Published in final edited form as:
Proteins. 2014 October ; 82(10): 2538–2551. doi:10.1002/prot.24617.

Structure-Based Simulations Reveal Concerted Dynamics of GPCR Activation

Nicholas Leioatts, **Pooja Suresh**, **Tod D. Romo**, and **Alan Grossfield**
Department of Biochemistry & Biophysics, University of Rochester Medical Center, Rochester, NY, 14642, USA

Alan Grossfield: alan_grossfield@urmc.rochester.edu

Abstract

G protein-coupled receptors (GPCRs) are a vital class of proteins that transduce biological signals across the cell membrane. However, their allosteric activation mechanism is not fully understood; crystal structures of active and inactive receptors have been reported, but the functional pathway between these two states remains elusive. Here, we employ structure-based (Gō-like) models to simulate activation of two GPCRs, rhodopsin and the β_2 adrenergic receptor (β_2AR). We used data-derived reaction coordinates that capture the activation mechanism for both proteins, showing that activation proceeds through quantitatively different paths in the two systems. Both reaction coordinates are determined from the dominant concerted motions in the simulations so the technique is broadly applicable. There were two surprising results. First, the main structural changes in the simulations were distributed throughout the transmembrane bundle, and not localized to the obvious areas of interest, such as the intracellular portion of helix 6. Second, the activation (and deactivation) paths were distinctly non-monotonic, populating states that were not simply interpolations between the inactive and active structures. These transitions also suggest a functional explanation for β_2AR's basal activity: it can proceed through a more broadly defined path during the observed transitions.

Keywords

G protein-coupled receptors; rhodopsin; adrenergic receptor; structural transitions; signal transduction

1 Introduction

G protein-coupled receptors (GPCRs) are a class of biomedically important membrane proteins with roles in diverse functions such as sight, muscle contraction, and gene transcription. They are also targeted by an estimated 25–50% of FDA approved drugs[1–3]. These membrane proteins function as allosteric transducers, responding to environmental stimuli and activating signal pathways inside the cell.

Supporting Information
Supporting information includes two movies, twelve additional figures, four tables, and additional methodological details.

Many class A GPCRs are hypothesized to contain a number of microswitches comprised of conserved structural motifs[4–10]. These include the CxWP motif in the protein's hydrophobic core, the E/DRY region on transmembrane helix 3 (TM3), and the NPxxY motif on helix 7 (TM7). The latter two features are both on the intracellular side of the receptor, adjacent to the G protein binding site. It has been suggested that these conserved regions play key roles in GPCR activation. The CxWP region is hypothesized to act as a toggle, or "transmission" switch, interacting directly with the ligand and coupling the binding pocket to the protein at large[8,11].

The outstanding efforts of many labs have yielded high resolution structures of these proteins, revealing the dominant conformations of both the active and inactive states for rhodopsin and the β_2 adrenergic receptor (β_2AR)[12–20]. These structures are vital to understanding GPCR signaling, and together with dynamics studies are advancing our understanding of GPCR function[8,9,21–31]. The emerging image is that functionally important regions of the protein are allosterically coupled to one another loosely[6,9,32], rather than in a specific, ordered, and linear cascade. While the end states of these motifs are experimentally characterized at high resolution[4–8,16,33–35], there is still much to learn about how these states interconvert.

Molecular simulation can provide insights into protein dynamics at unparalleled resolution; no experimental method can track the position and interactions of every atom with angstrom and picosecond resolution. However, current computational resources limit the amount of data that can be collected, often resulting in a paucity of sampling[22,36–38]. In order to obtain a statistically representative ensemble, a number of methods have been developed to allow faster sampling, such as replica exchange and metadynamics[39,40].

Alternatively, one can use a simpler, less computationally expensive model to describe the system. One such group of methods bases the forcefield itself on structural data. Structure-based models of dynamics construct a potential based on a particular set of initial coordinates. These models were pioneered by Gō and coworkers to study protein folding[41,42]. Since then, they have undergone many refinements aimed at better-characterizing biomolecules and extending their resolution[43–45]. These and similar models have also been used for studying functional transitions[46–54].

In the current study, we model the activation of two canonical GPCRs, rhodopsin and β_2AR, using the structure-based potential described by Whitford et al [44,45]. We use both experimentally motivated and data-derived reaction coordinates to describe the structural transitions. The biological meaning of the data-derived coordinates is discussed and these coordinates are used to quantify the differences between β_2AR and rhodopsin activation pathways.

2 Methods

2.1 Structure-based models

These simulations use a Gō-like force field, where the only favorable long-range interactions are those present in the native structure. As a result, the native state is stable even in the

absence of water or lipids. In this study, we used an atomistic variant of this method developed by Onuchic and coworkers[44,45] utilizing the web-based parametrization tools[55]. The Hamiltonian was described using the "Shadow Map" algorithm developed in Noel et al[56]. This model provided all-heavy-atom resolution and made extensive sampling feasible, allowing us to carefully quantify the statistical error in our dataset. We performed these calculations using two proteins, rhodopsin and β_2AR, each in both the active and inactive states; the structures used are identified in Table I. All coordinates were obtained from the protein data bank (www.pdb.org). There are several different structures of both proteins to choose from, so we chose active/inactive pairs that were most similar in terms of mutations and the regions experimentally resolved. In these simulations no attempt was made to model the membrane; rather, we relied solely on the conformation present in the crystal structures. The Hamiltonian was defined by the starting structure, as expected with a Gō-model and the protein did not unfold in any simulations.

2.1.1 Simulating structural transitions with Gō models—To simulate a structural transition, we began by parametrizing a Gō-like potential for the target state, e.g. the active form of rhodopsin. We ran simulations using that potential but starting from the other state, in this case the inactive structure. As a result, the protein efficiently undergoes a transition from the inactive to the active form. One important caveat is that these simulations are not in equilibrium, since the starting structure is unlikely and thus over represented statistically. This strategy is similar to that used by Sanbonmatsu, Onuchic, and coworkers in their work on the ribosome[57].

2.1.2 Simulation protocol—After the initial system setup, the protein underwent an energy minimization scheme using the steepest decent and then conjugate gradient methods, in order to reduce forces sufficiently and allow stable dynamics; this typically required 1500–3000 steps of minimization in total. We then ran 300,000 steps of dynamics using a timestep of 0.0001 and a temperature of 50; as previously noted,[44,45] the potential defined by SMOG does not use conventional units. Simulations were performed using gromacs 4.5.4[58]. For each system (β_2AR and rhodopsin) and each simulation type (activation and deactivation) an ensemble of 1,000 trajectories was produced, for a total of 4,000 trajectories.

2.2 Simulation analysis

Custom analysis tools were written using the LOOS framework (version 2.0.5, http://loos.sourceforge.net). LOOS is an extensible, object-oriented analysis toolkit designed to interface with the native file formats of all major simulation packages[59]. Principal components were calculated using singular value decomposition[60]. Data was plotted with gnuplot (version 4.6, http://www.gnuplot.info), structures and movies were made with LOOS and rendered in PyMol (version 1.6, http://www.pymol.org).

2.2.1 Comparing two eigensets—To quantify the similarity between two PCA results, we used the covariance overlap method (Equation 1), first derived by Hess[61]. This is a normalized metric for comparing two eigensets, where 0 indicates complete orthogonality and 1 indicates the data is identical. Importantly, the covariance overlap takes both the

eigenvectors (directions of motion) and the eigenvalues (amplitudes of motion) into account and has been used to compare long-timescale conformation changes[38,62–66]. It is defined here as $\Omega_{A,B}$ where *A* and *B* represent the two eigensets being compared:

$$\Omega_{A,B}=1-\left[\frac{\sum_{i}^{N}(\lambda_i^A+\lambda_i^B)-\sum_{i}^{N}\sum_{j}^{N}\sqrt{\lambda_i^A\lambda_j^B}(\vec{\nu}_i^A\cdot\vec{\nu}_j^B)^2}{\sum_{i}^{N}(\lambda_i^A+\lambda_i^B)}\right]^{1/2} \quad (1)$$

Here, λ represents a single eigenvalue associated with the eigenvector, ν. This value is summed over all N normal modes.

2.2.2 Contact-based reaction coordinate—In order to define a reaction coordinate, we looked specifically to the unique set of contacts that changed between the active and inactive crystal structures. A unique contact is defined as one that is present in one state but not the other. For the current study, we defined two residues to be in contact when the centroids of their sidechains were within 8 Å of each other. This analysis included all residues in the transmembrane bundle that were resolved in both the active and inactive crystal structures (see Table II). This set of contacts—both those broken and those formed—was monitored throughout each simulation. We constructed a matrix made from the timeseries of each contact throughout the simulation. The value of each cell was defined by:

$$u_{ij}(t)=\frac{1}{2}\tanh(r_{ij}(t)-8)+\frac{1}{2} \quad (2)$$

where r_{ij} is a the distance between residues *i* and *j*, at time *t*. The tanh function makes a smooth transition between formed (value ≈ 0) and broken (value ≈ 1) states. By contrast, if a simple binary function is used, the variance of the contact matrix is dominated by the fluctuations of residue pairs that sit near the 8 Å cutoff in one of the crystal structures. We then performed principal component analysis (PCA) in contact space using singular value decomposition, and used the first two principal components as our reaction coordinate.

