Energy landscapes reveal agonist's control of GPCR activation via microswitches

Oliver Fleetwood[a], **Pierre Matricon**[b], **Jens Carlsson**[b], **and Lucie Delemotte**[a]

[a]Science for Life Laboratory, Department of Applied Physics, KTH Royal Institute of Technology, Stockholm, Sweden; [b]Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

This manuscript was compiled on May 4, 2019

Abstract

Agonist binding to the extracellular part of G protein-coupled receptors (GPCRs) leads to conformational changes in the transmembrane region that activate cytosolic signalling pathways. Although high resolution structures of the inactive and active receptor states are available, the allosteric coupling that transmits the signal across the membrane is not fully understood. We calculated free energy landscapes of the β_2 adrenergic receptor using atomistic molecular dynamics simulations in an optimized string of swarms framework, which sheds new light on the roles of microswitches involved in activation. Contraction of the extracellular binding site in the presence of agonist is obligatorily coupled to conformational changes in a connector motif located in the core of the transmembrane region. In turn, the connector is probabilistically coupled to the conformation of the intracellular region: an active connector promotes desolvation of a buried solvent-filled cavity and a twist of the conserved NPxxY motif, which leads to a larger population of active-like states at the G protein binding site. This effect is further augmented by protonation of the strongly conserved Asp79, which locks the NPxxY motif and solvent cavity in active-like conformations. The agonist binding site hence communicates with the intracellular region via a cascade of locally connected switches and the free energy landscapes along these contributes to understanding of how ligands can stabilize distinct receptor states. We demonstrate that the developed simulation protocol is transferable to other class A GPCRs and anticipate that it will become a useful tool in design of drugs with specific signaling properties.

String method, Enhanced sampling, Activation mechanism, Collective variable, Free energy calculations, Supervised learning

Introduction

G protein-coupled receptors (GPCRs) are membrane proteins that activate cellular signaling pathways in response to extracellular stimuli. There are >800 GPCRs in the human genome (1) and these recognize a remarkably large repertoire of ligands such as neurotransmitters, peptides, proteins, and lipids. This large superfamily plays essential roles in numerous physiological processes and has become the most important class of drug targets (2). All GPCRs share a common architecture of seven transmembrane (TM) helices, which recognize the cognate ligand in the extracellular region and triggers intracellular signals via a more conserved cytosolic domain (Fig. 1) (3). GPCRs are inherently flexible proteins that exist in multiple conformational states; and drug binding alters the relative populations of these. Agonists will shift the equilibrium towards active-like receptor conformations, which promote binding of G proteins and other cytosolic proteins (e.g. arrestin), leading to initiation of signaling via multiple pathways. In the apo state, GPCRs can still access active-like conformations and thereby exhibit a small degree of signaling, which is referred to as basal activity.

Breakthroughs in structure determination of GPCRs during the last decade have provided insights into the process of activation at atomic resolution (Fig. 1). In particular, crystal structures of the β_2 adrenergic receptor (β_2AR) in both active and inactive conformational states (4–6) have revealed hallmarks of GPCR activation. The observations made for this prototypical receptor have recently been reinforced by crystal and cryogenic electron microscopy (cryo-EM) structures for other family members (7). The most prominent feature of GPCR activation is a large $\sim$ 1.1 nm outward movement of TM6 on the intracellular side (Fig. 1), which creates a large cavity for binding of cytosolic proteins. For many GPCRs, this conformational change disrupts the "ionic lock", a salt bridge between the conserved Glu268$^{6.30}$ (superscripts denote Ballesteros-Weinstein numbering (8)) and Arg131$^{3.50}$ (the R in the conserved DRY motif) that contributes to stabilization of the inactive state. Conserved changes in the extracellular part are more difficult to discern due to the lower sequence conservation in this region. In general, the structural changes are relatively subtle and only involve a small contraction of the orthosteric site (4–6). In the case of the β_2AR, the catechol group of adrenaline forms hydrogen bonds with Ser207$^{5.46}$, which leads to a $\sim$ 0.2 nm inward movement of TM5. These structural changes then propagate through the receptor via several conserved motifs. The rearrangement of TM5 influences a connector region (PI$^{3.40}$F$^{6.44}$ motif), which is in contact with the strongly conserved Asp79$^{2.50}$ and the NP$^{7.50}$xxY$^{7.53}$ motif via a network of ordered water molecules. A water-filled cavity surrounding Asp79$^{2.50}$ contributes to stabilizing a sodium ion in several crystal structures of the inactive state, e.g. for the β_1AR (9, 10). Upon activation, the cavity collapses, leading to dehydration and displacement of the sodium ion and potentially protonation of Asp79$^{2.50}$ (11, 12). Activation also involves a twist of the NPxxY motif, which reorients Tyr326$^{7.53}$ to a position where it can form a water-mediated interaction with Tyr219$^{5.58}$ (13) and thereby enable formation of the G protein binding site. Molecular understanding of the role of these individual microswitches in activation could guide the design of drugs with tailored signaling profiles.

The allosteric control of GPCR activation by extracellular ligands cannot be fully understood from the static structures captured by crystallography or cryo-EM. Mutagenesis and

L.D., J.C. and O.F. designed the project. O.F. developed and implemented the simulation protocol. O.F. and P.M. performed the simulations and analyzed data. All authors contributed to interpreting the results and writing the paper.

The authors declare no conflict of interest.

