bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Energy landscapes reveal agonist control of GPCR activation via microswitches

Oliver Fleetwood,[†] Pierre Matricon,[‡] Jens Carlsson,[‡] and Lucie Delemotte[*,†]

†*Science for Life Laboratory, Department of Applied Physics, KTH Royal Institute of Technology, Stockholm, Sweden*

‡*Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden*

E-mail: lucie.delemotte@scilifelab.se

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 several high resolution structures of the inactive and active receptor states are available, the atomistic details of the allosteric coupling that transmits the signal across the membrane are not fully understood. We calculated free energy landscapes of β_2 adrenergic receptor activation using atomistic molecular dynamics simulations in an optimized string of swarms framework, which sheds new light on the roles of microswitches in governing the equilibrium between conformational states. 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, a twist of the conserved NPxxY motif, and an interaction between two conserved tyrosines in transmembrane helices 5 and 7 (Y-Y motif),

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

which leads to a larger population of active-like states at the G protein binding site. This coupling is further augmented by protonation of the strongly conserved Asp79$^{2.50}$, which locks the solvent cavity, NPxxY, and Y-Y motifs in active-like conformations. The agonist binding site hence communicates with the intracellular region via a cascade of locally connected switches and characterizing the free energy landscapes along the conformation of these microswitches 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 the design of drugs with specific signaling properties.

Introduction

G protein-coupled receptors (GPCRs) are membrane proteins that activate cellular signaling pathways in response to extracellular stimuli. There are more than 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)[3,4]. 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 smaller 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,

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

crystal structures of the β_2 adrenergic receptor (β_2AR) in both active and inactive conformational states[5–14] 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[15]. The most prominent features of GPCR activation are a large $\sim 1.0 - 1.4$ nm outward movement of TM6 and a slight inward shift of TM7 on the intracellular side (Fig. 1), which create a large cavity for binding of cytosolic proteins.

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[5–7]. In the case of the β_2AR, the catechol group of adrenaline forms hydrogen bonds with Ser207$^{5.46}$ (superscripts denote Ballesteros-Weinstein numbering[16]), which leads to a ~ 0.2 nm inward bulge 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 highly conserved Asp79$^{2.50}$ and NP$^{7.50}$xxY$^{7.53}$ motif via a water-filled cavity. Upon activation, the cavity dehydrates and the NPxxY motif twists, which reorients Tyr326$^{7.53}$ to a position where it can form a water-mediated interaction with Tyr219$^{5.58}$ ("Y-Y motif")[17] and enables formation of the G protein binding site. Characterization of the role these individual microswitches play in determining the conformational ensemble of structures could guide the design of drugs with tailored signaling profiles.

Indeed, 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 spectroscopy studies[18–21] 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[22–30]. The

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

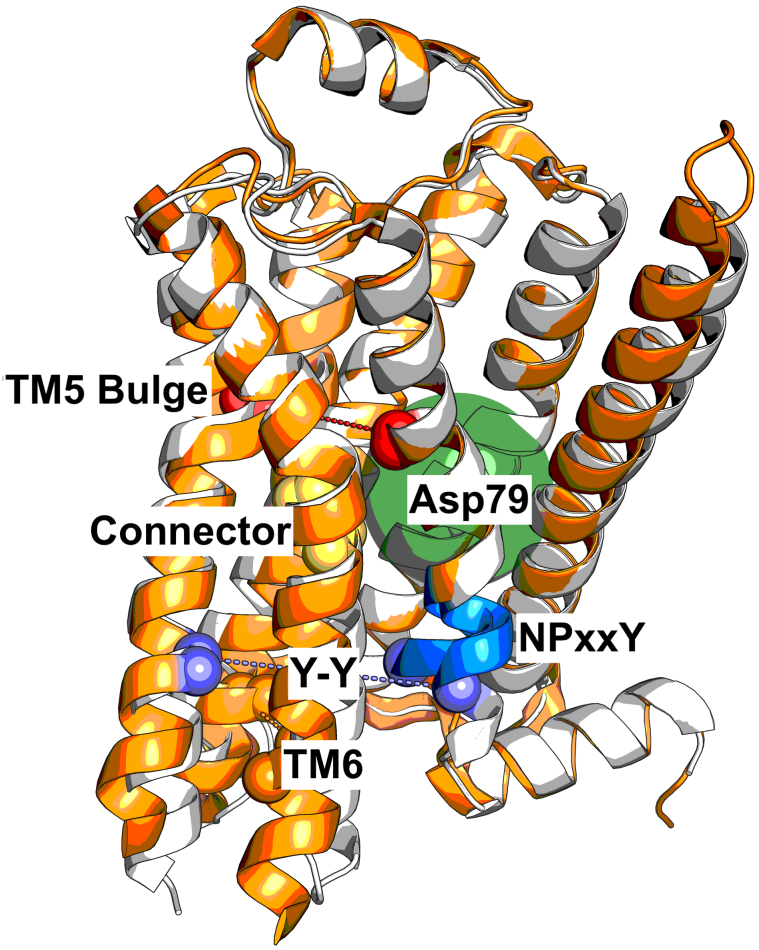


Figure 1: The activation mechanism of GPCRs involves a series of microswitches. Agonist binding leads to an inward bulge of TM5 (quantified as the distance between Ser207$^{5.46}$ and Gly315$^{7.41}$), which leads to a conformational change in the connector region (Ile121$^{3.40}$ and Phe282$^{6.44}$). The transmembrane cavity surrounding Asp79$^{2.50}$ dehydrates and the NPxxY motif twists upon activation, leading the Y-Y interaction to form. TM6 moves outwards to create the G protein binding site. The inactive (PDB ID 2RH1[5]) and active (PDB ID 3P0G[7]) structures are represented in white and orange, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

seminal paper by Dror et al[27] 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 active to inactive 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, on the other hand, provide a means to explore the conformational landscapes of proteins to a relatively low computational cost[31] and have been exploited to study the activation mechanism of GPCRs[22–24,26,28–30].