2.3 Statistical convergence

Using the principal components as reaction coordinates is only reasonable if they are robust across the dataset. In order to assess how many transition trajectories are required to confidently describe the pathway, we employed a bootstrap estimate of the variance in principal components. This was done by first varying the sample size from 5 to all 1,000 trajectories for a given system (see Section 2.1.2). For each sample, a set of 30 independent bootstrap datasets were created. Cartesian space PCA (using the alpha-carbons of the transmembrane helices) and contact space PCA were then performed on each of these sets. We then assessed the self-consistency of the modes produced as a function of data set size. This was done by computing the average absolute dot product of a single mode across all 30 bootstrap samples. The results, plotted in Supplementary Figures S1 and S3, show that PC1 was extremely well-defined in both the cartesian space and the contact space results; however, PC2 was heterogeneous in the cartesian space PCA, suggesting noise or

degeneracy in the paths. By contrast, both PCs were well-defined when using contact space PCA.

3 Results and Discussion

3.1 Dynamics of biologically conserved regions tell a conflicting story

We first set out to find a suitable reaction coordinate for activation by following the approach used by Dror et al.[67], analyzing the dynamics of two highly conserved regions: the NPxxY motif in transmembrane helix 7 (TM7), and the ionic lock (between TM3 and TM6). These regions are illustrated in Figure 1. Briefly, the tyrosine residue in the former motif is thought to disrupt the inter-helical hydrogen bonding network and stabilize the active conformation[4,15,16,19,68,69], while it has been suggested that R3.50 forms a salt bridge with E6.30 (the so-called ionic lock) that stabilizes the inactive state[70,71]. Both of these regions differ between the active and inactive forms of the receptor and have been implicated in the GPCR activation mechanism; however, their connection in the allosteric cascade is not fully understood[5–8,33,69,70].

Recently, Dror et al.[67] used analyses of these two regions as a reaction co-ordinate to describe β_2AR deactivation. They suggested deactivation occurs by a two-stage process, where the NPxxY region transitions to a more inactive-like structure first, followed by the ionic lock closing as TM6 moves toward the rest of the helical bundle. We performed a similar analysis on our structure-based simulations (Figure 2). Here, we measured both the NPxxY region RMSD from the inactive crystal structure and the distance between ionic lock residues as a function of time. The path taken by β_2AR during activation (Figure 2a) shows the NPxxY region beginning to deviate from its inactive form before the lock is fully open. This order was actually the opposite of that reported for β_2AR deactivation [67], but matched simulations of M2 muscarinic receptor activation performed using accelerated MD[72]. By contrast, the path taken in our structure-based simulationsof β_2AR deactivation (Figure 2b) is similar to that reported by Dror et al [67]. During deactivation, the NPxxY motif assumed a more inactive-like conformation before the ionic lock is formed. This was an encouraging result suggesting that this simple model produces dynamics similar to state-of-art all-atom results.

The difference in transition order (as seen between Figures 2a and 2b) suggests that the paths of activation and deactivation are not necessarily the same. There are two likely explanations for this difference. One may be a simple kinetic issue; the NPxxY transitions are local rearrangements and thus can occur quickly, while the ionic lock reports on the larger structural transition of a whole helix, and thus proceeds more slowly. Since our simulations begin in a non-equilibrium state, relaxation kinetics (as opposed to equilibrium fluctuations) control the pathway, and as such the paths can be different. Alternatively, it is possible that the activation and deactivation paths differ because different potentials are used in the two sets of simulations.

The results were quite different for rhodopsin (Figure 2c, d). The ionic lock made a clear transition in both activation and deactivation simulations (Panels c and d), but there was no

notable transition in the RMSD of the NPxxY region; in both the activation and deactivation simulations, the NPxxY motif prefers to stay near the inactive conformation.

The fact that this coordinate does not better capture the activation mechanism of rhodopsin was quite intriguing. While β_2AR and rhodopsin have similar structures, and this region is well-conserved, they perform diverse tasks and are activated in distinct ways. That these two regions worked well as a reaction coordinate for β_2AR, but not rhodopsin, and that β_2AR's activation path is qualitatively different from its deactivation path emphasizes the importance of choosing a reaction coordinate carefully and quantitatively analyzing the paths described. Proteins are high-dimension objects, and projecting their motions into a small number of reaction coordinates is both necessary and risky; a good choice will allow intuitive understanding of the results, while a poor choice may be actively misleading.

3.2 Principal component-based reaction coordinate reveals non-monotonic transition pathway

3.2.1 Transitions identified by cartesian PCA—Arguably the best way to be sure the reaction coordinate captures the information from the simulation is to derive the coordinate directly from the data. In this case, we used principal component analysis (PCA) to extract the most concerted motions in the simulations. We first applied PCA to the cartesian coordinates of the transmembrane c_α's[60,73–77]. Figure 3 shows the result of projecting the trajectories along the first two principal components with axes showing the unscaled displacement along those modes.

These results indicate that even when using a data-driven coordinate system —one designed to pick up only the largest concerted motions—the transition path is not a simple linear interpolation between the two crystal structures. Rather, in all cases PC2 overshoots the average value and then must backtrack as the protein progresses in PC1. This in turn shows the value of the simulations, revealing that structures populated during the transition may have characteristics that are not obvious from the end states alone.

One crucial question that must be answered is whether the data set is sufficiently large to allow reliable interpretation; while 1000 independent transitions seems like a large dataset, it is still important to test convergence quantitatively. We show the results in Supplementary Figure S1, where we used a bootstrap procedure to estimate how many trajectories are needed to produce consistent PCA results that can be used as meaningful reaction coordinates. Specifically, we randomly selected sets of trajectories (with replacement), computed the PCA for each set, and measured the mean absolute dot product between their first or second principal components. By tracking this quantity as a function of data set size, we can assess the reliability of the principal axes. For example, Figure S1a shows that for the β_2AR activation system, the absolute dot product converges (to ≈ 1) with fewer than 200 trajectories, indicating both eigenvectors are very well converged; similarly, Panel e shows that that variance has dropped below 10^{-4}; rhodopsin deactivation (Panels d and h) behaves similarly, while β_2AR deactivation (Panels b and f) converges somewhat more slowly. By contrast, the rhodopsin activation converges very slowly (Panels c and g), with the dot product remaining low and the variance high even for sets with 1000 trajectories. The poor convergence, specifically in rhodopsin activation, is possibly due to the fact that the active

structure is more open, and fluctuations after this transition are less well-defined. Thus, this reaction coordinate, while intuitive, may not be the best possible choice; we will explore a contact-based alternative in section 3.3.

3.2.2 Visualizing modes of motion—In order to better understand the PCA results, we mapped the first two modes back onto the protein structures for simulations of activation (Figure 4) and deactivation (Supplementary Figure S2); the arrows on the structure indicate the direction and relative magnitude of motion. PC1 (Figure 4a, b) showed a large displacement of helix 6 (orange helix in the upper left of each panel) outward in both proteins, as expected. By contrast, in PC2 (Figure 4c, d) TM6 makes a rotating motion in both β_2AR and rhodopsin, reminiscent to what was seen in previous all-atom work[60]. The amplitude of motion around the NPxxY region was markedly greater in β_2AR (Panels a and c) compared to rhodopsin (Panels b and d) for both PC1 and PC2. TM5 also made appreciable changes during rhodopsin activation, extending the intracellular portion of TM5 (Panels b and d, yellow helix). It is worth noting that all regions of the protein are affected. Results for deactivation were largely similar (as illustrated in Supplementary Figure S2).

3.3 Contact-based reaction coordinate reveals putative allosteric network

3.3.1 Transitions identified by contact space PCA—Because PC2 was poorly defined for rhodopsin activation when using standard cartesian-space PCA, we also applied an alternative procedure that is outlined in Section 2.2.2, performing PCA on time series of inter-residue contacts as instead of the coordinates. This analysis focuses on changes in topology as opposed to simply structure, and thus converges faster for flexible systems. Specifically, we calculated the time-progression of each unique contact to make a “transition matrix” where each cell is built using Equation 2. We then computed the principal components of this matrix using singular value decomposition and used the first two components as a reaction coordinate for each of the trajectories (Figure 5). As expected, this produced well-defined principal components; Supplementary Figure S3 shows that fewer than 200 trajectories would be necessary in all cases.

The contact-based reaction coordinates yielded similar results to those found using cartesian-space PCA. They identified a single, highly populated transition path for all four systems, characterized by deviations in the first two principal components. Here, we again observe that the trajectory of PC2 was non-monotonic for both proteins (Figure 3). In order to better understand why the system did not proceed directly to the final structure we analyzed the individual contacts more carefully.

3.3.2 Individual residues make non-monotonic transitions—Both PCA-based reaction coordinates revealed a path that was non-monotonic; the system “overshot” the minimum energy structure, particularly in PC2. This fact was especially evident in the contact-based reaction coordinate (Figure 5 and Supplementary Figure S4). The physical interpretation of this path is that some contacts that form (or break) during activation have a higher (lower) probability of being formed during the transition than they do in the end state. Such “backtracking” was noted before by Gosavi et al.[78,79] In any given trajectory, these

contacts regularly break and re-form (Supplementary Movie 1), so an average over many separate transitions is needed to reveal the pattern (Figure 6).

The non-monotonic motion in PC2 was most pronounced early in the trajectories — within the first 50,000 integration steps (or $\frac{1}{6}^{\text{th}}$ of the trajectory) (Supplementary Figure S4). This backtracking in PC2 suggests that there was some steric hindrance preventing the main transition identified by PC1. In this view PC2 serves primarily to relieve strain in the molecule by moving atoms out of the way and thus facilitating activation. This is most clear in Supplementary Figure S4A, which shows that PC2 in β_2AR activation has roughly the same value at the beginning and end of the simulation.

