JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

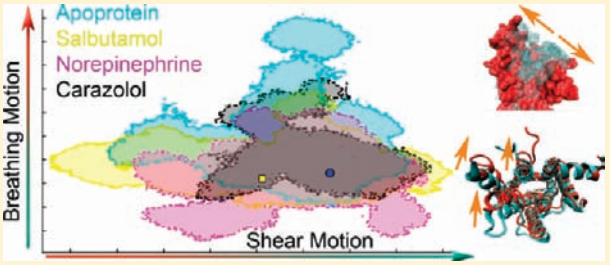
The Role of Conformational Ensembles in Ligand Recognition in G-Protein Coupled Receptors

Michiel J. M. Niesen, Supriyo Bhattacharya, and Nagarajan Vaidehi*

Department of Immunology, Beckman Research Institute of the City of Hope, Duarte, California 91010.

Supporting Information

ABSTRACT: G-protein coupled receptors (GPCRs) are allosteric membrane proteins mediating cellular signaling. GPCRs exhibit multiple inactive and active conformations, and the population balance between these conformations is altered upon binding of signaling molecules (or ligands). However, the nature of the conformational ensemble or the mechanism of the conformational transitions is not well understood. We present a multiscale computational approach combining a coarse-grained discrete conformational sampling method with fine-grained molecular dynamics investigating the effect of various ligands binding on the ensemble of conformations sampled by human β2-adrenergic receptor (β2AR). We show that the receptor, in the absence of any ligand, samples an extensive conformational space that includes breathing of the orthosteric ligand binding site and shear motion of the transmembrane helices 5 and 6 against the other helices. The shear motion is similar to the reorganization of the intracellular regions of TM3, TM5, and TM6 observed in the crystal structure of the active state of GPCRs. The binding of agonist norepinephrine or partial agonist salbutamol leads to the selection of a subset of conformations including active and inactive state conformations, while inverse agonist carazolol selects only inactive state conformations. The dynamics of water observed during the simulations provides an explanation for the conformational changes observed in the solution-based fluorescence spectroscopic measurements on agonist activated β2AR, which could not be explained by the agonist bound β2AR crystal structure. This study shows that the receptor activation depends on both the low energy states and the range of the conformations sampled by the receptor.



INTRODUCTION

The functionality of allosteric proteins to couple with various other proteins upon stimulation, stems from their capacity to sample and switch between various functional conformational states. There is considerable debate on the fundamental mechanisms underlying the function of allosteric proteins. There is growing evidence showing that allosteric proteins even in the absence of any stimulation such as ligand binding, sample an ensemble of conformational states, and that ligand binding largely leads to conformational selection from this ensemble.[1,2] G-protein coupled receptors (GPCRs) are seven-helical-transmembrane allosteric proteins that upon activation by extracellular signals, couple to the trimeric G-proteins or β-arrestin and transduce the signal from outside to inside the cell. GPCRs are activated by a variety of ligands ranging from photons to small molecules to large proteins. GPCRs are dynamic and known to adopt various active and inactive conformational states depending on the nature of the ligand that binds to the receptor, and the intracellular protein that couples to it.[3−5] The conformational changes effected by agonist and the G-protein binding lead to the activation of the receptor. Many GPCRs exhibit basal activity in the absence of any stimulation by ligands,[6] and therefore the speculation is that the conformational ensemble sampled by the receptor should contain the active state(s), even if the relative population of these states is low.

Understanding of the dynamics, the conformational ensemble, and the energy landscape that the GPCRs probe is crucial in translating the conformational sampling to functional efficacy.[7,8] This information is also vital in designing drugs with selectivity to a particular signaling pathway, known as "functional selectivity".[3,9,10] Activation of GPCRs occurs in a microsecond time scale, and a detailed mapping of the conformational transition pathway leading to activation is difficult to achieve with experiments because all of the intermediate states are short-lived. The leading questions are as follows: What does the conformational ensemble sampled by the receptor without any ligand (apoprotein) look like? Does ligand binding lead to conformational selection in GPCRs? Detailed all-atom molecular dynamics (MD) simulations are limited in the statistical information on conformational ensembles obtained from a single long MD trajectory due to the possible energy barriers between various intermediate states. Thus a multiscale simulation method that combines a coarse-grained technique sampling various kinetic states, with detailed all-atom simulations, is required for this task.

In this study we have used multiscale simulations on the human β2-adrenergic receptor (β2AR), a well-studied class A GPCR, and analyzed the conformation ensembles sampled by

Received: June 8, 2011
Published: July 18, 2011

the apoprotein, and with a full agonist norepinephrine, a partial agonist salbutamol, and inverse agonist carazolol bound. Starting from various receptor conformations sampled along previously calculated minimum energy pathways of activation for the β2AR[11] with the aforementioned ligands, we generated a swarm of all-atom MD simulation trajectories in explicit lipid and water totaling to ~5.8 μs.

Our results show that the apoprotein indeed samples a wider conformational space than when bound to any type of ligand. Ligand binding leads to selection of a subset of the apoprotein conformational ensemble. The crystal structure conformations of the active and inactive states of β2AR[12,13] are part of the conformational ensemble sampled by the apoprotein. However the relative population of these states is different when bound to full agonist, partial agonist, or inverse agonist. Agonist binding leads to multiple densely populated regions in the conformational ensemble compared to an inverse agonist. A combination of coarse-grain and fine-grain molecular dynamics studies suggest that the receptor activation not only depends on the lowest energy state sampled but also on the range of conformational sampling.

Principal component analysis (PCA) shows that the global movement in the apoprotein is dominated by the breathing motion of the orthosteric ligand binding site, and the shearing motion of helix 3, 5, and 6. The shear motion resembles the conformational changes observed in the crystal structure of the active state of β2AR.[12] Analysis of the interhelical contacts between well conserved (among class A GPCRs) residues, known as "functional microdomains"[14] in various receptor functional states show that these contacts function as tunable switches with continuous variation in distances rather than as a binary "on-off" switch upon activation.