In this work, we aimed to identify the most probable path describing the transition between inactive and active states of β_2AR and to characterize the full conformational ensemble along this pathway at an atomistic level of resolution. A new version of the string method with swarms of trajectories was first developed 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 ranges from weak to strong and is influenced by the protonation of Asp79$^{2.50}$. Finally, we demonstrated 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[32]. In this framework, the minimum free energy path

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in an 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 one-dimensional 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 CV set that embeds receptor activation by automatically inferring the inter-residue distances that are related to activation. Based on one of Dror et al.'s simulation trajectories of spontaneous deactivation of the β_2AR[27], we identified metastable states by clustering simulation configurations, followed by classification of these by training a fully connected neural network to identify states[33]. The most important input features for classification were identified via deep Taylor decomposition[34] 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). Encouragingly, this five-dimensional CV set captured the main degrees of freedom implicated in β_2AR activation, including the large outward movement of TM6 and a smaller inward shift of TM7 at the intracellular face of the receptor.

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 Dror et al.'s 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

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. Because we sought to only sample the vicinity of the most probably activation path, derivation of a converged free energy landscape required a mere total of ∼2-3 μs simulation time (Fig. S5, S6, S7, S8 and S9).

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 (SI methods and Fig. S10) and to derive the associated free energy landscape describing the outward movement of TM6 and inward shift of TM7 upon activation (Fig. 3).

For the apo receptor, one distinguishes three minima: one in the active region, one in the inactive one, and an intermediate one between these (Fig. 3a). 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 conformation of lower free energy (Fig. 3b). A number of characteristic variables (defined in Table S2) were calculated for the last iteration of the swarm of trajectories simulations. By localizing sudden shifts in the values of these parameters, we could pinpoint the location of important events in activation on the string (Fig. 3). In the absence of bound agonist, the connector assumes an active conformation in the early stages of activation. The Asp79$^{2.50}$ cavity then dehydrates, followed by the formation of the Y-Y interaction and NPxxY twist. Finally the TM5 bulge forms before the fully active conformation is reached, confirming an allosteric communication between the orthosteric site and intracellular region.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

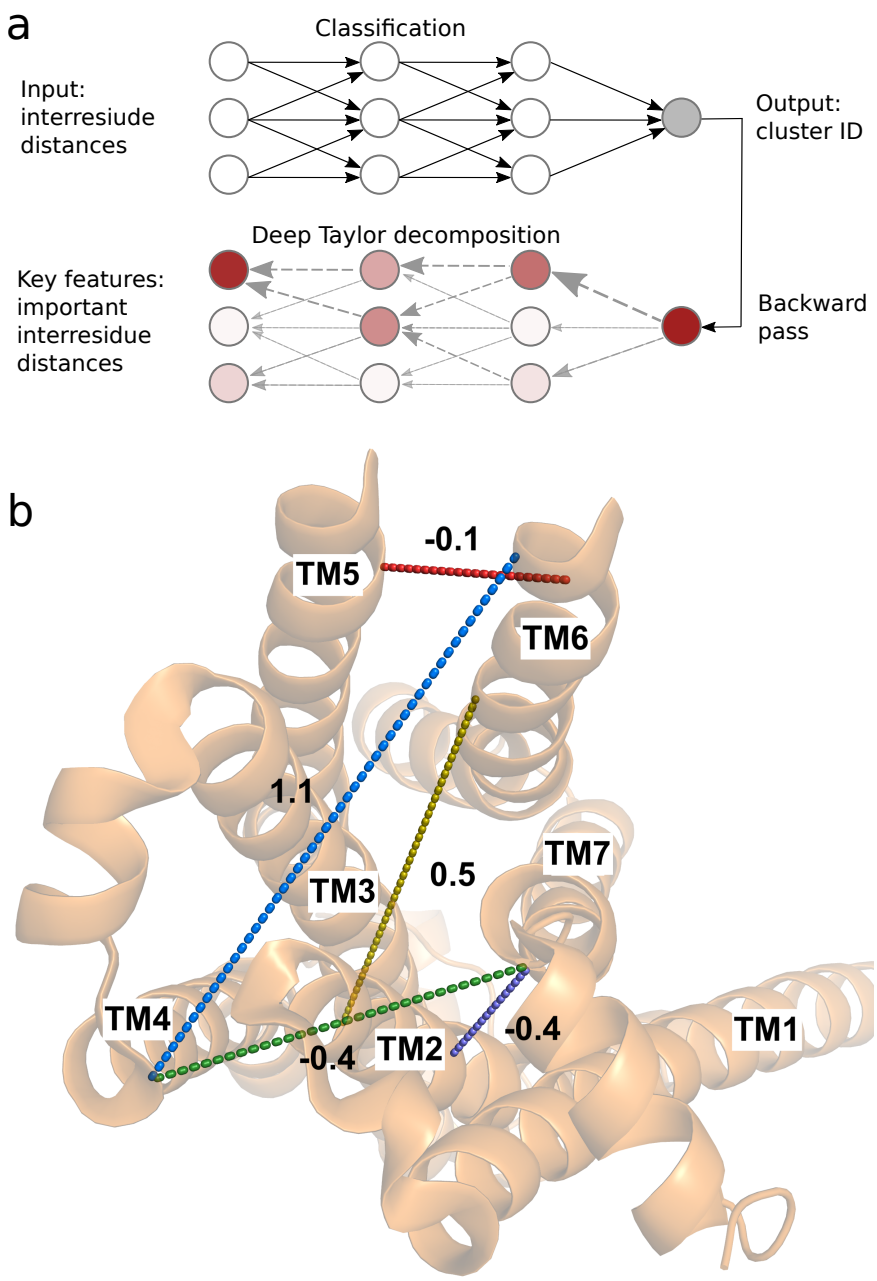


Figure 2: The five-dimensional collective variable (CV) space used to optimize the minimum free energy path was identified in a data-driven manner. (**a**) A fully connected neural network was trained to classify configurations in active, intermediate, and inactive metastable states (clusters). Deep Taylor decomposition was then used to identify the most important input inter-residue distances for the classification decision. The top ranked distances 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.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

