR and other GPCRs would be a relevant direction. In fact, this has been applied in our ongoing research and we already have some preliminary results that is not included in this dissertation. Similarly, with the first cryo-EM structure of allosteric modulator bound A_1AR[245], one can study the drug binding and unbinding mechanisms as well as explore related GPCR-G-protein interactions. With the advancement of new methods like LiGaMD, one can compute binding affinities of allosteric modulators at different subtypes of adenosine receptors (e. g. $A_{2A}AR$) to understand the mechanism of GPCR ligand selectivity.

Another important membrane protein enzyme I studied during my Ph.D. is ACE2 receptor, the functional receptor for SARS-COV-1/2. ACE2 is a protease that is inhibited by drug binding to the active site of the enzyme. The SARS-COV-2 receptor binding domain (RBD) binds to a different site of the receptor as compared to the active site. In my dissertation, I was able to elucidate the inhibitor drug binding pathways to ACE2 and the receptor dynamics during the process using LiGaMD simulations. However, I was always curious about the effects of active site inhibitor to the SARS-COV-2 (COVID-19) binding and functioning. This is because, there has been evidence of allosteric network between the active site and SARS-COV-1 RBD binding site. Hence, new drug development is important, especially those that can bind to the receptor and inhibit the coronavirus binding. Virtual screening in combination with retrospective docking is a good approach for computer aided drug design of the human ACE2 receptor. Another exciting direction in the project is using the complete structure of ACE2 for accelerated simulations and drug design. ACE2 is a transmembrane receptor with extracellular carboxy peptidase domain. During my Ph.D., I have only worked with the globular side of the receptor present in the extracellular environment and that has the catalytic site. The

transmembrane domain that is embedded in the lipid bilayer would provide stability to the receptor. Studying the ligand binding thermodynamics and kinetics, and protein structural dynamics with the complete structure of ACE2 receptor would be important in the field.

References

(1) Wallin, E.; Heijne, G. V. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science* **1998,** *7* (4), 1029.

(2) Terstappen, G. C.; Reggiani, A. In silico research in drug discovery. *Trends in pharmacological sciences* **2001,** *22* (1), 23.

(3) Arinaminpathy, Y.; Khurana, E.; Engelman, D. M.; Gerstein, M. B. Computational analysis of membrane proteins: the largest class of drug targets. *Drug discovery today* **2009,** *14* (23-24), 1130.

(4) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **2015,** *11* (8), 3584.

(5) Bhattarai, A.; Miao, Y. Gaussian Accelerated Molecular Dynamics for Elucidation of Drug Pathways. *Expert Opinion on Drug Discovery* **2018,** *13* (11), 1055.

(6) Pang, Y. T.; Miao, Y.; Wang, Y.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics in NAMD. *J Chem Theory Comput* **2017,** *13* (1), 9.

(7) Miao, Y.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Theory, Implementation and Applications. *Annual Reports in Computational Chemistry* **2017,** *13*, 231.

(8) Bhattarai, A.; Devkota, S.; Bhattarai, S.; Wolfe, M. S.; Miao, Y. Mechanisms of gamma-Secretase Activation and Substrate Processing. *ACS Cent Sci* **2020,** *6* (6), 969.

(9) Selkoe, D. J.; Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO molecular medicine* **2016,** *8* (6), 595.

(10) Panza, F.; Lozupone, M.; Logroscino, G.; Imbimbo, B. P. A critical appraisal of amyloid-beta-targeting therapies for Alzheimer disease. *Nat Rev Neurol* **2019,** *15* (2), 73.

(11) Molinuevo, J. L.; Ayton, S.; Batrla, R.; Bednar, M. M.; Bittner, T.; Cummings, J.; Fagan, A. M.; Hampel, H.; Mielke, M. M.; Mikulskis, A.et al. Current state of Alzheimer's fluid biomarkers. *Acta Neuropathol* **2018,** *136* (6), 821.

(12) Haass, C.; Kaether, C.; Thinakaran, G.; Sisodia, S. Trafficking and proteolytic processing of APP. *Cold Spring Harb Perspect Med* **2012,** *2* (5), a006270.

(13) Wolfe, M. S. Structure and Function of the gamma-Secretase Complex. *Biochemistry* **2019,** *58* (27), 2953.

(14) Steiner, H.; Fukumori, A.; Tagami, S.; Okochi, M. Making the final cut: pathogenic amyloid-beta peptide generation by gamma-secretase. *Cell Stress* **2018,** *2* (11), 292.

(15) Gu, Y.; Misonou, H.; Sato, T.; Dohmae, N.; Takio, K.; Ihara, Y. Distinct intramembrane cleavage of the beta-amyloid precursor protein family resembling gamma-secretase-like cleavage of Notch. *The Journal of biological chemistry* **2001,** *276* (38), 35235.

(16) Qi-Takahara, Y.; Morishima-Kawashima, M.; Tanimura, Y.; Dolios, G.; Hirotani, N.; Horikoshi, Y.; Kametani, F.; Maeda, M.; Saido, T. C.; Wang, R.et al. Longer forms of amyloid beta protein: implications for the mechanism of intramembrane cleavage by gamma-secretase. *J Neurosci* **2005,** *25* (2), 436.

(17) Takami, M.; Nagashima, Y.; Sano, Y.; Ishihara, S.; Morishima-Kawashima, M.; Funamoto, S.; Ihara, Y. gamma-Secretase: successive tripeptide and tetrapeptide release from the transmembrane domain of beta-carboxyl terminal fragment. *J Neurosci* **2009,** *29* (41), 13042.

(18) Bolduc, D. M.; Montagna, D. R.; Seghers, M. C.; Wolfe, M. S.; Selkoe, D. J. The amyloid-beta forming tripeptide cleavage mechanism of gamma-secretase. *eLife* **2016,** *5*, pii: e17578.

(19) Sato, T.; Dohmae, N.; Qi, Y.; Kakuda, N.; Misonou, H.; Mitsumori, R.; Maruyama, H.; Koo, E. H.; Haass, C.; Takio, K.et al. Potential link between amyloid beta-protein 42 and C-terminal