RESULTS

Previously we had used a coarse-grained conformational sampling method to calculate the minimum energy pathway leading to activation of β2AR in the presence of norepinephrine and salbutamol. In this study, we performed all-atom MD simulations for four different types of systems, namely the β2AR apoprotein (without any ligand), β2AR bound to norepinephrine, salbutamol, and carazolol. As described in the Methods section, the 85 starting conformations of β2AR for the four different all-atom MD simulations were taken from various points along the respective minimum energy pathways of activation previously calculated for norepinephrine and salbutamol bound β2AR.[11]

Principal Component Analysis and Global Receptor Motions in β2AR. We performed principal component analysis (PCA) on the aggregated MD trajectories, to analyze the nature of the global motions in the receptor when bound to different ligands. The most dominant global motion is along the principal component 1 (PC1). PC1 represents a reorganization of the TM helices in the intracellular (IC) region, with a shearing movement of TM helices 3, 5, and 6 about a core formed by helices 1, 2, 4, and 7 as shown in Figure 1a. This shear movement leads to partial opening of the IC region between TM helices 3, 5, and 6, similar to the type of movement observed in the crystal structure of activated β2AR,[12] but to a lesser extent observed here. Principal component 2 (PC2) represents opening and closing of the ligand-binding pocket as shown in Figure 1b and is dominant only in the apoprotein. Hereafter, we refer to PC1 as "shear motion" and PC2 as the "breathing motion" in the text and figures.

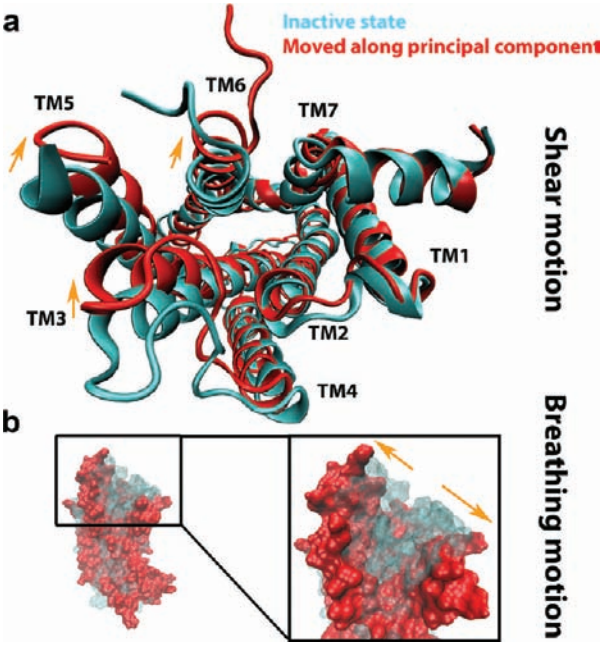


Figure 1. Representation of the first two principal components, the dominant motions, observed during the MD on human β2AR. (a) The intracellular view of the TM helices at the extremes of principal component 1. This is known as "shear motion" of the TM helices 3, 5, and 6 against the helices 1, 2, 4, and 7. (b) Breathing motion of the binding pocket; the transparent surface represents the closed state and the opaque surface represents the open state. This motion is captured by principal component 2.

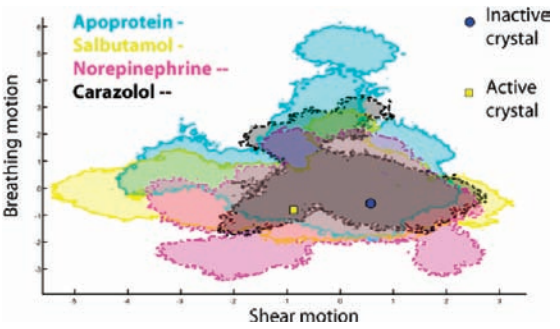


Figure 2. Conformational sampling along the first two principal coordinates with different ligands bound to the receptor. Crystal structures of carazolol bound (blue circle) (PDB: 2RH1) and agonist and nanobody bound (yellow square) (PDB: 3P0G) are also projected on this landscape as reference.

Conformation Selection by Ligands of β2AR. Projection of the MD trajectories on the two dominant principal components shows the extent of conformations sampled when the receptor is not bound to any ligand, and when bound to norepinephrine, salbutamol, or carazolol (Figure 2). We will refer to the carazolol−β2AR crystal structure[13] (pdb ID: 2RH1) as representative of the inactive state of the receptor, and the nanobody and agonist bound β2AR[12] as the active state (pdb ID: 3P0G). Figure 2 shows that the apoprotein (shown in cyan) samples a wider spread of conformations both in shearing and breathing modes, compared to the receptor bound to agonists and inverse agonist. The conformational space sampled by β2AR with bound

