# Exploring distinct binding site regions of $\beta_2$-adrenergic receptor via coarse-grained molecular dynamics simulations

**Sibel ÇAKAN,[1] Ebru Demet AKDOĞAN[2,*]**

[1]Graduate School of Computational Biology and Bioinformatics, Kadir Has University,
Cibali 34083, İstanbul, Turkey

[2]Department of Bioinformatics and Genetics, Faculty of Engineering and Natural Sciences,
Kadir Has University, Cibali 34083, İstanbul, Turkey

**Abstract:** $\beta_2$-Adrenergic receptor ($\beta_2$AR) is a G protein-coupled receptor that is highly flexible and able to recognize a wide range of ligands through its conformational variations. Active and inactive conformations revealed by recent crystallographic experiments do not provide a complete dynamic picture of the receptor, especially in the binding site. In this study, molecular dynamics (MD) simulation through a residue-based coarse-grained model is used as an alternative and efficient method to explore a wider conformational search space. The system was composed of $\beta_2$AR embedded into a 1-palmitoyl-2-oleoyl-phosphatidylcholine membrane bilayer with surrounding water. A total of 6 $\mu$s of simulation at constant NPT was performed for a system of 6868 coarse-grained beads. The system reached equilibrium at around 0.1 $\mu$s. The overall 3-dimensional structure was well preserved throughout the simulation. Local residue-based fluctuations were in good agreement with fully atomistic MD simulations. Four distinct snapshots were selected and reverse-mapped to all-atom representations with around 65,000 atoms. Each reverse-mapped system was later subjected to 100 ns of MD simulation for equilibration. Root mean square deviation clustering analysis yielded distinct receptor conformers for the binding site regions, which were suggested to be alternative representations of the binding pocket and thus were proposed as plausible targets in docking-based virtual screening experiments for the discovery of novel antagonists.

**Key words:** Coarse-grained, molecular dynamics, adrenergic receptor, binding site, 1-palmitoyl-2-oleoyl-phosphatidylcholine membrane

## 1. Introduction

Twenty-five percent of the eukaryotic genomes encode the membrane proteins that have significant roles such as transporting, signaling, and cell–cell interactions in biological cells.[1,2] They also constitute the largest class of drug targets, approximately 50% of all drugs in the market.[3] Despite their physiological and pharmaceutical importance, very few crystallographic structures are reported.[4] Still, it is difficult to understand the details of protein–membrane interactions via experimental methods alone. Thus, embedded proteins in a lipid bilayer are ideal systems for computer simulations.[5]

With current computational power, it is not possible to sample all intermediates of adrenergic receptors along the activation pathway that occur in millisecond time scales through traditional molecular dynamics (MD) simulations. Only initial stages of protein activation can be observed in atomistic-level simulations.[6−9]

*Correspondence: demet.akten@khas.edu.tr

On the other hand, coarse-grained (CG) models provide an efficient approach for increasing the time scale from nanoseconds to microseconds by grouping several atoms into 1 or 2 particles.[10,11] In a CG model, collective motions of the protein in larger time scales can be observed due to removal of high-frequency motions. CG models have been applied to different biomolecules such as lipids, membranes, proteins, and DNAs.[12−15]

For proteins, there exist several approaches of CG modeling that have different agreement between accuracy and transferability, with different degrees of independence from the reference structure. The earliest approaches of coarse-graining for proteins were the elastic network models and Go-like models, whose bias towards a reference structure makes them only weakly transferable to general dynamics studies.[16] In the 1970s, Levitt built a transferable CG model with a knowledge-based parameterization that inspired many successive studies.[17] The coarsest approach is the one-bead model that evolved from the Go-like models and includes more sophisticated potentials, but it still was not enough. To improve the specificity of the local interactions, an additional bead was added on the centroid of the side chain (2-bead model). Statistical analyses of the experimental structures were used for developing the force field of the 2-bead models.[18]

For lipids, Marrink et al. developed a CG model where 4 heavy atoms are represented on average by a single CG particle.[19,20] The model was employed to investigate the molecular basis of how the physicochemical properties of the phospholipid bilayer membrane affect self-assembly of visual rhodopsin molecules, which are members of the G protein-coupled receptor. They suggested that future application of the CG-MD method may contribute to a better understanding of the role of lipid diversity and protein structure in lipid-mediated protein–protein interactions.[21] In another work by Marrink and coworkers, a CG-MD simulation was performed for Kv1.2, a voltage-gated ion channel. The study revealed a possible gating mechanism.[22] Another CG-MD method was developed by Scott et al. for lipid bilayer self-assembly around membrane proteins. They accurately predicted the protein position in the bilayer with a range of different sizes and architectures.[23]

In this study, a CG model for a protein–lipid–water model developed by Shih et al. was used for studying the dynamics of $\beta_2$-adrenergic receptor ($\beta_2$AR) embedded in a lipid bilayer.[24] Shih's model originates from Marrink's CG lipid–water model, which was extended to include proteins as well. Shih's approach was successfully applied to study the self-assembly of lipoproteins.[25] The initial conformation of the receptor was taken from a previous all-atom MD simulation study[26] as the equilibrated system embedded in a 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) membrane and solvated with TIP3 water molecules. The protein was converted into a residue-based CG model where each residue has 2 interaction sites, 1 on backbone and 1 on the side chain.

In this study, the main goal of using a CG model was to obtain distinct conformers of $\beta_2$AR, especially at the binding site regions, by exploring a wider conformational search space. Experimentally revealed structures of the active and inactive states of the receptor show similar binding site regions,[27,28] although the bound agonists and antagonists have distinct chemical structures. However, it is crucial to have a distinct active or inactive state that would bind to certain agonists or antagonists more specifically. Our fully atomistic conformations obtained from reverse-mapping CG-MD snapshots show specific binding site regions that would be potential targets in docking experiments for the discovery of novel drug candidates.

## 2. Experimental

### 2.1. Coarse-graining the protein–lipid system

The initial conformation of the receptor was taken from an all-atom MD simulation trajectory obtained in a previous work[26] as the equilibrated snapshot embedded in a POPC membrane solvated with TIP3 water molecules. The protein and the lipids were first extracted from the system and converted into a residue-based CG model using the CG Tools Plugin of VMD.[29] The whole system was then solvated with CG water molecules using VMD's Solvate Module v1.2. To make the total charge of the system be zero, 13 sodium and 20 chlorine ions were added with a concentration of 0.154 M. The fully atomistic (FA) and CG models of the system, which consists of protein, membrane, water, and ions, are illustrated in Figure 1. The FA model with 68,001 atoms was represented with 6868 beads in the CG model. The periodic box was taken as 89 Å × 98 Å × 101 Å. Table 1 lists the number of atoms and the box dimensions for both the FA and CG models for comparison.
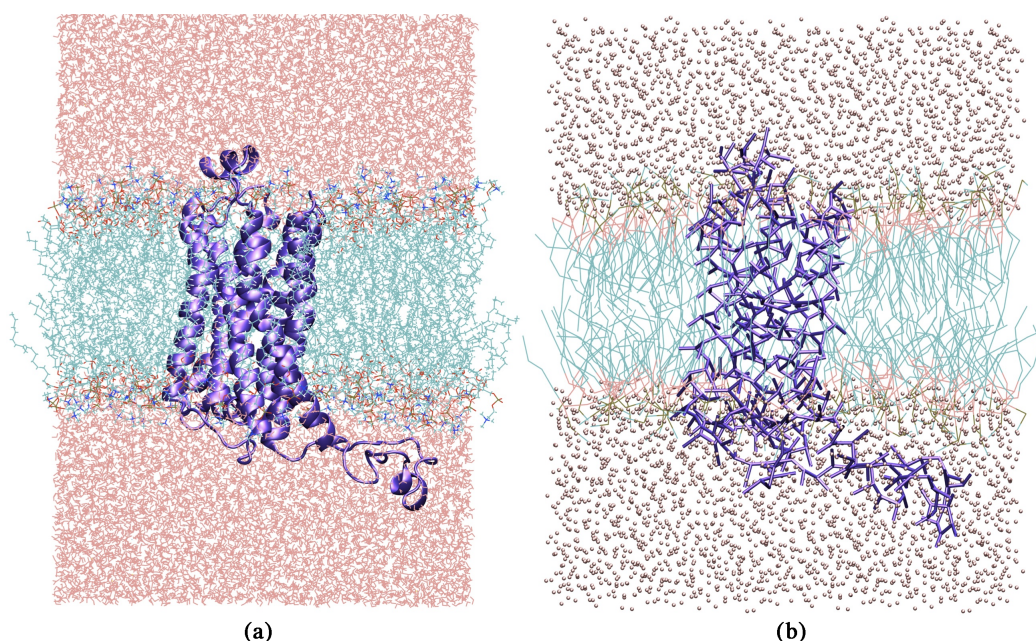