fragment gamma 49-99 of beta-amyloid precursor protein. *The Journal of biological chemistry* **2003,** *278* (27), 24294.
(20) Zhou, R.; Yang, G.; Guo, X.; Zhou, Q.; Lei, J.; Shi, Y. Recognition of the amyloid precursor protein by human gamma-secretase. *Science (New York, N.Y.)* **2019,** *363* (6428), eaaw0930
(21) Yang, G. H.; Zhou, R.; Zhou, Q.; Guo, X. F.; Yan, C. Y.; Ke, M.; Lei, J. L.; Shi, Y. G. Structural basis of Notch recognition by human gamma-secretase. *Nature* **2019,** *565* (7738), 192.
(22) Zhou, R.; Yang, G. H.; Guo, X. F.; Zhou, Q.; Lei, J. L.; Shi, Y. G. Recognition of the amyloid precursor protein by human gamma-secretase. *Science (New York, N.Y.)* **2019,** *363* (6428), 708.
(23) Hitzenberger, M.; Zacharias, M. Structural Modeling of gamma-Secretase A beta(n) Complex Formation and Substrate Processing. *Acs Chem Neurosci* **2019,** *10* (3), 1826.
(24) Hitzenberger, M.; Zacharias, M. gamma-Secretase Studied by Atomistic Molecular Dynamics Simulations: Global Dynamics, Enzyme Activation, Water Distribution and Lipid Binding. *Front Chem* **2019,** *6*, 640.
(25) Aguayo-Ortiz, R.; Chavez-Garcia, C.; Straub, J. E.; Dominguez, L. Characterizing the structural ensemble of gamma-secretase using a multiscale molecular dynamics approach. *Chem Sci* **2017,** *8* (8), 5576.
(26) Somavarapu, A. K.; Kepp, K. P. Membrane dynamics of γ-secretase provides a molecular basis for β-amyloid binding and processing. *Acs Chem Neurosci* **2017,** *8* (11), 2424.
(27) Lee, J. Y.; Feng, Z.; Xie, X. Q.; Bahar, I. Allosteric Modulation of Intact gamma-Secretase Structural Dynamics. *Biophys J* **2017,** *113* (12), 2634.
(28) Götz, A.; Mylonas, N.; Högel, P.; Silber, M.; Heinel, H.; Menig, S.; Vogel, A.; Feyrer, H.; Huster, D.; Luy, B. Modulating hinge flexibility in the APP transmembrane domain alters γ-secretase cleavage. *Biophysical journal* **2019,** *116* (11), 2103.
(29) Dehury, B.; Tang, N.; Kepp, K. P. Molecular dynamics of C99-bound γ-secretase reveal two binding modes with distinct compactness, stability, and active-site retention: implications for Aβ production. *Biochemical Journal* **2019,** *476* (7), 1173.
(30) Götz, A.; Högel, P.; Silber, M.; Chaitoglou, I.; Luy, B.; Muhle-Goll, C.; Scharnagl, C.; Langosch, D. Increased H-Bond Stability Relates to Altered ε-Cleavage Efficiency and Aβ Levels in the I45T Familial Alzheimer's Disease Mutant of APP. *Scientific reports* **2019,** *9* (1), 1.
(31) Gertsik, N.; Ende, C. W. A.; Geoghegan, K. F.; Nguyen, C.; Mukherjee, P.; Mente, S.; Seneviratne, U.; Johnson, D. S.; Li, Y. M. Mapping the Binding Site of BMS-708163 on gamma-Secretase with Cleavable Photoprobes. *Cell Chem Biol* **2017,** *24* (1), 3.
(32) Petit, D.; Hitzenberger, M.; Lismont, S.; Zoltowska, K. M.; Ryan, N. S.; Mercken, M.; Bischoff, F.; Zacharias, M.; Chávez-Gutiérrez, L. Extracellular interface between APP and Nicastrin regulates Aβ length and response to γ-secretase modulators. *The EMBO journal* **2019,** *38* (12), e101494.
(33) Hitzenberger, M.; Zacharias, M. Uncovering the Binding Mode of γ-Secretase Inhibitors. *Acs Chem Neurosci* **2019,** *10* (8), 3398.
(34) Bolduc, D. M.; Montagna, D. R.; Gu, Y.; Selkoe, D. J.; Wolfe, M. S. Nicastrin functions to sterically hinder gamma-secretase-substrate interactions driven by substrate transmembrane domain. *Proceedings of the National Academy of Sciences of the United States of America* **2016,** *113* (5), E509.
(35) Miao, Y. Acceleration of Biomolecular Kinetics in Gaussian Accelerated Molecular Dynamics. *Journal of Chemical Physics* **2018,** *149* (7), 072308.
(36) Miao, Y.; McCammon, J. A. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Molecular simulation* **2016,** *42* (13), 1046.

(37) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *Journal of Chemical Physics* **2004,** *120* (24), 11919.
(38) Voter, A. F. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters* **1997,** *78* (20), 3908.
(39) Shen, T. Y.; Hamelberg, D. A statistical analysis of the precision of reweighting-based simulations. *Journal of Chemical Physics* **2008,** *129* (3), 034103.
(40) Wang, J.; Miao, Y. Mechanistic Insights into Specific G Protein Interactions with Adenosine Receptors. *The Journal of Physical Chemistry B* **2019,** *123* (30), 6462.
(41) Miao, Y.; McCammon, J. A. Mechanism of the G-Protein Mimetic Nanobody Binding to a Muscarinic G-Protein-Coupled Receptor. *Proceedings of the National Academy of Sciences of the United States of America* **2018,** *115* (12), 3036.
(42) Bhattarai, A.; Wang, J.; Miao, Y. G-Protein-Coupled Receptor-Membrane Interactions Depend on the Receptor Activation State. *J Comput Chem* **2019,** *41*, 460.
(43) East, K. W.; Newton, J. C.; Morzan, U. N.; Narkhede, Y. B.; Acharya, A.; Skeens, E.; Jogl, G.; Batista, V. S.; Palermo, G.; Lisi, G. P. Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J Am Chem Soc* **2020,** *142* (3), 1348.
(44) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Central Science* **2019,** *5* (4), 651.
(45) Kamp, F.; Winkler, E.; Trambauer, J.; Ebke, A.; Fluhrer, R.; Steiner, H. Intramembrane proteolysis of β-amyloid precursor protein by γ-secretase is an unusually slow process. *Biophysical journal* **2015,** *108* (5), 1229.
(46) Langosch, D.; Scharnagl, C.; Steiner, H.; Lemberg, M. Understanding intramembrane proteolysis: from protein dynamics to reaction kinetics. *Trends in biochemical sciences* **2015,** *40* (6), 318.
(47) Bai, X.-c.; Rajendra, E.; Yang, G.; Shi, Y.; Scheres, S. H. Sampling the conformational space of the catalytic subunit of human γ-secretase. *eLife* **2015,** *4*, e11182.
(48) Esler, W. P.; Das, C.; Wolfe, M. S. Probing pockets S2-S4' of the gamma-secretase active site with (hydroxyethyl)urea peptidomimetics. *Bioorg Med Chem Lett* **2004,** *14* (8), 1935.
(49) Cai, T.; Yonaga, M.; Tomita, T. Activation of γ-secretase trimming activity by topological changes of transmembrane domain 1 of presenilin 1. *Journal of Neuroscience* **2017,** *37* (50), 12272.
(50) Takagi-Niidome, S.; Sasaki, T.; Osawa, S.; Sato, T.; Morishima, K.; Cai, T.; Iwatsubo, T.; Tomita, T. Cooperative roles of hydrophilic loop 1 and the C-terminus of presenilin 1 in the substrate-gating mechanism of γ-secretase. *Journal of Neuroscience* **2015,** *35* (6), 2646.
(51) Sato, C.; Takagi, S.; Tomita, T.; Iwatsubo, T. The C-terminal PAL motif and transmembrane domain 9 of presenilin 1 are involved in the formation of the catalytic pore of the γ-secretase. *Journal of Neuroscience* **2008,** *28* (24), 6264.
(52) Wolfe, M. S. Intramembrane proteolysis. *Chem Rev* **2009,** *109* (4), 1599.
(53) Wu, Z.; Yan, N.; Feng, L.; Oberstein, A.; Yan, H.; Baker, R. P.; Gu, L.; Jeffrey, P. D.; Urban, S.; Shi, Y. Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* **2006,** *13* (12), 1084.
(54) Feng, L.; Yan, H.; Wu, Z.; Yan, N.; Wang, Z.; Jeffrey, P. D.; Shi, Y. Structure of a site-2 protease family intramembrane metalloprotease. *Science (New York, N.Y.)* **2007,** *318* (5856), 1608.
(55) Li, X.; Dang, S.; Yan, C.; Gong, X.; Wang, J.; Shi, Y. Structure of a presenilin family intramembrane aspartate protease. *Nature* **2013,** *493* (7430), 56.