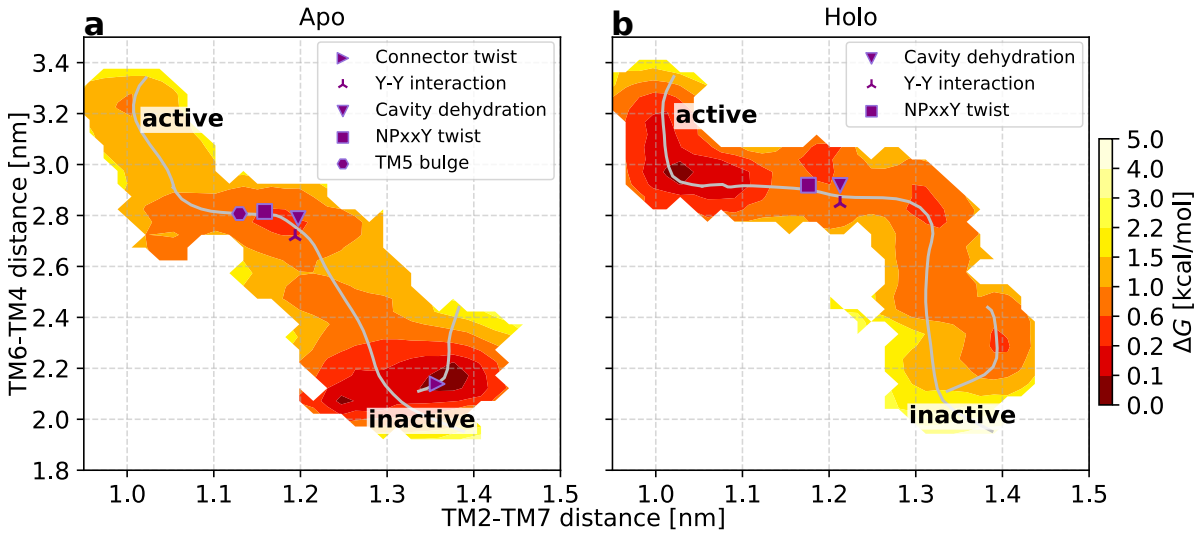


Figure 3: The free energy landscapes are projected onto the first two CVs used to optimize the minimum free energy path. The minimum free energy pathways were optimized in the absence (**a**) and in the presence of a bound agonist (**b**). Characteristic events occurring during activation (see Table S2 for their definition) are located on the string. The active and inactive labels mark the regions close to the active and inactive structures 3P0G[7] and 2RH1[5].

Binding of the agonist ligand modifies the order in which the helices rearrange, as can be seen when projecting the minimum free energy path along various CVs (Fig. 3 and S5). The most probable activation pathway begins with an outward movement of TM6, followed by the inward shift of TM7. The connector and the bulge in TM5 are locked in their active-like states in the presence of agonist, which enables large outward movements of TM6. Similarly to the apo case, to reach the fully active conformation, the Asp79$^{2.50}$ cavity dehydrates and the Y-Y interaction forms shortly before the NPxxY twists.

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.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The roles of conserved microswitches were assessed by comparing free energy landscapes in the absence and presence of a bound agonist. We first evaluated how the TM5 bulge, 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 the TM5 bulge in the orthosteric site (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 simulations. The 0.5 nm outward movement of TM6 at the minimum of the landscape (Fig. 4a,b) 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[19,35].

Propagation of activation through microswitches

Structural changes in the orthosteric site of the β_2AR (Fig. 4s) have been proposed to propagate towards the intracellular part via a "connector" centered around Ile121$^{3.40}$ and Phe282$^{6.44}$[7] (Fig. 4t). Whereas we found the TM5 bulge and outward movement of TM6 to be loosely coupled, the free energy landscapes demonstrated that changes in the orthosteric site has a direct influence on the connector region (Fig. 4d,e). In the absence of agonist, both inactive and active conformations of the connector were populated and the connector region could assume an active-like state even if the TM5 bulge was inactive (Fig. 4d). In contrast, agonist binding resulted in a single free energy minimum where both the TM5 bulge and the connector were constrained to their active conformations (Fig. 4e).