E-mail: lucie.delemotte@scilifelab.se and jens.carlsson@icm.uu.se

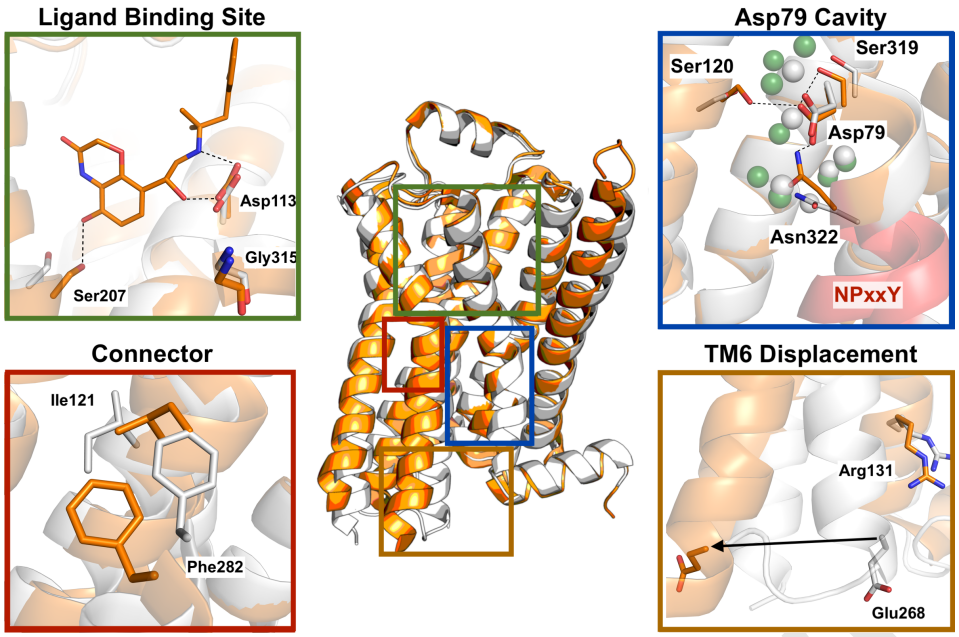

Fig. 1. The activation mechanism of GPCRs involves a series of microswitches. Agonist binding leads to a contraction of the binding site (quantified as the distance between Ser207$^{5.46}$ and Gly315$^{7.41}$), which leads to a conformational change in the connector region (Ile121$^{3.40}$ and Phe282$^{6.44}$). The transmembrane cavity containing Asp79$^{2.50}$ residue dehydrates and the NPxxY motif twists upon activation. Finally, TM6 moves outwards to create the G protein binding site, potentially disrupting the salt bridge between Arg131$^{3.50}$ and Glu268$^{6.30}$ ("ionic lock"). The inactive (PDB ID 2RH1) and active (PDB ID 3P0G) structures are represented in white and orange, respectively. The conserved NPxxY motif is highlighted in red. Important residues and the agonist are represented as sticks. Water molecules in the β_2AR and the β_1AR inactive structures (PDB IDs 2RH1 and 4BVN) are represented as white and green spheres, respectively.

spectroscopy studies (14–17) have suggested that the efficacy of a ligand is determined by a complex interplay between different microswitches and population of distinct states lead to specific functional outcomes. Molecular dynamics (MD) simulations are well suited to study the conformational landscape of GPCRs as this method can provide an atomistic view of the flexible receptor in the presence of membrane, aqueous solvent, and ligands. The seminal paper by Dror et al (18) provided insights into the activation mechanism of the β_2AR by monitoring how a crystallized active state conformation relaxed to an inactive state upon removal of the intracellular binding partner, a G protein mimicking nanobody, using MD simulations. Although key conformational changes involved in the transition from inactive to active conformations were identified in these simulations, this approach has inherent limitations. Indeed, understanding the roles of different microswitches in activation and the strength of the coupling between these requires mapping of the relevant free energy landscapes, which are still too costly to calculate using brute-force MD simulations. Enhanced sampling methods provide a means to explore the conformational landscapes of proteins to a relatively low computational cost (19). In this work, we aimed to identify the most probable path describing the transition between inactive and active states of β_2AR in the presence or absence of a bound agonist and optimized a version of the string method with swarms of trajectories for this purpose. The free energy landscapes associated with β_2AR activation revealed that whereas agonist binding is only loosely connected to outward movement of TM6, the coupling between microswitch pairs in the transmembrane region range from weak to strong and is influenced by the protonation of Asp79$^{2.50}$. Finally, we demonstrate that our approach can be transferred to study free energy landscapes of any class A GPCR at a modest computational cost.

Results

Optimizing the enhanced sampling protocol for increased efficiency. We aimed to compute the most probable transition pathway linking the inactive and active states of β_2AR and the relative free energy of the states lining this pathway. For this purpose we chose the string of swarms method (20). In this framework, the minimum free energy path in a N-dimensional collective variable (CV) space is estimated iteratively from the drift along the local gradient of the free energy landscape (Fig. S1). From each point on the string, a number of short trajectories are launched (a swarm), which enables us to calculate the drift. The string is then updated considering this drift and reparametrized to ensure full sampling of the configurational space along the pathway. Convergence is reached when the string diffuses around an equilibrium position. The method allows to sample a high-dimensional space at a relatively inexpensive computational cost since it only samples along the path of interest.

We characterized the pathway linking equilibrated conformations originating from active and inactive structures of the the β_2AR (PDB IDs 3P0G and 2RH1, respectively), adding two short strings spanning the active and in the inactive regions to increase sampling of the end state environments (see Material and Methods and SI Methods). First, we characterized a five dimensional CV set that embeds receptor activation by analyzing a MD simulation trajectory of spontaneous deactivation of the β_2AR (18). To do so, we identified metastable states by clustering simulation configurations, and classified those by training a fully connected neural network to identify states. The most important input values for classification were identified via deep Taylor decomposition and taken as CVs (Fig. 2 and S2). The set of CVs we inferred was a network of interatomic distances between all seven TM helices (Fig. 2). This CV set incorporates many known degrees of freedom im-

plicated in β_2AR activation, including the TM3-TM6 distance and a CV coupling the extra- and intracellular domains.

To speed up convergence of the string optimization, we initiated our string simulation from a rough guess of the minimum free-energy landscape. The latter was obtained by estimating the density of points from the Dror et al. trajectory in this CV space (Fig. S3). We also introduced algorithmic improvements to the string of swarms method: we adaptively chose the number of trajectories in a swarm, gradually increased the number of points on the string and introduced a reparametrization algorithm that improves performance as well as promotes exchanges of configurations between adjacent string points (Supplementary Methods, Fig. S1, S4). We carried out 300 iterations of string optimization for each system, considering a number of points on the string ranging from 20 to 43 and swarms consisting of 16 to 32 10 ps trajectories. Derivation of one free energy landscape requires a total of ∼2-3 μs simulation time.

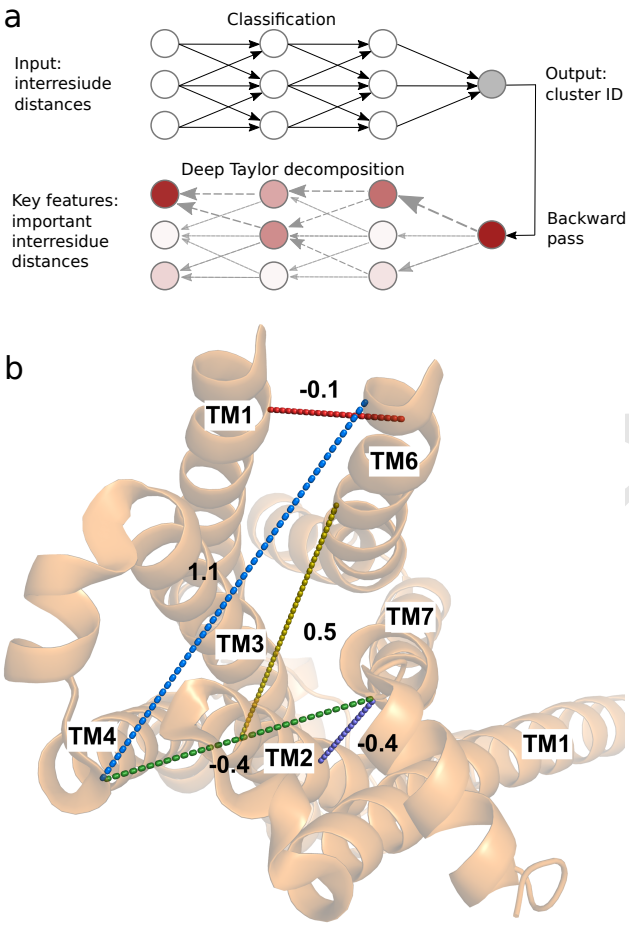

Fig. 2. The five-dimensional collective variable (CV) space used to optimize the minimum free energy path was identified in a data-driven manner. (**a**) A fully connected neural network was trained to classify configurations in active, intermediate, and inactive metastable states (clusters). Deep Taylor decomposition was then used to identify the most important input inter-residue distances for the classification decision and the top ones were used as CVs. (**b**) The five CVs used in this work projected onto an intracellular view of the active crystal structure (PDB ID 3P0G). The CVs corresponding to distances TM2-TM7, TM6-TM4, TM7-TM4, TM3-TM6 and TM6-TM5 defined in Table S1, are shown as purple, blue, green, yellow and red dashed lines, respectively. The change of these distance CVs from the inactive to the active state structures is reported in nm.