To better understand these transitions, we plot the state (formed or broken) of individual contacts for the first third of each trajectory; after that, nearly all trajectories had reached their target state and were fluctuating about it. To reduce noise, we present the average time series for each unique contact in Figure 6, considering all 1000 activation trajectories (results for deactivation are plotted in Supplementary Figure S5).

Figure 6a, b highlights two contacts randomly chosen from our rhodopsin activation simulations. Panel a shows a contact that transitions monotonically: once that contact forms it remains formed throughout the simulation. By contrast, the plot in Figure 6b shows a contact that forms only briefly and is then re-broken. The curves are colored by transition value (Equation 2), where formed contacts are red and broken contacts are black. In Figure 6c, d this analysis is expanded to all contacts (with color still representing transition value), but here the y-axis contains the set of all unique contacts formed (top plot) or broken (bottom plot). In both β_2AR and rhodopsin activation, there is “checkering” in the x-axis, indicating multiple contacts make “backtracking” transitions to their starting orientation (similar to those contacts plotted in 6b). These contacts were spread throughout the protein and not localized to any specific region as can be seen in Supplementary Figure S6. Details for deciding which contacts are non-monotonic, or backtrack are provided in the Supplementary Information. That residue contacts re-form and re-break shows there is some topological frustration[80] in the transitions, despite the fact that no such frustration is built into the model itself. This underscores the need for dynamics data to complement structural information.

To better understand the complex mechanisms underlying protein function, it will likely be necessary to study their dynamic properties directly. This is particularly important for GPCRs, where many allosteric ligands modulate activity in different ways; recent simulations suggest that activation occurs via distinct pathways when different ligands are bound[28,31]. The method presented here not only identified a non-trivial reaction coordinate, but it also resulted in a putative network of interactions that are vital to activation. This network was defined in terms of a reduced dataset that allowed quantitative comparison between the four different systems studied (see Section 3.4). The problem that remains is how to biologically interpret the reduced set of motions described in contact space.

3.3.3 Interpreting the contact-based coordinate—To visualize changes in contact space, we mapped the principal components back onto the structures themselves, with bonds

showing those contacts crucial to activation. Specifically, we drew bonds between the c_α's of residues making unique contacts, using the thickness of the bonds to indicate the variance of that contact within the component shown. For clarity, we only show those contacts that cumulatively constitute half of the amplitude of each principal component. In addition to activation (Figure 7), results obtained for deactivation (Supplementary Figure S7) and for PC2 (Supplementary Figures S8 and S9) are available in the supplementary information.

Interestingly, these contacts were spread throughout the proteins and not limited to the obvious areas of change, such as the cytoplasmic face and TM6. Figure 7 revealed a network of contacts that may be fundamentally important to GPCR signaling. These bonds show how the whole protein is linked in contact-space, and their presence in the most concerted principal component suggests that an allosteric connection exists between the hydrophobic ligand binding pocket and the cytoplasmic G protein binding site. This demonstrates the power of letting the data speak for itself, rather than constraining analysis to a preconceived reaction coordinate.

In β_2AR, we identified a cluster of contacts near the G protein binding site that is broken on activation (Figure 7a). This included contacts between TM6 and the NPxxY motif in TM7. However, contacts are simultaneously formed between the more extracellular portions of TM6 and TM7. At the same time, multiple contacts between the intracellular portions of TM5 and TM6 form (Figure 7b). This mode gives a potential mechanism explaining how the ligand, located in a pocket toward the extracellular portion of the protein, can alter the structure and dynamics of the intracellular face.

In rhodopsin, the contacts were also spread over the whole surface of the protein. A cluster on the extracellular side of helices 5, 6 and 7 was identified with contacts both breaking and forming (Figure 7c, d). Notably, there is precedent for extracellular motion in rhodopsin, and others have suggested this motion may impact the conformation of the extracellular loops, specifically during activation[81]. In the future, it would be interesting to experimentally test the importance of these contacts using mutagenesis or crosslinking experiments.

Another surprising result was that there were many more contacts contributing to the first principal component in rhodopsin than in β_2AR (Supplementary Table Si). This corroborates an observation we noted previously: although dark-state rhodopsin has a vanishingly small basal activity, it was much more flexible than β_2AR in all-atom simulations[60]. In the current study, rhodopsin transitions involved more contacts than β_2AR, but β_2AR populated a broader range of structures during transitions (Figure 5a, b vs. 5c, d).

3.4 Proteins activate via quantitatively different pathways

In both cartesian and contact-based results, the path traversed by rhodopsin was narrower than in β_2AR; this can be seen by visual inspection of Figures 3 and 5. The highly-populated endpoint was also broader in β_2AR than in rhodopsin. Together, this suggests that β_2AR activation may proceed through a more diverse ensemble of states and that its active ensemble may be more expansive. This makes functional sense as β_2AR has detectable basal activity, while rhodopsin acts like a switch, with essentially no activity when its ligand is bound in the 11-*cis* form. Using both data-derived PCA-based reaction coordinates, we

quantitatively compared the transition paths using previously established methods for comparing eigensets from biomolecular covariance matrices[36,38,61,62].

To make this comparison we used the covariance overlap (Equation 1). For each iteration two trajectories were selected at random and compared. This process was repeated 100 times for each data point in figures 8 and 9. In those figures we report the average value ± standard deviation. These samples were also used to test for statistical significance in the difference between pathways using a Welch's T-test.

3.4.1 How similar are the paths taken by β_2AR and rhodopsin?—First, we compared transition paths taken by β_2AR to those taken by rhodopsin (Figure 8), for both activation (blue) and deactivation (red) simulations. For each dataset, trajectories were randomly chosen and compared using the covariance overlap (Equation 1). This analysis was performed using cartesian PCA results (contact-based results (8b). Additional details of this analysis are provided in the Supplemental Methods (Section S1.3). Both panels follow the same format, with comparisons between activation simulations in blue and comparisons between deactivation simulations in red. We first compared sets of β_2AR simulations to themselves (β_2AR) by selecting 100 pairs of β_2AR activation eigensets and computing the covariance overlap for each pair. Similarly, the second bar shows a comparison of eigensets obtained from rhodopsin activation simulations (rhodopsin). Finally, we compared the paths from these two proteins to one another, creating pairs by pulling an eigenset from each ensemble (cross). The cross comparison was significantly lower than either self comparison (as measured using Welch's T-test, at a significance level of 0.05). This demonstrates that the proteins take statistically different paths. The p-values between the overlaps are listed in Supplementary Table Siii).

We also found that β_2AR-β_2AR overlaps were lower than the rhodopsin-rhodopsin overlaps, suggesting more variance in β_2AR pathways. This correlates well with the broader β_2AR transition as seen in Figure 5, reinforcing the idea that there is a larger variance in β_2AR paths. As mentioned before, this larger variance may be indicative of β_2AR's ability to activate spontaneously.

The covariance overlaps are systematically lower for the contact-space PCA (Figure 8b) than for cartesian-space (Figure 8a). This is expected, because the dimensionality of the contact-space analysis is significantly higher, and our previous work showed that the covariance overlap decreases systematically with increased matrix size[65]; this is discussed further in the supplementary information (Section S1.3).

3.4.2 How similar are activation and deactivation?—Next, we compared the pathways of activation and deactivation for each pro-tein (Figure 9). This was done separately for β_2AR (blue) and rhodopsin (red). When comparing two β_2AR activation simulations (activation) or two deactivation simulations (deactivation) the calculated overlap was similar, indicating that the activation and deactivation paths have similar heterogeneities. However, when comparing an activation simulation to a deactivation simulation (cross), the drop in covariance overlap was statistically significant for both

proteins; thus, in these calculations, deactivation is not a simple reversal of the activation path. This is consistent with our previous qualitative analysis in Section 3.1 and Figure 2.

As experimental structures for other GPCRs become available it will be important to quantitatively identify their activation pathways, especially in the context of related proteins. Also, as more computational resources become available similar analyses should be performed using more rigorous models. In particular, the quantitative assessments of statistical error discussed above will be critical in figuring out precisely how much sampling is needed to get reliable results. Our data indicates that using the contact-space PCA requires roughly 200 trajectories to be fully converged, so at present this is probably beyond the reach of even the most powerful brute-force calculation[67], although alternative sampling strategies may be plausible[31].

4 Conclusion

We used Gō-like models to simulate activation of two model GPCRs, rhodopsin and the β_2 adrenergic receptor. While an experimentally motivated reaction coordinate captured β_2AR transitions, it proved inadequate when applied to rhodopsin. We therefore introduced a novel data-derived coordinate based on the unique set of contacts that change between two states of a molecule. Using this coordinate, we were able to track the changes in activation of both proteins. These results were corroborated by a standard cartesian-based PCA analysis. We showed that transition paths taken by the two proteins were statistically different, indicating the diversity of GPCR activation. In addition, β_2AR transitions were more heterogeneous, suggesting a possible explanation for β_2AR's basal activity. In the future, it will be important to test this observation on other GPCRs as more active-state crystal structures are published. Using a general reaction coordinate that does not assume biological knowledge *a priori* will likely be an important means to differentiate the *ensemble* dynamics of proteins, and will be important for fully understanding the mechanism of their activation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Jessi Leioatts for her critical review and suggestions to the manuscript. Computational resources were generously provided by the University of Rochester Center for Integrated Research Computing. This work was supported by an Institutional Ruth L. Kirschstein National Research Service Award (GM068411, to NL) and NIH project number GM095496 (AG).