**Figure 1.** $\beta_2$ AR embedded in a POPC membrane and surrounded with water in a) FA and b) CG representations.

**Table 1.** Periodic box dimension and the total number of atoms in each component of the system composed of protein, lipid, waters, and ions in FA and CG models.

| System | Periodic box dimension (Å) | Proteins | Lipids | Waters | Ions | Total number of atoms |
|--------|----------------------------|----------|--------|--------|------|-----------------------|
| FA | 86 × 86 × 100 | 5055 | 20,770 | 42,135 | 41 | 68,001 |
| CG | 89 × 98 × 101 | 607 | 1860 | 4368 | 33 | 6868 |

A total of 6 $\mu$s of CG-MD simulation was performed in NPT ensemble at 1 atm and 310 K, with a time step of 5 fs using the Nanoscale Molecular Dynamics (NAMD) v2.7 simulation tool. The system was gradually heated from 30 K to target temperature to raise the kinetic energy with minimal conformational changes. The system was then minimized for 30,000 steps through the conjugate gradient algorithm to eliminate the steric clashes between system components. Langevin dynamics with a damping coefficient of 5 ps$^{-1}$ was used

for maintaining the temperature constancy; constant pressure was maintained using a Langevin Nosè–Hoover piston with a period of 1000 fs and a decay time of 500 fs. Furthermore, 1-2 exclusion was used for nonbonded interactions with a cut-off value of 12 Å with shifting starting at 9 Å.

The 6-$\mu$s-long CG-MD trajectory, which consists of 5929 frames, was clustered based on the backbone atom's root mean square deviation (RMSD) value using the k-means algorithm implemented in the kclust module of the Multiscale Modeling Tools of the Structural Biology Tool Set.[30] Clusterings were based on 4 different regions of the protein: the core (protein excluding ICL3), the intracellular loop 3 (ICL3), the transmembrane, and the binding site, as illustrated in Figure 2. From each clustering, a snapshot, which was nearest to the cluster's centroid (average structure) from the most populated cluster, was selected for the reverse-mapping procedure.
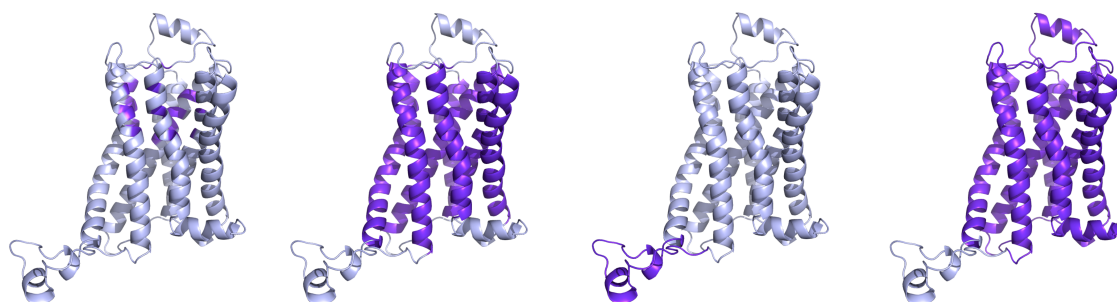


**Figure 2.** The 4 regions of the receptor: a) binding site, b) transmembrane, c) ICL3, and d) core, all colored in purple.

The CG Tools Plugin of VMD was used for reverse-mapping, which recovers all of the atomistic details. The protein's CG model was first extracted from the system and reverse-mapped alone. The protein was then embedded into a POPC bilayer membrane. The VMD Membrane Builder Plugin was used to generate the cell membrane. The POPC doubled-layered cell membrane was built with a 100-Å thickness in the direction of the z-axis. The membrane's x and y dimensions, which are also the periodic box dimensions in the same direction, were set according to the protein's size. To illustrate, minimum and maximum x coordinates of the protein were determined as –53.6 Å and –0.9 Å, which makes the x dimension of the protein approximately 55 Å. As the minimum distance between the protein and the boundary of the periodic cell was set to 15 Å, the periodic box dimension in the x direction was set to 85 Å (=55 + 15 + 15). The box dimension in the y direction was determined similarly. Once the protein and the membrane were aligned, the system was solvated with a thickness of 15 Å TIP3 water molecules in both directions of the z-axis. Table 2 displays the dimensions of the protein and the periodic box for all 4 reverse-mapped systems (abbreviated as RM1, RM2, RM3, and RM4) in detail. Finally, each system was ionized with Na$^+$ and Cl$^-$ ions for a concentration of 0.154 mol/L, to make the total charge of the system equal to zero. The total number of atoms in each system is given in Table 2.

**Table 2.** System size, periodic box dimension, and the total number of atoms in each component of the system composed of protein, lipid, waters, and ions in all 4 reverse-mapped (RM) models.

| Model | Protein's dimensions ($\text{Å}^3$) | Box dimensions ($\text{Å}^3$) | Protein | Lipid | Water | Na$^+$/Cl$^-$ | Total |
|-------|-------------|-------------|---------|-------|-------|---------|-------|
| RM1 | $55 \times 59 \times 68$ | $85 \times 89 \times 100$ | 5055 | 18760 | 40311 | 39/46 | 64211 |
| RM2 | $55 \times 56 \times 69$ | $85 \times 86 \times 100$ | 5055 | 19028 | 41247 | 40/47 | 65417 |
| RM3 | $54 \times 59 \times 68$ | $84 \times 89 \times 100$ | 5055 | 19430 | 42174 | 41/48 | 66748 |
| RM4 | $54 \times 61 \times 65$ | $84 \times 91 \times 100$ | 5055 | 20100 | 43584 | 42/49 | 68830 |

## 2.2. Reverse-mapping back to all-atom representation

Each system was subjected to 3 preparation phases, which included melting the lipids and constraining and then releasing the protein. The melting phase of the lipids consists of 1000 steps of energy minimization followed by 0.5 ns of MD simulation while fixing the protein, water, ions, and lipid head groups in space and allowing the lipid tails to move freely. As a result, a more realistic fluid-like, disordered cell membrane was created. In the second stage, the system was allowed to move, except that the protein's motion was severely constrained. Finally, the constraint on the protein was removed and the whole system was allowed to move freely. At each stage, 1000 steps of energy minimization using the conjugate gradient algorithm were followed by 0.5 ns of MD simulation using the NAMD v2.7 software tool. Following these preparation steps, 100 ns of MD simulation was conducted for each RM system for further equilibration and production stages.