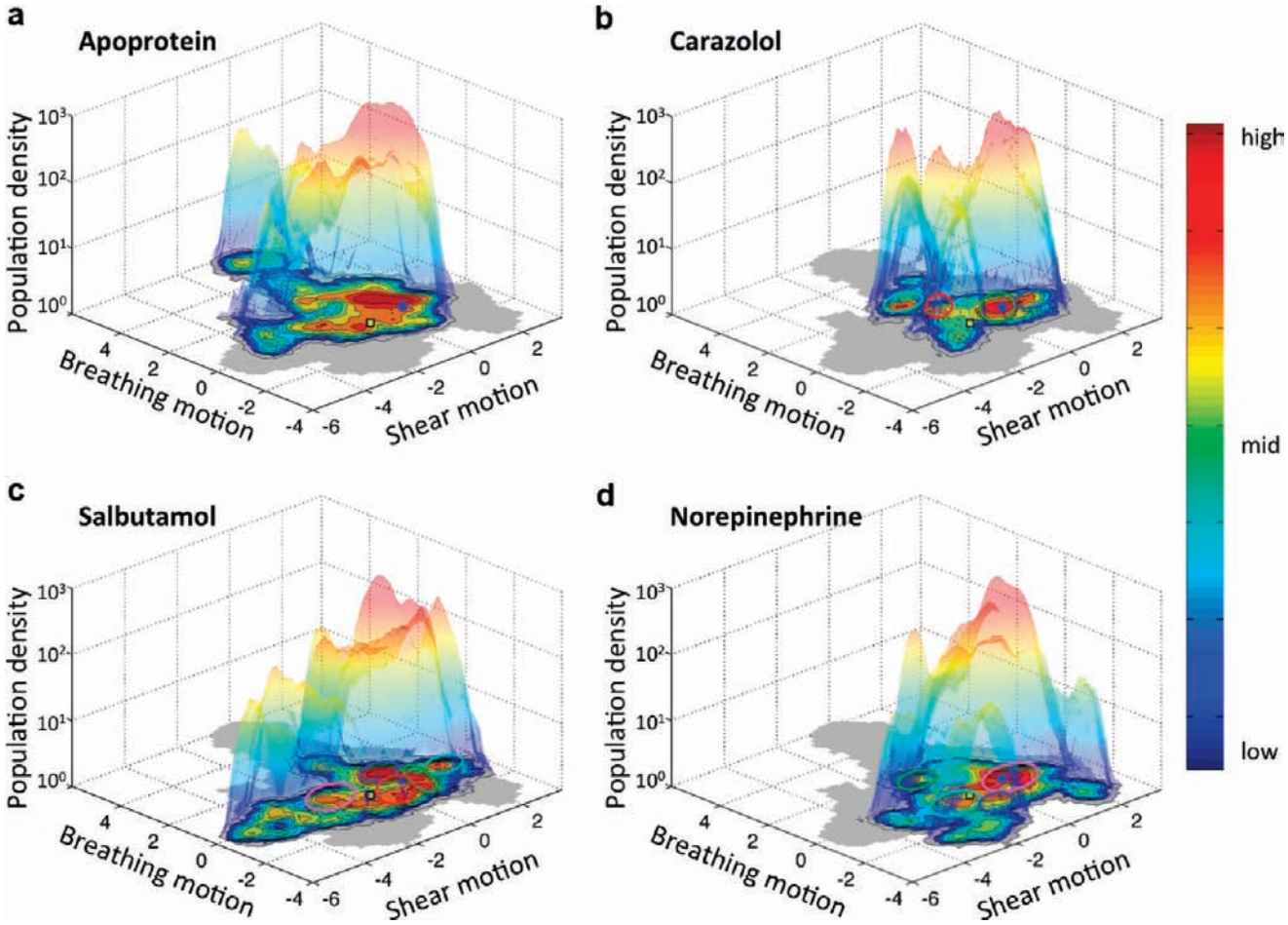


Figure 3. Population density (varies high to low red to blue) overlay of the conformational sampling of various ligand-bound receptors. Circles indicate densely populated clusters found in each case. The blue dot is the projection of the location of the inactive crystal structure (pdb ID: 2RH1) on this PC landscape, and the yellow square is the projection of the location of the active state crystal structure (pdb ID: 3P0G). The gray shaded region represents the conformational space accessed by the receptor over all simulations, with different ligands bound, combined.

agonists (Figure 2; full agonist norepinephrine shown in pink, partial agonist salbutamol shown in yellow) is a smaller subset of the space sampled by the apoprotein. Agonist binding retains a large range in the shear motion, while shrinking the breathing motion of the binding site. Binding of the inverse agonist carazolol to β2AR leads to shrinkage in both the breathing mode of the binding site and the shear motion of the intracellular part of the TM helices. Thus agonist binding leads to conformational selection that preserved the shear mode but reduced the breathing mode, and inverse agonist binding leads to selection of conformations that show reduction both in shear and breathing modes. There are conformations sampled by norepinephrine bound β2AR that are ligand induced (see the nonoverlapping regions between the pink and pale blue regions in Figure 2). However, the population density of these conformations is relatively low as we demonstrate in the next section. Projection of the conformational ensembles in the principal components 3 and 4 also show similar conformational selection by ligands (Supporting Information, Figure S1).

Agonists Sample Multiple Low Energy States Compared to the Inverse Agonist. Figures 3A to 3D show the calculated population density of the various conformational states sampled during the dynamics of the apoprotein, and with various ligands bound. The red regions in these figures are regions of high population and the blue regions are of low population. The gray shaded region at the floor of the three-dimensional plot represents the combined conformational space sampled by the apoprotein and all the ligand bound receptors. It is evident from Figure 3 that the total populated regions cover a larger area in the apoprotein compared to when ligands are bound. The apoprotein samples the widest range of densely populated ensembles, followed by both norepinephrine and salbutamol, while carazolol is the most selective (Supporting Information, Figure S2). For each ligand-bound receptor, the densely populated regions have been clustered (shown as circles in Figure 3), using population density cutoff as described in the Methods section. Population of these clusters and coordinate rmsd (backbone atoms in the TM region) between the average structure in each densely populated region and the crystal structures of inactive, and active β2AR have been calculated and shown in Table S2 of the Supporting Information.

Both norepinephrine- and salbutamol-bound receptor states show a wider range of densely populated clusters (red regions in Figure 3 enclosed by circles) more than that of carazolol. The relative free energies of the various clusters calculated from the population densities (as shown in Supporting Information, Figure S3) also confirm that the agonists sample more densely populated conformations than the inverse agonist.

