Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Cloud-based simulations on Google Exacycle reveal ligand-modulation of GPCR activation pathways

Kai J. Kohlhoff[#1,4,*], **Diwakar Shukla**[#1,2], **Morgan Lawrenz**[#2], **Gregory R. Bowman**[2], **David E. Konerding**[4], **Dan Belov**[4], **Russ B. Altman**[1,3,*], and **Vijay S. Pande**[2,*]

[1]Department of Bioengineering, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA.

[2]Department of Chemistry, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA.

[3]Department of Genetics, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA.

[4]Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA.

[#] These authors contributed equally to this work.

Abstract

Simulations can provide tremendous insight into atomistic details of biological mechanisms, but micro- to milliseconds timescales are historically only accessible on dedicated supercomputers. We demonstrate that cloud computing is a viable alternative, bringing long-timescale processes within reach of a broader community. We used Google's Exacycle cloud computing platform to simulate 2 milliseconds of dynamics of the β2 adrenergic receptor — a major drug target G protein-coupled receptor (GPCR). Markov state models aggregate independent simulations into a single statistical model that is validated by previous computational and experimental results. Moreover, our models provide an atomistic description of the activation of a GPCR, revealing multiple activation pathways. Agonists and inverse agonists interact differently with these pathways, with profound implications for drug design

G protein-coupled receptors (GPCRs) are a family of membrane-bound alpha-helical proteins that regulate a large variety of physiological processes by transmitting signals from extracellular binding of diverse ligands to intracellular signaling molecules. These proteins are exceedingly prominent drug targets, responsible for at least one third of all marketable drugs and half of the total market volume for pharmaceuticals (1). The β_2-adrenergic

*Correspondence to: kohlhoff@google.com, russ.altman@stanford.edu, pande@stanford.edu..

Using Google's Exacycle cloud computing platform an unprecedented 2 milliseconds of molecular dynamics of a major drug target G protein-coupled receptor was simulated. Markov state models aggregate independent simulations into a statistical model that provides the first atomistic description of GPCR ligand-modulated activation pathways, with profound implications for drug design.

receptor (β_2AR) is implicated in type-2 diabetes, obesity, and asthma, and is a member of the class A, rhodopsin-like GPCRs. These proteins share a highly conserved motif of seven transmembrane helices connected by three extracellular and three intracellular loops. β_2AR is experimentally well studied, and high-resolution X-ray structures of both the inactive (1) and several active states (2, 3) have been determined in recent years. However, despite this rapid progress toward understanding of these important molecules, little is known about the mechanisms by which small molecules modulate their activity.

Molecular dynamics (MD) simulations have already begun to provide insights into the underlying dynamics and structural ensembles of GPCRs (4–8). However, many phenomena of interest still remain out of reach. For example, one recent study used special-purpose hardware (9) to reach an unprecedented total simulation time of several hundred μs (5). These results provided insights into the mechanism of deactivation but were unable to capture activation. Moreover, it remains unclear how to make further advances, particularly for researchers without access to such specialized hardware.

To capture the mechanism of β_2AR activation, we have followed an alternative approach to MD in which we extended the principles behind the volunteer distributed computing platform Folding@home (10) to cloud computing more broadly. Specifically, we ran tens of thousands of independent simulations on Google Exacycle(11), a cloud computing initiative that provides an interface for running compute jobs directly on Google's production infrastructure. Markov state models (MSMs) were then used to stitch together these massively parallel simulations into a single statistical model, capturing rare events on timescales far longer than those reached by any individual simulation (12–14). Our approach reproduces a variety of previous experimental and computational results, including mutual information networks of correlated residues, and we explain how key structural elements change along ligand-modulated activation pathways. Moreover, we show that the MSMs can improve our understanding of drug efficacy at GPCR receptors and can be incorporated into an effective structure-based drug design approach.

Results

Using our cloud-based approach, we simulated 2.15 milliseconds of β_2AR dynamics. Simulations were initiated from both an inactive (PDB 2RH1) (1) and active (PDB 3P0G) (2) crystal structure of β_2AR. We also ran simulations in the presence of two ligands—the partial inverse agonist carazolol and the full agonist BI-167107—to understand how these small molecules alter the behavior of β_2AR. We find that activation and deactivation proceed through multiple pathways and typically visit metastable intermediate states. Our MSMs provide a human readable view of how ligands modulate the complex conformational landscape of β_2AR and improve performance of computer-aided drug design approaches. More generally, our cloud-based approach should be a powerful and broadly available tool for studying many biological systems.