## 3. Results and discussion

To determine the equilibrium and production stages of the CG-MD simulation, the RMSD from the initial structure was determined using backbone atoms only, as shown in Figure 3. Three different RMSD profiles were calculated, 1 for the whole receptor, 1 for the core region, and 1 for the ICL3 region. For the RMSD profile of the whole receptor, the structural alignment was based on the whole structure (named RMSD All Fit All), whereas the RMSD calculations of the core and the ICL3 region were based on the alignment of the core region (named RMSD Core Fit Core and RMSD ICL3 Fit Core). Clearly, the receptor reaches equilibrium at a very early stage, which is around 100 ns. The ICL3 region is the most mobile part of the system, and thus fluctuates at around 8 Å throughout the trajectory.
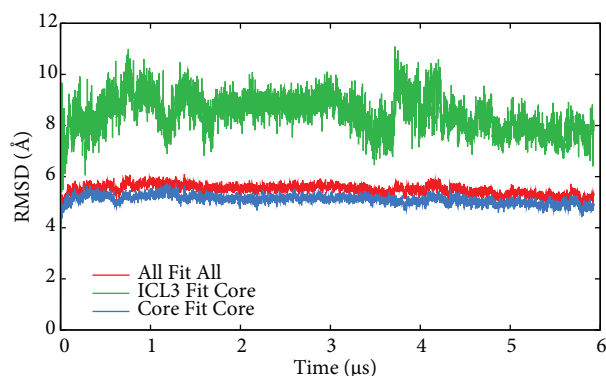


**Figure 3.** RMSD profiles for different regions of the receptor, with respect to the initial frame (see text for explanation).

The CG model's mobility during 6 $\mu$s of simulation was expressed by the root mean square fluctuation (RMSF) profile, which was compared to the RMSF profile of a 0.8-$\mu$s-long FA MD simulation[26] as shown in Figure 4. Accordingly, the highest fluctuations of the FA model, which are observed for the residues in the loop regions, especially the ICL3 region, are not well pronounced in the CG model. While the RMSF value of ICL3 is up to 15 Å in the FA model, it only reaches 4 Å in the CG model. Still, the degrees of fluctuations in the transmembrane regions of the 2 models agree well.

The CG-MD trajectory was clustered based on the RMSD value of 4 different regions (core, intracellular loop 3 (ICL3), transmembrane, and binding site). For each case, the cluster, which contains the highest number of snapshots, was determined. A representative structure was then selected as the conformation closest to the

average structure of that cluster. All 4 representatives of the CG models were reverse-mapped and subjected to equilibration as discussed in Section 2. The RMSD profile during equilibration was calculated for the core and ICL3 regions after aligning each snapshot to the initial conformation based on the core region (see Figure 5). In all 4 reverse-mapped FA (RM-FA) short simulations, the core region starts to deviate from the initial structure and reaches up to 4 Å within the first 10 ns. Afterwards, the core region remains stable throughout the simulation, especially with fewer fluctuations during the last 40 ns. The ICL3 region's deviation from the initial frame increases up to 10 Å within 10 ns and fluctuates significantly throughout the simulation. The RMSF profiles of the 4 RM-FA simulations are illustrated in Figure 6 alongside the 100-ns-long FA simulation (identified as FAS1), which was originally constructed as FA. The highest fluctuations in all 4 RM-FA profiles were observed for the loop regions, especially for ICL3 as the most mobile region in RM1, RM2, and RM3. The overall shape and the intensity of fluctuations were in accordance with FA simulation.
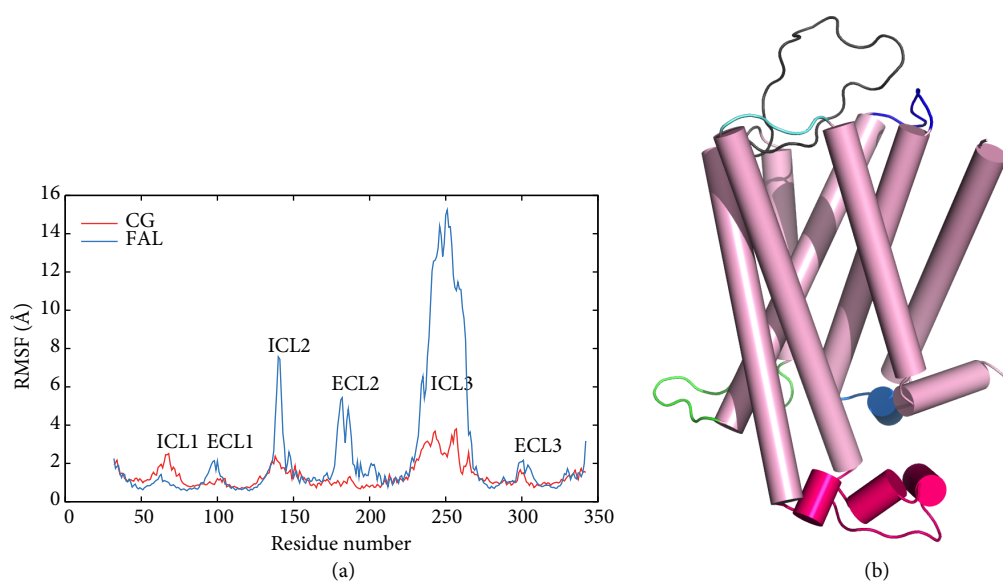


**Figure 4.** a) RMSF profiles of CG and FA models during 6 $\mu$s and 0.8 $\mu$s of simulation, respectively, with b) cartoon representation of $\beta_2$AR indicating the loop regions.
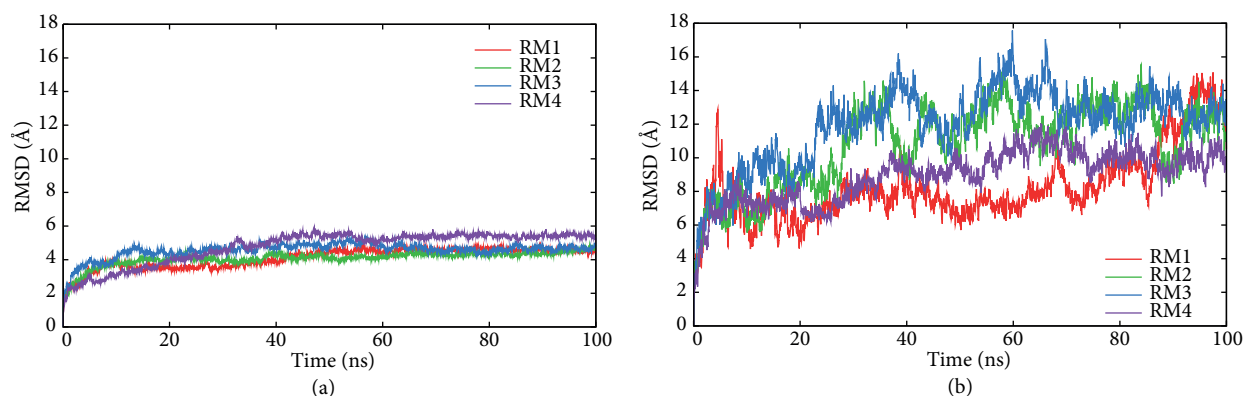