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

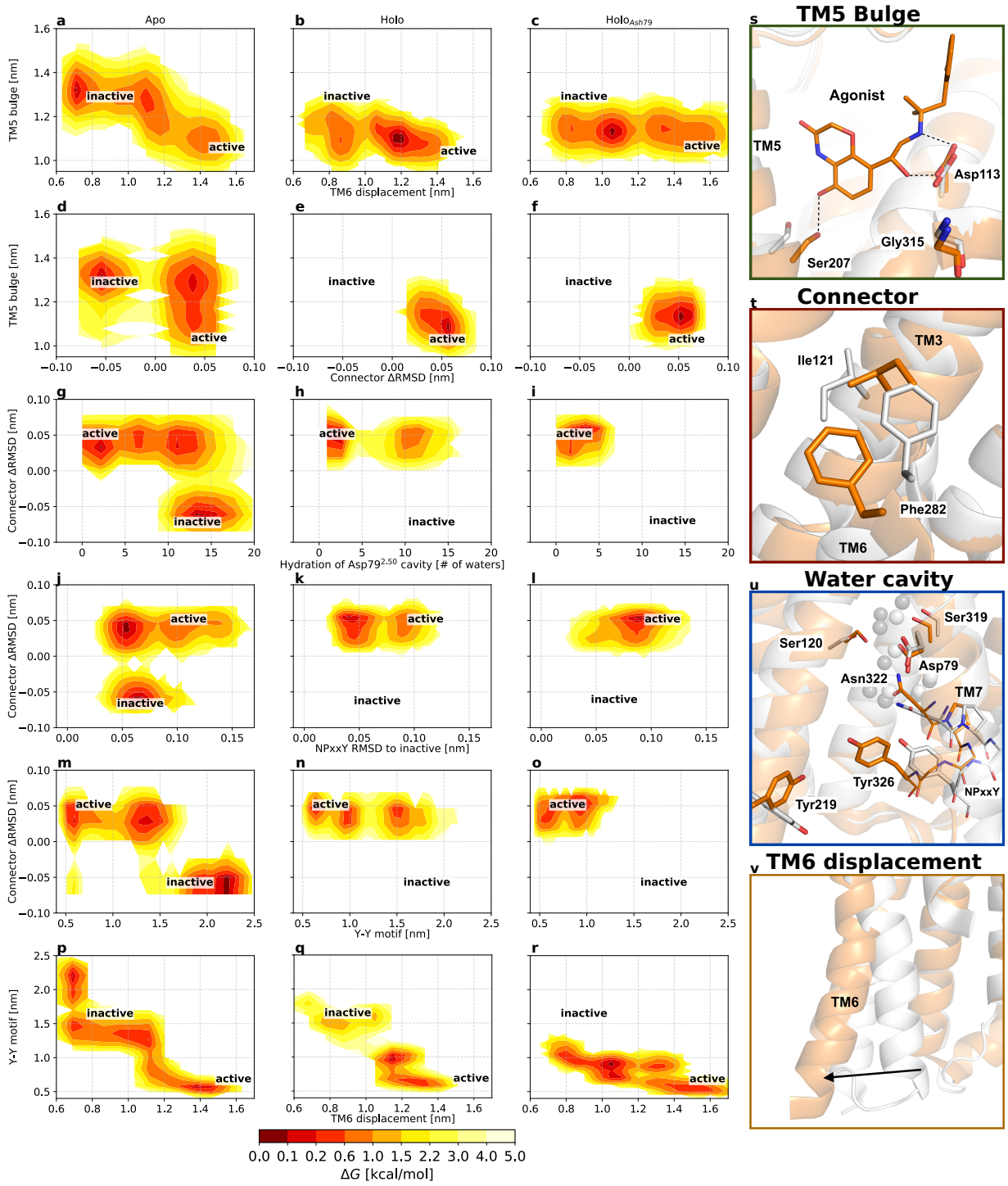


Figure 4: Free energy landscapes projected along variables of interest highlight changes in the pairwise coupling of microswitches following binding of an agonist ligand (middle column) and protonation of conserved residue Asp79$^{2.50}$ (right column). The free energy landscapes are projected along (**a-c**) The TM5 bulge (distance between Ser207$^{5.46}$ and Gly315$^{7.41}$, representing the ligand binding site contraction) and the distance between Leu272$^{6.34}$ and Arg131$^{3.50}$, representing the outward movement of TM6; (**d-f**) The TM5 bulge (distance between Ser207$^{5.46}$ and Gly315$^{7.41}$) and difference between the RMSD of Ile121$^{3.40}$ and Phe282$^{6.44}$ heavy atoms to the active and inactive crystal structures[5,7], representing the connector region ΔRMSD; (**g-i**) the connector region ΔRMSD and the number of water molecules within 0.8 nm from Asp79$^{2.50}$, representing the hydration of the Asp79$^{2.50}$ cavity; (**j-l**) the connector region ΔRMSD and the RMSD of the NPxxY motif relative to the inactive structure 2RH1; (**m-o**) the connector region ΔRMSD and the distance between the two tyrosines implicated in the Y-Y interaction (Y-Y motif). (**p-r**) the Y-Y motif distance 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 S1 for microswitch definitions. (**s-v**) Vignettes showing the conformation of the different microswitches in the active and inactive structures 3P0G (orange) and 2RH1 (white): (**s**) the TM5 bulge shifting Ser207$^{5.46}$ inward, (**t**) the connector region containing Ile121$^{3.40}$ and Phe282$^{6.44}$, (**u**) the Asp79$^{2.50}$ cavity, NPxxY motif, and the two tyrosines Tyr219$^{5.58}$-Tyr326$^{7.53}$ of the Y-Y motif, and (**v**) the outward movement of TM6 upon activation.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

From the connector region, activation is propagated via several conserved motifs (Fig. 4g,h,j,k,m,n and u)[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 cavity surrounding Asp79$^{2.50}$, to the conformation of the NPxxY motif, and to the conformation of the Y-Y motif. In the apo state, an inactive connector region was tightly coupled to a hydrated cavity with $\sim 10-17$ waters (Fig. 4g) and an inactive NPxxY motif (Fig. 4j). 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[36]. In contrast, the active conformation of the connector was compatible with both the fully hydrated cavity and a desolvated state with only 4 to 5 water molecules as well as an inactive and active NPxxY motif. Interestingly, the connector configuration was more tightly coupled to the Y-Y motif. An inactive connector was coupled to a broken Y-Y interaction whereas it was in intermediate state or fully formed with an active connector (Fig. 4m).

The free energy landscapes suggested that binding of agonist reduce the number of inactive-like states of the Asp79$^{2.50}$ cavity, NPxxY, Y-Y motifs (Fig. 4h,k,n) as well as the orientation of the Met82$^{2.53}$ (Fig. S9), a residue located one helix turn above Asp79$^{2.50}$ that has been the focus of NMR studies[37]. For the Asp79$^{2.50}$ cavity and NPxxY motif, both active and inactive conformations were accessible in the presence of agonist, but the active conformations were more favored energetically (Fig. S11). The presence of agonist resulted in an energy landscape with one active-like and two intermediate distances of the Y-Y motif. The two latter minima were shifted by approximately 0.1 nm towards the active conformation compared to the apo state (Fig. 4n).

The final combination of microswitches connected the Y-Y motif to the motion of TM6 (Fig. 4p,q). In the apo state, there was a relatively tight coupling between the Y-Y interaction and the TM3-TM6 distance, with several metastable states lining the minimum free energy pathway between inactive and active conformations (Fig. 4p). Ligand binding did not alter the coupling between these two microswitches, but generally tilted the free energy

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

landscape towards more active states along both of these dimensions (Fig. 4q).

Having computed the free energy landscapes in absence and presence of ligand thus enables us to characterize the involvement of each microswitch in the transmission of allosteric coupling between the orthosteric ligand and the intracellular partner binding sites.