Markov state models (MSMs) predict ligand-specific intermediate states in activation dynamics

To elucidate the mechanism of receptor activation, we built kinetic network MSMs from our data set. MD simulations describe intrinsic receptor dynamics in atomistic detail, while a MSM provides a summarized view of the ensemble of spontaneous fluctuations exhibited by the molecule at equilibrium (15). This helps to identify key conformational states of the receptor, and to quantify the state thermodynamic populations and the kinetics of state transitions. We built 3000 state MSMs using clustering along four structural metrics that represent the key differences between the active (2, 3) and inactive(1) crystal structures of β_2AR (see Methods) and map out the transitions between all states. Other structural metrics not highlighted in the main text, including dynamics of the extracellular loops and movement of all transmembrane helices, are discussed in the Supplementary Information.

Our MSM provides unique and novel insight into ligand-modulated activation of β_2AR for the first time. Using the MSM, we generated 150 μs activation trajectories (see Methods). In Figure 1b-d, these trajectories are shown as projections along the four described structural criteria for receptor with agonist BI-167107 (Figure 1b) and inverse agonist carazolol (Figure 1c) bound, as well as the apo-receptor (Figure 1d). Agonist bound simulations stabilize active-like conformations (boxed regions) for between 1.5-5.25 μs, while inverse agonist and apo simulations do not sample active state conformations along all structural metrics. The differences in activation dynamics are consistent with the biological function of the ligands that we have simulated, and confirm observables from previous simulations(5).

Figure 2a shows the network representation of the 3000 state MSM built from the simulations of agonist bound receptor. Such a detailed picture of β_2AR kinetics is useful for delineating activation pathways at atomistic resolution, which is outside the limits of temporal and spatial resolution of most experimental techniques. MSMs also provide a way to simplify this picture by discarding fast conformational dynamics to obtain a human comprehensible model of receptor dynamics comprised of lumped, kinetically relevant states. This lumping procedure was applied to reduce the 3000 state model to a simplified 10 state model of β_2AR dynamics (Figure 2b). This macro state model of β_2AR reveals two highly connected states, which are identified as inactive state (I) and the intermediate state (R') and several states with fewer connections, including the active state (R*). The intermediate state (R') is found to be different for the agonist and the inverse agonist (see Supplementary Figures S7, S8 and S9). For agonist bound receptor, the key structural difference between the inactive and intermediate states is the flipping of the I121-F282 residues in the connector region (shown in Fig 1a), whereas for the inverse agonist, the twisting of the Helix 7 NPxxY region residues distinguishes the two states. The intermediate state for the apo-receptor is the same as the intermediate state in the MSM of the agonist bound receptor.

β_2AR activation occurs through many parallel pathways

Transition path theory(16–18) analysis of the MSMs indicates a multitude of pathways with similar flux (see Supplementary Figure S10). Projection of these pathways along three of the four criteria used to construct the MSMs and described in Figure 1 allows the simplest view

of the predominant pathways for activation. In Figure 2c, unique, high flux pathways are shown. We employed experimentally validated mutual information analysis (See Supplementary Figure S11 and Supplementary Tables S1-S3) to find correlated residue networks that differ for simulations with different ligands and for the apo receptor (Figure 2d). The agonist (BI-167107) strengthens correlations between extracellular and intracellular residue groups, while inverse agonist (carazolol) disconnects these groups and apo receptor displays only indiscriminate correlations. These different networks of correlated residues give rise to the distinct activation pathways shown in Figure 2c. We map out the structural changes along the highest flux pathways in Figure 3 and highlight the most general trends for activation.