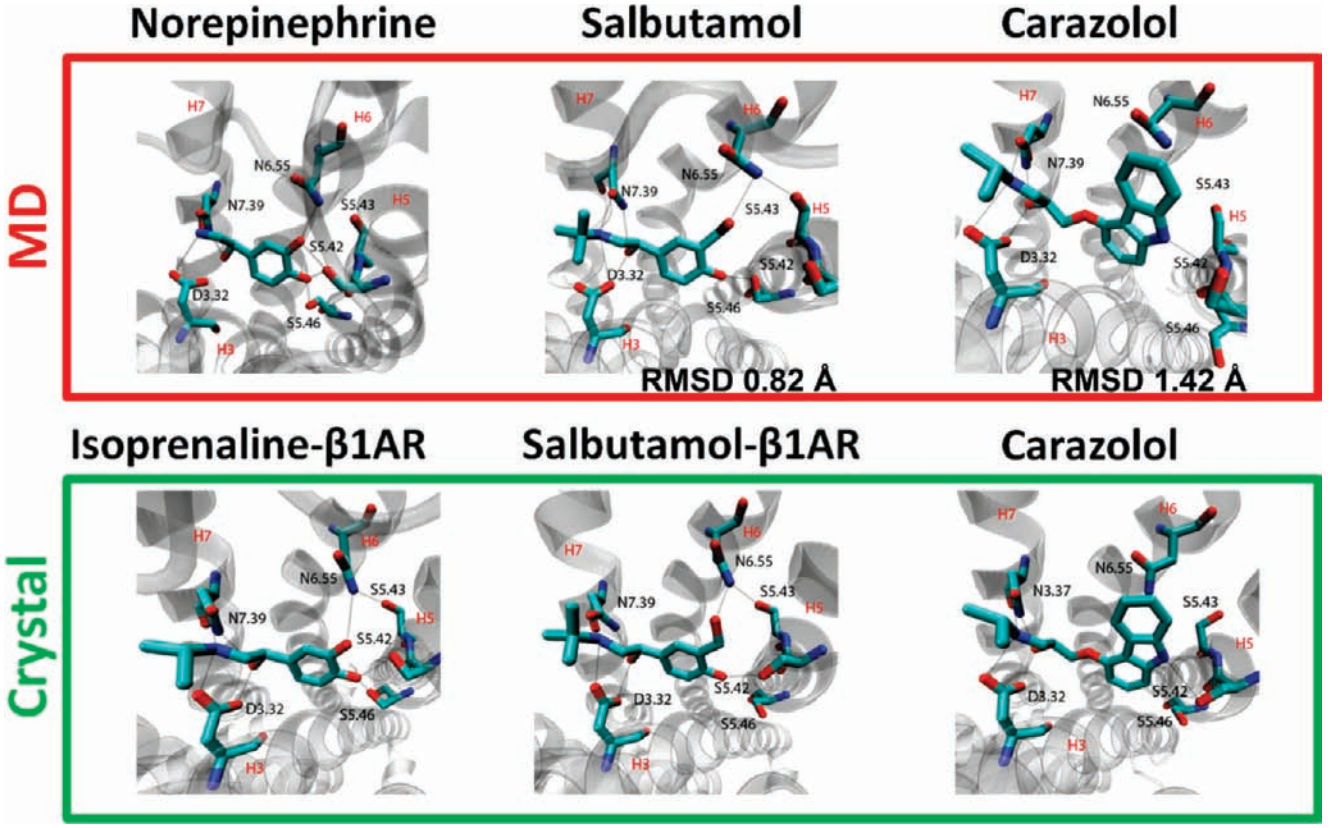


Figure 4. The most common ligand poses, as defined by ligand−receptor contacts, found during MD correspond well with known crystal structures. The rmsd values between the representative ligand pose from MD and the ligand pose in the corresponding crystal structure is also displayed. The rmsd between ligands was calculated after aligning the receptor using PyMol.

Agonist-Bound β2AR Crystal Structures Are a Part of the Ensemble of Conformational States. Carazolol stabilizes a densely populated ensemble of states near the inactive crystal structure as seen in Figure 3B. The average structure in this ensemble is 1.1 Å rmsd (backbone atoms in the TM region) (Table S1) from the inactive crystal structure and 2.2 Å from the active state crystal structure of β2AR (Figure 3B). Partial agonist salbutamol and agonist norepinephrine stabilize more than one densely populated conformational cluster (shown in Figures 3C and 3D, respectively). One of the densely populated norepinephrine bound ensembles is located close to the inactive crystal structure (1.1 Å rmsd). The average structure from another densely populated ensemble for the norepinephrine bound β2AR (marked as the cyan circle in Figure 3D) is 2.0 Å rmsd from the active state crystal structure of β2AR (shown as yellow square in Figure 3D). Partial agonist salbutamol also shows multiple densely populated regions, one of these near the crystal structure of inactive β2AR (1.2 Å rmsd). Although salbutamol bound β2AR ensemble conformations are separated from the active state crystal structure of β2AR by a backbone rmsd of 2.1 Å, they show the intracellular movements observed during activation as shown in Supporting Information, Figure S4. However this IC opening is not as wide as in the active state crystal structure of β2AR[12] or the opsin crystal structure.[15] It is closer in magnitude to the motion observed in the agonist-bound crystal structure of the adenosine receptor.[16]

Agonists Are Dynamic in Their Binding Sites Compared to Inverse Agonist. To analyze the diversity in ligand poses sampled during the MD simulations, we clustered distinct ligand poses by receptor ligand contacts, in the most densely populated receptor cluster for each type of ligand. We have compared the ligand poses to the corresponding crystal structures. The crystal structures used in this comparison are of carazolol in β2AR (pdb ID: 2RH1), salbutamol in β1AR[17] (pdb ID: 2Y04), and isoprenaline[17] (pdb ID: 2Y03) bound β1AR. We used salbutamol in β1AR since there is no crystal structure of β2AR with salbutamol-bound. The structure of isoprenaline in β1AR was used since it is close to norepinephrine. The agonist in the crystal structure of β2AR is not similar to norepinehrine or salbutamol, it does not have a catechol group and it is a much longer agonist, therefore we resorted to the crystal structures of β1AR for comparing the ligand binding sites. These comparisons will give further validation for our calculations. It should be noted that although the agonist is bound, the β1AR receptors in these crystal structures are still in the inactive state since the thermostable mutant used for crystallization has been designed to stabilize the inactive state of the receptor. Therefore the difference between the inverse agonist and agonist-bound crystal structures of β1AR is only in the ligand binding site.[17]