References

1. Bahar I, Lezon TR, Bakan A, Shrivastava IH. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem Rev. 2010; 110:1463–1497. [PubMed: 19785456]
2. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? Nat Rev Drug Discov. 2006; 5:993–996. [PubMed: 17139284]
3. Wootten D, Christopoulos A, Sexton PM. Emerging paradigms in gpcr allostery: implications for drug discovery. Nat Rev Drug Discov. 2013; 12:630–644. [PubMed: 23903222]

4. Hofmann KP, Scheerer P, Hildebrand PW, Choe HW, Park JH, Heck M, Ernst OP. A g protein-coupled receptor at work: the rhodopsin model. Trends Biochem Sci. 2009; 34:540–552. [PubMed: 19836958]
5. Nygaard R, Frimurer TM, Holst B, Rosenkilde MM, Schwartz TW. Ligand binding and micro-switches in 7tm receptor structures. Trends Pharmacol Sci. 2009; 30:249–259. [PubMed: 19375807]
6. Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A, Filipek S. Action of molecular switches in gpcrs–theoretical and experimental studies. Curr Med Chem. 2012; 19:1090–1109. [PubMed: 22300046]
7. Schwartz TW, Frimurer TM, Holst B, Rosenkilde MM, Elling CE. Molecular mechanism of 7tm receptor activation–a global toggle switch model. Annu Rev Pharmacol Toxicol. 2006; 46:481–519. [PubMed: 16402913]
8. Shi L, Liapakis G, Xu R, Guarnieri F, Ballesteros JA, Javitch JA. Beta2 adrenergic receptor activation. modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. J Biol Chem. 2002; 277:40989–40996. [PubMed: 12167654]
9. Rosenbaum DM, Rasmussen SGF, Kobilka BK. The structure and function of g-protein-coupled receptors. Nature. 2009; 459:356–363. [PubMed: 19458711]
10. Mertz B, Struts AV, Feller SE, Brown MF. Molecular simulations and solid-state nmr investigate dynamical structure in rhodopsin activation. Biochim Biophys Acta. 2012; 1818:241–251. [PubMed: 21851809]
11. Deupi X, Standfuss J. Structural insights into agonist-induced activation of g-protein-coupled receptors. Curr Opin Struct Biol. 2011; 21:541–551. [PubMed: 21723721]
12. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Trong IL, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A g protein-coupled receptor. Science. 2000; 289:739–745. [PubMed: 10926528]
13. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2. 2 a crystal structure. J Mol Biol. 2004; 342:571–583. [PubMed: 15327956]
14. Salom D, Lodowski DT, Stenkamp RE, Trong IL, Golczak M, Jastrzebska B, Harris T, Ballesteros JA, Palczewski K. Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. Proc Natl Acad Sci U S A. 2006; 103:16123–16128. [PubMed: 17060607]
15. Choe HW, Kim YJ, Park JH, Morizumi T, Pai EF, Krau N, Hofmann KP, Scheerer P, Ernst OP. Crystal structure of metarhodopsin ii. Nature. 2011
16. Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD, Schertler GFX. The structural basis of agonist-induced activation in constitutively active rhodopsin. Nature. 2011
17. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic g protein-coupled receptor. Science. 2007; 318:1258–1265. [PubMed: 17962520]
18. Wacker D, Fenalti G, Brown MA, Katritch V, Abagyan R, Cherezov V, Stevens RC. Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by x-ray crystallography. J Am Chem Soc. 2010; 132:11443–11445. [PubMed: 20669948]
19. Rasmussen SGF, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK. Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. Nature. 2011; 469:175–180. [PubMed: 21228869]
20. Rasmussen SGF, Devree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK. Crystal structure of the β_2 adrenergic receptor-gs protein complex. Nature. 2011
21. Provasi D, Artacho MC, Negri A, Mobarec JC, Filizola M. Ligand-induced modulation of the free-energy landscape of g protein-coupled receptors explored by adaptive biasing techniques. PLoS Comput Biol. 2011; 7:e1002193. [PubMed: 22022248]
22. Grossfield A. Recent progress in the study of g protein-coupled receptors with molecular dynamics computer simulations. Biochim Biophys Acta. 2011; 1808:1868–1878. [PubMed: 21443858]

23. Lee JY, Lyman E. Agonist dynamics and conformational selection during microsecond simulations of the a2a adenosine receptor. Biophys J. 2012; 102:2114–2120. [PubMed: 22824275]
24. Kobilka B, Schertler GFX. New g-protein-coupled receptor crystal structures: insights and limitations. Trends Pharmacol Sci. 2008; 29:79–83. [PubMed: 18194818]
25. Johnston JM, Filizola M. Showcasing modern molecular dynamics simulations of membrane proteins through g protein-coupled receptors. Curr Opin Struct Biol. 2011; 21:552–558. [PubMed: 21764295]
26. Shan J, Khelashvili G, Mondal S, Mehler EL, Weinstein H. Ligand-dependent conformations and dynamics of the serotonin 5-ht(2a) receptor determine its activation and membrane-driven oligomerization properties. PLoS Comput Biol. 2012; 8:e1002473. [PubMed: 22532793]
27. Struts AV, Salgado GFJ, Martnez-Mayorga K, Brown MF. Retinal dynamics underlie its switch from inverse agonist to agonist during rhodopsin activation. Nat Struct Mol Biol. 2011; 18:392–394. [PubMed: 21278756]
28. Li J, Jonsson AL, Beuming T, Shelley JC, Voth GA. Ligand-dependent activation and deactivation of the human adenosine a(2a) receptor. J Am Chem Soc. 2013; 135:8749–8759. [PubMed: 23678995]
29. Lodowski DT, Angel TE, Palczewski K. Comparative analysis of gpcr crystal structures. Photochem Photobiol. 2009; 85:425–430. [PubMed: 19192200]
30. Leioatts N, Mertz B, Martnez-Mayorga K, Romo TD, Pitman MC, Feller SE, Grossfield A, Brown MF. Retinal ligand mobility explains internal hydration and reconciles active rhodopsin structures. Biochemistry. 2014
31. Kohlhoff KJ, Shukla D, Lawrenz M, Bowman GR, Konerding DE, Belov D, Altman RB, Pande VS. Cloud-based simulations on google exacycle reveal ligand modulation of gpcr activation pathways. Nat Chem. 2014; 6:15–21. [PubMed: 24345941]
32. Kobilka BK, Deupi X. Conformational complexity of g-protein-coupled receptors. Trends Pharmacol Sci. 2007; 28:397–406. [PubMed: 17629961]
33. Farrens DL, Altenbach C, Yang K, Hubbell WL, Khorana HG. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. Science. 1996; 274:768–770. [PubMed: 8864113]
34. Altenbach C, Kusnetzow AK, Ernst OP, Hofmann KP, Hubbell WL. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. Proc Natl Acad Sci U S A. 2008; 105:7439–7444. [PubMed: 18490656]
35. Cai K, Klein-Seetharaman J, Hwa J, Hubbell WL, Khorana HG. Structure and function in rhodopsin: effects of disulfide cross-links in the cytoplasmic face of rhodopsin on transducin activation and phosphorylation by rhodopsin kinase. Biochemistry. 1999; 38:12893–12898. [PubMed: 10504260]
36. Grossfield A, Feller SE, Pitman MC. Convergence of molecular dynamics simulations of membrane proteins. Proteins: Struct, Funct, Bioinf. 2007; 67:31–40.
37. Neale C, Bennett WFD, Tieleman DP, Poms R. Statistical convergence of equilibrium properties in simulations of molecular solutes embedded in lipid bilayers. Journal of Chemical Theory and Computation. 2011; 7:4175–4188.
38. Romo TD, Grossfield A. Block covariance overlap method and convergence in molecular dynamics simulation. J Chem Theory Comput. 2011; 7:2464–2472.
39. Swendsen, Wang. Replica monte carlo simulation of spin glasses. Phys Rev Lett. 1986; 57:2607–2609. [PubMed: 10033814]
40. Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci U S A. 2002; 99:12562–12566. [PubMed: 12271136]
41. Taketomi H, Ueda Y, Gō N. Studies on protein folding, unfolding and fluctuations by computer simulation. i. the effect of specific amino acid sequence represented by specific inter-unit interactions. Int J Pept Protein Res. 1975; 7:445–459. [PubMed: 1201909]
42. Gō N, Taketomi H. Respective roles of short- and long-range interactions in protein folding. Proc Natl Acad Sci U S A. 1978; 75:559–563. [PubMed: 273218]

43. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? an investigation for small globular proteins. J Mol Biol. 2000; 298:937–953. [PubMed: 10801360]
44. Whitford PC, Sanbonmatsu KY, Onuchic JN. Biomolecular dynamics: order-disorder transitions and energy landscapes. Rep Prog Phys. 2012; 75:076601. [PubMed: 22790780]
45. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. Proteins. 2009; 75:430–441. [PubMed: 18837035]
46. Whitford PC, Miyashita O, Levy Y, Onuchic JN. Conformational transitions of adenylate kinase: switching by cracking. J Mol Biol. 2007; 366:1661–1671. [PubMed: 17217965]
47. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci U S A. 2003; 100:12570–12575. [PubMed: 14566052]
48. Miyashita O, Wolynes PG, Onuchic JN. Simple energy landscape model for the kinetics of functional transitions in proteins. J Phys Chem B. 2005; 109:1959–1969. [PubMed: 16851180]
49. Koga N, Takada S. Folding-based molecular simulations reveal mechanisms of the rotary motor f1-atpase. Proc Natl Acad Sci U S A. 2006; 103:5367–5372. [PubMed: 16567655]
50. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. Proc Natl Acad Sci U S A. 2006; 103:11844–11849. [PubMed: 16877541]
51. Jamros MA, Oliveira LC, Whitford PC, Onuchic JN, Adams JA, Jennings PA. Substrate-specific reorganization of the conformational ensemble of csk implicates novel modes of kinase function. PLoS Comput Biol. 2012; 8:e1002695. [PubMed: 23028292]
52. Best RB, Chen YG, Hummer G. Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. Structure. 2005; 13:1755–1763. [PubMed: 16338404]
53. Maragakis P, Karplus M. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. J Mol Biol. 2005; 352:807–822. [PubMed: 16139299]
54. Chu JW, Voth GA. Coarse-grained free energy functions for studying protein conformational changes: a double-well network model. Biophys J. 2007; 93:3860–3871. [PubMed: 17704151]
55. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN. Smog@ctbp: simplified deployment of structure-based models in gromacs. Nucleic Acids Res. 2010; 38:W657–W661. [PubMed: 20525782]
56. Noel JK, Whitford PC, Onuchic JN. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. J Phys Chem B. 2012; 116:8692–8702. [PubMed: 22536820]
57. Whitford PC, Geggier P, Altman RB, Blanchard SC, Onuchic JN, Sanbonmatsu KY. Accommodation of aminoacyl-trna into the ribosome involves reversible excursions along multiple pathways. RNA. 2010; 16:1196–1204. [PubMed: 20427512]
58. Pronk S, Pll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. Gromacs 4. 5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013; 29:845–854. [PubMed: 23407358]
59. Romo TD, Grossfield A. LOOS: an extensible platform for the structural analysis of simulations. Conf Proc IEEE Eng Med Biol Soc. 2009; 2009:2332–2335. [PubMed: 19965179]
60. Romo TD, Grossfield A, Pitman MC. Concerted interconversion between ionic lock substates of the beta(2) adrenergic receptor revealed by microsecond timescale molecular dynamics. Biophys J. 2010; 98:76–84. [PubMed: 20074514]
61. Hess B. Convergence of sampling in protein simulations. Phys Rev E Stat, Nonlinear Soft Matter Phys. 2002; 65:031910.
62. Faraldo-Gomez JD, Forrest LR, Baaden M, Bond PJ, Domene C, Patargias G, Cuthbertson J, Sansom MSP. Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. Proteins: Struct, Funct, Bioinf. 2004; 57:783–791.

63. Micheletti C, Carloni P, Maritan A. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and gaussian models. Proteins: Struct, Funct, Bioinf. 2004; 55:635–645.
64. Pang A, Arinaminpathy Y, Sansom MSP, Biggin PC. Comparative molecular dynamics–similar folds and similar motions? Proteins: Struct, Funct, Bioinf. 2005; 61:809–822.
65. Leioatts N, Romo TD, Grossfield A. Elastic network models are robust to variations in formalism. J Chem Theory Comput. 2012; 8:2424–2434. [PubMed: 22924033]
66. Seckler JM, Leioatts N, Miao H, Grossfield A. The interplay of structure and dynamics: Insights from a survey of hiv-1 reverse transcriptase crystal structures. Proteins. 2013
67. Dror RO, Arlow DH, Maragakis P, Mildorf TJ, Pan AC, Xu H, Borhani DW, Shaw DE. Activation mechanism of the 2-adrenergic receptor. Proc Natl Acad Sci U S A. 2011; 108:18684–18689. [PubMed: 22031696]
68. Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP. Crystal structure of the ligand-free g-protein-coupled receptor opsin. Nature. 2008; 454:183–187. [PubMed: 18563085]
69. Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP, Ernst OP. Role of the conserved npxxy(x)5,6f motif in the rhodopsin ground state and during activation. Proc Natl Acad Sci U S A. 2003; 100:2290–2295. [PubMed: 12601165]
70. Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SG, Shi L, Gether U, Javitch JA. Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. J Biol Chem. 2001; 276:29171–29177. [PubMed: 11375997]
71. Vogel R, Mahalingam M, Ldeke S, Huber T, Siebert F, Sakmar TP. Functional role of the “ionic lock”–an interhelical hydrogen-bond network in family a heptahelical receptors. J Mol Biol. 2008; 380:648–655. [PubMed: 18554610]
72. Miao Y, Nichols SE, Gasper PM, Metzger VT, McCammon JA. Activation and dynamic network of the m2 muscarinic receptor. Proc Natl Acad Sci U S A. 2013; 110:10982–10987. [PubMed: 23781107]
73. Brooks BR, Jane3zi3c D, Karplus M. Harmonic analysis of large systems. i. methodology. J Comput Chem. 1995; 16:1522–1542.
74. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. Proteins: Struct, Funct, Bioinf. 1993; 17:412–425.
75. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins. 1991; 11:205–217. [PubMed: 1749773]
76. Horiuchi T, Gō N. Projection of monte carlo and molecular dynamics trajectories onto the normal mode axes: human lysozyme. Proteins. 1991; 10:106–116. [PubMed: 1896424]
77. Case DA. Normal mode analysis of protein dynamics. Current Opinion in Structural Biology. 1994; 4:285–290.
78. Gosavi S, Chavez LL, Jennings PA, Onuchic JN. Topological frustration and the folding of interleukin-1 beta. J Mol Biol. 2006; 357:986–996. [PubMed: 16469330]
79. Gosavi S, Whitford PC, Jennings PA, Onuchic JN. Extracting function from a beta-trefoil folding motif. Proc Natl Acad Sci U S A. 2008; 105:10384–10389. [PubMed: 18650393]
80. Shea JE, Onuchic JN, Brooks C 3rd. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment b of protein a. Proc Natl Acad Sci U S A. 1999; 96:12512–12517. [PubMed: 10535953]
81. Ahuja S, Hornak V, Yan ECY, Syrett N, Goncalves JA, Hirshfeld A, Ziliox M, Sakmar TP, Sheves M, Reeves PJ, Smith SO, Eilers M. Helix movement is coupled to displacement of the second extracellular loop in rhodopsin activation. Nat Struct Mol Biol. 2009; 16:168–175. [PubMed: 19182802]
82. Ballesteros, JA.; Weinstein, H. [19] integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in g protein-coupled receptors. In: Sealfon, SC., editor. Receptor Molecular Biology. Academic Press; p. 366-428.volume 25 of Methods in Neurosciences

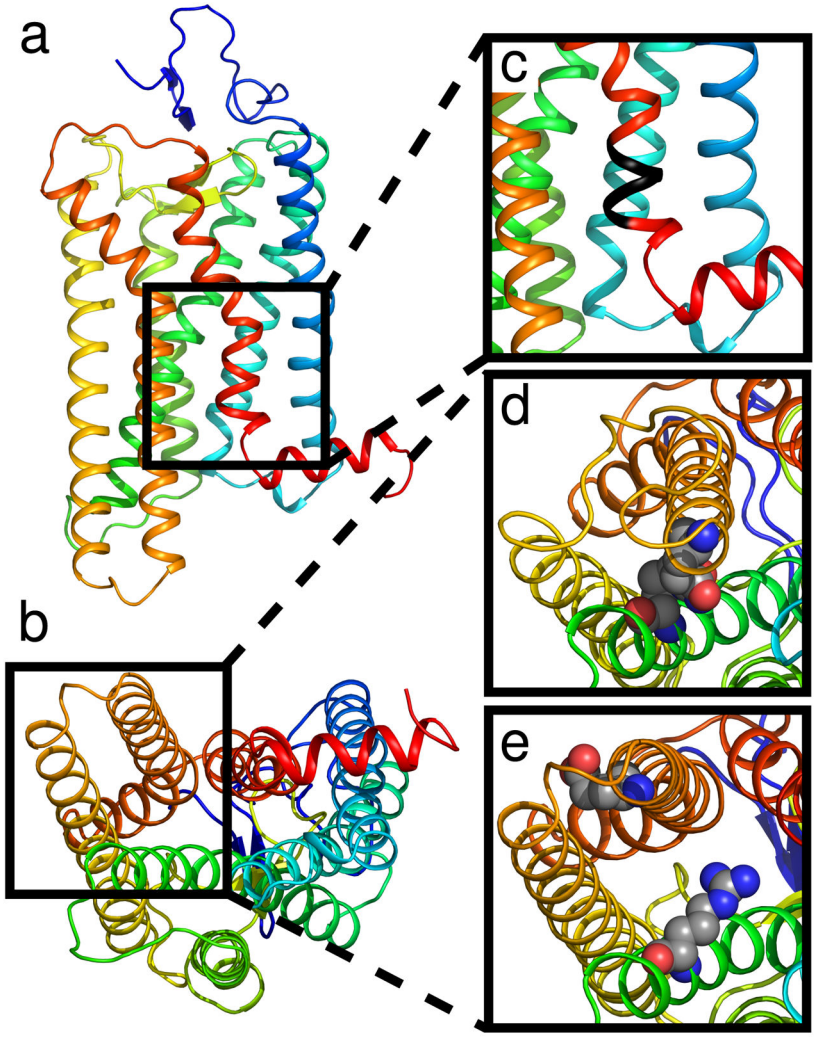

Figure 1.