Figure 3 (panel 1) shows inactive conformations for all structural metrics. The Helix 3-6 distance is ~8 Å, with the ionic lock formed for 40-50% of the time in our simulations. The NPxxY region is in an extended alpha helix, and the inactive connector, I121 and F282, is stabilized by M215 and M279, respectively. These methionines undergo significant conformational changes during activation in our simulations (Supplementary Figure S17), as seen in crystal structures(2, 3) and NMR experiments(8, 19). Shown in panel 2, activation occurs initially with Helix 6 moving away from Helix 3, and M279 moves outward towards its active crystal pose. For different pathways, this helix change occurs prior to other metric changes, or more gradually, concomitant with NPxxY and connector changes (Figure 2c). In panel 3, the active NPxxY conformation is stabilized for both ligands by N322 hydrogen bonding to S120 and D79 – the latter aspartic acid is known to affect binding of agonists and G protein transfer(20) (Supplementary Figure S18). For inverse agonist simulations, the NPxxY Y326 forms stable hydrogen bonds with Y219 (Supplementary Figure S18), which points between Helix 3 and 6. Agonist bound receptor also samples close contact of these residues, but they are rarely in hydrogen-bonding range in active states. A partial connector active transition is observed in the panel 3 snapshot from the alternate inverse agonist pathway with the slowest connector flip (red path in Figure 2c), while the agonist bound snapshot shows a complete connector flip along with H5 bulge formation. For both ligands, the methyl group of M215 and the F208 phenyl group make close contacts with F282 in active receptor states (Supplementary Figure S17). In Panel 4, active state ligand interactions are shown, where agonist hydrogen bonds with S207, while inverse agonist hydrogen bonds instead to S203, located half a helix turn above S207. Both serines are key functional residues(21), but the S207 hydrogen bond could account, in part, for agonist stabilization of active conformations seen in Figure 1 that can be selected by G proteins. This effect is also seen in the binding site connections to the connector and to intracellular regions that are unique to the mutual information network for agonist bound simulations (Figure 2d).

Small molecule docking to β₂AR activation pathways enriches diverse GPCR ligand chemotypes

MSM states from high flux activation pathways identified in this study were targeted with small molecule docking of a database of β₂AR agonists, antagonists, and decoys(22) with the program Surflex(23, 24). For both agonists and antagonists, docking to the MSM states along activation pathways gives high Area Under the Receiver Operating Characteristic

Curve (AUC) values (Supplementary Figure S22), which evaluates selection of true ligands from decoys. These results are a statistically significant improvement over results from docking to the active (3P0G) and inactive (2RH1) crystal structures and to randomly selected snapshots from long-timescale, agonist bound β_2AR deactivation simulations(5). Next, we show that docking to MSM states expands the chemical space of our docking results, an essential advantage in docking approaches(25). Top scoring ligands for each MSM state were clustered according to their shape and chemistry, revealing a diversity of chemotypes that are highly ranked, or enriched, differentially by MSM states along the activation pathways. Examples of chemotypes that would have been undiscovered by virtual screen docking to the crystal structures alone or without knowledge of the full activation pathway are shown in Figure 4, with a full description in Supplementary Figures S23, S24. These results underscore the utility of MSMs for picking functional intermediate GPCR states that have different estimated affinities for known ligand chemotypes. Knowledge of this correspondence between ligand type and receptor conformation may be fruitful for future drug design efforts and can give testable predictions for ligands that may isolate rare intermediate conformations of receptors.

Discussion

The existence of parallel paths, comprising many intermediates, has reshaped the paradigm of protein folding, and led to the formulation of Markov state models as a new scheme to conceptualize such complex phenomena. Our results indicate that GPCR conformational dynamics is also well represented within the theoretical framework of MSMs, in which the receptor exists in multiple discrete conformational states with active and inactive states connected via multiple pathways. Furthermore, we show that ligands act by modulating the receptor dynamics to prefer different pathways. In contrast, methods that reduce simulation data to a single reaction coordinate by discarding dynamics along other independent coordinates may simplify the description of the conformational change, but could miss key features of the mechanism, including alternate pathways and kinetically trapped states. (26).

This study emphasizes the need to go beyond simple, few state models to capture the hidden complexity of GPCR activation. In particular, drug discovery efforts have focused on screening available GPCR crystal structures and fail to find leads with diverse efficacy profiles, beyond antagonists and inverse agonists. (27) Using MSMs to incorporate the rich structural data reported here into this process creates an opportunity for developing drugs that interact more closely with diverse receptor states(28), for overall increased efficacy and specificity.

The unprecedented millisecond simulation time scales presented here for GPCR activation require computing architectures capable of such extensive sampling. Cloud computing provides a promising new avenue to tackle these types of questions more routinely, leaving the complexity of the underlying infrastructure hidden and enabling a larger proportion of time spent on science rather than administrative tasks. Our work on Google's Exacycle platform demonstrates that large-scale exploratory analysis in the cloud can deliver new insight into biological problems. Existing commercial cloud computing initiatives enable researchers to migrate their own computational needs to the cloud in a scalable, elastic,

efficient, and easily accessible computational manner. The impact of this commodity computing approach is magnified when matched with novel algorithms, such as MSMs, that can extract the most pertinent information for researchers.