Figure 4 compares the binding site of the most dominant ligand pose from the densely populated ensemble for each ligand-bound receptor (MD panel in Figure 4) to the corresponding crystal structures (crystal panel). There were 46 out of 47 (98%) ligand residue contacts for norepinephrine compared to the crystal structure with isoprenaline, 39 out of 50 (78%) contacts for salbutamol, observed in the crystal structures captured in the predicted binding site obtained from the simulations. It should be noted that the crystal structures of β1AR with

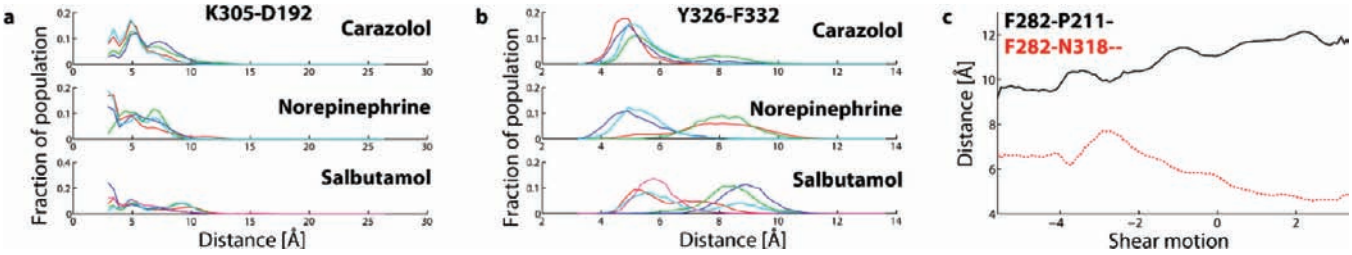


Figure 5. Dynamics of important interhelical contacts of the receptor for each ligand. Colors used correspond to the highly populated clusters as identified in Figure 3. (a) Salt bridge between K305$^{7.32}$ on TM7 and D192 on EC2. (b) Hydrophobic interaction between Y326$^{7.53}$ and F332 on TM8. (c) Change in rotamer of F282$^{6.45}$, flipping from TM7 (N318$^{7.47}$) toward TM5 (P211$^{5.50}$).

agonists bound were published after our simulations were performed. All the ligands make contacts with D113$^{3.32}$ on TM3 and N312$^{7.39}$ on TM7 through their positively charged protonated amine ion and the β-hydroxyl group, respectively. Here we have used GPCR specific residue numbering.[18] These residue contacts are common for many ligands (agonists, antagonists alike) in β2AR and the closely related β1AR. Agonists salbutamol and norepinephrine make contacts with the cluster of three serines on TM5, namely S203$^{5.42}$, S204$^{5.43}$, and S207$^{5.46}$. The hydrogen bonds with the serines are dynamic as they are made and broken during the MD simulations. We find that both norepinephrine and salbutamol are dynamic inside the binding site showing several distinct ligand poses, whereas the inverse agonist carazolol, samples just one pose and makes contact with S203$^{5.42}$ only and not the other two serines (see Figure 4). Supporting Information, Figure S5 shows various ligand poses observed in the MD simulations for the three ligands bound to β2AR.

Dynamics of Functional Microdomains. It is postulated that interhelical residue contacts between highly conserved residues across class A GPCRs, known as "functional microdomains" could serve as conformational switches for activation of the receptor. Here we have characterized the dynamic nature of some of the well-studied conformational switches in class A GPCR structures.[5,14,19,20] Supporting Information, Figure S6 shows the interhelical contacts studied here. Figure 5 shows the population of each of the microdomain distances in the receptor conformations extracted from the most populated ensembles for each ligand-receptor pair (as clustered in Figure 3). Bokoch *et al.* used solid state NMR measurements to show that agonist binding leads to weakening of the salt bridge between K305$^{7.32}$ and D192^{ECL2} in TM7 and extracellular loop 2, respectively.[21] Our MD results show that this salt bridge is indeed weakened and is dynamic in the simulations of β2AR with norepinephrine or salbutamol (Figure 5A). Another interhelical hydrophobic contact between Y326$^{7.53}$ and F332 on helix 8 that is formed in the inactive state of β2AR[13] is clearly broken in the active state.[12] The crystal structure of the active state shows that the side chain conformation of the aromatic residue F282$^{6.45}$ rotates, adopting a different side chain conformation closer to TM5 rather than close to that of TM7 as in the inactive state. In our simulations we see that this distance between F282$^{6.45}$ and TM7 (N318$^{7.47}$) increases (Figure 5B) when the shear motion of TM3, TM5, and TM6 occurs. The side chain rotamer of F282$^{6.45}$ moves closer to P211$^{5.50}$ on TM5. The so-called "ionic lock", a salt bridge between R131$^{3.50}$ and E268$^{6.30}$ found in the inactive state conformation of rhodopsin, is dynamic in the carazolol-bound β2AR ensembles (Supporting Information, Figure S7). It is made and broken even in the inactive state unlike rhodopsin. Previous MD simulations on the crystal structure of carazolol bound β2AR also showed that the ionic lock is dynamic, thus explaining the basal activity of β2AR.[22] Importantly, we observe that the interhelical conformational switches exist in multistate rather than a two-state "on-and off" switch as previously thought,[20] with the contact between Y326$^{7.53}$ and F332 as the only exception.