**Figure 5.** RMSD profiles of the 4 reverse-mapped models during 100 ns of FA MD simulation: a) Core Fit Core and b) ICL3 Fit Core (see text for explanation).
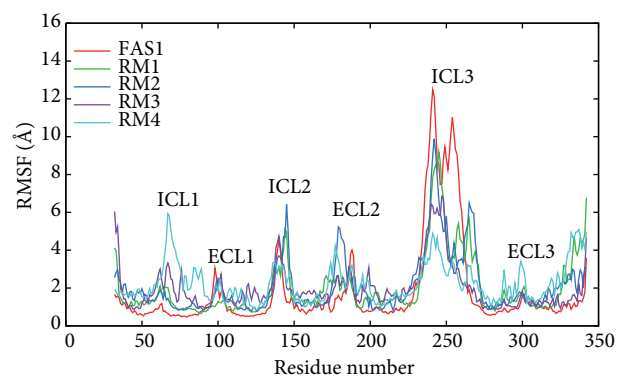
**Figure 6.** RMSF profiles of 4 reverse-mapped models each for 100-ns MD simulation. The RMSD profile of a FA model for 100-ns MD simulation is given for comparison.
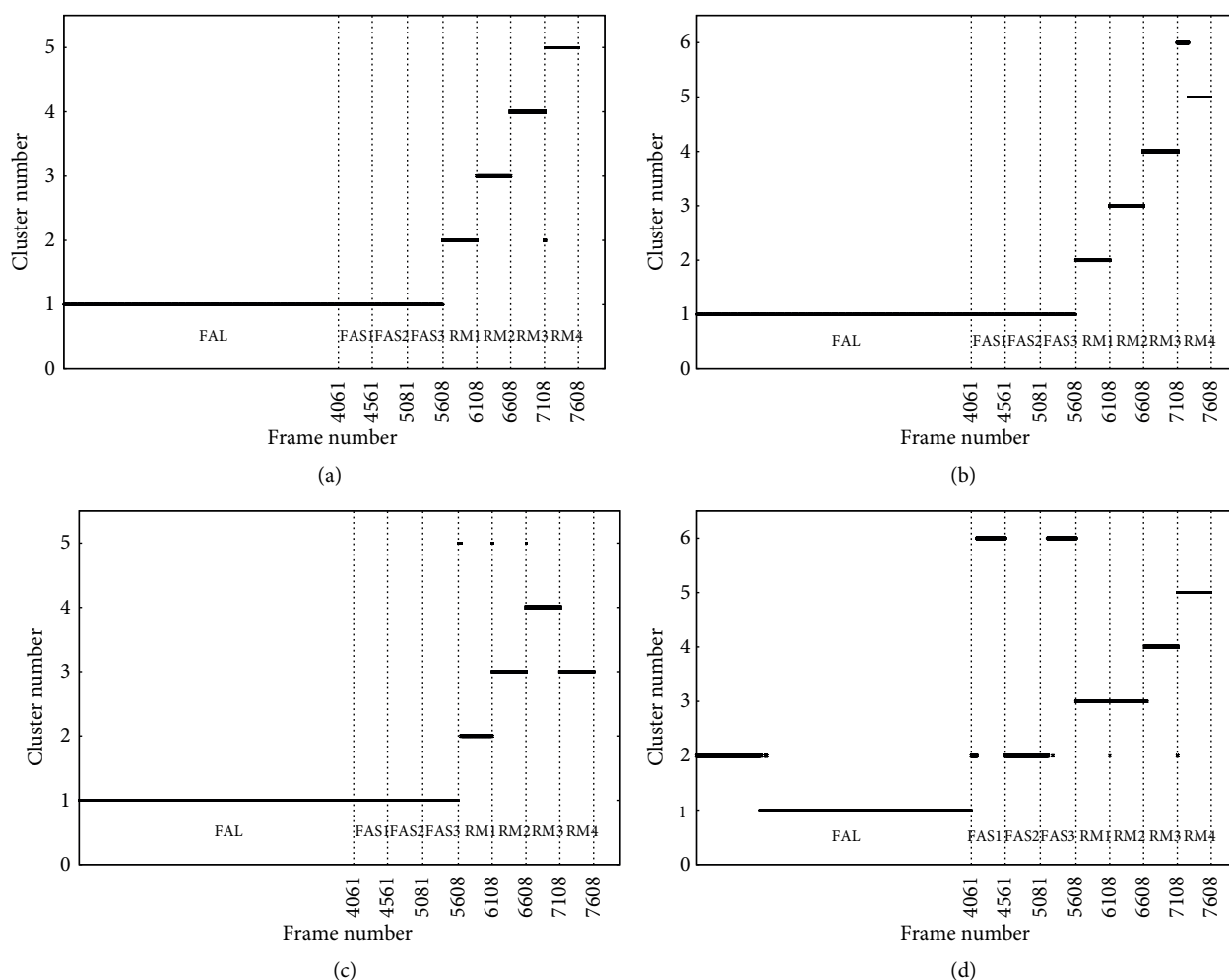


**Figure 7.** Cluster profile of all FA simulations (FA and RM-FA) based on a) core region, b) transmembrane region, c) binding site region, and d) ICL3 region, with RMSD cut-off values of 5 Å, 3 Å, 2.8 Å, and 5.9 Å, respectively.

To reveal distinct conformers of $\beta_2$ AR, all RM and FA simulation trajectories were clustered via the kclust algorithm using the Multiscale Modeling Tools of the Structural Biology Tool Set. All MD snapshots were first aligned to the initial state of the receptor for the 800-ns-long FA simulation. They were then clustered based on 4 different regions of the receptor, core, transmembrane, binding site, and ICL3 regions for different RMSD thresholds. The cluster profiles shown in Figure 7 illustrate each MD snapshot with a frame number versus the cluster number in which the snapshot was found. The RMSD threshold values for the core, ICL3, transmembrane, and binding-site regions were set to 5 Å, 5.9 Å, 3 Å, and 2.8 Å, respectively, such that they would produce 5 or 6 clusters in total. The conformations that fall into the same cluster are similar to each other with an RMSD value equal to or less than the threshold RMSD value.

In all 4 cluster profiles, the 800-ns-long FA simulation (FAL) with 4061 snapshots is shown in the first range [1–4061] of the x-axis, followed by 3 FA simulations of 100 ns in length (FAS1, FAS2, and FAS3), each having around 500 conformers and thus occupying the intervals of [4062–4561], [4562–5081], and [5082–5608]. The last 4 intervals belong to 100-ns-long RM-FA simulations (RM1, RM2, RM3, and RM4) and are located at the [5609–6108], [6109–6608], [6609–7108], and [7109-7608] ranges on the x-axis.

In Figure 7a, the clustering was performed based on the core region with an RMSD threshold value of 5 Å. Clearly, all FA simulation snapshots fell into the same cluster, while each 500 RM-FA simulation snapshot was grouped into distinct clusters. Similar profiles were obtained after clustering all simulation snapshots based on transmembrane, binding site, and ICL3 regions, as well (see Figures 7b–7d). Each cluster in all 4 profiles represents a unique conformation of the receptor, which was observed for all 4 RM-FA models. On the other hand, no distinctiveness was observed for FA models for which all MD snapshots fell into the same cluster.

The distinct structure of the receptor, especially in the binding site region, may either indicate a different conformational rearrangement of the receptor or a structural deformation in some part of the receptor. The energy profiles indicate that RM-FA simulations are as stable as FA simulations. Furthermore, RMSF profiles indicate a structural mobility, which agrees well with the original FA models. Still, it would be difficult to conclude the nature of the distinctiveness by looking only at the energy and RMSF profiles. Thus, a structural comparison is mandatory.