(56) Wang, Y.; Zhang, Y.; Ha, Y. Crystal structure of a rhomboid family intramembrane protease. *Nature* **2006,** *444* (7116), 179.
(57) Cho, S.; Baker, R. P.; Ji, M.; Urban, S. Ten catalytic snapshots of rhomboid intramembrane proteolysis from gate opening to peptide release. *Nat Struct Mol Biol* **2019,** *26* (10), 910.
(58) Yang, G.; Zhou, R.; Zhou, Q.; Guo, X.; Yan, C.; Ke, M.; Lei, J.; Shi, Y. Structural basis of Notch recognition by human γ-secretase. *Nature* **2019,** *565* (7738), 192.
(59) Miao, Y. Acceleration of biomolecular kinetics in Gaussian accelerated molecular dynamics. *The Journal of chemical physics* **2018,** *149* (7), 072308.
(60) Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Gohlke, H. Amber 14, University of California, San Francisco. **2014**.
(61) Kobayashi, C.; Jung, J.; Matsunaga, Y.; Mori, T.; Ando, T.; Tamura, K.; Kamiya, M.; Sugita, Y. GENESIS website. *GENESIS 1*, 0.
(62) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; McCammon, J. A. Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation* **2014,** *10* (7), 2677.
(63) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *Journal of Chemical Physics* **2007,** *127* (15), 155102.
(64) Roux, B. The Calculation of the Potential of Mean Force Using Computer-Simulations. *Comput Phys Commun* **1995,** *91* (1-3), 275.
(65) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **2018,** *46* (W1), W296.
(66) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph Model* **1996,** *14* (1), 33.
(67) Vanommeslaeghe, K.; MacKerell, A. D., Jr. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochimica et biophysica acta* **2014**, DOI:10.1016/j.bbagen.2014.08.004 10.1016/j.bbagen.2014.08.004.
(68) Kappel, K.; Miao, Y.; McCammon, J. A. Accelerated Molecular Dynamics Simulations of Ligand Binding to a Muscarinic G-protein Coupled Receptor. *Quarterly Reviews of Biophysics* **2015,** *48* (04), 479.
(69) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013,** *9* (7), 3084.
(70) Bhattarai, A.; Devkota, S.; Do, H. N.; Wang, J.; Bhattarai, S.; Wolfe, M. S.; Miao, Y. Mechanism of Tripeptide Trimming of Amyloid β-Peptide 49 by γ-Secretase. *Journal of the American Chemical Society* **2022,** *144* (14), 6215.
(71) Weller, J.; Budson, A. Current understanding of Alzheimer's disease diagnosis and treatment. *F1000Research* **2018,** *7*, F1000 Faculty Rev.
(72) Funamoto, S.; Tagami, S.; Okochi, M.; Morishima-Kawashima, M. Seminars in cell & developmental biology, 2020; p 64.
(73) Hemming, M. L.; Elias, J. E.; Gygi, S. P.; Selkoe, D. J. Proteomic Profiling of gamma-Secretase Substrates and Mapping of Substrate Requirements. *Plos Biol* **2008,** *6* (10), 2314.
(74) Guner, G.; Lichtenthaler, S. F. The substrate repertoire of gamma-secretase/presenilin. *Semin Cell Dev Biol* **2020,** *105*, 27.
(75) Gu, Y.; Misonou, H.; Sato, T.; Dohmae, N.; Takio, K.; Ihara, Y. Distinct intramembrane cleavage of the β-amyloid precursor protein family resembling γ-secretase-like cleavage of Notch. *Journal of Biological Chemistry* **2001,** *276* (38), 35235.

(76) Selkoe, D. J.; Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO molecular medicine* **2016,** *8* (6), 595.
(77) Tanzi, R. E. The genetics of Alzheimer disease. *Cold Spring Harbor perspectives in medicine* **2012,** *2* (10), a006296.
(78) Bolduc, D. M.; Montagna, D. R.; Seghers, M. C.; Wolfe, M. S.; Selkoe, D. J. The amyloid-beta forming tripeptide cleavage mechanism of γ-secretase. *eLife* **2016,** *5*, e17578.
(79) Zhou, R.; Yang, G.; Guo, X.; Zhou, Q.; Lei, J.; Shi, Y. Recognition of the amyloid precursor protein by human γ-secretase. *Science (New York, N.Y.)* **2019,** *363* (6428), eaaw0930.
(80) Dehury, B.; Somavarapu, A. K.; Kepp, K. P. A computer-simulated mechanism of familial Alzheimer's disease: Mutations enhance thermal dynamics and favor looser substrate-binding to γ-secretase. *Journal of Structural Biology* **2020,** *212* (3), 107648.
(81) Dehury, B.; Tang, N.; Mehra, R.; Blundell, T. L.; Kepp, K. P. Side-by-side comparison of Notch- and C83 binding to γ-secretase in a complete membrane model at physiological temperature. *RSC Advances* **2020,** *10* (52), 31215.
(82) Chen, S.-Y.; Zacharias, M. How mutations perturb γ-secretase active site studied by free energy simulations. *Acs Chem Neurosci* **2020,** *11* (20), 3321.
(83) Soto-Ospina, A.; Araque Marín, P.; Bedoya, G.; Sepulveda-Falla, D.; Villegas Lanau, A. Protein Predictive Modeling and Simulation of Mutations of Presenilin-1 Familial Alzheimer's Disease on the Orthosteric Site. *Frontiers in molecular biosciences* **2021,** *8*, 387.
(84) Orzeł, U.; Jakowiecki, J.; Młynarczyk, K.; Filipek, S. The Role of Cholesterol in Amyloidogenic Substrate Binding to the γ-Secretase Complex. *Biomolecules* **2021,** *11* (7), 935.
(85) Devkota, S.; Williams, T. D.; Wolfe, M. S. Familial Alzheimer's disease mutations in amyloid protein precursor alter proteolysis by γ-secretase to increase amyloid β-peptides of≥ 45 residues. *Journal of Biological Chemistry* **2021,** *296*, 100281.
(86) Wang, J.; Miao, Y. Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *The Journal of Chemical Physics* **2020,** *153* (15), 154109.
(87) Xue, Y.; Yuwen, T.; Zhu, F.; Skrynnikov, N. R. Role of electrostatic interactions in binding of peptides and intrinsically disordered proteins to their folded targets. 1. NMR and MD characterization of the complex between the c-Crk N-SH3 domain and the peptide Sos. *Biochemistry* **2014,** *53* (41), 6473.
(88) Li, Y.-M.; Lai, M.-T.; Xu, M.; Huang, Q.; DiMuzio-Mower, J.; Sardana, M. K.; Shi, X.-P.; Yin, K.-C.; Shafer, J. A.; Gardell, S. J. Presenilin 1 is linked with γ-secretase activity in the detergent solubilized state. *Proceedings of the National Academy of Sciences* **2000,** *97* (11), 6138.
(89) Herl, L.; Thomas, A. V.; Lill, C. M.; Banks, M.; Deng, A.; Jones, P. B.; Spoelgen, R.; Hyman, B. T.; Berezovska, O. Mutations in amyloid precursor protein affect its interactions with presenilin/γ-secretase. *Molecular and Cellular Neuroscience* **2009,** *41* (2), 166.
(90) Lichtenthaler, S. F.; Wang, R.; Grimm, H.; Uljon, S. N.; Masters, C. L.; Beyreuther, K. Mechanism of the cleavage specificity of Alzheimer's disease γ-secretase identified by phenylalanine-scanning mutagenesis of the transmembrane domain of the amyloid precursor protein. *Proceedings of the National Academy of Sciences* **1999,** *96* (6), 3053.
(91) Suárez-Calvet, M.; Belbin, O.; Pera, M.; Badiola, N.; Magrané, J.; Guardia-Laguarta, C.; Muñoz, L.; Colom-Cadena, M.; Clarimón, J.; Lleó, A. Autosomal-dominant Alzheimer's disease mutations at the same codon of amyloid precursor protein differentially alter Aβ production. *Journal of neurochemistry* **2014,** *128* (2), 330.
(92) Dubois, B.; Feldman, H. H.; Jacova, C.; Hampel, H.; Molinuevo, J. L.; Blennow, K.; DeKosky, S. T.; Gauthier, S.; Selkoe, D.; Bateman, R. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *The Lancet Neurology* **2014,** *13* (6), 614.