Impact of Asp$79^{2.50}$ protonation

Protonation of two ionizable residues, Asp$79^{2.50}$ and Asp$130^{3.49}$, has been proposed to be involved in GPCR activation[27,38,39]. In particular, MD simulations have indicated that Asp$79^{2.50}$, 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[38,40]. As a previous simulation study of the β_2AR showed that Asp$79^{2.50}$ (but not Asp$130^{3.49}$) may alter the activation pathway[27], we repeated the calculations of the minimum free energy pathway of activation with this residue in its protonated (neutral) form (Fig. S7).

The free energy landscapes describing changes in the orthosteric site were similar to those obtained in simulations with Asp$79^{2.50}$ ionized (Fig. 4c,f). There was a weak coupling between the TM5 bulge and the intracellular region, with two major energy wells describing the conformation of TM6. Compared to the agonist-bound receptor with Asp$79^{2.50}$ ionized, the minima were shifted further towards active-like conformations for the protonated state (Fig. 4c). The TM5 bulge remained strongly coupled to conformational changes observed in the connector region irrespective of the ionization state of Asp$79^{2.50}$ (Fig. 4e,f). The largest effects of Asp$79^{2.50}$ protonation were observed for the hydrated cavity surrounding this residue (Fig. 4h,i), NPxxY (Fig. 4k,l), and Y-Y motif (Fig. 4n,o): whereas the free energy landscapes showed that both active- and inactive-like conformations of these switches were populated in simulations with ionized Asp$79^{2.50}$, the protonated state only resulted in energy wells close to the active conformation (Fig. 4i,l,o). It was also evident that TM6 was stabilized in more active-like conformations by the protonated Asp$79^{2.50}$ (Fig. 4r).

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Interestingly, intermediate states in which the Y-Y motif, as well as the NPxxY, adopted an active-like conformation and a low TM6 displacement were also populated (Fig. 4r and S11). Such an intermediate state has been observed, albeit rarely, in simulations considering a protonated Asp79$^{2.50}$[27].

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. 5). 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 active NPxxY motif appeared for the protonated Asp79$^{2.50}$, which was not favored energetically in the simulations of the ionized form (Fig. S8k,l and S11) 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. 5)[41,42]. Our protocol thus allowed us to sample metastable states not yet resolved by experimental structure determination that may play a role in activation.

Transferability of the methodology to other GPCRs

Efficient characterization of free energy landscapes with the string method relies on selection of appropriate CVs, which is 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. 6). Strikingly, the active and inactive structures clearly separate in two distinct clusters. This indicates that these CVs can describe the activation of many class A GPCRs and the protocol presented herein can be applied to other family members.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

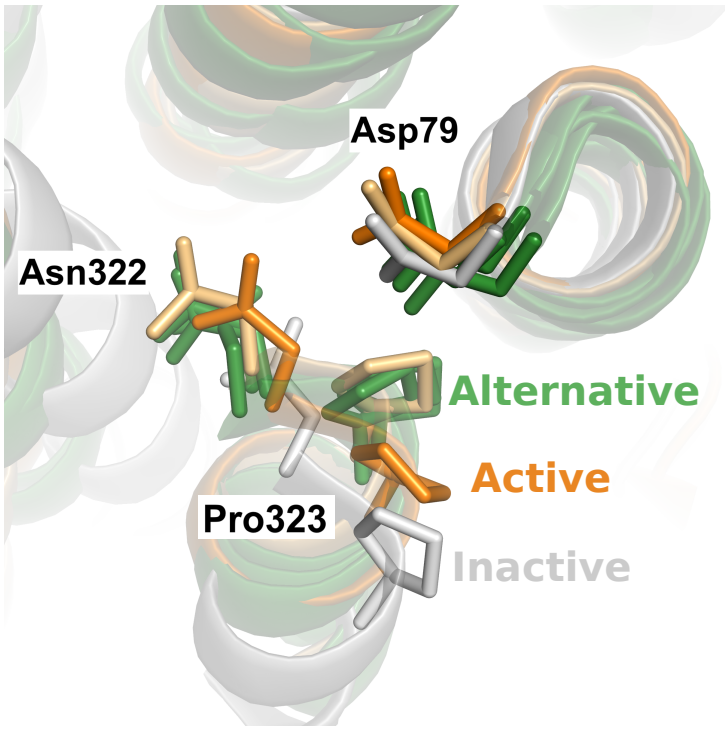


Figure 5: An alternative conformation of the NPxxY was identified along the most probable pathway calculated with a protonated Asp79$^{2.50}$. Representative simulation snapshots of active-like states of the β_2AR with Asp79$^{2.50}$ ionized and protonated are shown in orange and light-orange, respectively. Representative simulation snapshots of inactive-like state of the β_2ARis shown in white. The structural comparison highlights the resemblance between an alternative conformation of the NPxxY motif (light orange) that is favored by Asp79$^{2.50}$ protonation in the simulations and observed in three crystal structures of other class A GPCRs. Structures of the other GPCRs are shown in green (PDB IDs: 6DRX[41] - Agonist-bound 5-HT$_{2B}$, 3QAK[43] - Agonist-bound A$_{2A}$AR, and 6DO1[44] - Angiotensin II type 1 receptor in an active conformation).