For structural comparison, the crystal structure (PDB id: 2RH1) was aligned based on the core region to one of the snapshots of the RM4 simulation, which was chosen as the representative conformation closest to the average structure of the most populated cluster. As shown in Figure 8, the RMSD value was calculated as 5.47 Å for the whole structure. Each helix in the snapshot was also aligned to its corresponding helix in the crystal structure in order to see the change in the secondary structure. Except helices 1, 2, and 5 (TMI, TMII, and TMV), all 4 helices of the RM4 model were in good agreement with the crystal structure with acceptable RMSD values. TMI, TMII, and TMV could not keep their original structures, which broke around the middle. On the other hand, TMIII, TMIV, TMVI, and TMVII were well preserved with RMSD values smaller than 2.5 Å. Corresponding helical alignments performed for RM1, RM2, and RM3 models are illustrated in supplementary Figure 9. In nature, it is known that the helical motif of the 7-helix transmembrane receptor is well preserved to perform its signaling function. The secondary structural deformations observed in our simulation likely originated from the force field defined for the CG model of the membrane–protein system. On the other hand, the degree of these deformations does not disturb the final state of the binding site since the overall 3-dimensional structure of the receptor was well maintained.
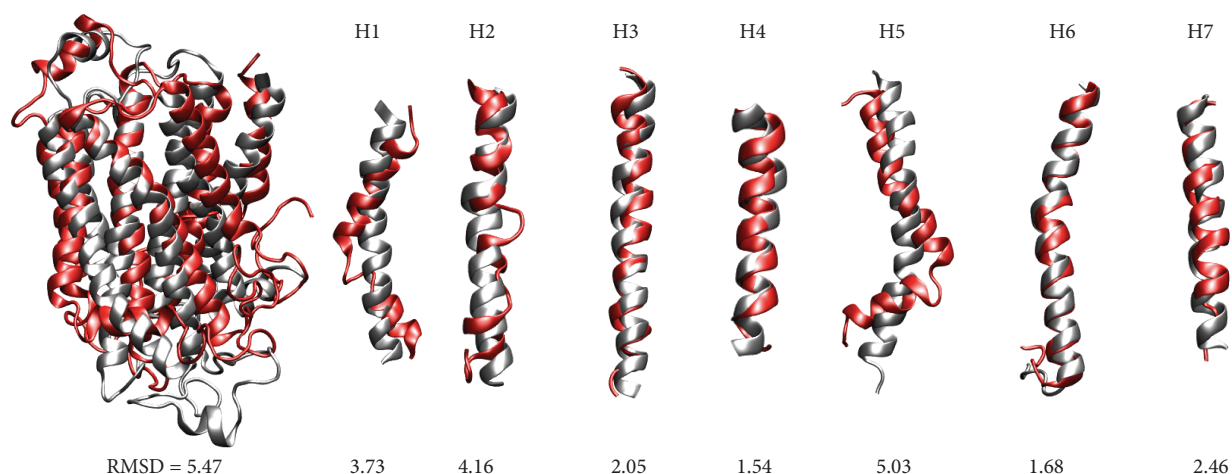
H1  H2  H3  H4  H5  H6  H7

RMSD = 5.47    3.73    4.16    2.05    1.54    5.03    1.68    2.46

**Figure 8.** Alignment of each helix in RM4 model (in red) with the corresponding helix in the crystal structure (gray). The 2 receptors on the left are aligned based on transmembrane region.
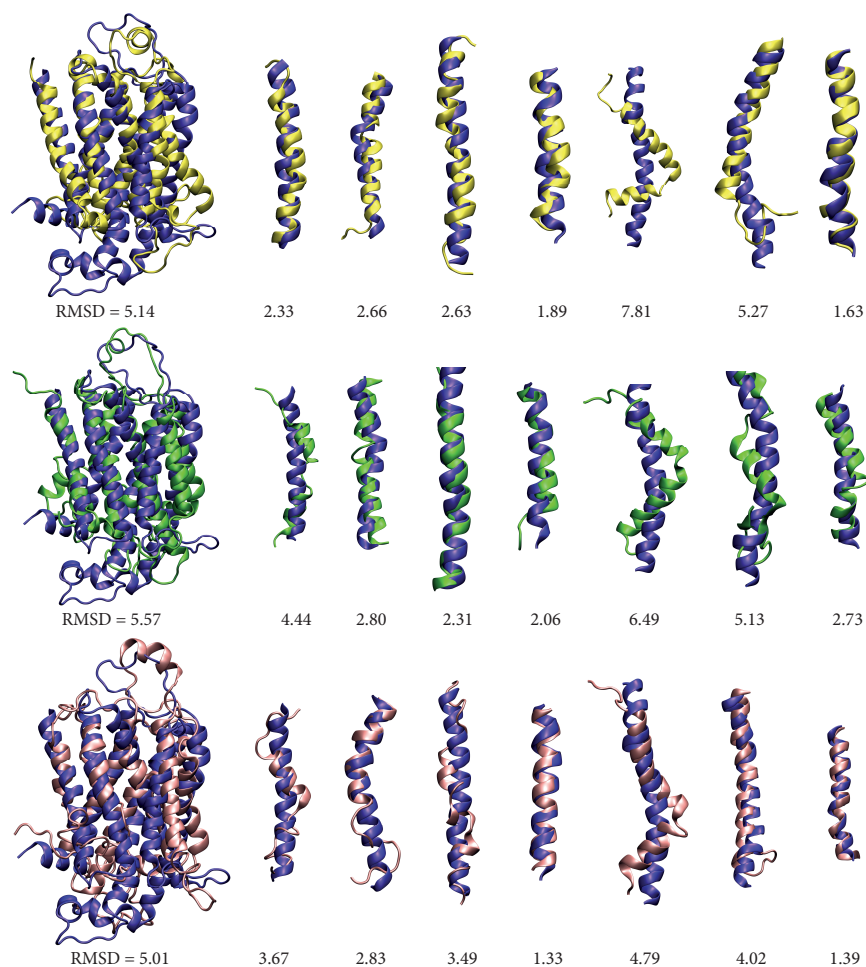


RMSD = 5.14    2.33    2.66    2.63    1.89    7.81    5.27    1.63

RMSD = 5.57    4.44    2.80    2.31    2.06    6.49    5.13    2.73

RMSD = 5.01    3.67    2.83    3.49    1.33    4.79    4.02    1.39

**Figure 9.** Alignment of each helix in a) RM1 (in yellow), b) RM2 (in green), and c) RM3 (in pink) models with the corresponding helix in the crystal structure (in dark blue; PDB id: 2RH1). The 2 receptors on the left are aligned based on transmembrane region.

In order to monitor the secondary structural variation with respect to time, the structure type in the RM4 model was plotted as a function of time, as shown in Figure 10a. The color code explains the structure type as: 0 = T (turn), 1 = C (coil), 2 = B (isolated bridge), 3 = E (beta sheet), 4 = H (alpha helix), 5 = G (3-10 helix), and 6 = I (pi helix). The secondary structure profile of the 800-ns-long FA simulation is given in Figure 10b for comparison. The structural deformations of helices could be clearly seen in the RM-FA simulation. In particular, TMII has some serious deformations represented by a pi helix and a turn type. TMV also has a few residues around the middle region turning into a turn type colored in blue. The profile clearly reveals that most of the deformations already existed at the beginning of the simulation. Thus, helices most likely deformed during CG-MD simulation. The MD simulation of the RM-FA models simply could not fix the deformations in the secondary structure due to the necessity of large energy jumps; however, it was able to preserve the main tertiary contacts between helices and those between the receptor and its surroundings.
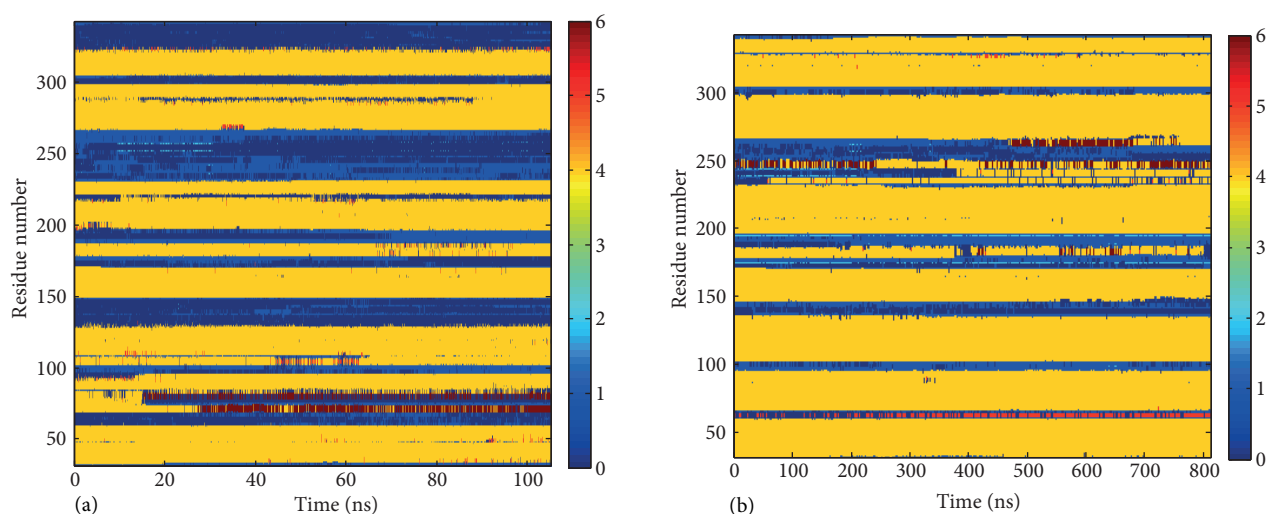