(93) Quintero-Monzon, O.; Martin, M. M.; Fernandez, M. A.; Cappello, C. A.; Krzysiak, A. J.; Osenkowski, P.; Wolfe, M. S. Dissociation between the Processivity and Total Activity of gamma-Secretase: Implications for the Mechanism of Alzheimer's Disease-Causing Presenilin Mutations. *Biochemistry* **2011,** *50* (42), 9023.
(94) Fernandez, M. A.; Klutkowski, J. A.; Freret, T.; Wolfe, M. S. Alzheimer Presenilin-1 Mutations Dramatically Reduce Trimming of Long Amyloid beta-Peptides (A beta) by gamma-Secretase to Increase 42-to-40-Residue A beta. *Journal of Biological Chemistry* **2014,** *289* (45), 31043.
(95) Yang, G.; Zhou, R.; Guo, X.; Yan, C.; Lei, J.; Shi, Y. Structural basis of gamma-secretase inhibition and modulation by small molecule drugs. *Cell* **2021,** *184* (2), 521.
(96) Güner, G.; Lichtenthaler, S. F. Seminars in Cell & Developmental Biology, 2020.
(97) Tamaoka, A.; Odaka, A.; Ishibashi, Y.; Usami, M.; Sahara, N.; Suzuki, N.; Nukina, N.; Mizusawa, H.; Shoji, S. i.; Kanazawa, I. APP717 missense mutation affects the ratio of amyloid beta protein species (A beta 1-42/43 and a beta 1-40) in familial Alzheimer's disease brain. *Journal of Biological Chemistry* **1994,** *269* (52), 32721.
(98) Case, D. A.; Belfon, K.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D.; Cheatham, T.; Cruzeiro, V. W. D.; Darden, T.; Duke, R. E.; Giambasu, G. Amber 2020. **2020**.
(99) Jacobson, K. A.; Gao, Z.-G. Adenosine receptors as therapeutic targets. *Nature Reviews Drug Discovery* **2006,** *5*, 247.
(100) Fernandis, A. Z.; Wenk, M. R. Membrane lipids as signaling molecules. *Current opinion in lipidology* **2007,** *18* (2), 121.
(101) Manna, M.; Nieminen, T.; Vattulainen, I. Understanding the Role of Lipids in Signaling Through Atomistic and Multiscale Simulations of Cell Membranes. *Annual review of biophysics* **2019,** *48*.
(102) Chattopadhyay, A.; Raghuraman, H. Application of fluorescence spectroscopy to membrane protein structure and dynamics. *Curr Sci* **2004,** *87* (2), 175.
(103) CHATTOPADHYAY, A. Experimental and Computational Approaches to Study Membranes and Lipid–-Protein Interactions. *Computational Biophysics of Membrane Proteins* **2016**, 137.
(104) Chattopadhyay, A.; Haldar, S. Dynamic insight into protein structure utilizing red edge excitation shift. *Accounts of chemical research* **2013,** *47* (1), 12.
(105) Chattopadhyay, A.; Jafurulla, M. In *Biochemical Roles of Eukaryotic Cell Surface Macromolecules*; Springer, 2015.
(106) Pucadyil, T. J.; Chattopadhyay, A. Cholesterol modulates ligand binding and G-protein coupling to serotonin1A receptors from bovine hippocampus. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2004,** *1663* (1-2), 188.
(107) Saxena, R.; Chattopadhyay, A. Membrane cholesterol stabilizes the human serotonin1A receptor. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2012,** *1818* (12), 2936.
(108) Casiraghi, M.; Damian, M.; Lescop, E.; Point, E.; Moncoq, K.; Morellet, N.; Levy, D.; Marie, J.; Guittet, E.; Banères, J.-L. Functional modulation of a G protein-coupled receptor conformational landscape in a lipid bilayer. *Journal of the American Chemical Society* **2016,** *138* (35), 11170.
(109) Manna, M.; Mukhopadhyay, C. Cholesterol driven alteration of the conformation and dynamics of phospholamban in model membranes. *Physical Chemistry Chemical Physics* **2011,** *13* (45), 20188.
(110) Muth, S.; Fries, A.; Gimpl, G. Cholesterol-induced conformational changes in the oxytocin receptor. *Biochemical Journal* **2011,** *437* (3), 541.
(111) Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V.-P.; Chien, E. Y.; Velasquez, J.; Kuhn, P.; Stevens, R. C. A specific cholesterol binding site is established by the 2.8 Å structure of the human β2-adrenergic receptor. *Structure* **2008,** *16* (6), 897.