Minimum free energy pathway of β_2AR activation. We derived the most probable transition path (Fig. S5) between active and inactive states of β_2AR in the presence and absence of an agonist ligand. The swarm trajectories allowed us to compute transitions between discrete states in the vicinity of the most probable transition path and derive the associated free energy landscape (Fig. S6).

For the apo receptor, one distinguishes three minima, one in the active region, one in the inactive one, and an intermediate one between them (Fig. 3a and S7). As anticipated, regions close to the inactive endpoint are stabilized relative to the other two states. Binding of an agonist changes the number of minima to four and shifts the relative stability of states, making regions close to the active endpoint of lower free energy (Fig. 3b). Binding of the agonist ligand also modifies the order in which the helices rearrange, as can be seen when projecting the minimum free energy path along these various CVs (Fig. 3 and S5).

A number of characteristic variables (defined in Table S1) were calculated for the last iteration of the swarm of trajectories simulations. By localizing a sudden shift in these parameters' values, we could pinpoint the location of characteristic events on the string (Table S2, Fig. 3). In the absence of bound ligand, a sodium ion exits the cavity via the extracellular side early during activation and the other microswitches (Asp79$^{2.50}$ cavity dehydration, NPxxY twist) flip at the same time as the ion leaves. In the presence of the agonist ligand on the other hand, the ionic lock first breaks. This occurs between the region occupied by the crystal structure (PDB ID 2RH1) and the inactive state basin (Fig. 3b), in agreement with the observation that the ionic lock is not always formed in the inactive conformation (15, 21). Then the Asp79$^{2.50}$ cavity dehydrates shortly before the NPxxY twists to an active-like conformation.

Coupling between orthosteric and G-protein binding sites. As the final configurations of the string are at equilibrium in all dimensions, the trajectories from the last iterations can be used to compute the free energy landscape as a function of any variables (SI methods). This allowed a detailed analysis of how conformational changes induced by agonist binding propagate through the receptor to the G protein binding site. The roles of conserved microswitches were assessed by comparing free energy landscapes in the presence and absence of a bound agonist. We first evaluated how the distance between Ser207$^{5.46}$ and Gly315$^{7.41}$, which reflects how the binding site contracts upon activation, influences the intracellular distance between TM6 and TM3 (Fig. 4a,b). In the absence of bound agonist, the receptor accessed both active and inactive conformations of the binding site, with the minimum of the free energy located close to the inactive state distance between TM3 and TM6 . The TM3-TM6 distance could be as large as 1.5 nm when the ligand binding site was in the active conformation, an observation compatible with basal activity. Agonist binding led to the stabilization of a contracted conformation of the binding site, corresponding to an inward movement of Ser207$^{5.46}$ (Fig. 4b). Although both inactive and active conformations remained accessible in the presence of the ligand, the minimum of the TM3-TM6 distance was shifted towards a more active-like state. A remarkable long-range allosteric coupling (>2 nm) between the orthosteric and G protein binding sites was hence captured by our calculations. The 0.5 nm outward movement

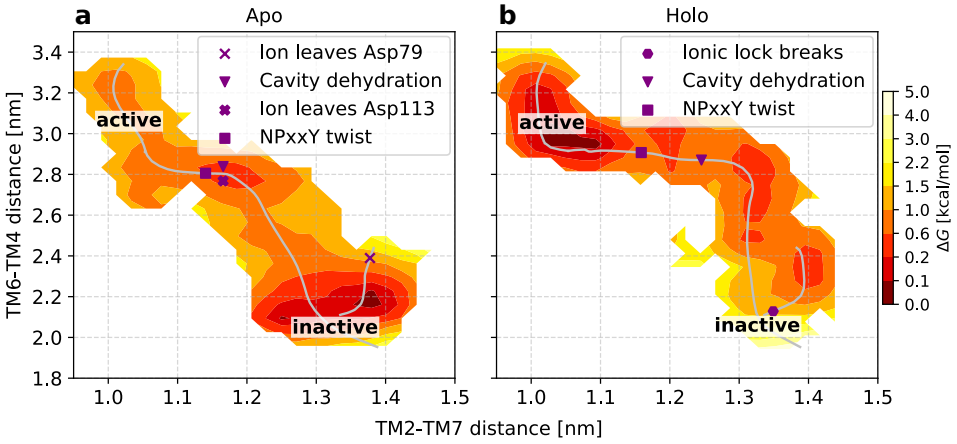

Fig. 3. The free energy landscapes are projected onto the first two CVs used to optimize the minimum free energy path. The minimum free energy pathways were optimized in the absence (**a**) and in the presence of a bound agonist (**b**). Characteristic events occurring during activation (see Table S1-S2 for their definition) are located on the string.

of TM6 at the minimum of the landscape is smaller than that observed in active crystal structures, in agreement with experiments demonstrating that the fully active conformation can only be stabilized in the presence of an intracellular partner (15, 22).

Propagation of activation through microswitches. Structural changes in the orthosteric site of the β_2AR have been proposed to propagate towards the intracellular part via a connector region centered around Ile121$^{3.40}$ and Phe282$^{6.44}$ (6). Whereas we found the contraction of the orthosteric site and the outward movement of TM6 to be loosely coupled, the free energy landscapes demonstrated that changes in the conformation of Ser207$^{5.46}$ has a strong influence on the connector region (Fig. 4d,e). In the absence of agonist, both inactive and active conformations of the connector were populated (Fig. 4d). In contrast, agonist binding resulted in a single free energy minimum where both Ser207$^{5.46}$ and the connector were constrained to their active conformations (Fig. 4e).