Role of Water in the Dynamics of β2AR. Using fluorescence-labeled experiments on purified β2AR, Swaminath et al. showed that the IC edge of TM6 shows changes in conformation and is more solvent-exposed upon full agonist binding.[23] Yao and coworkers also showed a red shift (indicating a conformational change) accompanied by a decrease in fluorescence intensity in purified β2AR reconstituted into lipid HDL particles, upon exposure to an agonist.[4] However, the recently published crystal structure of an irreversible agonist FAUC50-bound β2AR shows no such conformational changes in the IC edge of TM6 where the fluorescence label was placed.[24] To examine this ambiguity, we calculated the solvent accessibility (number of water molecules within 4 Å) around C265 (located in the IC region of TM6) to examine the changes in solvent exposure of the vicinity of this residue that was labeled in the fluorescence experiments. Figure 6a shows the regions of the norepinephrine-bound receptor that are more (or less) hydrated than the apoprotein, during the MD simulations. As expected, ligand binding leads to removal of some of the waters from the binding site. This is quantified in the average number of water molecules near D113$^{3.32}$, which is higher in the apoprotein compared to in the ligand-bound structures as shown in Figure 6b. The IC regions around residue C265 are more water-exposed during the dynamics with the agonist-bound β2AR compared to those in simulations with apoprotein or inverse-agonist bound. Figure 6c shows that the average number of water molecules within 4 Å of residues near C265 increases markedly upon agonist binding compared to that of the apoprotein or carazolol bound β2AR. The receptor dynamics leading to more water exposure of C265 explains the fluorescence intensity changes upon agonist binding observed in the fluorescence experiments in solution.

■ DISCUSSION

Principal component analysis of the aggregated MD trajectories indicates that a large conformational space is accessible to the apoprotein, both in the opening of the binding site as well as reorganization of the TM helices even without a ligand being present. This is similar to the reorganization of the intracellular regions of TM3, TM5, and TM6 observed in the crystal structure of the opsin apoprotein.[15] The shearing motion that we observe could modulate G-protein binding and is therefore indicative of

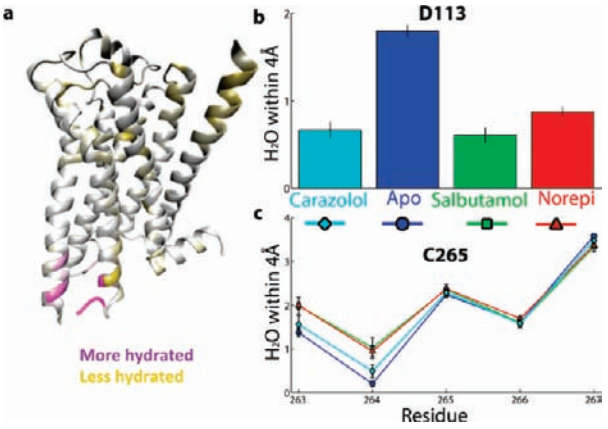


Figure 6. (a) Representation of more hydrated (magenta) and less hydrated (yellow) regions in the receptor upon agonist binding. (b) Residue D113$^{3.32}$ in the ligand binding site is significantly less hydrated in the presence of any ligand as compared to to that in the presence of apoprotein. (c) Residues 263 and 264 in the proximity of C265 at the bottom of TM6 are significantly more hydrated when bound to an agonist as compared to when bound to an inverse agonist or the apoprotein. Significant difference means p-value < 0.05 when cross-comparing the ensemble of simulations for each ligand−receptor combination.

the basal activity of the receptor. The large breathing motion could facilitate the diffusion of ligands into the binding cavity. Agonist binding retains a large range in the shear motion, while shrinking the breathing motion of the binding site. This is in agreement with the experimental observation from spin labeling experiments,[25] fluorescence spectroscopy,[23] X-ray crystallography,[16] and solid state NMR[19] on class A GPCRs that show agonist binding leads to reorganization of the intracellular domain of TM3, TM5, and TM6. Binding of the inverse agonist carazolol leads to the reduction of both breathing motion of the binding pocket and shear motion at the intracellular interface. This is in keeping with the reduced basal activity of the carazolol-bound β2AR.[13] The dynamics of the cytoplasmic interface as exemplified in the shear motion could facilitate G-protein coupling in the apoprotein and agonist-bound receptors, whereas limiting this motion could inhibit receptor signaling while the receptor is bound to the inverse agonist. The cellular environment could further modulate the extent of the shearing and breathing motions leading to different levels of coupling of G-protein to the receptor.

In the conformational space of the PCs, the inverse agonist carazolol stabilizes the fewest receptor conformations (Supporting Information, Figure S2) with the most populated conformational cluster centered on the inactive crystal (Figure 3 and Supporting Information, Table S2), whereas the agonist norepinephrine and partial agonist salbutamol stabilize a wider range of conformations that can be divided into several densely populated clusters. It is interesting to note that one of these densely populated norepinephrine-bound clusters is still located near the inactive crystal structure (1.1 Å rmsd), demonstrating that the agonist-bound receptor still favorably samples the inactive state conformation. This is also reflected in the recent crystal structure of the agonist-bound β2AR (covalently bound agonist without nanobody) that shows no movement in the IC region of TM5 and -6.[24] Both norepinephrine and salbutamol-bound receptors sample ensembles of conformations that show partial opening of the IC interface (average rmsd from active β2AR crystal $\approx$ 2 Å) facilitated by outward tilt of TM6. This tilting movement of TM6 is less severe compared to that of the crystal structure of active β2AR and is more similar to the tilt observed in the agonist-bound adenosine receptor crystal structure. The extent of TM6 movement is different in the crystal structures of opsin, adenosine receptor with agonist bound, and the active state crystal structure of β2AR, underlining the fact that TM6 is dynamic. It is possible that the fully active receptor conformation featuring greater cytoplasmic displacement of TM6 is stabilized in β-ARs only in the presence of G-protein.

We have demonstrated that a multiscale simulation method with a coarse-grained sampling method mapping the minimum energy pathway in going from the inactive to the active state followed by all-atom MD simulations is required for comprehensive conformational sampling of GPCRs. A single long time scale all-atom MD simulation starting from the inactive conformational state of a GPCR does not provide a comprehensive sampling of states because of the possible activation energy barriers in going between the intermediate kinetic states.