Methods

Simulation systems

Membrane-aligned crystal structures for PDB IDs 2RH1 and 3P0G were extracted from the OPM database (29). Both the inactive (PDB 2RH1) and the active (PDB 3P0G) crystal structure were simulated as ligand-free apo structure as well as receptor bound to the partial inverse agonist carazolol and the full agonist BI-167107(1, 2). Residues in transmembrane helices, and helix centers are defined as follows (with residue numbers in brackets): Helix 1 (29-60), Helix 2 (67-96), Helix 3 (103-136), Helix 4 (147-171), Helix 5 (197-229), Helix 6 (267-298), Helix 7 (305-328) according to (30). The crystal structures lack residues at N- and C-termini that were not resolved during crystallography. For 2RH1, our structures entail residues 30-342, and for 3P0G residues 23-344. Both have gaps in the sequence, where the intracellular loop 3 (ICL3) between helices 5 and 6 is replaced in 2RH1 and 3P0G with T4-lysozyme and a nanobody, respectively. These residues are 231-262 for 2RH1, and 228-264 for 3P0G. Since β_2AR remains functional even in the absence of ICL3, most of the simulations were done without this loop. One system was added in which apo-2RH1 was complemented by a model of ICL3 obtain from SuperLooper(31) for a total of 7 simulated systems.

All amino acids were protonated according to their pKa at neutral pH. We note that titratable residues D79 (2.50) and D130 (3.50) have been proposed to change protonation state during activation, and that acidic pH increases β_2AR basal activity by approximately twofold(6, 32, 33). Recent long time-scale simulations varied the pH of these residues and found that while the rate of deactivation increased by twofold for charged D130, the mechanism outlined for the receptor was largely unchanged (5). This study also focused its analysis on the neutral pH charge of β_2AR. All molecular dynamics simulations were performed with the Gromacs 4.5.3 molecular dynamics package (34), compiled as 64-bit Linux binaries from unmodified source code. Each system was simulated in a POPC lipid bilayer and solvated in TIP3P water molecules. Comprehensive details on system setup and simulation parameters are provided in the SI.

MD Simulations on Exacycle

Many large data centers are built with sufficient redundancies to handle peaks in user activity and cluster or fiber failures while maintaining low-latency responsiveness, leaving excess cycles during normal operation. Exacycle provides a platform for running massively parallel jobs on spare capacity of Google's production infrastructure. Initially, a limited number of grants were issued for select eScience projects (11). Exacycle was created as an alternative approach to large scale scientific computing, replacing complex parallelization of jobs across highly interconnected, specialized computing hardware by embarrassingly parallel jobs on highly heterogeneous commodity hardware. In principle, this enables the use of an arbitrary number of distributed processors. For MD, we replaced individual long-

timescale trajectories by stochastic sampling across a large number of trajectories. This allowed us to simulate β_2AR, with almost 60,000 atoms, at a rate exceeding 1 ms per month with a peak throughput of 80 μs a day. For comparison, the special-purpose Anton supercomputer achieves up to 16.4 μs of chemical time per day, interpolated from Figure 5 in (9).

All simulations were generated as sets of Exacycle work units. For each β_2AR system, 10,000-20,000 trajectories were simulated in three rounds with inter-round Markov State Model (MSM)-guided uniform sampling(12, 35) to determine new starting structures. Average trajectory lengths per round correspond to 5 ns, 12.5 ns, and 12.5 ns, respectively. Apo states in rounds 1 and 2 were run in a set of 20,000 trajectories, and 10,000 trajectories for non-apo states. The first round was started from equilibrated crystal structures with a reference temperature of 300 K. In between rounds, each set of trajectories was sampled at 0.5 ns and clustered according to the MSM criteria described below into 1000 microstates with MSMBuilder 2 (36) from which new starting structures were chosen uniformly for the next round. All systems were heated to a reference temperature of 330 K for the third round to achieve improved sampling. The higher temperature was chosen to increase the likelihood of rare events occurring while decreasing the impact of local minima. This is based on the assumption that while we are likely to observe a shift in kinetics, the thermodynamics of the GPCRs during this last stage would be preserved, since according to our observations changes in stabilizing enthalpic and entropic factors are dominated by subtle local conformational changes and remain small. Whereas significantly more entropically stabilized states might see a positive population shift in comparison with more enthalpically stabilized states as temperature increases, we expect this to be a negligible effect between the similarly stabilized active and inactive states. The final set of trajectories indicated convergence by comparing the conformational landscapes along the order parameters shown in Supplementary Figures S1-S4. Thus, for analysis, only data from the final round were used, equating to 125 μs aggregate molecular dynamics data each, or 750 μs total, subsampled at a rate of 0.5 ns for a total of about 250,000 molecular structures per simulated system.