From the connector region, activation is propagated via several conserved motifs (Fig. 4g,h,j,k,m and n) (3). To investigate the communication between microswitches in the core of the TM region, we analyzed if the connector was coupled to solvation of the water network surrounding Asp79$^{2.50}$ and conformation of the NPxxY motif. In the apo state, an inactive connector region was tightly coupled to a hydrated cavity with $\sim 10-17$ waters (Fig. 4g). This result is consistent with a high-resolution crystal structure of an inactive β_1 adrenergic receptor (PDB ID 4BVN), in which a solvent network in this region was identified (9). The active connector was compatible with both the fully hydrated cavity and a desolvated state with up to $\sim 4-5$ water molecules. The free energy landscapes also suggested that the more dehydrated cavity, which resembled that observed in the active β_2AR conformation, was slightly more favored in the presence of agonist (Fig. 4g). The active connector resulted in two minima for the NPxxY motif and the corresponding receptor structures resembled the active and inactive structures in this region (Fig. 4j). Agonist binding favored the dehydration of the cavity but only resulted in a small perturbation of the free energy landscapes along the NPxxY RMSD dimension (Fig. 4h,k). In contrast to the connector, the Asp79$^{2.50}$ cavity and NPxxY motif were hence loosely coupled to the orthosteric site. However, It should be noted that the connector switch influenced the NPxxY motif via the hydration state of the Asp79$^{2.50}$ cavity. An inactive connector could be coupled to both a fully hydrated (inactive) Asp79$^{2.50}$ cavity and inactive NPxxY motif in the simulation of the apo receptor. The active connector, on the other hand, allowed the receptor to access both the inactive and active conformations of these two switches. The final combination of microswitches connected the NPxxY region to the motion of TM6 (Fig. 4m,n). A larger TM3-TM6 distance was clearly favored for an active-like NPxxY motif and several metastable states lined the minimum free energy pathway between inactive and active conformations. The locations of minima in the free energy landscape were only slightly perturbed by ligand binding, but the barriers between the states were reduced for the receptor-agonist complex (Fig. 4n).

Several structures of class A GPCRs solved in the inactive state have revealed a sodium ion bound to the conserved Asp79$^{2.50}$ (Fig. 5a) (10). Sodium binding to this residue has also been studied for several class A GPCRs with simulations (24–26). To investigate potential interactions between Asp79$^{2.50}$ and sodium ions, which were randomly added at physiological concentration to the simulation system, we calculated the free energy landscape as a function of TM6 displacement and the distance between Asp79$^{2.50}$ and the closest sodium ion (Fig. 5b, c). In the apo form, five meta-stable states were identified (Fig. 5b). In the active-like conformation of TM6, the closest sodium interacted with a specific site in the second extracellular loop. Notably, sodium ions have been confirmed to bind in this pocket in crystal structures of adrenergic receptors (Fig. 5a) (23). In an intermediate conformation of TM6, the closest sodium was either bound to the second extracellular loop or descended into the binding site and formed a salt bridge to Asp113$^{3.32}$. Finally, in the completely inactive state of TM6, the closest sodium ion either remained bound to Asp113$^{3.32}$ or was located in the Asp79$^{2.50}$ cavity. Sodium was hence only present in the Asp79$^{2.50}$ cavity when TM6 had completely relaxed to an inactive conformation and even small increases of the TM3-TM6 distance were incompatible with ion binding to this site. In the agonist-bound receptor, sodium remained strongly bound to the second extracellular loop irrespective of the TM3-TM6 distance (Fig. 5c). This likely results from access to the binding site via the extracellular side being blocked by the bound ligand (11, 24, 26). Spontaneous Na^+ binding to appropriate protein regions further confirms