Overview of conserved regions. Rhodopsin is shown as a cartoon using rainbow coloring (a) in the plane of the membrane and (b) from the cytoplasmic side. (c) Location of the NPxxY motif in TM7 is highlighted in black. The ionic lock between TM3 and TM6 is shown in its (d) closed and (e) opened state.

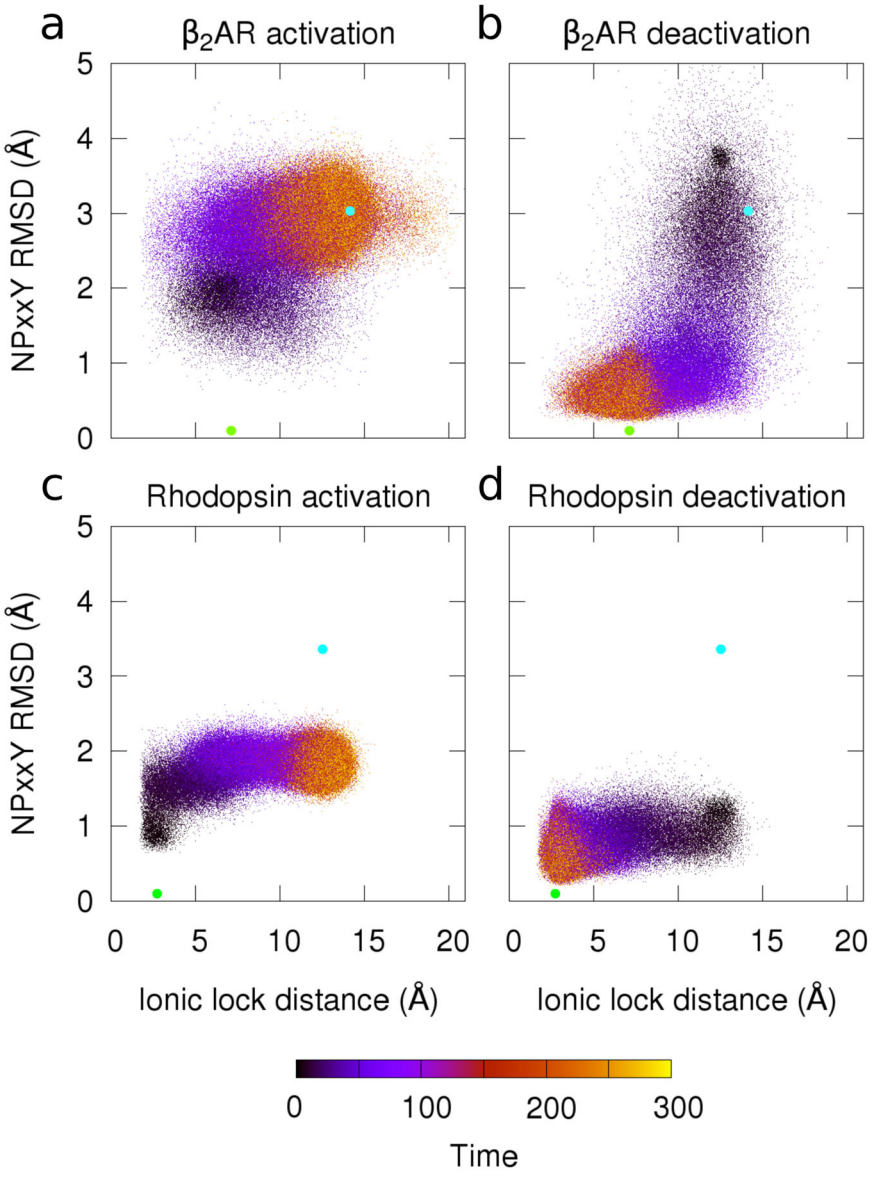

Figure 2.

Transitions in cytoplasmic motifs during activation and deactivation. Plots show the RMSD to the inactive NPxxY region (using the inactive structures in Table I). RMSD was calculated using backbone heavy atoms (y-axis). This is plotted with the distance between ionic lock residues (R3.50 to E6.30) using the minimum distance between heavy atoms (x-axis). Data color indicates the trajectory time (arbitrary units). The active (cyan circle) and inactive (green circle) crystal structures are plotted for reference. 1,000 simulations are shown for each of the four systems studied: (a) β_2AR activation, (b) β_2AR deactivation, (c) rhodopsin activation, and (d) rhodopsin deactivation.

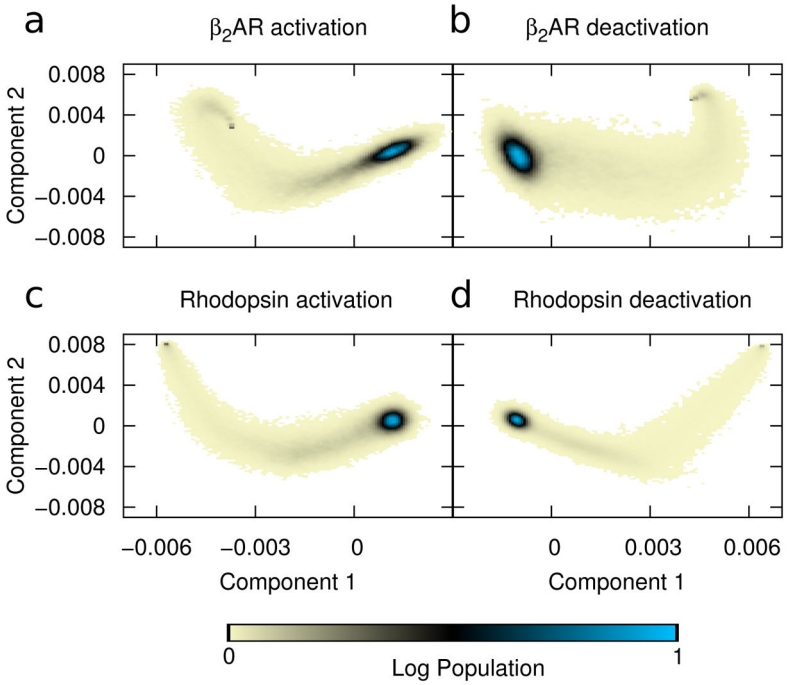

Figure 3.

Principal component-based reaction coordinate. PCA of the covariance matrix in cartesian space is used to define the principal axes. Color shows the log-scale population from all simulations projected onto principal components one (x-axis) and two (y-axis) for all 1,000 trajectories in each ensemble. Displacement units are arbitrary. Data is shown for simulations of (a) β_2AR activation, (b) β_2AR deactivation, (c) rhodopsin activation, and (d) rhodopsin deactivation. In all cases the peak population density indicates the final structure.

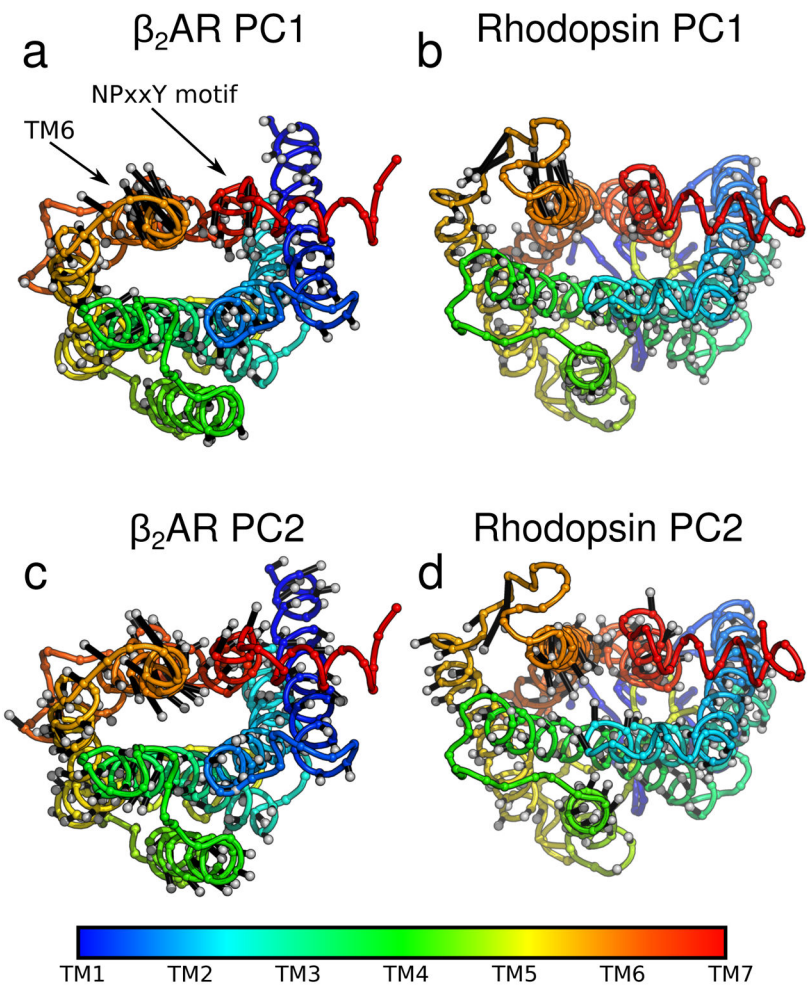

Figure 4.