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

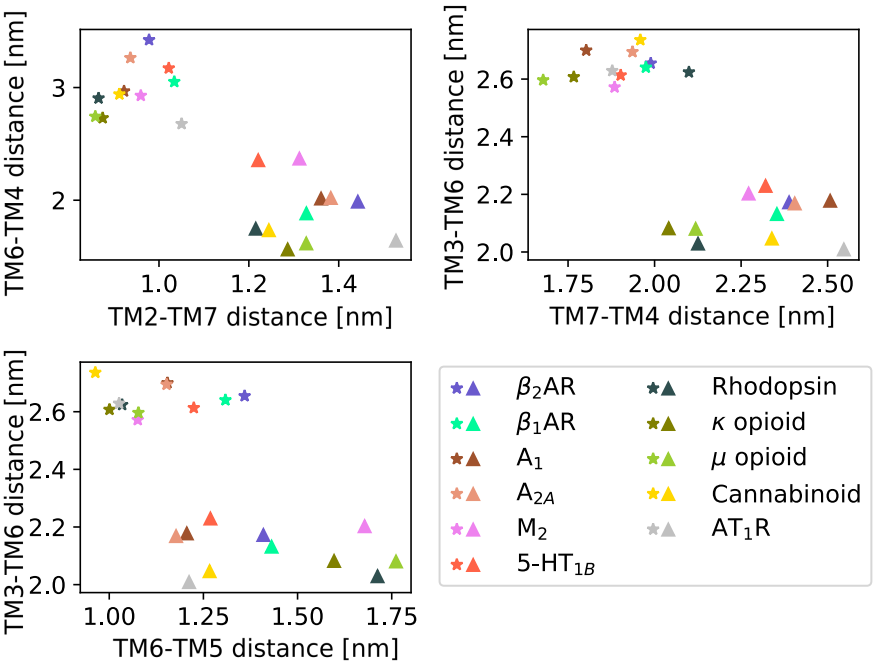


Figure 6: Active and inactive class A GPCR structures cluster into two distinct groups in the 5-dimensional CV space used in this work. β_2AR and ten active (stars) and inactive (triangles) structures of other class A GPCRs are projected onto the five original CVs. Details of the mapping can be found 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.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

Experimental structures of the β_2AR in inactive and active conformations provide a basis for molecular understanding of GPCR signaling. However, it has become increasingly clear that the static structures do not capture all functionally relevant states involved in activation of these molecular machines[22,23,28,29]. In a pioneering study, Dror et al. gained insights into the activation pathway of the β_2AR from long-timescale MD simulations[27]. 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 on methodologies suggested in previous molecular simulation studies[23,25,26,29] to further allow to assess the impact of agonist binding on microswitches involved in activation.

Our work recapitulates a number of key findings of previous enhanced sampling[23,25,26,28,29] and long-timescale scale simulations of GPCRs[27]. In agreement with previous work, our simulations revealed a complex conformational landscape of the β_2AR with only weak coupling between the othosteric site and G protein binding site. Through the analysis of the free energy landscapes we could quantify the coupling between spatially connected microswitches in the transmembrane region, which ranged from strong to relatively weak and was influenced by protonation of Asp79$^{2.50}$. Our study also further illustrates that the energy landscapes depend on the variables chosen to project the conformational states. This is not an artifact of the protocol but rather an inherent limitation of dimensionality reduction. In particular, it is clear that considering only three states along the activation path (an active, intermediate and inactive state) does not allow to capture the complexity of the conformational changes induced by ligand binding[22,23,28,29]. The present work, through the resolution of the complete conformational ensemble lining the most probable activation pathway, can aid in determining the placement of spectroscopic probes to monitor the distribution of specific microswitches under different conditions.

Our work shows agreement with spectroscopy studies although a quantitative compar-

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ison remains challenging: 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[45], NMR spectroscopy experiments have demonstrated that agonist binding stabilizes other conformations in certain parts of the TM region, e.g. close to Met215$^{5.54}$ and Met82$^{2.53}$[37]. For example, the region surrounding Met82$^{2.53}$, located one helical turn above Asp79$^{2.50}$, was shown to adopt three conformations in the prescence of a neutral antagonist (with similar signalling properties as the apo receptor). A bound agonist, on the other hand, restricted this region to a single active-like state[35,37]. Similar to these experiments, our free energy landscapes demonstrate that several conformations of the connector, Asp79$^{2.50}$ cavity and the orientation of Met82$^{2.53}$ are available in the apo condition (Fig. S9d-f and S8g-i,m-o). Especially, Fig. S9d-f shows three conformational Met82$^{2.53}$ states of comparable free energy for the apo receptor, whereas the landscape is significantly shifted to favour a single active state conformation for the agonist bound receptor. Protonation of Asp79$^{2.50}$ further increases this shift. 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.

The rapidly increasing number of crystal structures provides insights into how different GPCRs, despite recognizing disparate ligands and sharing low sequence identity in the orthosteric site, can result in very similar conformational changes. The bulge of TM5, which we found to be strongly coupled to the connector region, is is also present in active-state structures of other GPCRs[46,47]. Whereas this conformational change is the result of a direct interaction with TM5 in the case of the β_2AR, agonist interactions with other helices appear to indirectly influence the same region in the case of the μ opioid and A_{2A} adenosine receptor[46,47]. In other cases, the connector region was found to have a similar arrangement

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in the active and inactive state (e.g. the M_2 muscarinic receptor)[48]. In such cases, receptor activation may be controlled by direct modulation of the Asp79$^{2.50}$ cavity, e.g. the via the conserved Trp286$^{6.48}$. 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[21,49,50]. 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[38,40]. In turn, the protonated Asp79$^{2.50}$ stabilizes a structure of the NPxxY 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[51]. 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[9]. 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 than plain MD simulations[27,28], and comparable to other enhanced sampling methods such as GaMD[22]. 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

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 1 or 2 dimensions, as is done in most CV-based methods such as umbrella sampling or metadynamics[52].

On the other hand, a well-known limitation with the string with swarms 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. 6) 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[53].

Despite the major progress in structural biology for GPCRs, many aspects of receptor function remain misunderstood. 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, poten-

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

tially making it possible to design ligands with biased signalling properties[21]. The method is equally well suited to study the effect of allosteric modulators and the influence of different lipidic environments[54]. 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.

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[45]. 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[55] and the TIP3P water model[56]. The protein was inserted in a POPC[57] bilayer and solvated in explicit solvent. Na^+ and Cl^- ions were added at 0.15M concentration. System preparation was performed using CHARMM-GUI[58]. MD simulations were run with GROMACS 2016.5[59] patched with plumed 2.4.1[60] under a 1 atm pressure and at 310.15 K temperature.