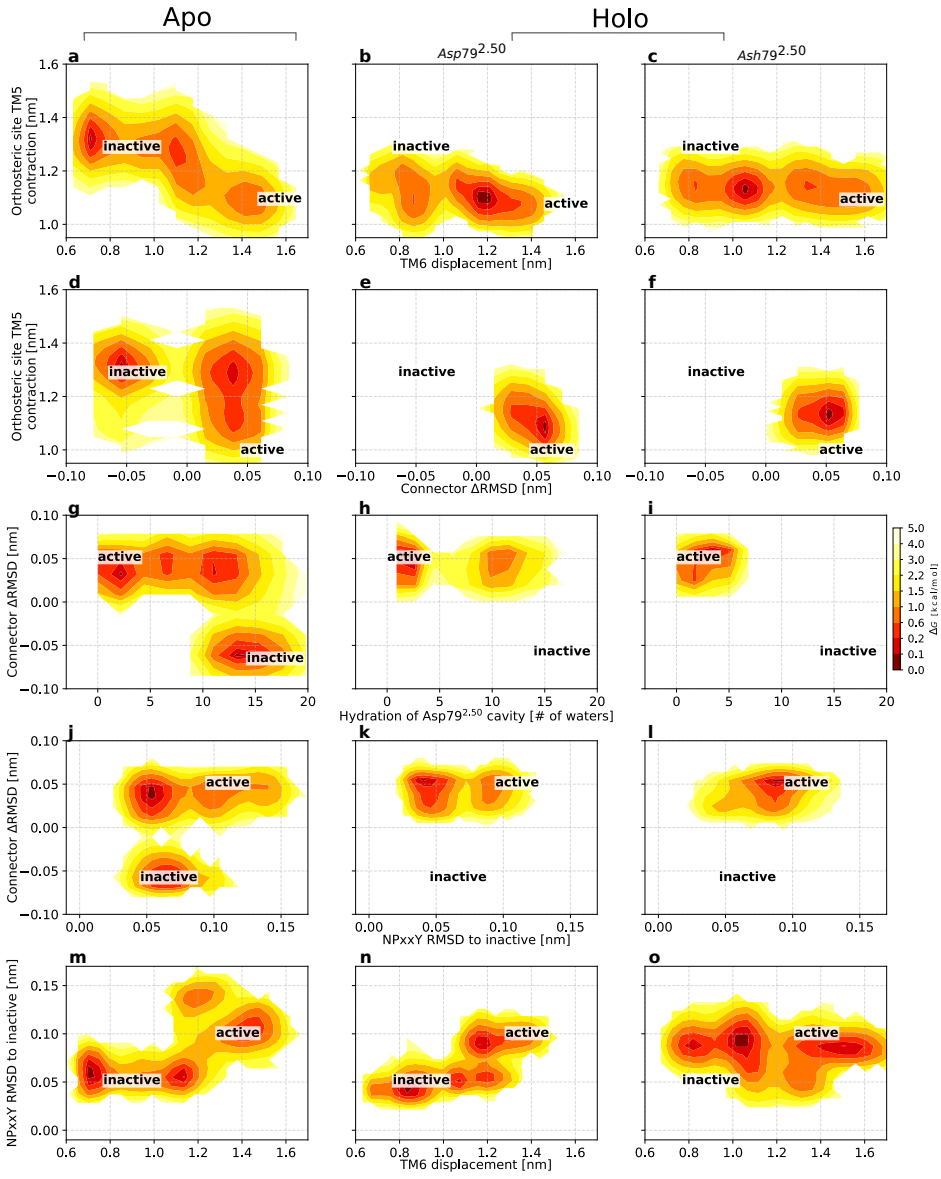

Fig. 4. Free energy landscapes projected along variables of interest highlight changes in the pairwise coupling of microswitches following binding of an agonist ligand (middle column) and protonation of conserved residue Asp$79^{2.50}$ (right column). The free energy landscapes are projected along (**a-c**) the distance between Ser$207^{5.46}$ and Gly$315^{7.41}$, representing the ligand binding site contraction and the distance between Leu$272^{6.34}$ and Arg$131^{3.50}$, representing the outward movement of TM6; (**d-f**) the distance between Ser$207^{5.46}$ and Gly$315^{7.41}$ and difference between the RMSD of Ile$121^{3.40}$ and Phe$282^{6.44}$ heavy atoms to the active and inactive crystal structures, representing the connector region $\triangle$RMSD; (**g-i**) the connector region $\triangle$RMSD and the number of water molecules within 0.8 nm from Asp$79^{2.50}$, representing the hydration of the Asp$79^{2.50}$ cavity; (**j-l**) the connector region $\triangle$RMSD and the RMSD of the NPxxY motif relative to the inactive state; (**m-o**) the NPxxY motif RMSD and the displacement of TM6. Active and inactive state regions are labeled for each variable pair. Regions of low free energy are shown in red and of high free energy in light yellow. Free energies are reported in kcal/mol. See Table S1for microswitch definitions.

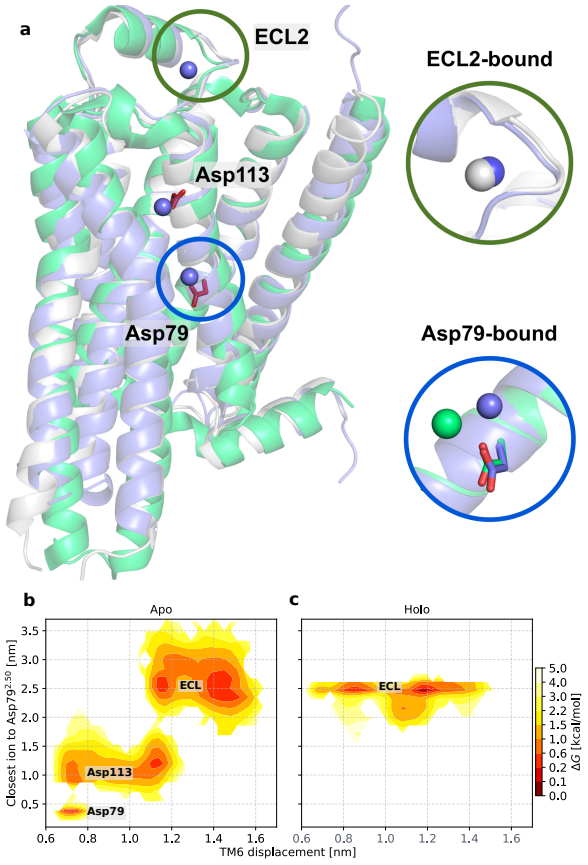

Fig. 5. A sodium ion can bind to three sites in a state-dependent manner. (**a**) The inactive β_1AR (PDB code: 4BVN (9)), β_2AR(PDB code: 4LDE (23)) structures and representative inactive state simulation snapshot are shown as green, white and blue ribbons, respectively. Representative positions of sodium ions are shown as blue spheres (the coordinates of the top ion bound to the extracellular loop was extracted from the holo trajectory in an active-like conformation, while the two lower positions were extracted from apo, inactive state simulations). In the insets, a comparison to the position of the Na^+ ions found in the active β_2AR structure (white sphere) and in the inactive β_1AR structure (green sphere) is provided. (**b, c**) Free energy landscapes along the TM6 displacement and the closest distance between Asp79$^{2.50}$ and a sodium ion for the apo and holo simulations.

the relevance of the conformational sampling enabled by the our computational protocol.

Impact of Asp79$^{2.50}$ protonation. Several previous studies have suggested that Asp79$^{2.50}$, the most conserved residue among class A GPCRs, has a pK_a value close to physiological pH and that the ionization state of this residue changes upon activation (11, 12). To assess the role of Asp79$^{2.50}$ in receptor activation, we repeated the calculations of the minimum free energy pathway of activation with this residue in its protonated (neutral) form.

The free energy landscapes describing changes in the orthosteric site were similar to those obtained in simulations with Asp79$^{2.50}$ ionized (Fig. 4c,f and Fig. S8). There was a weak coupling between the orthosteric site and the intracellular region, with two major energy wells describing the conformation of TM6. Compared to the agonist-bound receptor with Asp79$^{2.50}$ ionized, the minima were shifted further towards active-like conformations for the protonated state (Fig. 4c). The inward movement of Ser207$^{5.46}$ upon agonist binding remained strongly coupled to conformational changes observed in the connector region irrespective of the ionization state of Asp79$^{2.50}$ (Fig. 4e,f). The largest effects of Asp79$^{2.50}$ protonation were observed for the hydrated cavity surrounding this residue (Fig. 4h,i) and the NPxxY motif (Fig. 4k,l): whereas the free energy landscapes showed that both active- and inactive-like conformations of the NPxxY motif and cavity were populated in simulations with ionized Asp79$^{2.50}$, the protonated state resulted in a single energy well close to the active conformation for both these microswitches (Fig. 4j,l). It was also evident that TM6 was stabilized in more active-like conformations by the protonated Asp79$^{2.50}$ (Fig. 4n,o).

Comparison of representative structures from the simulations of ionized and protonated Asp79$^{2.50}$ in active-like states revealed that two distinct conformations of the NPxxY motif were obtained (Fig. 6). The simulations carried out with ionized Asp79$^{2.50}$ favored structures that were more similar to the crystal structure of the active β_2AR. An alternative conformation of the NPxxY motif appeared for the protonated Asp79$^{2.50}$, which was not favored energetically in the simulations of the ionized form (Fig. S9k,l). Although this conformation of the NPxxY motif did not match any β_2AR crystal structure, it was strikingly similar to conformations observed in crystal structures of other class A GPCRs in either agonist-bound (serotonin 5-HT_{2B} and A_{2A} adenosine receptors) or active (angiotensin II type 1) conformations (Fig. 6) (27, 28). Our protocol thus allowed us to sample metastable states that were never captured by structural methods. This indicates that the computational methodology we present here can likely be applied to several members of the class A GPCR superfamily.