Small molecule docking and chemotypes clustering

MSM states from TPT pathways with flux > 30% of the maximum flux were selected for a retrospective small molecule docking screen. For the agonist bound MSM, this corresponds to 20 MSM states, for the inverse agonist bound 43 states, and for the apo, 102 MSM states – for a total of ~2.9 million docking calculations. We also dock to both the active (3P0G) and inactive (2RH1) crystal structures, and to 20 randomly selected snapshots from previously performed long-timescale agonist bound GPCR deactivation simulations(5). All snapshots were aligned to the same active crystal structure (3P0G) before docking. The Surflex(23, 24) docking program was used to dock compounds from the GPCR Decoy and Ligand Database(22), which is comprised of ~200 known GPCR agonists and antagonists, and ~8000 structurally dissimilar but property matched decoys drawn from the compound library ZINC(37). Ligands were prepared with Surflex protonation tools, and OpenEye Omega(38) was used to enumerate stereoisomers for compounds with up to 4 chiral centers. Ligands were docked and scored with 20 snapshots from each MSM state, recording the

highest docking score. For computing receiver operating characteristic (ROC) plots, the best scoring stereoisomer was assigned to the ligand.

The top 10% scoring true ligands (decoys were excluded) were selected from each MSM state, resulting in 3300 enriched compounds for each agonist and antagonist docking set. These compounds were clustered by their chemotype with a k-centers algorithm, evaluated by Tanimoto values from unaligned (docked conformation preserved) shape and chemistry overlap computations with ROCS(39). A cutoff for the clustering was selected as the value separating the top 5% of all Tanimotos computed for all compound pair overlaps, to give 935 clusters for antagonist chemotypes and 497 agonist chemotypes (including different stereoisomers).

Analysis Software

Before analysis, several million files containing trajectory segments and related data were retrieved, merged, filtered, and pre-processed on Google's infrastructure using a combination of FlumeJava (40), MapReduce (41), and Bigtable (42). Coordinates for protein and ligand were stored in GROMACS .xtc format as well as converted to a binary format suitable for Dremel (43). VMD(44), GROMACS(34), and Dremel were used to analyse geometric properties such as interatomic distances and charge distributions. Markov State Models were generated with MSMBuilder 2 (36), which was also used for transition path theory analysis(18, 45). Surflex 2.6(24) and OMEGA 2.4.6 were used for small molecule docking and ROCS 3.1.2 was used to generate overlap scores for chemotype clustering.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References and Notes

1. Cherezov V, et al. High-Resolution Crystal Structure of an Engineered Human β2-Adrenergic G Protein–Coupled Receptor. Science. 2007; 318:1258–1265. [PubMed: 17962520]

2. Rasmussen SGF, et al. Structure of a nanobody-stabilized active state of the β2 adrenoceptor. Nature. 2011; 469:175–180. [PubMed: 21228869]

3. Rasmussen SGF, et al. Crystal structure of the β2 adrenergic receptor-Gs protein complex. Nature. 2011; 477:549–555. [PubMed: 21772288]

4. Dror RO, et al. Identification of two distinct inactive conformations of the β2-adrenergic receptor reconciles structural and biochemical observations. Proc. Natl. Acad. Sci. U.S.A. 2009; 106:4689–4694. [PubMed: 19258456]

5. Dror RO, et al. Activation mechanism of the β2-adrenergic receptor. Proc. Natl. Acad. Sci. U.S.A. 2011; 108:18684–18689. [PubMed: 22031696]

6. Vanni S, Neri M, Tavernelli I, Rothlisberger U. Predicting Novel Binding Modes of Agonists to β Adrenergic Receptors Using All-Atom Molecular Dynamics Simulations. PLoS Comput. Biol. 2011; 7:e1001053. [PubMed: 21253557]

7. Ivetac A, Andrew McCammon J. Mapping the Druggable Allosteric Space of G-Protein Coupled Receptors: a Fragment-Based Molecular Dynamics Approach. Chem. Biol. Drug Des. 2010; 76:201–217. [PubMed: 20626410]

8. Nygaard R, et al. The Dynamic Process of β2-Adrenergic Receptor Activation. Cell. 2013; 152:532–542. [PubMed: 23374348]

9. Shaw DE, et al. Anton, a special-purpose machine for molecular dynamics simulation. Commun. ACM. 2008; 51:91–97.

10. Shirts M, Pande VS. Screen Savers of the World Unite. Science. 2000; 290:1903–1904. [PubMed: 17742054]

11. Hellerstein JL, Kohlhoff KJ, Konerding DE. Science in the Cloud: Accelerating Discovery in the 21st Century. IEEE Internet Comput. 2012; 16:64–68.