Direction of the most concerted motions in activation. Principal components (from the analysis described in Section 3.2.1) are plotted as silver-tipped black vectors extending from c_α's on protein cartoons. Vector length is proportional to each residue's contribution to the principal component. The initial protein structure (the inactive state) is shown with the NPxxY motif and TM6 labeled. The spectrum bar shows the residues sequence number. Data is shown for (a) PC1 ofβ_2AR, (b) PC1 of rhodopsin, (c) PC2 of β_2AR, and (d) PC2 of rhodopsin.

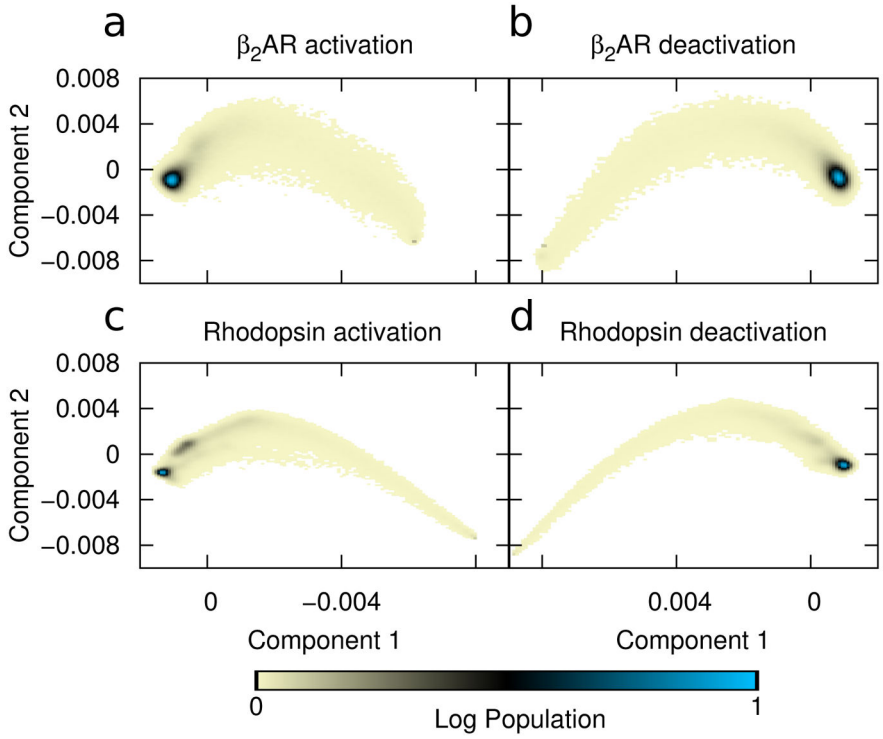

Figure 5.

Contact-based general reaction coordinate. Displacement along PC1 and PC2 of the "transition matrix" (Section 2.2.2) is shown. Color indicates the relative population of these displacements (log scale) across all trajectories. Displacements are given in arbitrary units. Data is shown for simulations of (a) β_2AR activation, (b) β_2AR deactivation, (c) rhodopsin activation, and (d) rhodopsin deactivation.

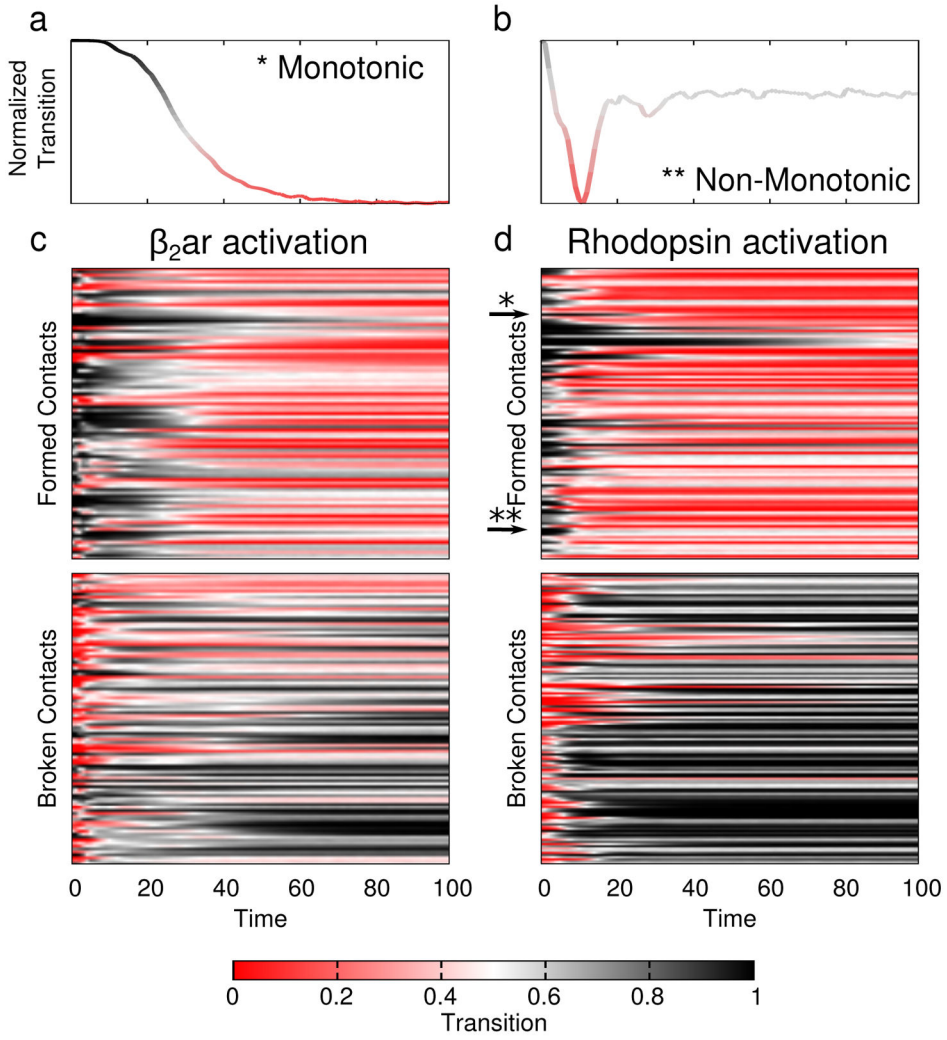

Figure 6.

Contact transitions during activation. The average transition value (color) across all 1,000 trajectories is plotted using Equation 2. Red indicates a formed contact and black one that is broken. (a and b) Single contacts selected from rhodopsin activation illustrate monotonic and non-monotonic transitions. (a) Example of a monotonic transition. (b) An example contact that makes a backward transition. (c and d) Each unique contact (y-axis) is shown as a separate row. The first 100 frames (x-axis) are shown for (c) β_2AR activation and (d) rhodopsin activation.

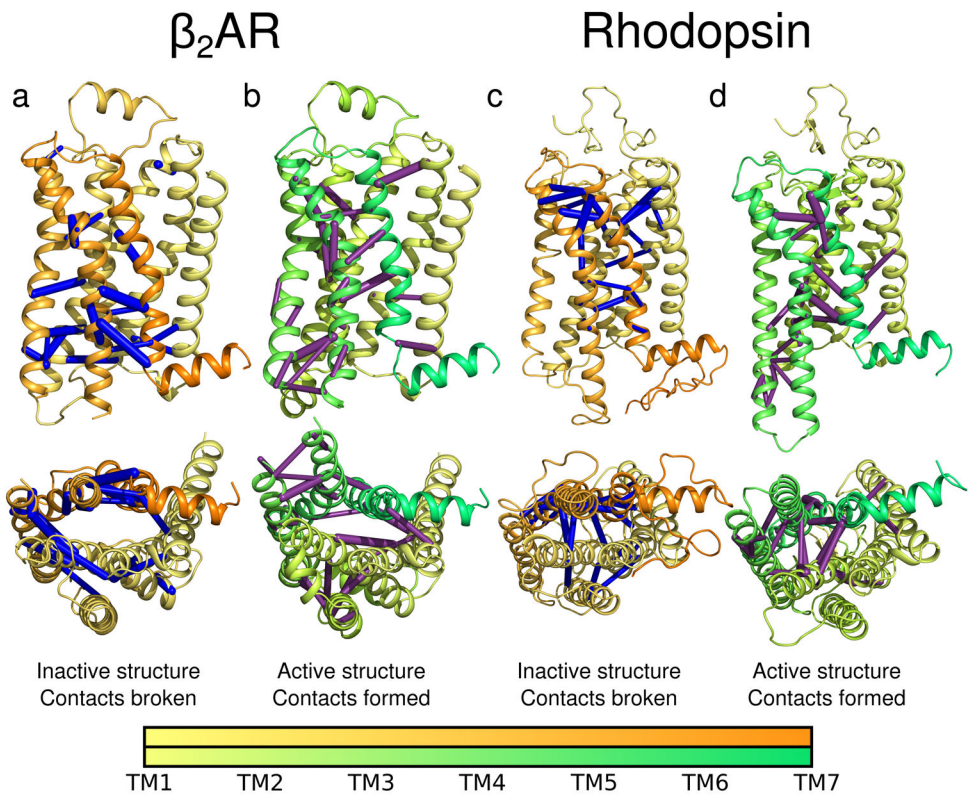

Figure 7.