**Figure 10.** Secondary structure variation along the trajectory for a) RM4 model and b) FA model.

To reveal the exact time at which these structural deformations occurred, the RMSD profiles of each helix in the reverse-mapped structures were calculated based on the crystal structure (PDB id: 2RH1), as illustrated in Figure 11. Accordingly, TMV and TMVI had the highest initial RMSD values near 4 Å in all 4 RM models. Deformation of TMV was similar in all 4 RM-FA simulations. Just after reverse-mapping, TMV's RMSD value was around 4 Å, and towards the end of the simulation, it increased up to 6 Å. Generally, the initial RMSD value determines the final state of the helical structure. If the initial RMSD is greater than 2.5 Å, it becomes more difficult to recover the structural motif during the simulation. However, exceptionally, the RMSD value of TMVI decreased from around 4 Å to around 2 Å in the RM3 and RM4 simulations.

To trace back to the deformation, the RMSD profile of each helix was calculated during CG simulation after aligning each helix with its corresponding helix in the crystal structure. As shown in Figure 12, the highest RMSD values were observed in TMV and TMVI. The increase in the RMSD value for all helices occurred in early stages of the CG simulation and remained nearly stable afterwards, except a slight decrease from 5 Å to 4 Å observed for TMV in the last 2 $\mu$s of the simulation. It is evident that the force field that was redefined for the CG model is not sufficient to represent the interaction network within a helical structure. On the other hand, it is most likely that selecting a snapshot from the CG-MD trajectory that had deviated the least from the crystal structure would result in less severe helical deformations upon reverse-mapping.
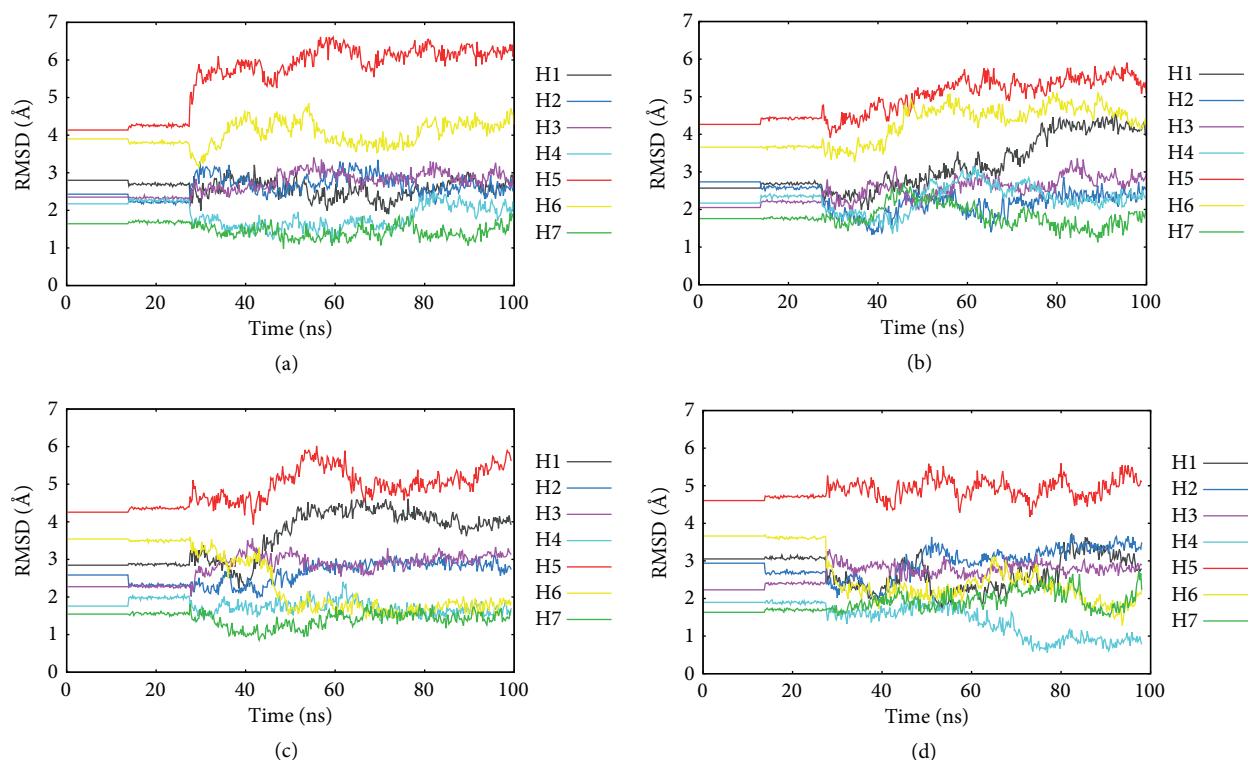
**Figure 11.** RMSD profiles of each helix during 100 ns of simulation for a) RM1, b) RM2, c) RM3, and d) RM4 models.
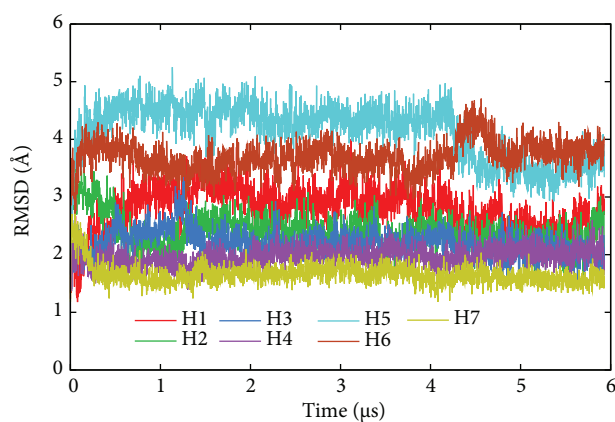