To identify CVs, we performed clustering[61] on the frames from unrestrained MD simulation trajectory of β_2AR (condition A in Dror et al.). The CVs were selected by training a multilayer perceptron classifier[62] using as input all the inter-residue distances and as output the cluster ID, followed by using Deep Taylor decomposition[34] 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[5]

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and 3P0G[7]. The initial path for simulations Holo_t was guessed using data from Dror et al.: a rough estimate of the free energy landscape was calculated from the probability density landscape estimated using the Scikit-learn[62] 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 $\text{Holo}_{Ash79,t}$. All active (Apo_a, Holo_a and $\text{Holo}_{Ash79,a}$) and inactive (Apo_i, Holo_i and $\text{Holo}_{Ash79,i}$) substrings were initiated as straight paths between the endpoints. Even though the system was initiated from the nanobody bound structure 3P0G, the active state substring sampled conformations closer to the G protein bound structure 3SN6 (Fig. S12)[6]. 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 S7) and posterior distribution of free energy profiles given the data was narrow (Fig. S6).

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[63] along a given variable (Fig. S10). 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: $\rho_i T_{ij} = \rho_j T_{ji}$. A Metropolis Markov chain Monte Carlo (MCMC) method was used to sample over the posterior distribution of transition matrices, given the unregularized elements of T_{ij}[64], and thereby obtain a distribution of free energy profiles for ρ (Fig. S6, S8 and S9). All code to run the simulations and reproduce the results in this paper is available for download[65].

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgement

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.

Supporting Information Available

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Fredriksson, R.; Lagerström, M. C.; Lundin, L.-G.; Schiöth, H. B. *Mol. Pharmacol.* **2003**, *63*, 1256–1272.

(2) Hauser, A. S.; Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B.; Gloriam, D. E. *Nat Rev Drug Discov* **2017**, *16*, 829.

(3) Weis, W. I.; Kobilka, B. K. *Annu. Rev. Biochem.* **2018**, *87*, 897–919.

(4) Manglik, A.; Kruse, A. C. *Biochemistry* **2017**, *56*, 5628–5634.

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

(6) Rasmussen, S. G. F. et al. *Nature* **2011**, *477*, 549–555.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(7) Rasmussen, S. G.; Choi, H.-J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; DeVree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S. *Nature* **2011**, *469*, 175.

(8) Masureel, M. et al. *Nature Chemical Biology* **2018**, *14*, 1059–1066.

(9) Staus, D. P. et al. *Nature* **2016**, *535*, 448–452.

(10) Liu, X.; Ahn, S.; Kahsai, A. W.; Meng, K.-C.; Latorraca, N. R.; Pani, B.; Venkatakrishnan, A. J.; Masoudi, A.; Weis, W. I.; Dror, R. O.; Chen, X.; Lefkowitz, R. J.; Kobilka, B. K. *Nature* **2017**, *548*, 480–484.

(11) Ring, A. M.; Manglik, A.; Kruse, A. C.; Enos, M. D.; Weis, W. I.; Garcia, K. C.; Kobilka, B. K. *Nature* **2013**, *502*, 575–579.

(12) Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V.-P.; Chien, E. Y.; Velasquez, J.; Kuhn, P.; Stevens, R. C. *Structure* **2008**, *16*, 897–905.

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

(14) Rasmussen, S. G. F.; Choi, H.-J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S.; Edwards, P. C.; Burghammer, M.; Ratnala, V. R. P.; Sanishvili, R.; Fischetti, R. F.; Schertler, G. F. X.; Weis, W. I.; Kobilka, B. K. *Nature* **2007**, *450*, 383–387.

(15) García-Nafría, J.; Tate, C. G. *Mol. Cell. Endocrinol.* **2019**, *488*, 1–13.

(16) Ballesteros, J. A.; Weinstein, H. *Methods in Neurosciences*; Elsevier, 1995; pp 366–428.

(17) Latorraca, N. R.; Venkatakrishnan, A.; Dror, R. O. *Chem. Rev.* **2016**, *117*, 139–155.

(18) Lamichhane, R.; Liu, J. J.; Pljevaljcic, G.; White, K. L.; van der Schans, E.; Katritch, V.; Stevens, R. C.; Wüthrich, K.; Millar, D. P. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 14254–14259.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(19) Manglik, A.; Kim, T. H.; Masureel, M.; Altenbach, C.; Yang, Z.; Hilger, D.; Lerch, M. T.; Kobilka, T. S.; Thian, F. S.; Hubbell, W. L.; Prosser, R. S.; Kobilka, B. K. *Cell* **2015**, *161*, 1101–1111.

(20) Gregorio, G. G.; Masureel, M.; Hilger, D.; Terry, D. S.; Juette, M.; Zhao, H.; Zhou, Z.; Perez-Aguilar, J. M.; Hauge, M.; Mathiasen, S. *Nature* **2017**, *547*, 68.

(21) Picard, L.-P.; Schonegge, A.-M.; Bouvier, M. *ACS Pharmacol Transl Sci* **2019**,

(22) Miao, Y.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 12162–12167.

(23) Bhattacharya, S.; Vaidehi, N. *J. Am. Chem. Soc.* **2010**, *132*, 5205–5214.

(24) Niesen, M. J. M.; Bhattacharya, S.; Vaidehi, N. *J. Am. Chem. Soc.* **2011**, *133*, 13197–13204.

(25) Shan, J.; Khelashvili, G.; Mondal, S.; Mehler, E. L.; Weinstein, H. *PLoS Comput. Biol.* **2012**, *8*, e1002473.

(26) Li, J.; Jonsson, A. L.; Beuming, T.; Shelley, J. C.; Voth, G. A. *J. Am. Chem. Soc.* **2013**, *135*, 8749–8759.

(27) Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 18684–18689.

(28) Kohlhoff, K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. *Nat. Chem.* **2014**, *6*, 15.

(29) Provasi, D.; Artacho, M. C.; Negri, A.; Mobarec, J. C.; Filizola, M. *PLoS Comput. Biol.* **2011**, *7*, e1002193.

(30) Tikhonova, I. G.; Selvam, B.; Ivetac, A.; Wereszczynski, J.; McCammon, J. A. *Biochemistry* **2013**, *52*, 5593–5603.