We have shown that the receptor without any ligand shows dominant global motion in the breathing of the orthosteric ligand binding site and also shearing motion of TM helices 3, 5, and 6. This shear mode resembles the conformational changes that lead to receptor activation as observed in the crystal structure of the active state of β2AR. The binding of the full agonist norepinephrine, and partial agonist salbutamol, leads to conformational selection of a subset of conformations sampled by the apoprotein. This subset of conformations show reduced binding site breathing motion, but retain the range of motion in the shear mode. The binding of the inverse agonist carazolol leads to reduced breathing and shear modes. Agonist binding leads to multiple densely populated conformational states, one of which is the inactive state of the receptor. The densely populated regions of the agonist-bound structures compare well to those of the crystal structures of agonist-bound β1AR, which have been published subsequently. Specifically, conformations from the densely populated ensemble show binding sites similar to the crystal structure of the agonist bound β1AR. Using MD simulations we were able to explain the dynamical conformational changes observed in the solution-based fluorescence spectroscopic measurements on agonist activated β2AR, which could not be explained by the agonist bound β2AR crystal structure. These results illustrate that receptor activation is determined not only by the lowest energy conformational ensemble but also by the range of conformational states explored in this ensemble. The conformational switches leading to activation in the functional microdomains are multistate switches rather than two-state switches as thought previously. As more crystal structures of GPCRs emerge, this work sets the framework for using multiscale methods for analysis of activation in other GPCRs.

METHODS

Clustering of Receptor Conformations from Coarse-Grain Simulations. β2AR conformations from the coarse grained Monte Carlo simulations starting from the inactive state crystal structure of β2AR, from our previous study[11] were clustered using the k-means clustering algorithm in Matlab. To study the activation of agonist-bound β2AR, we calculated the minimum energy pathway between the inactive and active (agonist-stabilized states) β2AR conformations along the coarse-grained binding energy landscape obtained from the LITiCon

method.[11] The minimum energy pathway was computed using a biased Monte Carlo (MC) procedure, where the acceptance of each MC move was slightly biased toward the final active state, to ensure efficient convergence. The receptor conformations sampled along the activation pathway using the MC scheme were clustered and used as starting conformations for the fine-grained MD simulations. More details on the method is given in the Supporting Information. Once all coarse-grained conformations were clustered, one representative conformation (nearest to cluster center) from each cluster was selected as the starting conformation for the MD simulations.

MD Simulation. The receptor conformations obtained from the coarse-grained simulations were solvated in explicit palmitoyl oleoyl phosphatidyl choline (POPC) lipid and water. The lipid was packed around the protein using the inflategro package in GROMACS.[26] SPC water molecules[27] were added on both sides of the lipid bilayer, and four chloride ions were added in the water by displacing random water molecules to neutralize the system. The final systems were minimized by steepest descent energy minimization with a maximum force of 1000 kJ/mol/nm as convergence criterion; protein and ligand were kept fixed using position restraints. MD simulations on β2AR in a POPC lipid bilayer were performed using GROMACS 4.0.5[26] and the GROMOS96 53a6 Forcefield[28] extended with Berger lipid parameters.[29] Short-range nonbonded interactions were truncated at 1.2 nm, with the neighbor list updated every 10 fs. To account for the cutoff in the van der Waals interactions, long-range dispersion correction was applied to energy and pressure terms. Long-range electrostatics was calculated using the smooth particle mesh Ewald (PME) method. Bonds were constrained using the LINCS algorithm to allow for a simulation step size of 2 fs. For continuity, periodic boundary conditions were applied in every dimension.

Each system was equilibrated by performing 100 ps of MD at 310 K using a NVT ensemble followed by 5 ns of MD under NPT conditions with a pressure of 1 bar. The velocity rescaling thermostat was used for temperature coupling during the equilibration, and a Parrinello–Rahman barostat[30] was used for pressure coupling. The protein and ligand were kept in place during these equilibration steps using position restraints of 1000 kJ/mol/nm^2. After the systems were equilibrated at the correct temperature and pressure, MD simulations of up to 100 ns were performed for each of the 85 unique initial conformations, using a NVT ensemble with a Nosé–Hoover thermostat.[31] Data were not collected for the first 5 ns of MD.

Principal Component Analysis of MD Trajectories. The most important receptor motions can be described in just a few principal modes.[32,33] To study the dominant conformational changes sampled during MD, principal component analysis (PCA) was performed on the net conformational ensemble obtained by combining the individual MD trajectories using GROMACS. Only backbone atoms in TM2–7 (Supporting Information, Table S1) were included in the analysis, this to reduce noise by the highly flexible loops and TM1. Eigenvalues and eigenvectors were calculated from the covariance matrix, using the g_covar command in GROMACS. The calculated eigenvalues directly correspond to the fraction of the variance in receptor conformation represented by the corresponding eigenvector.[33] Quantification of the variance in receptor conformation covered by the most dominant PCs is given in Supporting Information Figure S8. To determine the major domain motions, conformational changes along the top four principal components (PC) were analyzed. The first two PCs described the breathing and shearing motions as discussed in the "results and discussion" section. Ligand binding leads to conformational selection in the third and the fourth PCs as shown in Supporting Information Figure S1.

Identification of dominant ligand poses in the PC coordinates: Within the conformational ensemble obtained for each ligand-bound receptor, common ligand poses were identified based on contact patterns observed between polar ligand atoms and polar protein atoms in the binding site. Here, a contact is defined by a distance cutoff of 3.5 Å. We ranked ligand-protein contacts by occurrence and selected common contacts. Thus, we can distinguish among many similar ligand poses that are close in rmsd but differ in protein–ligand contacts.

Analysis of Microdomains and Water. Microdomains were analyzed by inter-residue distances extracted every 10 ps from the MD trajectories. A figure of the microdomains analyzed is given in Supporting Information Figure S6. To determine the level of hydration or water exposure for each chosen residue, we calculated the average number of water molecules within 4 Å of each residue in the receptor; this was done using the g_rdf function in GROMACS.