(112) Zocher, M.; Zhang, C.; Rasmussen, S. G.; Kobilka, B. K.; Müller, D. J. Cholesterol increases kinetic, energetic, and mechanical stability of the human β2-adrenergic receptor. *Proceedings of the National Academy of Sciences* **2012,** *109* (50), E3463.
(113) Gimpl, G. Interaction of G protein coupled receptors and cholesterol. *Chemistry and physics of lipids* **2016,** *199*, 61.
(114) Gimpl, G.; Burger, K.; Fahrenholz, F. Cholesterol as modulator of receptor function. *Biochemistry* **1997,** *36* (36), 10959.
(115) Oates, J.; Watts, A. Uncovering the intimate relationship between lipids, cholesterol and GPCR activation. *Current opinion in structural biology* **2011,** *21* (6), 802.
(116) Paila, Y. D.; Chattopadhyay, A. The function of G-protein coupled receptors and membrane cholesterol: specific or general interaction? *Glycoconjugate journal* **2009,** *26* (6), 711.
(117) Paila, Y. D.; Jindal, E.; Goswami, S. K.; Chattopadhyay, A. Cholesterol depletion enhances adrenergic signaling in cardiac myocytes. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2011,** *1808* (1), 461.
(118) Pontier, S. M.; Percherancier, Y.; Galandrin, S.; Breit, A.; Galés, C.; Bouvier, M. Cholesterol-dependent Separation of the β2-Adrenergic Receptor from Its Partners Determines Signaling Efficacy INSIGHT INTO NANOSCALE ORGANIZATION OF SIGNAL TRANSDUCTION. *Journal of biological chemistry* **2008,** *283* (36), 24659.
(119) Pucadyil, T. J.; Chattopadhyay, A. Role of cholesterol in the function and organization of G-protein coupled receptors. *Progress in lipid research* **2006,** *45* (4), 295.
(120) Hedger, G.; Sansom, M. S. Lipid interaction sites on channels, transporters and receptors: recent insights from molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2016,** *1858* (10), 2390.
(121) Inagaki, S.; Ghirlando, R.; White, J. F.; Gvozdenovic-Jeremic, J.; Northup, J. K.; Grisshammer, R. Modulation of the interaction between neurotensin receptor NTS1 and Gq protein by lipid. *Journal of molecular biology* **2012,** *417* (1-2), 95.
(122) Oates, J.; Faust, B.; Attrill, H.; Harding, P.; Orwick, M.; Watts, A. The role of cholesterol on the activity and stability of neurotensin receptor 1. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2012,** *1818* (9), 2228.
(123) Dawaliby, R.; Trubbia, C.; Delporte, C.; Masureel, M.; Van Antwerpen, P.; Kobilka, B. K.; Govaerts, C. Allosteric regulation of G protein–coupled receptor activity by phospholipids. *Nature chemical biology* **2016,** *12* (1), 35.
(124) Sengupta, D.; Kumar, G. A.; Prasanna, X.; Chattopadhyay, A. In *Computational Biophysics of Membrane Proteins*, 2016.
(125) Bruzzese, A.; Gil, C.; Dalton, J. A.; Giraldo, J. Structural insights into positive and negative allosteric regulation of a G protein-coupled receptor through protein-lipid interactions. *Scientific reports* **2018,** *8* (1), 4456.
(126) Neale, C.; Herce, H. D.; Pomes, R.; García, A. E. Can specific protein-lipid interactions stabilize an active state of the beta 2 adrenergic receptor? *Biophysical journal* **2015,** *109* (8), 1652.
(127) Ballesteros, J. A.; Weinstein, H. In *Methods in Neurosciences*; Sealfon, S. C., Ed.; Academic Press, 1995; Vol. 25.
(128) Salas-Estrada, L. A.; Leioatts, N.; Romo, T. D.; Grossfield, A. Lipids alter rhodopsin function via ligand-like and solvent-like interactions. *Biophysical journal* **2018,** *114* (2), 355.
(129) Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E. Activation mechanism of the β2-adrenergic receptor. *Proceedings of the National Academy of Sciences* **2011,** *108* (46), 18684.
(130) Latorraca, N. R.; Venkatakrishnan, A.; Dror, R. O. GPCR dynamics: structures in motion. *Chemical reviews* **2016,** *117* (1), 139.

(131) Song, W.; Yen, H.-Y.; Robinson, C. V.; Sansom, M. S. State-dependent lipid interactions with the A2a receptor revealed by MD simulations using in vivo-mimetic membranes. *Structure* **2019,** *27* (2), 392.
(132) Israelewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics and Modelling* **2001,** *19* (1), 13.
(133) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977,** *23* (2), 187.
(134) Miao, Y.; McCammon, J. A. Graded activation and free energy landscapes of a muscarinic G-protein–coupled receptor. *Proceedings of the National Academy of Sciences* **2016,** *113* (43), 12162.
(135) Miao, Y.; McCammon, J. A. In *Annual reports in computational chemistry*; Elsevier, 2017; Vol. 13.
(136) Sibener, L. V.; Fernandes, R. A.; Kolawole, E. M.; Carbone, C. B.; Liu, F.; McAffee, D.; Birnbaum, M. E.; Yang, X.; Su, L. F.; Yu, W. Isolation of a structural mechanism for uncoupling T cell receptor signaling from peptide-MHC binding. *Cell* **2018,** *174* (3), 672.
(137) Miao, Y.; McCammon, J. A. Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. *Proceedings of the National Academy of Sciences* **2018,** *115* (12), 3036.
(138) Salawu, E. O. The Impairment of TorsinA's Binding to and Interactions With Its Activator: An Atomistic Molecular Dynamics Study of Primary Dystonia. *Frontiers in molecular biosciences* **2018,** *5*.
(139) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proceedings of the National Academy of Sciences* **2017,** *114* (28), 7260.
(140) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Central Science* **2019**.
(141) Draper-Joyce, C. J.; Khoshouei, M.; Thal, D. M.; Liang, Y.-L.; Nguyen, A. T.; Furness, S. G.; Venugopal, H.; Baltos, J.-A.; Plitzko, J. M.; Danev, R. Structure of the adenosine-bound human adenosine A 1 receptor–G i complex. *Nature* **2018,** *558* (7711), 559.
(142) Cheng, R. K.; Segala, E.; Robertson, N.; Deflorian, F.; Doré, A. S.; Errey, J. C.; Fiez-Vandal, C.; Marshall, F. H.; Cooke, R. M. Structures of human A1 and A2A adenosine receptors with xanthines reveal determinants of selectivity. *Structure* **2017,** *25* (8), 1275.
(143) Glukhova, A.; Thal, D. M.; Nguyen, A. T.; Vecchio, E. A.; Jörg, M.; Scammells, P. J.; May, L. T.; Sexton, P. M.; Christopoulos, A. Structure of the adenosine A1 receptor reveals the basis for subtype selectivity. *Cell* **2017,** *168* (5), 867.
(144) Perly, B.; Smith, I. C.; Jarrell, H. C. Acyl chain dynamics of phosphatidylethanolamines containing oleic acid and dihydrosterculic acid: deuteron NMR relaxation studies. *Biochemistry* **1985,** *24* (17), 4659.
(145) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **2010,** *31* (4), 671.
(146) Nguyen, A. T.; Vecchio, E. A.; Thomas, T.; Nguyen, T. D.; Aurelio, L.; Scammells, P. J.; White, P. J.; Sexton, P. M.; Gregory, K. J.; May, L. T.et al. Role of the Second Extracellular Loop of the Adenosine A1 Receptor on Allosteric Modulator Binding, Signaling, and Cooperativity. *Molecular pharmacology* **2016,** *90* (6), 715.