(31) Harpole, T. J.; Delemotte, L. *Biochim Biophys Acta Biomembr* **2018**, *1860*, 909–926.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(32) Pan, A. C.; Sezer, D.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3432–3440.

(33) Fleetwood, O.; Kasimova, M. A.; Westerlund, A. M.; Delemotte, L. **2019**,

(34) Montavon, G.; Lapuschkin, S.; Binder, A.; Samek, W.; Müller, K.-R. *Pattern Recognit.* **2017**, *65*, 211–222.

(35) Nygaard, R. et al. *Cell* **2013**, *152*, 532–542.

(36) Miller-Gallacher, J. L.; Nehme, R.; Warne, T.; Edwards, P. C.; Schertler, G. F.; Leslie, A. G.; Tate, C. G. *PLoS One* **2014**, *9*, e92727.

(37) Kofuku, Y.; Ueda, T.; Okude, J.; Shiraishi, Y.; Kondo, K.; Maeda, M.; Tsujishita, H.; Shimada, I. *Nature Communications* **2012**, *3*.

(38) Ranganathan, A.; Dror, R. O.; Carlsson, J. *Biochemistry* **2014**, *53*, 7283–7296.

(39) Arnis, S.; Fahmy, K.; Hofmann, K. P.; Sakmar, T. P. *Journal of Biological Chemistry* **1994**, *269*, 23879–23881.

(40) Vickery, O. N.; Carvalheda, C. A.; Zaidi, S. A.; Pisliakov, A. V.; Katritch, V.; Zachariae, U. *Structure* **2018**, *26*, 171–180.e2.

(41) McCorvy, J. D.; Wacker, D.; Wang, S.; Agegnehu, B.; Liu, J.; Lansu, K.; Tribo, A. R.; Olsen, R. H. J.; Che, T.; Jin, J.; Roth, B. L. *Nat. Struct. Mol. Biol.* **2018**, *25*, 787–796.

(42) Wingler, L. M.; McMahon, C.; Staus, D. P.; Lefkowitz, R. J.; Kruse, A. C. *Cell* **2019**, *176*, 479–490.e12.

(43) Xu, F.; Wu, H.; Katritch, V.; Han, G. W.; Jacobson, K. A.; Gao, Z.-G.; Cherezov, V.; Stevens, R. C. *Science* **2011**, *332*, 322–327.

(44) Wingler, L. M.; McMahon, C.; Staus, D. P.; Lefkowitz, R. J.; Kruse, A. C. *Cell* **2019**, *176*, 479–490.e12.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(45) Rosenbaum, D. M. et al. *Nature* **2011**, *469*, 236–240.

(46) Huang, W. et al. *Nature* **2015**, *524*, 315–321.

(47) Lebon, G.; Warne, T.; Edwards, P. C.; Bennett, K.; Langmead, C. J.; Leslie, A. G. W.; Tate, C. G. *Nature* **2011**, *474*, 521–525.

(48) Kruse, A. C. et al. *Nature* **2013**, *504*, 101–106.

(49) Fenalti, G.; Giguere, P. M.; Katritch, V.; Huang, X.-P.; Thompson, A. A.; Cherezov, V.; Roth, B. L.; Stevens, R. C. *Nature* **2014**, *506*, 191–196.

(50) Eddy, M. T.; Lee, M.-Y.; Gao, Z.-G.; White, K. L.; Didenko, T.; Horst, R.; Audet, M.; Stanczak, P.; McClary, K. M.; Han, G. W.; Jacobson, K. A.; Stevens, R. C.; Wüthrich, K. *Cell* **2018**, *172*, 68–80.e12.

(51) Liu, J. J.; Horst, R.; Katritch, V.; Stevens, R. C.; Wuthrich, K. *Science* **2012**, *335*, 1106–1110.

(52) Pietrucci, F. *Reviews in Physics* **2017**, *2*, 32–45.

(53) Pan, A. C.; Roux, B. *J. Chem. Phys.* **2008**, *129*, 064107.

(54) Song, W.; Yen, H.-Y.; Robinson, C. V.; Sansom, M. S. *Structure* **2019**, *27*, 392–403.e3.

(55) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. *Nat. Methods* **2016**, *14*, 71.

(56) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(57) Klauda, J. B.; Venable, R. M.; Freites, J. A.; OConnor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell Jr, A. D.; Pastor, R. W. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(58) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y. *J. Chem. Theory Comput.* **2015**, *12*, 405–413.

(59) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1*, 19–25.

(60) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

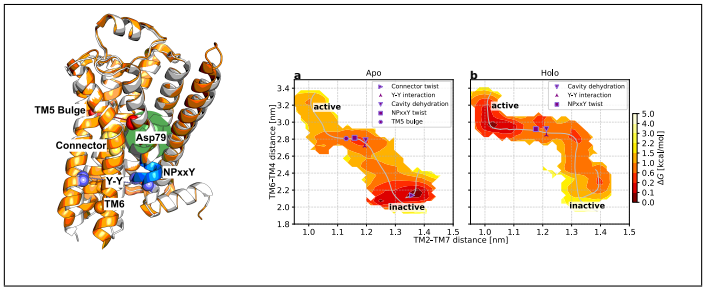
(61) Westerlund, A. M.; Delemotte, L. *PLoS Comput. Biol.* **2018**, *14*, e1006072.

(62) Pedregosa, F. et al. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(63) Lev, B.; Murail, S.; Poitevin, F.; Cromer, B. A.; Baaden, M.; Delarue, M.; Allen, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E4158–E4167.

(64) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. *J. Chem. Theory Comput.* **2011**, *7*, 3412–3419.

(65) Fleetwood, O. Code to study GPCRs using the String Method with swarms of trajectories. `https://github.com/delemottelab/gpcr-string-method-2019`, 2019; `https://doi.org/10.5281/zenodo.2662201`.

bioRxiv preprint doi: https://doi.org/10.1101/627026; this version posted September 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Graphical TOC Entry