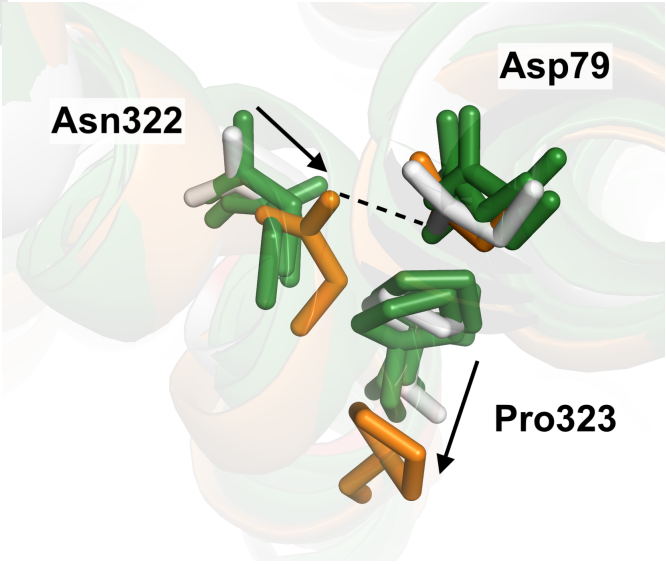

Fig. 6. An alternative conformation of the NPxxY was identified along the most probable pathway calculated with a protonated Asp79$^{2.50}$. Representative simulation snapshots of active-like states of the β_2AR with Asp79$^{2.50}$ ionized and neutral are shown in orange and white, respectively. Structures of other GPCRs captured in agonist-bound or active states are shown in green (PDB IDs: 6DRX (27) - Agonist-bound 5-HT_{2B}, 3QAK (29) - Agonist-bound A_{2A}AR, and 6DO1 (30) - Angiotensin II type 1 receptor in an active conformation). The structural comparison highlights the resemblance between an alternative conformation of the NPxxY motif that is favored by Asp79$^{2.50}$ protonation and crystal structures of other class A GPCRs. Arrows indicate structural differences between the simulation snapshots with Asp79$^{2.50}$ neutral and ionized.

Transferability of the methodology to other GPCRs. Efficient characterization of free energy landscapes with the string method relies on selection of appropriate CVs, a non-trivial task. Here, CVs were derived from a conventional MD trajectory of β_2AR deactivation in a data-driven fashion. Considering that the conformational changes involved in class A GPCR activation are largely conserved (3), we explored the possibility of transferring the CVs to the conformational sampling of other GPCRs. We mapped the CVs identified for β_2AR to ten other class A GPCRs with active and inactive structures available (Fig. 7). Strikingly, the active and inactive structures clearly separate in two distinct clusters. This indicates that these CVs can describe the activation of the entire family of class A GPCRs and the protocol presented herein can be applied to these other receptors.

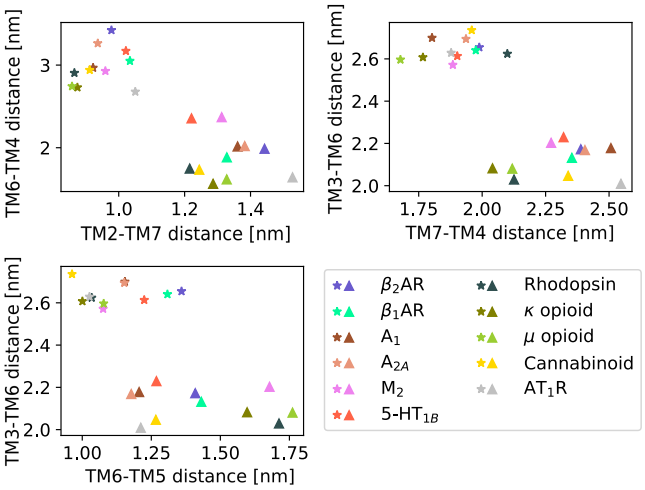

Fig. 7. Active and inactive class A GPCR structures cluster into two distinct groups in the 5-dimensional CV space used in this work. β_2AR and ten other class A GPCRs active (stars) and inactive (triangles) structures are projected onto the five original CVs. Details of the mapping can be found in Table S3. Their clustering into two groups highlights that the activation mechanism of all class A GPCRs can likely by described by the CVs determined herein.

Discussion

Crystallographic structures of the β_2AR in inactive and active conformations provides a basis for molecular understanding of GPCR signaling. However, it has become increasingly clear that these static structures do not capture all functionally relevant states involved in activation of these molecular machines. In a pioneering study, Dror et al. gained insights into the activation pathway of the β_2AR from a large number of long-timescale MD simulations (18). Despite this computational tour-de-force enabled by the development of hardware specialized for MD, these simulations did not allow to quantify the accessibility of different conformational states. The approach proposed in this work builds on the data generated by Dror et al. and further allows to assess the impact of agonist binding on microswitches involved in activation.

Despite some differences in simulation setups, our work recapitulates several key findings of the work by Dror et al (18). In agreement with the long time-scale simulations, we find that the orthosteric and G protein-binding site are loosely coupled. However, only through the analysis of the free energy landscapes could we determine that coupling between spatially connected microswitches ranged from very strong to relatively weak and was influenced by protonation of Asp79$^{2.50}$. Our study also illustrates that the energy landscapes depend on the variables chosen to project the conformational states. This is not an artefact of the protocol but rather an inherent limitation of dimensionality reduction, which has specific implications for experimental design: depending on the placement of spectroscopic probes, one may only be able to resolve a subset of available states. In particular, it is now clear that considering only three states along the activation path (an active, intermediate and inactive state) will not allow to capture the complexity of the conformational changes induced by ligand binding.

One of the unresolved questions regarding GPCR activation is how the orthosteric site communicates with microswitches buried in the TM region. Whereas the β_2AR crystallized in complex with agonists in the absence of intracellular partner (e.g. G protein or G protein-mimicking nanobody) are similar to identical to those determined with antagonists (31), NMR spectroscopy experiments have demonstrated that agonist binding does stabilize other conformations in certain parts of the TM region, e.g. close to Met215$^{5.54}$ and Met82$^{2.53}$. For example, the region surrounding Met82$^{2.53}$, located one helical turn above Asp79$^{2.50}$, was shown to adopt two conformations in the absence of orthosteric ligand. A bound agonist, on the other hand, restricted this region to a single active-like state (22). Similar to these experiments, our free energy landscapes demonstrate that several conformations of the connector and Asp79$^{2.50}$ cavity are available in the apo condition. In the presence of agonist, the connector is locked in a single state and a desolvated state of the Asp79$^{2.50}$ cavity is stabilized, which creates a more active-like receptor conformation in the vicinity of Met82$^{2.53}$. In agreement with the NMR data, we also find that the agonist cannot stabilize the fully active conformation of the receptor and that TM6 accesses several intermediate conformations that are distinct from those observed in crystal structures.

Several recent experimental studies have demonstrated that Asp79$^{2.50}$ and residues forming the hydrated TM cavity play an important role in signalling and can even steer activation via G protein-dependent and G protein-independent pathways (17, 32, 33). One mechanism by which Asp79$^{2.50}$ could control the receptor conformation is via its protonation state. Agonist binding destabilizes the water network in the solvated TM cavity, which may lead to a larger population of protonated Asp79$^{2.50}$ and disrupt binding of sodium to this pocket (11, 12). In turn, the protonated Asp79$^{2.50}$ stabilizes a structure of the NPxxY motif that has been observed for other class A GPCRs crystallized in active and active-like states, suggesting that this alternative conformation of TM7 may be relevant for function. For example, NMR experiments have shown that agonists that preferentially signal via arrestin mainly affect the conformation of TM7 (34). Interestingly, the ionized and protonated forms of Asp79$^{2.50}$ also stabilize different TM6 conformations, which could change the intracellular interface that interacts with G proteins and arrestins. These results, combined with the fact Asp79$^{2.50}$ is the most conserved residue in the class A receptor family, support that this region is a central hub for controlling class A receptor activation.