**Figure 12.** RMSD profiles of each helix during the 6-$\mu$s CG simulation.

To reveal the distinct conformers of the binding site region obtained from reverse-mapped structures, 5 representative structures taken from 5 different clusters were aligned to the crystal structure of the inactive state (colored in gray; PDB id: 2RH1[3]) as shown in Figure 13. Clustering was based on binding site regions using all snapshots of the simulations including 1 FAL, 3 FAS, and 4 RM-FA models. For a RMSD cut-off value of 2.8 Å, 5 clusters were obtained as illustrated in Figure 7c. The representative structures were chosen as the conformations that were closest to the centroid (average structure) of the cluster. For each structure, an RMSD value of the binding site region with respect to the crystal structure was calculated as 1.3 Å, 3.0 Å, 3.5 Å, 3.7 Å, and 3.5 Å for 5 clusters. The first RMSD value of 1.3 Å corresponds to the original atomistic model (FAL),

whereas the other 4 RMSD values around 3 Å correspond to RM-FA models. To compare the binding pockets of the snapshots with that of the crystal structure, the partial inverse agonist carazolol was shown in its bound form to the crystal structure as well. In Figure 13b, the binding pocket of RM-FA is completely closed up with carazolol inside, whereas the binding pockets of the other 3 RM-FA structures are half open and comparable to the inactive state. The distance between Ser 207 and Asp 113, which is known to be around 11–12 Å in crystal form of the inactive state and 8–10 Å in the active state, is determined as 13.7 Å for FAL and as 13.9 Å, 11.4 Å, 13.8 Å, and 13.6 Å for the RM-FA structures in Figure 13. Ser 207 and Asp 113 are 2 anchor residues situated on each side of the binding site and interacting with hydrophobic and polar groups of the ligand, respectively. Except for the second RM-FA model, all the other 3 have more open binding sites with distance values above 13 Å. In all RM-FA models, there exists a considerable amount of space in the binding site where a ligand can easily bind without any steric clash with the protein (see Figure 14 for an alternative representation of the binding pocket). This indicates that despite a few structural deformations on helices, the RM-FA models have a well-structured binding cavity with distance values comparable to the crystal structure.
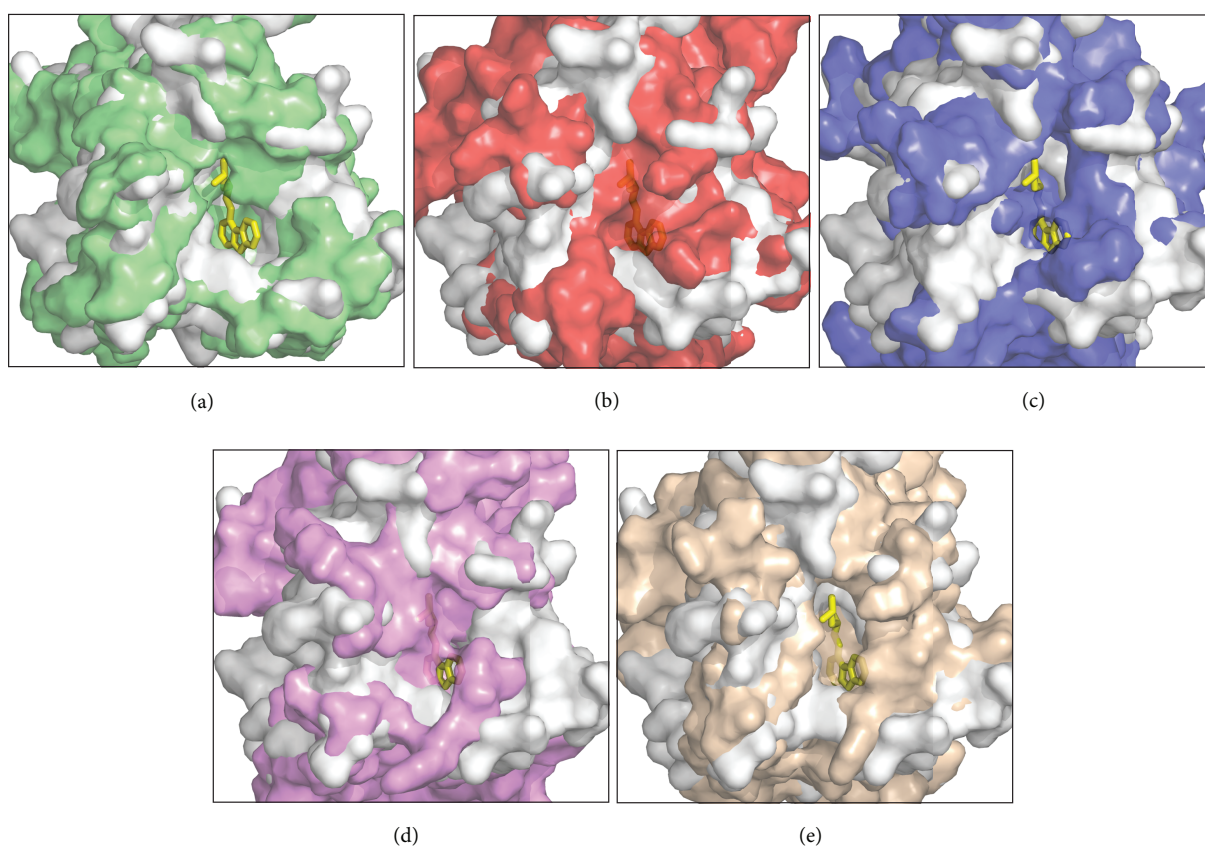


(a)  (b)  (c)

(d)  (e)

**Figure 13.** Five representative structures taken from each cluster in Figure 7c, which are closest to their cluster's centroid and aligned to 2RH1 (shown in light gray): a) cluster #1 (FAL), b) cluster #2 (RM1, frame #5950), c) cluster #3 (RM2, frame #6310), d) cluster #4 (RM3, frame #6784), and e) cluster #5 (RM1, frame #5635). Carazolol is shown in yellow in the stick representation in all 5 subfigures.
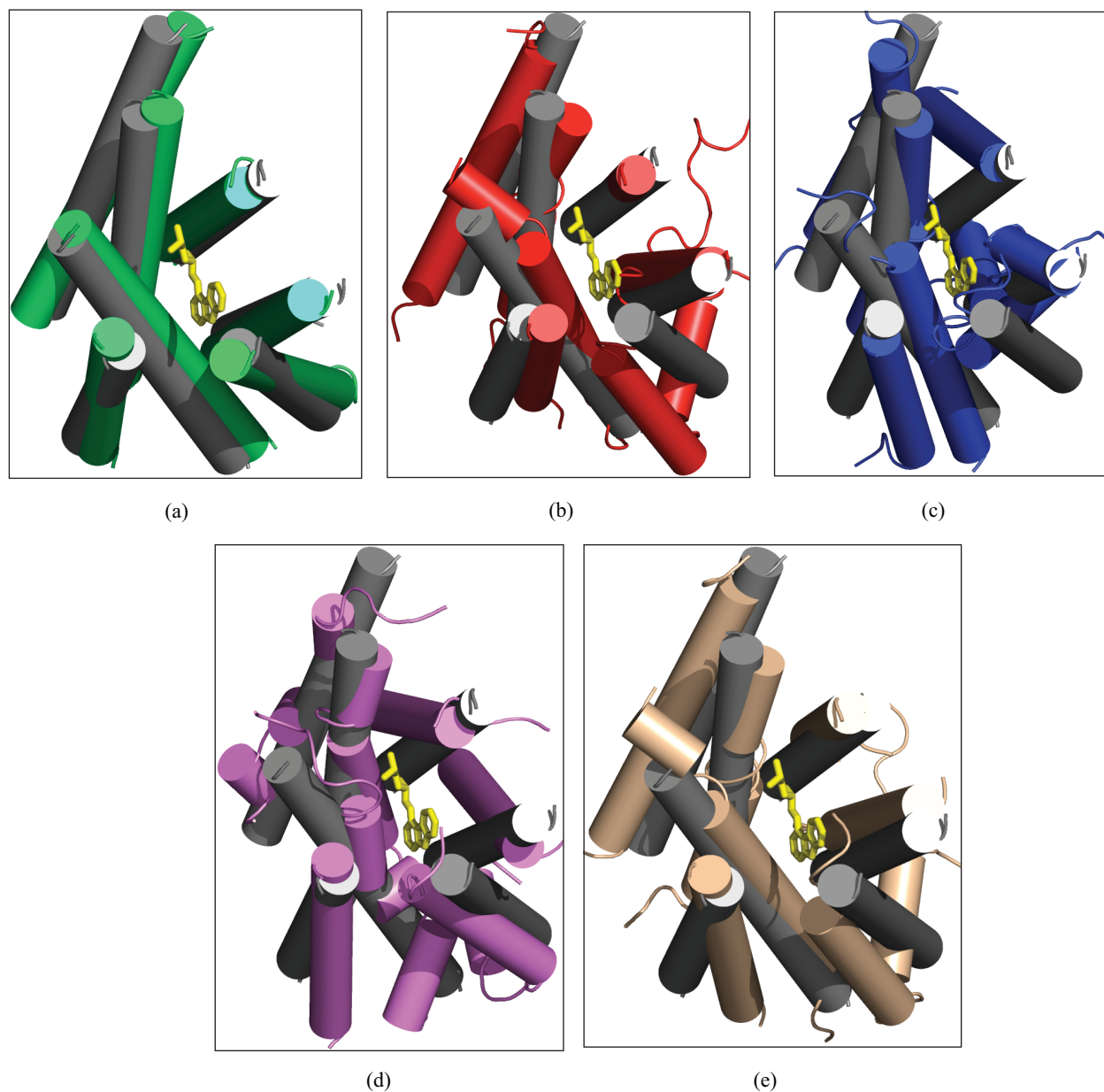
**Figure 14.** Five representative structures taken from each cluster in Figure 7c, which are closest to their cluster's centroid and aligned to 2RH1 (shown in light gray): a) cluster #1 (FAL), (b) cluster #2 (RM1, frame #5950), c) cluster #3 (RM2, frame #6310), d) cluster #4 (RM3, frame #6784), and e) cluster #5 (RM1, frame #5635). Carazolol is shown in yellow in the stick representation in all 5 subfigures.