12. Bowman GR, Huang X, Pande VS. Using generalized ensemble simulations and Markov state models to identify conformational states. Methods. 2009; 49:197–201. [PubMed: 19410002]

13. Senne M, Trendelkamp-Schroer B, Mey ASJS, Schütte C, Noé F. EMMA: A Software Package for Markov Model Building and Analysis. J. Chem. Theory Comput. 2012; 8:2223–2238. [PubMed: 26588955]

14. Noé F, Fischer S. Transition networks for modeling the kinetics of conformational change in macromolecules. Current Opinion in Structural Biology. 2008; 18:154–162. [PubMed: 18378442]

15. Bowman GR, Geissler PL. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. Proc. Natl. Acad. Sci. U.S.A. 2012; 109:11681–11686. [PubMed: 22753506]

16. Vanden-Eijnden WE,E. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. Annu. Rev. Phys. Chem. 2010; 61:391–420. [PubMed: 18999998]

17. Metzner P, Schütte C, Vanden-Eijnden E. Transition Path Theory for Markov Jump Processes. Multiscale Modeling & Simulation. 2009; 7:1192–1219.

18. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. Proc. Natl. Acad. Sci. U.S.A. 2009; 106:19011–19016. [PubMed: 19887634]

19. Kofuku Y, et al. Efficacy of the β2-adrenergic receptor is determined by conformational equilibrium in the transmembrane region. Nature Commun. 2012; 3:1045. [PubMed: 22948827]

20. Strader CD, et al. Conserved aspartic acid residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. J. Biol. Chem. 1988; 263:10267–10271. [PubMed: 2899076]

21. Liapakis G, et al. The Forgotten Serine A Critical Role for Ser-2035.42 in Ligand Binding to and Activation of the β2 Adrenergic Receptor. J. Biol. Chem. 2000; 275:37779–37788. [PubMed: 10964911]

22. Gatica EA, Cavasotto CN. Ligand and decoy sets for docking to G protein-coupled receptors. J Chem Inf Model. 2012; 52:1–6. [PubMed: 22168315]

23. Jain AN. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. J. Comput. Aided Mol. Des. 2007; 21:281–306. [PubMed: 17387436]

24. Spitzer R, Jain AN. Surflex-Dock: Docking benchmarks and real-world application. J Comput Aided Mol Des. 2012; 26:687–699. [PubMed: 22569590]

25. Shoichet BK. Virtual screening of chemical libraries. Nature. 2004; 432:862–865. [PubMed: 15602552]

26. Lane TJ, Shukla D, Beauchamp KA, Pande VS. To Milliseconds and Beyond: Challenges in the Simulation of Protein Folding. Curr. Opin. Struc. Bio. 2013; 23:58–65.

27. Shoichet BK, Kobilka BK. Structure-based drug screening for G-protein-coupled receptors. Trends Pharmacol. Sci. 2012; 33:268–272. [PubMed: 22503476]

28. Schames JR, et al. Discovery of a novel binding trench in HIV integrase. J. Med. Chem. 2004; 47:1879–1881. [PubMed: 15055986]

29. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. OPM: orientations of proteins in membranes database. Bioinformatics. 2006; 22:623–625. [PubMed: 16397007]

30. Rosenbaum DM, et al. Structure and function of an irreversible agonist-β2 adrenoceptor complex. Nature. 2011; 469:236–240. [PubMed: 21228876]

31. Hildebrand PW, et al. SuperLooper--a prediction server for the modeling of loops in globular and membrane proteins. Nucleic Acids Res. 2009; 37:W571–W574. [PubMed: 19429894]

32. Ballesteros JA, et al. Activation of the β2-Adrenergic Receptor Involves Disruption of an Ionic Lock between the Cytoplasmic Ends of Transmembrane Segments 3 and 6. J. Biol. Chem. 2001; 276:29171–29177. [PubMed: 11375997]

33. Ghanouni P, et al. The Effect of pH on β2 Adrenoceptor Function Evidence For Protonation-dependent Activation. J. Biol. Chem. 2000; 275:3121–3127. [PubMed: 10652295]

34. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput. 2008; 4:435–447. [PubMed: 26620784]