(147) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005,** *26* (16), 1781.
(148) Guixa-González, R.; Rodriguez-Espigares, I.; Ramírez-Anguita, J. M.; Carrio-Gaspar, P.; Martinez-Seara, H.; Giorgino, T.; Selent, J. MEMBPLUGIN: studying membrane complexity in VMD. *Bioinformatics* **2014,** *30* (10), 1478.
(149) Wang, J.; Miao, Y. Mechanistic Insights into Specific G Protein Interactions with Adenosine Receptors Revealed by Accelerated Molecular Simulations. *bioRxiv* **2019**, 541250.
(150) Miao, Y.; Bhattarai, A.; Nguyen, A. T.; Christopoulos, A.; May, L. T. Structural Basis for Binding of Allosteric Drug Leads in the Adenosine A 1 Receptor. *Scientific reports* **2018,** *8* (1), 16836.
(151) Miao, Y.; Nichols, S. E.; Gasper, P. M.; Metzger, V. T.; McCammon, J. A. Activation and dynamic network of the M2 muscarinic receptor. *Proceedings of the National Academy of Sciences* **2013,** *110* (27), 10982.
(152) Bhattarai, A.; Wang, J.; Miao, Y. Retrospective ensemble docking of allosteric modulators in an adenosine G-protein-coupled receptor. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2020,** *1864* (8), 129615.
(153) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discovery* **2002,** *1* (9), 727.
(154) Lappano, R.; Maggiolini, M. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nat. Rev. Drug Discovery* **2011,** *10* (1), 47.
(155) Jacobson, K. A. New paradigms in GPCR drug discovery. *Biochem. Pharmacol.* **2015,** *98* (4), 541.
(156) Jacobson, K. A.; Gao, Z.-G. Adenosine receptors as therapeutic targets. *Nat. Rev. Drug Discovery* **2006,** *5* (3), 247.
(157) Bruns, R. F.; Fergus, J. H. Allosteric enhancement of adenosine A1 receptor binding and function by 2-amino-3-benzoylthiophenes. *Mol. Pharmacol.* **1990,** *38* (6), 939.
(158) Bruns, R. F.; Fergus, J. H.; Coughenour, L. L.; Courtland, G. G.; Pugsley, T. A.; Dodd, J. H.; Tinney, F. J. Structure-activity relationships for enhancement of adenosine A1 receptor binding by 2-amino-3-benzoylthiophenes. *Mol. Pharmacol.* **1990,** *38* (6), 950.
(159) Romagnoli, R.; Baraldi, P. G.; Ijzerman, A. P.; Massink, A.; Cruz-Lopez, O.; Lopez-Cara, L. C.; Saponaro, G.; Preti, D.; Aghazadeh Tabrizi, M.; Baraldi, S.et al. Synthesis and Biological Evaluation of Novel Allosteric Enhancers of the A1 Adenosine Receptor Based on 2-Amino-3-(4'-Chlorobenzoyl)-4-Substituted-5-Arylethynyl Thiophene. *J. Med. Chem.* **2014,** *57* (18), 7673.
(160) Baraldi, P. G.; Zaid, A. N.; Lampronti, I.; Fruttarolo, F.; Pavani, M. G.; Tabrizi, M. A.; Shryock, J. C.; Leung, E.; Romagnoli, R. Synthesis and biological effects of a new series of 2-amino-3-benzoylthiophenes as allosteric enhancers of A1-adenosine receptor. *Bioorg. Med. Chem. Lett.* **2000,** *10* (17), 1953.
(161) Tranberg, C. E.; Zickgraf, A.; Giunta, B. N.; Luetjens, H.; Figler, H.; Murphree, L. J.; Falke, R.; Fleischer, H.; Linden, J.; Scammells, P. J.et al. 2-Amino-3-aroyl-4,5-alkylthiophenes: Agonist Allosteric Enhancers at Human A1 Adenosine Receptors. *J. Med. Chem.* **2002,** *45* (2), 382.
(162) Aurelio, L.; Valant, C.; Flynn, B. L.; Sexton, P. M.; Christopoulos, A.; Scammells, P. J. Allosteric Modulators of the Adenosine A1 Receptor: Synthesis and Pharmacological Evaluation of 4-Substituted 2-Amino-3-benzoylthiophenes. *J. Med. Chem.* **2009,** *52* (14), 4543.
(163) Aurelio, L.; Valant, C.; Figler, H.; Flynn, B. L.; Linden, J.; Sexton, P. M.; Christopoulos, A.; Scammells, P. J. 3- and 6-Substituted 2-amino-4,5,6,7-tetrahydrothieno[2,3-c]pyridines as A1 adenosine receptor allosteric modulators and antagonists. *Bioorg. Med. Chem.* **2009,** *17* (20), 7353.
(164) Aurelio, L.; Christopoulos, A.; Flynn, B. L.; Scammells, P. J.; Sexton, P. M.; Valant, C. The synthesis and biological evaluation of 2-amino-4,5,6,7,8,9-hexahydrocycloocta[b]thiophenes

as allosteric modulators of the A1 adenosine receptor. *Bioorg. Med. Chem. Lett.* **2011,** *21* (12), 3704.

(165) Valant, C.; Aurelio, L.; Devine, S. M.; Ashton, T. D.; White, J. M.; Sexton, P. M.; Christopoulos, A.; Scammells, P. J. Synthesis and Characterization of Novel 2-Amino-3-benzoylthiophene Derivatives as Biased Allosteric Agonists and Modulators of the Adenosine A1 Receptor. *J. Med. Chem.* **2012,** *55* (5), 2367.

(166) Hill, S. J.; May, L. T.; Kellam, B.; Woolard, J. Allosteric interactions at adenosine A1 and A3 receptors: new insights into the role of small molecules and receptor dimerization. *Br. J. Pharmacol.* **2014,** *171* (5), 1102.

(167) Lütjens, H.; Zickgraf, A.; Figler, H.; Linden, J.; Olsson, R. A.; Scammells, P. J. 2-Amino-3-benzoylthiophene Allosteric Enhancers of A1 Adenosine Agonist Binding: New 3, 4-, and 5-Modifications. *J. Med. Chem.* **2003,** *46* (10), 1870.

(168) Li, X.; Conklin, D.; Pan, H.-L.; Eisenach, J. C. Allosteric Adenosine Receptor Modulation Reduces Hypersensitivity Following Peripheral Inflammation by a Central Mechanism. *J. Pharmacol. Exp. Ther.* **2003,** *305* (3), 950.

(169) Childers, S. R.; Li, X.; Xiao, R.; Eisenach, J. C. Allosteric modulation of adenosine A1 receptor coupling to G-proteins in brain. *J. Neurochem.* **2005,** *93* (3), 715.

(170) Rahuel, J.; Rasetti, V.; Maibaum, J.; Rüeger, H.; Göschke, R.; Cohen, N.; Stutz, S.; Cumin, F.; Fuhrer, W.; Wood, J. Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chem. Biol.* **2000,** *7* (7), 493.