■ ASSOCIATED CONTENT

Supporting Information. Complete refs 12, 21, and 24. Methods for the coarse-grained simulations, selection of initial conformations, generation of initial conformations in explicit solvent and calculation of ligand-protein contacts. Tables (1) defining the residues in the protein included in PCA and rmsd analysis, and (2) providing further information on densely populated conformational clusters. Figures showing (1) PC 3 and 4, (2) conformational selectivity of ligands, (3) relative free energy landscapes for ligand bound receptor, (4) comparison of the IC movement in MD with crystal structures, (5) various ligand poses found in our simulations, (6) location of the functional microdomains that were analyzed, (7) DRY salt bridge behavior, (8) quantification of PC contributions, (9) clustering of the coarse-grained simulations and (10), depiction of the criterion used to determine clustering was sufficient. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
nvaidehi@coh.org

■ ACKNOWLEDGMENT

This research was supported in part by Boehringer-Ingelheim and by the National Science Foundation through TeraGrid resources under Grant Number TG-MCB110030. We thank Dr. Christofer Tautermann for helpful discussions.

■ REFERENCES

(1) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789.
(2) Csermely, P.; Palotai, R.; Nussinov, R. *Trends Biochem. Sci.* **2010**, *35*, 539.
(3) Urban, J. D.; Clarke, W. P.; von Zastrow, M.; Nichols, D. E.; Kobilka, B.; Weinstein, H.; Javitch, J. A.; Roth, B. L.; Christopoulos, A.; Sexton, P. M.; Miller, K. J.; Spedding, M.; Mailman, R. B. *J. Pharmacol. Exp. Ther.* **2007**, *320*, 1.
(4) Yao, X. J.; Velez Ruiz, G.; Whorton, M. R.; Rasmussen, S. G.; DeVree, B. T.; Deupi, X.; Sunahara, R. K.; Kobilka, B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 9501.
(5) Bhattacharya, S.; Hall, S. E.; Li, H.; Vaidehi, N. *Biophys. J.* **2008**, *94*, 2027.
(6) Kobilka, B. K.; Deupi, X. *Trends Pharmacol. Sci.* **2007**, *28*, 397.
(7) Vaidehi, N.; Kenakin, T. *Curr. Opin. Pharmacol.* **2010**, *10*, 775.
(8) Deupi, X.; Kobilka, B. K. *Physiology (Bethesda)* **2010**, *25*, 293.
(9) Kenakin, T. *Mol. Pharmacol.* **2007**, *72*, 1393.
(10) Leduc, M.; Breton, B.; Gales, C.; Le Gouill, C.; Bouvier, M.; Chemtob, S.; Heveker, N. *J. Pharmacol. Exp. Ther.* **2009**, *331*, 297.

(11) Bhattacharya, S.; Vaidehi, N. *J. Am. Chem. Soc.* **2010**, *132*, 5205.
(12) Rasmussen, S. G.;; Choi, H. J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; Devree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S.; Schnapp, A.; Konetzki, I.; Sunahara, R. K.; Gellman, S. H.; Pautsch, A.; *Nature* 2011, *469*, 175.
(13) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. *Science* **2007**, *318*, 1258.
(14) Khelashvili, G.; Grossfield, A.; Feller, S. E.; Pitman, M. C.; Weinstein, H. *Proteins* **2009**, *76*, 403.
(15) Park, J. H.; Scheerer, P.; Hofmann, K. P.; Choe, H. W.; Ernst, O. P. *Nature* **2008**, *454*, 183.
(16) Xu, F.; Wu, H.; Katritch, V.; Han, G. W.; Jacobson, K. A.; Gao, Z. G.; Cherezov, V.; Stevens, R. C. *Science* **2011**, *332*, 322.
(17) Warne, T.; Moukhametzianov, R.; Baker, J. G.; Nehme, R.; Edwards, P. C.; Leslie, A. G.; Schertler, G. F.; Tate, C. G. *Nature* **2011**, *469*, 241.
(18) Ballesteros, J. A.; Weinstein, H. *Methods Neurosci.* **1995**, *25*, 366.
(19) Ahuja, S.; Smith, S. O. *Trends Pharmacol. Sci.* **2009**, *30*, 494.
(20) Yao, X.; Parnot, C.; Deupi, X.; Ratnala, V. R.; Swaminath, G.; Farrens, D.; Kobilka, B. *Nat Chem Biol* **2006**, *2*, 417.
(21) Bokoch, M. P.; Zou, Y.; Rasmussen, S. G.; Liu, C. W.; Nygaard, R.; Rosenbaum, D. M.; Fung , J. J.; Choi, H. J.; Thian, F. S.; Kobilka, T. S.; Puglisi, J. D.; Weis, W. I.; Pardo, L.; Prosser, R. S. *Nature* 2010, *463*, 108.
(22) Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M. O.; Piana, S.; Shaw, D. E.; et al. *Proc Natl Acad Sci U S A* **2009**, *106*, 4689.
(23) Swaminath, G.; Xiang, Y.; Lee, T. W.; Steenhuis, J.; Parnot, C.; Kobilka, B. K. *J. Biol. Chem.* **2004**, *279*, 686.
(24) Rosenbaum, D. M.; Zhang, C.; Lyons, J. A.; Holl, R.; Aragao, D.; Arlow, D. H.; Rasmussen, S. G.; Choi, H. J.; Devree, B. T.; Sunahara, R. K.; Chae, P. S.; Gellman, S. H.; Dror, R. O.; Shaw, D. E.; Weis, W. I.; et al. *Nature* 2011, *469*, 236.
(25) Farrens, D. L.; Altenbach, C.; Yang, K.; Hubbell, W. L.; Khorana, H. G. *Science* **1996**, *274*, 768.
(26) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701.
(27) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J.Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Reidel: Dordrecht, The Netherlands, 1981.
(28) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656.
(29) Berger, O.; Edholm, O.; Jahnig, F. *Biophys. J.* **1997**, *72*, 2002.
(30) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.
(31) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.
(32) Yang, L. W.; Eyal, E.; Chennubhotla, C.; Jee, J.; Gronenborn, A. M.; Bahar, I. *Structure* **2007**, *15*, 741.
(33) Yang, L.; Song, G.; Carriquiry, A.; Jernigan, R. L. *Structure* **2008**, *16*, 321.