With the protocol developed herein, we sampled enough transitions along the activation path to obtain free energy profiles of GPCR activation by accumulating a few microseconds of total simulation time. Compared to regular MD simulations, the optimised string of swarms method can thus provide reliable energetic insights using 1-2 orders magnitude less simulation time (18, 35). From a practical point of view, the short trajectories in the swarms of trajectories method are easy to run in parallel with minimal communication overhead even in a heterogeneous computational environment. An important consideration that has guided our choice of enhanced sampling methodology is that the method has the advantage to function well in high dimensional space, i.e. with many CVs. This is because we only optimize the one-dimensional string, instead of opting to sample the entire landscape spanned by the CVs. This means we can utilize a high-dimensional CV space, thus alleviating the need to reduce the dimensionality of the conformational landscape to 2 or 3 dimensions, as is done in most CV-based methods such as umbrella sampling or metadynamics (36).

On the other hand, a well-known limitation with the swarms of trajectories method is that it only guarantees to converge to the most probable path closest to the initial path guess, and not necessarily the globally most probable path. The naive assumption of a straight initial path is not guaranteed to converge to the latter. Here we have proposed to alleviate this shortcoming by exploiting previous knowledge of the activation pathway and deriving an initial guess of the pathway likely to be close to the globally most probable path. An initial pathway can also be transferred from a similar system (as revealed in Fig. 7) or inferred from available experimental data. If multiple pathways are nevertheless expected, the protocol presented herein provides the tools necessary to compare them: the swarms from separate string simulations can be included in the same transition matrix and be used to compute a single free energy landscape. It is also worth noting that the relative free energies of the endpoint conformations (evaluated by integration over the endpoints basins) do not depend on the transition pathway and should anyways be estimated correctly. The protocol is also applicable to complex transitions involving many intermediate states: in such case, one may launch multiple strings to explore different parts of the activation path, and let every substring converge separately, eventually combining the transitions derived from them to yield a single free energy landscape. Finally, we note that the string of swarms method can also be used as a complementary method to instantiate Markov State Models (MSM) simulations (37).

Despite the major progress in structural biology for GPCRs, many aspects of receptor function are not well understood. Insights from atomistic MD simulations will continue to be valuable tools for interpreting experimental data. We expect our methodology to allow further insights into how binding of agonists influences the conformational landscape, potentially making it possible to design ligands with biased signalling properties (17). The method is equally well suited to study the effect of allosteric modulators and the influence of different lipidic environments. As the same approach can be transferred to other class A GPCRs, future applications will shed light on the common principles of activation as well as on the details that give each receptor a unique signalling profile, paving the way to the design of more effective drugs.

Materials and Methods

All swarm of trajectories simulations were instantiated with the coordinates from the 3P0G structure (the first two simulation systems in Table S4). The Asn187Glu mutation in 3P0G was reverted and Glu122$^{3.41}$ was protonated due to its localization in a particularly hydrophobic pocket, as has been common practice in other simulations of this receptor (31). Residues His172$^{4.64}$ as well as His178$^{4.70}$ were protonated at the epsilon position, in order to face negatively charged residues Glu107$^{3.26}$ and Glu180 respectively. The systems were parametrized using the CHARMM36m force field (38) and the TIP3P water model (39). The protein was inserted in a POPC (40) bilayer and solvated in explicit solvent. Na^+ and Cl^- ions were added at 0.15M concentration. System preparation was performed using CHARMM-GUI (41). MD simulations were run with GROMACS 2016.5 (42) patched with plumed 2.4.1 (43) under a 1 atm pressure and at 310.15 K.

To identify CVs, we performed clustering (44) on the frames from unrestrained MD simulation trajectory of β_2AR (condition A in (18)). The CVs were selected by training a multilayer perceptron classifier (45) using as input all the inter-residue distances and as output the cluster ID, followed by using Deep Taylor decomposition (46) to find key distances that could discriminate between clusters.

The endpoints of the main strings describing the transition between inactive and active states (subscript $_t$) were fixed to the output coordinates of equilibrated structures 2RH1 (4) and 3P0G (6). The initial path for simulations Holo$_t$ was guessed using data from (18): a rough estimate of the free energy landscape was calculated from the probability density landscape estimated using the Scikit-learn (45) kernel density estimator with automatic bandwidth detection (Fig. S3). Two additional short strings were set up to increase sampling in the active (subscript $_a$) and inactive (subscript $_i$) regions. The average, partially converged path between iterations 20-30 from Holo$_t$ was used as input path for simulations Apo$_t$ and Holo$_{Ash79,t}$. All active (Apo$_a$, Holo$_a$ and Holo$_{Ash79,a}$) and inactive (Apo$_i$, Holo$_i$ and Holo$_{Ash79,i}$) substrings were initiated as straight paths between the endpoints. The swarms of trajectories simulations with optimizations (see SI and Fig. S4) were run for 300 iterations, at which point the strings had not changed on average for many iterations (Fig. S5 and S8) and posterior distribution of free energy profiles given the data was small (Fig. S7).

By discretizing the system into microstates, or bins, it is possible to use the short trajectories from the swarms to create a transition matrix and derive the free energy distribution of the system (47) along some variable (Fig. S6). In practice, The transition probabilities T_{ij} of the transition matrix T can be estimated from the normalized number of transitions, N_{ij}, from bin i to bin j: $T_{ij} = N_{ij} / \sum_k N_{ik}$. The transition matrix of a physical system at equilibrium is constrained by detailed balance, such that for the stationary probability distribution, ρ: $\rho_i T_{ij} = \rho_j T_{ji}$. Metropolis Markov chain Monte Carlo (MCMC) was used to sample over the posterior distribution of transition matrices, given the unregularized elements of T_{ij} (48), and thereby obtain a distribution of free energy profiles for ρ (Fig. S7, S9 and S10). All code to run the simulations and reproduce the results in this paper has been published online (49).

ACKNOWLEDGMENTS. This work was supported by grants from the Gustafsson Foundation and Science for Life Laboratory to JC and LD. The work was also supported by grants from the Swedish Research Council (2017-4676) and the Swedish strategic research program eSSENCE to JC. The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at PDC Centre for High Performance Computing (PDC-HPC). The authors thank D.E. Shaw Research for providing access to their MD trajectories.

1. Fredriksson R, Lagerström MC, Lundin LG, Schiöth HB (2003) The g-protein-coupled receptors in the human genome form five main families. phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* 63(6):1256–1272.
2. Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE (2017) Trends in gpcr drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* 16(12):829.