## 4. Conclusion

CG simulations conducted at 310 K yielded a less mobile receptor in all its regions compared to a FA simulation. The RMSF profile clearly indicates more flexibility at the loop regions relative to the transmembrane region. RMSD clusterings of CG-MD trajectory based on core, transmembrane, and binding site regions do not reveal structural diversity, with only 1 large cluster that constitutes 75% of all snapshots. The analysis of 4 reverse-mapped structures that were subjected to 100 ns of MD simulation shows that the RMSF values agree well

with those obtained from the 100-ns-long FA MD simulations. Their energy values were stable throughout the simulation. The overall 3-dimensional structure of the receptor was maintained in all 4 reverse-mapped models, yet some deformations were observed, especially in TMV and TMVI. These deformations occurred at the early stages of the CG simulations. The reverse-mapped simulations simply could not recover the structural motif due to high-energy barriers.

Despite these helical deformations, the 3-dimensional structure of the binding site remained stable throughout the simulation. For each RM-FA simulation, the RMSD profile of the binding site region with respect to the crystal structure of the inactive state (PDB id: 2RH1) showed that stability has been reached at around 4–4.5 Å. Furthermore, RMSD clustering of all RM snapshots with FA snapshots revealed distinct clusters for RM conformations, each representing a unique state of the binding pocket. Thus, the structural diversity presented by coarse-graining followed by reverse-mapping was found to be satisfactory. This methodology can be used as an alternative for exploring a wider conformational search space, especially for large membrane systems, that can be easily trapped in one of the local minima due to high-energy barriers. However, the newly presented conformations of the binding pocket need to be further tested in virtual screening experiments to reveal their capability of extracting the known agonists and antagonists. In addition, as a future work, the force field representing the protein's internal energy should be adjusted to avoid the helical deformations.

## Acknowledgment

## References

1. Wallin, E.; von Heijne, G. *Protein Sci.* **1998**, *7*, 1029–1038.

2. Nilsson, J.; Persson, B.; von Heijne, G. *Proteins* **2005**, *60*, 606–616.

3. Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. *Science* **2007**, *318*, 1258–1265.

4. White, S. H. *Protein Sci.* **2004**, *13*, 1948–1949.

5. Lindahl, E.; Sansom, M. S. P. *Curr. Opin. Struc. Biol.* **2008**, *18*, 425–431.

6. Spijker, P.; van Hoof, B.; Debertrand, M.; Markvoort, A. J.; Vaidehi, N.; Hilbers, P. A. J. *Int. J. Mol. Sci.* **2010**, *11*, 2393–2420.

7. Vaidehi, N.; Floriano, W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P.; Choi, E. J.; Zamanakos, G.; Goddard, W.A. 3rd. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12622–12627.

8. Freddolino, P. L.; Kalani, M. Y. S.; Vaidehi, N.; Floriano, W. B.; Hall, S. E.; Trabanino, R. J.; Kam, V. W. T.; Goddard, W. A. 3rd. *Proc. Natl. Acad. Sci.* **2004**, *101*, 2736–2741.

9. Kalani, M. Y. S.; Vaidehi, N.; Freddolino, P. L.; Hall, S. E.; Trabanino, R. J.; Floriano, W. B.; Kam, V. W. T.; Kalani, M. A.; Goddard, W. A. 3rd. *Proc. Natl. Acad. Sci.* **2004**, *101*, 3815–3820.

10. Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *Biophys. J.* **2004**, *87*, 2107–2115.

11. Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. *J. Struct. Biol.* **2006**, *157*, 593–605.

12. Treptow, W.; Marrink, S. J.; Tarek, M. *J. Phys. Chem. B Letters* **2008**, *112*, 3277–3282.

13. Jogini, V.; Roux, B. *Biophys. J.* **2007**, *93*, 3070–3082.

14. Freites, J. A.; Tobias, D. J; White, S. H. *Biophys. J.* **2006**, *91*, L90–L92.

15. DeMille, R. C; Cheatham, T. E.; Molinero, V. *J. Phys. Chem. B* **2011**, *115*, 132–142.

16. Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.

17. Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59–107.

18. Bahar, I.; Jernigan, R. *J. Mol. Biol.* **1997**, *266*, 195–214.

19. Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.

20. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

21. Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. *J. Am. Chem. Soc.* **2007**, *129*, 10126–10132.

22. Treptow, W.; Marrink, S. J.; Mounir, T. *J. Phys. Chem. B* **2008**, *112*, 3277–3282.

23. Scott, K. A.; Bond, P. J.; Ivetac, A.; Chetwynd, A. P.; Khalid, S.; Sansom, M. S. P. *Structure* **2008**, *16*, 621–630.

24. Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K. *J. Phys. Chem. B* **2006**, *110*, 3674–3684.

25. Shih, A.; Freddolino, P. L.; Arkhipov, A.; Schulten, K. *J. Struct. Biol.* **2007**, *157*, 579–592.

26. Özcan, Ö. Exploring the Intrinsic Dynamics of Human Beta-2 Adrenergic G-Protein Coupled and Its Potential Use in Computational Drug Design Studies. MSc thesis, Boğaziçi University, İstanbul, 2011.

27. Wacker, D.; Fenalti, G.; Brown, M. A.; Katritch, V.; Abagyan, R.; Cherezov, V.; Stevens, R. C. *J. Am. Chem. Soc.* **2010**, *132*, 11443–11445.

28. Rosenbaum, D. M.; Zhang, C.; Lyons, J. A.; Holl, R.; Aragao, D.; Arlow, D. H.; Rasmussen, S. G. F.; Choi, H. J.; Devree, B. T.; Sunahara, R. K.; et al. *Nature* **2011**, *469*, 236–240.

29. Humprey, A.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

30. Feig, M. J.; Karanicolas, J.; Brooks, C. L. *J. Mol. Graph. Model.* **2004**, *22*, 377–395.