35. Voelz VA, Bowman GR, Beauchamp KA, Pande VS. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1−39). J. Am. Chem. Soc. 2010; 132:1526–1528. [PubMed: 20070076]

36. Beauchamp KA, et al. MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. J. Chem. Theory Comput. 2011; 7:3412–3419. [PubMed: 22125474]

37. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A Free Tool to Discover Chemistry for Biology. J. Chem. Inf. Model. 2012; 52:1757–1768. [PubMed: 22587354]

38. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. J. Chem. Inf. Model. 2010; 50:572–584. [PubMed: 20235588]

39. ROCS, version 3.1.2. OpenEye Scientific Software, Inc.; Santa Fe, NM: 2011.

40. Chambers, C., et al. PLDI '10. ACM; New York, NY, USA: 2010. Proceedings of the 2010 ACM SIGPLAN conference on Programming language design and implementation; p. 363-375.

41. Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. Commun. ACM. 2008; 51:107–113.

42. Chang F, et al. Bigtable: A Distributed Storage System for Structured Data. ACM Trans. Comput. Syst. 2008; 26(4):1–4. 26.

43. Melnik S, et al. Dremel: interactive analysis of web-scale datasets. Proc. VLDB Endow. 2010; 3:330–339.

44. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J. Mol. Graph. 1996; 14:33–38. [PubMed: 8744570]

45. Berezhkovskii A, Hummer G, Szabo A. Reactive flux and folding pathways in network models of coarse-grained protein dynamics. J. Chem. Phys. 2009; 130:205102–205102–5. [PubMed: 19485483]
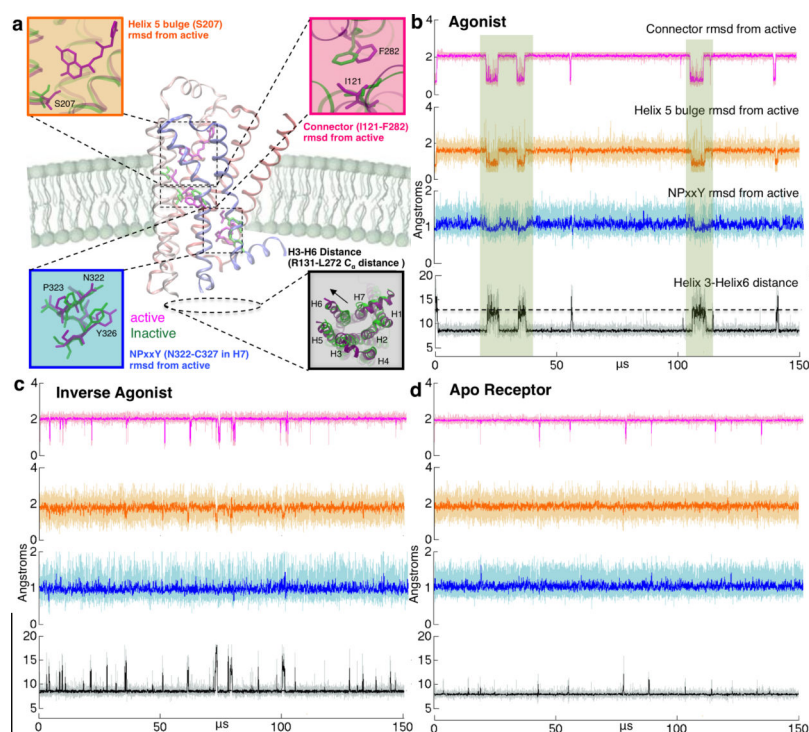
Figure 1. MSM activation trajectories on the sub-millisecond timescale