(171) Wang, J.; Bhattarai, A.; Ahmad, W. I.; Farnan, T. S.; John, K. P.; Miao, Y. In *GPCRs*; Jastrzebska, B.;Park, P. S. H., Eds.; Academic Press, 2020.

(172) Kaldor, S. W.; Kalish, V. J.; Davies, J. F.; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.et al. Viracept (Nelfinavir Mesylate, AG1343): A Potent, Orally Bioavailable Inhibitor of HIV-1 Protease. *J. Med. Chem.* **1997,** *40* (24), 3979.

(173) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. Method for Including the Dynamic Fluctuations of a Protein in Computer-Aided Drug Design. *J. Phys. Chem. A* **1999,** *103* (49), 10213.

(174) Evangelista Falcon, W.; Ellingson, S. R.; Smith, J. C.; Baudry, J. Ensemble Docking in Drug Discovery: How Many Protein Configurations from Molecular Dynamics Simulations are Needed To Reproduce Known Ligand Binding? *J. Phys. Chem. B* **2019,** *123* (25), 5189.

(175) Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018,** *114* (10), 2271.

(176) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme. *J. Am. Chem. Soc.* **2002,** *124* (20), 5632.

(177) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* **2004,** *47* (8), 1879.

(178) Amaro, R. E.; Schnaufer, A.; Interthal, H.; Hol, W.; Stuart, K. D.; McCammon, J. A. Discovery of drug-like inhibitors of an essential RNA-editing ligase in Trypanosoma brucei. *Proc. Natl. Acad. Sci. USA* **2008,** *105* (45), 17278.

(179) Haynes, K. M.; Abdali, N.; Jhawar, V.; Zgurskaya, H. I.; Parks, J. M.; Green, A. T.; Baudry, J.; Rybenkov, V. V.; Smith, J. C.; Walker, J. K. Identification and Structure–Activity Relationships of Novel Compounds that Potentiate the Activities of Antibiotics in Escherichia coli. *J. Med. Chem.* **2017,** *60* (14), 6205.

(180) Kapoor, K.; McGill, N.; Peterson, C. B.; Meyers, H. V.; Blackburn, M. N.; Baudry, J. Discovery of Novel Nonactive Site Inhibitors of the Prothrombinase Enzyme Complex. *J. Chem. Inf. Model.* **2016,** *56* (3), 535.
(181) Velazquez, H. A.; Riccardi, D.; Xiao, Z.; Quarles, L. D.; Yates, C. R.; Baudry, J.; Smith, J. C. Ensemble docking to difficult targets in early-stage drug discovery: Methodology and application to fibroblast growth factor 23. *Chem. Biol. Drug Des.* **2018,** *91* (2), 491.
(182) Evangelista, W.; Weir, R. L.; Ellingson, S. R.; Harris, J. B.; Kapoor, K.; Smith, J. C.; Baudry, J. Ensemble-based docking: From hit discovery to metabolism and toxicity predictions. *Bioorg. Med. Chem.* **2016,** *24* (20), 4928.
(183) Pi, M.; Kapoor, K.; Ye, R.; Smith, J. C.; Baudry, J.; Quarles, L. D. GPCR6A Is a Molecular Target for the Natural Products Gallate and EGCG in Green Tea. *Mol. Nutr. Food Res.* **2018,** *62* (8), 1700770.
(184) He, H.; Weir, R. L.; Toutounchian, J. J.; Pagadala, J.; Steinle, J. J.; Baudry, J.; Miller, D. D.; Yates, C. R. The quinic acid derivative KZ-41 prevents glucose-induced caspase-3 activation in retinal endothelial cells through an IGF-1 receptor dependent mechanism. *PLOS ONE* **2017,** *12* (8), e0180808.
(185) Huang, X.-P.; Karpiak, J.; Kroeze, W. K.; Zhu, H.; Chen, X.; Moy, S. S.; Saddoris, K. A.; Nikolova, V. D.; Farrell, M. S.; Wang, S.et al. Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. *Nature* **2015,** *527* (7579), 477.
(186) Miao, Y.; Goldfeld, D. A.; Moo, E. V.; Sexton, P. M.; Christopoulos, A.; McCammon, J. A.; Valant, C. Accelerated structure-based design of chemically diverse allosteric modulators of a muscarinic G protein-coupled receptor. *Proc. Natl. Acad. Sci. USA* **2016,** *113* (38), E5675.
(187) Romagnoli, R.; Baraldi, P. G.; Tabrizi, M. A.; Gessi, S.; Borea, P. A.; Merighi, S. Allosteric Enhancers of A1 Adenosine Receptors: State of the Art and New Horizons for Drug Development. *Curr. Med. Chem.* **2010,** *17* (30), 3488.
(188) Nguyen, A. T. N.; Vecchio, E. A.; Thomas, T.; Nguyen, T. D.; Aurelio, L.; Scammells, P. J.; White, P. J.; Sexton, P. M.; Gregory, K. J.; May, L. T.et al. Role of the Second Extracellular Loop of the Adenosine A1 Receptor on Allosteric Modulator Binding, Signaling, and Cooperativity. *Mol. Pharmacol.* **2016,** *90* (6), 715.
(189) Nguyen, A. T. N.; Baltos, J.-A.; Thomas, T.; Nguyen, T. D.; Muñoz, L. L.; Gregory, K. J.; White, P. J.; Sexton, P. M.; Christopoulos, A.; May, L. T. Extracellular Loop 2 of the Adenosine A1 Receptor Has a Key Role in Orthosteric Ligand Affinity and Agonist Efficacy. *Mol. Pharmacol.* **2016,** *90* (6), 703.
(190) Miao, Y.; Bhattarai, A.; Nguyen, A. T. N.; Christopoulos, A.; May, L. T. Structural Basis for Binding of Allosteric Drug Leads in the Adenosine A1 Receptor. *Sci. Rep.* **2018,** *8* (1), 16836.
(191) Bhattarai, A.; Wang, J.; Miao, Y. G-Protein-Coupled Receptor–Membrane Interactions Depend on the Receptor Activation State. *J. Comput. Chem.* **2019,** *0* (0).
(192) Wang, J.; Miao, Y. Mechanistic Insights into Specific G Protein Interactions with Adenosine Receptors. *J. Phys. Chem. B* **2019,** *123* (30), 6462.
(193) Vincenzi, F.; Targa, M.; Romagnoli, R.; Merighi, S.; Gessi, S.; Baraldi, P. G.; Borea, P. A.; Varani, K. TRR469, a potent A1 adenosine receptor allosteric modulator, exhibits anti-nociceptive properties in acute and neuropathic pain models in mice. *Neuropharmacology* **2014,** *81*, 6.
(194) Kimatrai-Salvador, M.; Baraldi, P. G.; Romagnoli, R. Allosteric modulation of A1-adenosine receptor: a review. *Drug Discovery Today: Technol.* **2013,** *10* (2), e285.
(195) Jacobson, K. A.; Gao, Z.-G.; Göblyös, A.; Ijzerman, A. P. In *Adv. Pharmacol.*; Jacobson, K. A.;Linden, J., Eds.; Academic Press, 2011; Vol. 61.