Contacts characterizing activation. The contacts whose transitions contributed the most to PC1 (i.e. those accounting for 50% of PC1 displacement) are mapped onto cartoons of β_2AR and rhodopsin as bonds between c_α's. The bond thickness is proportional to the contribution made to the principal component. Bonds broken during a activation are shown in *blue*, those formed are shown in *purple*. Cartoons of inactive protein are colored *yellow to orange*, the active proteins are colored *yellow to green*. The spectrum bars indicate residue numbers. Panels (a) and (b) show data for β_2AR activation simulations. Panels (c) and (d) show data for rhodopsin activation simulations.

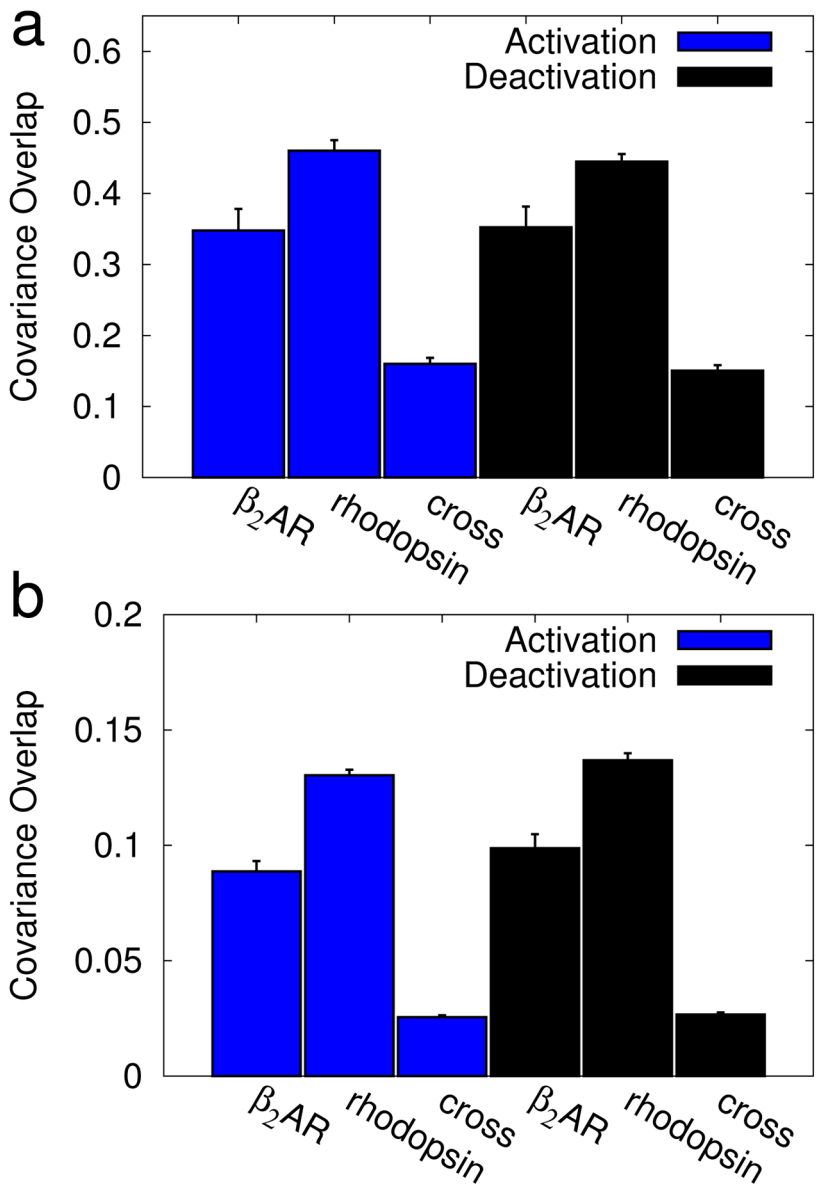

Figure 8.

Similarity between β_2AR and rhodopsin transitions. The average covariance overlap[61] is shown for a random subset of 100 trajectories to compare the pathway taken by β_2AR to that taken by rhodopsin. The error bars represent the standard deviation. Data is shown for overlaps between (a) cartesian-space PCA and (b) contact-based PCA. Within each panel results are split into activation simulations (*blue bars*) and deactivation simulations (*red bars*). The average overlaps were calculated between: a pair of β_2AR simulations (β_2*AR*), a pair of rhodopsin simulations (*rhodopsin*), and from a β_2AR simulation to a rhodopsin simulation (*cross*). The difference in overlap between β_2*AR* simulations and a *cross*-comparison, and between *rhodopsin* simulations and a *cross*-comparison were found to be statistically significant ($p < 0.05$); p-values are listed in Table Sii.

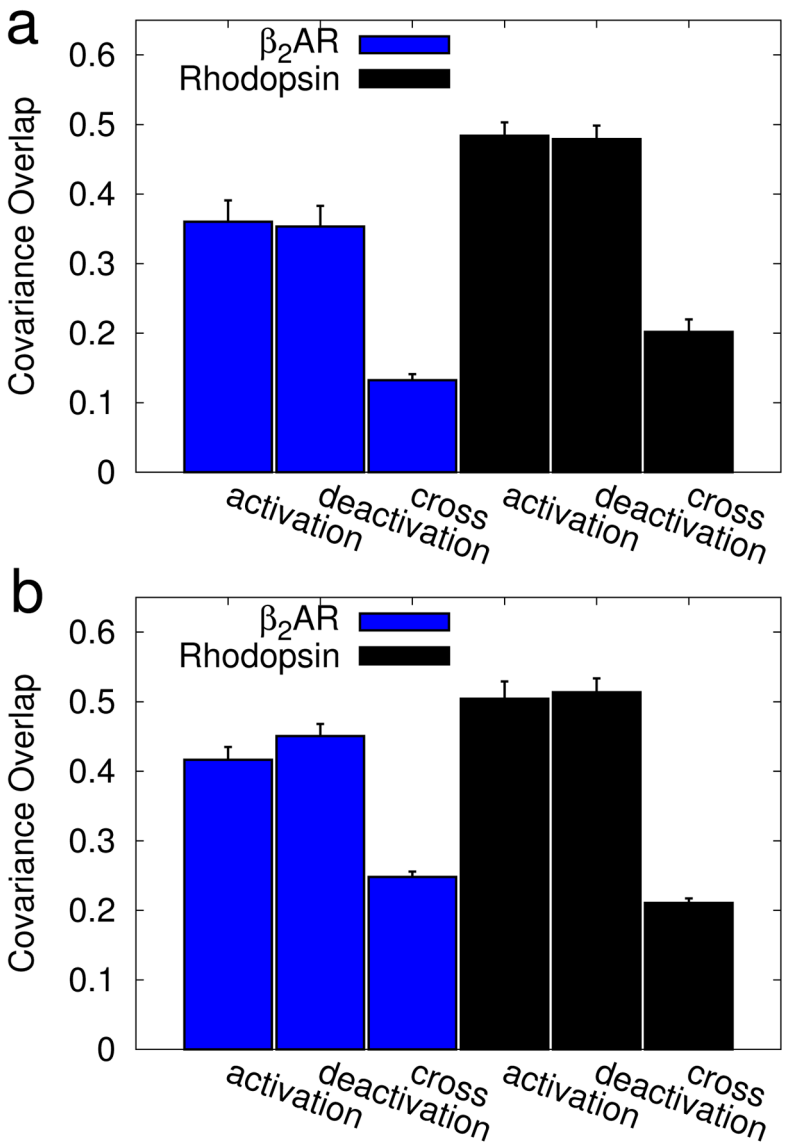

Figure 9.

Similarity between activation and deactivation pathways. The average covariance overlap[61] is shown for a subset of 100 trajectories, chosen randomly. The error bars represent the standard deviation. Results are shown for (a) covariance-based PCA and (b) contact-based PCA. Data is split into β_2AR simulations (*blue bars*) and rhodopsin simulations (*red bars*). The bars represent overlaps between: a pair of activation simulations (*activation*), a pair of deactivation simulations (*deactivation*), and from an activation simulation to a deactivation simulation (*cross*). The difference in overlap between *activation* simulations and a *cross*-comparison, and between *deactivation* simulations and a *cross*-comparison were found to be statistically significant ($p < 0.05$); p-values are listed in Table Siii.

Table I

PDB accession numbers of structures used.

Protein	Inactive	Active
Rhodopsin	1U19[13]	3PXO[15]
*β*2AR	3NY8[18]	3P0G[19]

Table II

GPCR transmembrane regions used in calculations. Superscripts use the notation developed by Ballestoros and Weinstein[82].

Protein	TM1	TM2	TM3	TM4	TM5	TM6	TM7
β_2^{AR}	$W^{1.31}$-$Q^{1.59}$	$V^{2.38}$-$M^{2.67}$	$N^{3.22}$-$I^{3.54}$	$N^{4.40}$-$Q^{4.62}$	$N^{5.35}$-$K^{5.66}$	$K^{6.29}$-$I^{6.60}$	$E^{7.33}$-$R^{7.55}$
rhodopsin	$W^{1.29}$-$K^{1.59}$	$P^{2.38}$-$H^{2.67}$	$P^{3.22}$-$V^{3.54}$	$N^{4.40}$-$V^{4.62}$	$N^{5.35}$-$Q^{5.71}$	$T^{6.25}$-$T^{6.60}$	$I^{7.33}$-$M^{7.56}$