3. Weis WI, Kobilka BK (2018) The molecular basis of g protein–coupled receptor activation. *Annu. Rev. Biochem.* 87:897–919.
4. Cherezov V, et al. (2007) High-resolution crystal structure of an engineered human β2-adrenergic g protein–coupled receptor. *Science* 318(5854):1258–1265.
5. Rasmussen SGF, et al. (2011) Crystal structure of the β2 adrenergic receptor–gs protein complex. *Nature* 477(7366):549–555.
6. Rasmussen SG, et al. (2011) Structure of a nanobody-stabilized active state of the β2 adrenoceptor. *Nature* 469(7329):175.
7. García-Nafría J, Tate CG (2019) Cryo-EM structures of GPCRs coupled to gs, gi and go. *Mol. Cell. Endocrinol.* 488:1–13.
8. Ballesteros JA, Weinstein H (1995) [19] integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in g protein-coupled receptors in *Methods in Neurosciences*. (Elsevier), pp. 366–428.
9. Miller-Gallacher JL, et al. (2014) The 2.1 åresolution structure of cyanopindolol-bound β1-adrenoceptor identifies an intramembrane na+ ion that stabilises the ligand-free receptor. *PLoS One* 9(3):e92727.
10. Katritch V, et al. (2014) Allosteric sodium in class a gpcr signaling. *Trends Biochem. Sci.* 39(5):233–244.
11. Vickery ON, et al. (2018) Intracellular transfer of na+ in an active-state g-protein-coupled receptor. *Structure* 26(1):171–180.e2.
12. Ranganathan A, Dror RO, Carlsson J (2014) Insights into the role of asp$79^{2.50}$ in β2 adrenergic receptor activation from molecular dynamics simulations. *Biochemistry* 53(46):7283–7296.
13. Latorraca NR, Venkatakrishnan A, Dror RO (2016) Gpcr dynamics: structures in motion. *Chem. Rev.* 117(1):139–155.
14. Lamichhane R, et al. (2015) Single-molecule view of basal activity and activation mechanisms of the g protein-coupled receptor β2ar. *Proc. Natl. Acad. Sci. U.S.A.* 112(46):14254–14259.
15. Manglik A, et al. (2015) Structural insights into the dynamic process of β2-adrenergic receptor signaling. *Cell* 161(5):1101–1111.
16. Gregorio GG, et al. (2017) Single-molecule analysis of ligand efficacy in β2ar–g-protein activation. *Nature* 547(7661):68.
17. Picard LP, Schonegge AM, Bouvier M (2019) Structural insight into g protein-coupled receptor signaling efficacy and bias between gs and β-arrestin. *ACS Pharmacol Transl Sci.*
18. Dror RO, et al. (2011) Activation mechanism of the β2-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* 108(46):18684–18689.
19. Harpole TJ, Delemotte L (2018) Conformational landscapes of membrane proteins delineated by enhanced sampling molecular dynamics simulations. *Biochim Biophys Acta Biomembr* 1860(4):909–926.
20. Pan AC, Sezer D, Roux B (2008) Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* 112(11):3432–3440.
21. Dror RO, et al. (2009) Identification of two distinct inactive conformations of the β2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. U.S.A.* 106(12):4689–4694.
22. Nygaard R, et al. (2013) The dynamic process of β2-adrenergic receptor activation. *Cell* 152(3):532–542.
23. Ring AM, et al. (2013) Adrenaline-activated structure of β2-adrenoceptor stabilized by an engineered nanobody. *Nature* 502(7472):575–579.
24. Selent J, Sanz F, Pastor M, Fabritiis GD (2010) Induced effects of sodium ions on dopaminergic g-protein coupled receptors. *PLoS Comput. Biol.* 6(8):e1000884.
25. Hu X, et al. (2019) Kinetic and thermodynamic insights into sodium ion translocation through the μ-opioid receptor from molecular dynamics and machine learning analysis. *PLoS Comput. Biol.* 15(1):e1006689.
26. Selvam B, Shamsi Z, Shukla D (2018) Universality of the sodium ion binding mechanism in class a g-protein-coupled receptors. *Angew. Chem.* 130(12):3102–3107.
27. McCorvy JD, et al. (2018) Structural determinants of 5-HT2b receptor activation and biased agonism. *Nat. Struct. Mol. Biol.* 25(9):787–796.
28. Wingler LM, McMahon C, Staus DP, Lefkowitz RJ, Kruse AC (2019) Distinctive activation mechanism for angiotensin receptor revealed by a synthetic nanobody. *Cell* 176(3):479–490.e12.
29. Xu F, et al. (2011) Structure of an agonist-bound human a2a adenosine receptor. *Science* 332(6027):322–327.
30. Wingler LM, McMahon C, Staus DP, Lefkowitz RJ, Kruse AC (2019) Distinctive activation mechanism for angiotensin receptor revealed by a synthetic nanobody. *Cell* 176(3):479–490.
31. Rosenbaum DM, et al. (2011) Structure and function of an irreversible agonist-β2 adrenoceptor complex. *Nature* 469(7329):236–240.
32. Fenalti G, et al. (2014) Molecular control of δ-opioid receptor signalling. *Nature* 506(7487):191–196.
33. Eddy MT, et al. (2018) Allosteric coupling of drug binding and intracellular signaling in the a2a adenosine receptor. *Cell* 172(1-2):68–80.e12.
34. Liu JJ, Horst R, Katritch V, Stevens RC, Wuthrich K (2012) Biased signaling pathways in β2-adrenergic receptor characterized by 19f-nmr. *Science* 335(6072):1106–1110.
35. Kohlhoff KJ, et al. (2014) Cloud-based simulations on google exacycle reveal ligand modulation of gpcr activation pathways. *Nat. Chem.* 6(1):15.
36. Pietrucci F (2017) Strategies for the exploration of free energy landscapes: Unity in diversity and challenges ahead. *Reviews in Physics* 2:32–45.
37. Pan AC, Roux B (2008) Building markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* 129(6):064107.
38. Huang J, et al. (2016) Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* 14(1):71.
39. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79(2):926–935.
40. Klauda JB, et al. (2010) Update of the charmm all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* 114(23):7830–7843.
41. Lee J, et al. (2015) Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *J. Chem. Theory Comput.* 12(1):405–413.
42. Abraham MJ, et al. (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1:19–25.
43. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) Plumed 2: New feathers for an old bird. *Comput. Phys. Commun.* 185(2):604–613.
44. Westerlund AM, Delemotte L (2018) Effect of ca2+ on the promiscuous target-protein binding of calmodulin. *PLoS Comput. Biol.* 14(4):e1006072.
45. Pedregosa F, et al. (2011) Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12:2825–2830.
46. Montavon G, Lapuschkin S, Binder A, Samek W, Müller KR (2017) Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern Recognit.* 65:211–222.
47. Lev B, et al. (2017) String method solution of the gating pathways for a pentameric ligand-gated ion channel. *Proc. Natl. Acad. Sci. U.S.A.* 114(21):E4158–E4167.
48. Beauchamp KA, et al. (2011) Msmbuilder2: modeling conformational dynamics on the picosecond to millisecond scale. *J. Chem. Theory Comput.* 7(10):3412–3419.
49. Fleetwood O (2019) Code to study gpcrs using the string method with swarms of trajectories (https://github.com/delemottelab/gpcr-string-method-2019).