(196) Figler, H.; Olsson, R. A.; Linden, J. Allosteric Enhancers of A₁ Adenosine Receptors Increase Receptor-G Protein Coupling and Counteract Guanine Nucleotide Effects on Agonist Binding. *Mol. Pharmacol.* **2003,** *64* (6), 1557.
(197) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012,** *55* (14), 6582.
(198) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785.
(199) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **2016,** *11* (5), 905.
(200) Ravindranath, P. A.; Forli, S.; Goodsell, D. S.; Olson, A. J.; Sanner, M. F. AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS Comput. Biol.* **2015,** *11* (12), e1004586.
(201) Pearlman, D. A.; Charifson, P. S. Improved scoring of ligand− protein interactions using OWFEG free energy grids. *J. Med. Chem.* **2001,** *44* (4), 502.
(202) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015,** *11* (8), 3584.
(203) Miao, Y.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. *Annu. Rep. Comput. Chem.* **2017,** *13*, 231.
(204) Miao, Y. Acceleration of biomolecular kinetics in Gaussian accelerated molecular dynamics. *J. Chem. Phys.* **2018,** *149* (7), 072308.
(205) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004,** *120* (24), 11919.
(206) Shen, T.; Hamelberg, D. A statistical analysis of the precision of reweighting-based simulations. *J. Chem. Phys.* **2008,** *129* (3), 034103.
(207) D.A. Case, D. S. C., T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2018), AMBER 2018, University of California, San Francisco.
(208) Pang, Y. T.; Miao, Y.; Wang, Y.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics in NAMD. *J. Chem. Theory Comput.* **2017,** *13* (1), 9.
(209) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014,** *10* (7), 2677.
(210) Hamelberg, D.; de Oliveira, C. A.; McCammon, J. A. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J. Chem. Phys.* **2007,** *127* (15), 155102.
(211) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995,** *91* (1-3), 275.
(212) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33.
(213) Vanommeslaeghe, K.; MacKerell, A. D., Jr. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta.* **2015,** *1850* (5), 861.

(214) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010,** *31* (4), 671.
(215) Kappel, K.; Miao, Y.; McCammon, J. A. Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. *Q. Rev. Biophys.* **2015,** *48* (4), 479.
(216) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781.
(217) Roe, D. R.; Cheatham, T. E., 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9* (7), 3084.
(218) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998,** *19* (14), 1639.
(219) Wong, C. F. Flexible receptor docking for drug discovery. *Expert Opin. Drug Discov.* **2015,** *10* (11), 1189.
(220) Bhattarai, A.; Pawnikar, S.; Miao, Y. Mechanism of ligand recognition by human ACE2 receptor. *The journal of physical chemistry letters* **2021,** *12* (20), 4814.
(221) Gross, L. Z. F.; Sacerdoti, M.; Piiper, A.; Zeuzem, S.; Leroux, A. E.; Biondi, R. M. ACE2, the Receptor that Enables the Infection by SARS-CoV-2: Biochemistry, Structure, Allostery and Evaluation of the Potential Development of ACE2 Modulators. **2020**.
(222) Gheblawi, M.; Wang, K.; Viveiros, A.; Nguyen, Q.; Zhong, J.-C.; Turner, A. J.; Raizada, M. K.; Grant, M. B.; Oudit, G. Y. 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* **2020,** *126* (10), 1456.
(223) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)* **2020,** *367* (6485), 1444.
(224) Patel, V. B.; Zhong, J.-C.; Grant, M. B.; Oudit, G. Y. Role of the ACE2/angiotensin 1–7 axis of the renin–angiotensin system in heart failure. *Circulation research* **2016,** *118* (8), 1313.
(225) Perlot, T.; Penninger, J. M. ACE2–From the renin–angiotensin system to gut microbiota and malnutrition. *Microbes and infection* **2013,** *15* (13), 866.
(226) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **2020**.
(227) Towler, P.; Staker, B.; Prasad, S. G.; Menon, S.; Tang, J.; Parsons, T.; Ryan, D.; Fisher, M.; Williams, D.; Dales, N. A. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *Journal of Biological Chemistry* **2004,** *279* (17), 17996.
(228) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.-Y. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* **2020**.
(229) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020,** *581* (7807), 221.
(230) Song, W.; Gui, M.; Wang, X.; Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS pathogens* **2018,** *14* (8), e1007236.
(231) Li, F.; Li, W.; Farzan, M.; Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science (New York, N.Y.)* **2005,** *309* (5742), 1864.

(232) Huentelman, M. J.; Zubcevic, J.; Hernandez Prada, J. A.; Xiao, X.; Dimitrov, D. S.; Raizada, M. K.; Ostrov, D. A. Structure-based discovery of a novel angiotensin-converting enzyme 2 inhibitor. *Hypertension* **2004,** *44* (6), 903.
(233) 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.; Ostrov, D. A. Structure-based identification of small-molecule angiotensin-converting enzyme 2 activators as novel antihypertensive agents. *Hypertension* **2008,** *51* (5), 1312.
(234) Nami, B.; Ghanaeian, A.; Ghanaeian, K.; Nami, N. The effect of ACE2 inhibitor MLN-4760 on the interaction of SARS-CoV-2 spike protein with human ACE2: a molecular dynamics study. **2020**.
(235) 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.; Stewart, M. Substrate-based design of the first class of angiotensin-converting enzyme-related carboxypeptidase (ACE2) inhibitors. *Journal of the American Chemical Society* **2002,** *124* (40), 11852.
(236) Miao, Y.; Bhattarai, A.; Wang, J. Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. *Journal of Chemical Theory and Computation* **2020,** *16* (9), 5526.
(237) Donoghue, M.; Hsieh, F.; Baronas, E.; Godbout, K.; Gosselin, M.; Stagliano, N.; Donovan, M.; Woolf, B.; Robison, K.; Jeyaseelan, R. A novel angiotensin-converting enzyme–related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circulation research* **2000,** *87* (5), e1.
(238) Wade, R. C.; Gabdoulline, R. R.; Ludemann, S. K.; Lounnas, V. Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (11), 5942.
(239) Huang, Y. M.; McCammon, J. A.; Miao, Y. Replica Exchange Gaussian Accelerated Molecular Dynamics: Improved Enhanced Sampling and Free Energy Calculation. *J Chem Theory Comput* **2018,** *14* (4), 1853.
(240) Oshima, H.; Re, S.; Sugita, Y. Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules. *J Chem Theory Comput* **2019,** *15* (10), 5199.
(241) Plattner, N.; Noe, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun* **2015,** *6*, 7653.
(242) Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *Journal of the American Chemical Society* **2002,** *124* (20), 5632.
(243) Amaro, R. E.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J Comput Aid Mol Des* **2008,** *22* (9), 693.
(244) Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophysical Journal* **2018,** *114* (10), 2271.
(245) Draper-Joyce, C. J.; Bhola, R.; Wang, J.; Bhattarai, A.; Nguyen, A. T.; O'Sullivan, K.; Chia, L. Y.; Venugopal, H.; Valant, C.; Thal, D. M. Positive allosteric mechanisms of adenosine A1 receptor-mediated analgesia. *Nature* **2021,** *597* (7877), 571.