(a) The key differences between active and inactive structures of β_2AR. The root mean squared deviation (RMSD) of the connector region (I121-F282), the H5 bulge (S207-F208) RMSD from active crystal structure (3P0G), the distance between Helix 3 and Helix 6 (measured as R131-L272 distance) and the RMSD of NPxxY region (N322-C327) in Helix 7 from the active crystal structure (3P0G). Active crystal structure values are marked with dashed lines. 150 μs MSM activation trajectories were built using the data from simulations of GPCR bound to (b) agonist BI-167107, (c) inverse agonist carazolol as well as (d) the apo receptor. Agonist bound simulations stabilize active-like conformations (highlighted with boxes) throughout the trajectory and deactivate in 2.5 ± 0.05 μs. Inverse agonist bound simulations deactivate from active states slightly faster, at 1.3 ± 0.1 μs and Apo simulations deactivate in 0.67 ± 0.02 μs (See Supplementary Figures S5 and S6). These timescales corroborate previous simulations of deactivation (5) and experiments when keeping in mind that only G protein binding can truly stabilize the active states.(30)
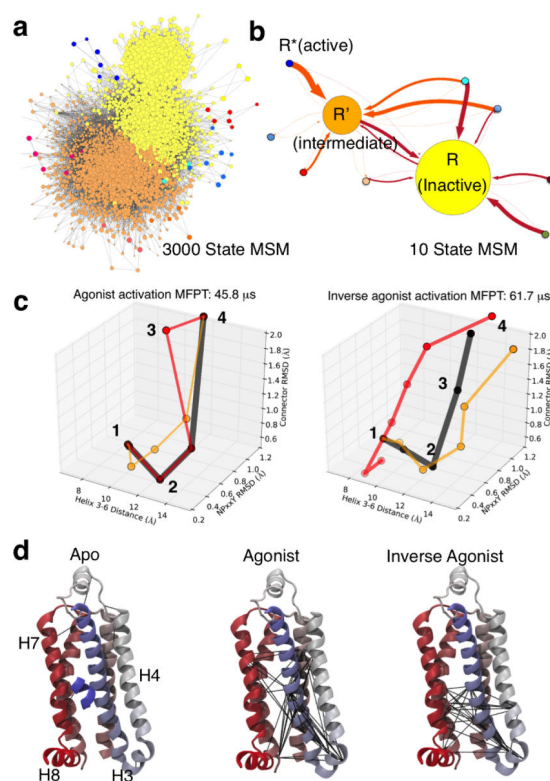
Figure 2. Markov State models and high flux activation pathways for agonist and inverse agonist bound simulations

(a) Network representation of the 3000 state MSM built from the simulations of agonist bound GPCR with each circle representing an individual conformational state. (b) 10 State MSMs build from the 3000 state MSMs using spectral clustering methods to identify kinetically relevant states. The circles in the 3000 state MSM are colored according to their membership in the coarse-grained 10 state MSM. The weight of arrow indicates the transition probability between states. (c) Pathways are shown as states (circles) connected along the 3-D reaction coordinate used, in part, to build the MSM. Pathway connections are scaled by the path flux relative to the highest flux in black; for inverse agonist pathways, red is 61% and orange is 51% of the max; for agonist red: 48% and orange: 35%. (d) Mutual information networks of dynamically correlated residues. Black lines indicate connected residue pairs, and only helices 3-7 are shown in the image for clarity. Agonist bound simulations reveal a network of residues that connect the extra and intra-cellular parts of the receptor to stabilize active states, whereas inverse agonist eliminates these connections and blocks activation.
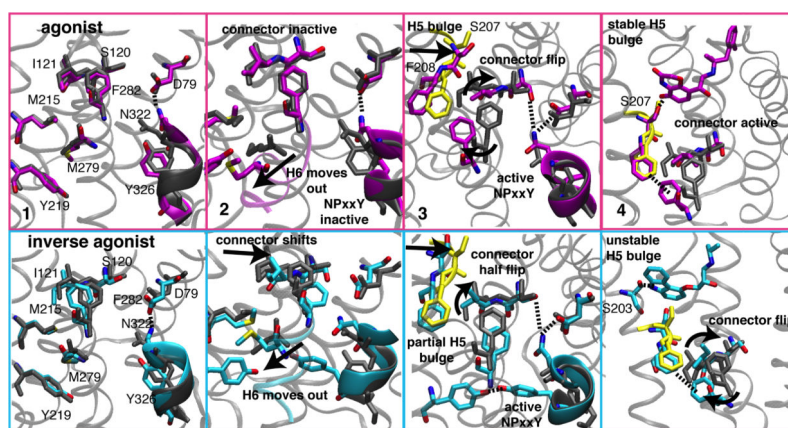
Figure 3. Structural details of the activation pathways

Representative structures from states along the activation pathways in Figure 2c labeled by number in four panels. The transition from inactive conformations in panel 1 proceed via Helix 3-6 outward movement (Panel 2), the switching of M215 interactions from connector I121 to F282 as these residues flip conformation (Panel 3), and changes in the H5 region around F208, S207, and S203 (Panel 4) that form ligand-mediated interactions for stabilizing active conformations that can be selected by G proteins. Residues in grey are the aligned inactive crystal (2RH1) conformations; residues in yellow are from the active crystal structure (3P0G).
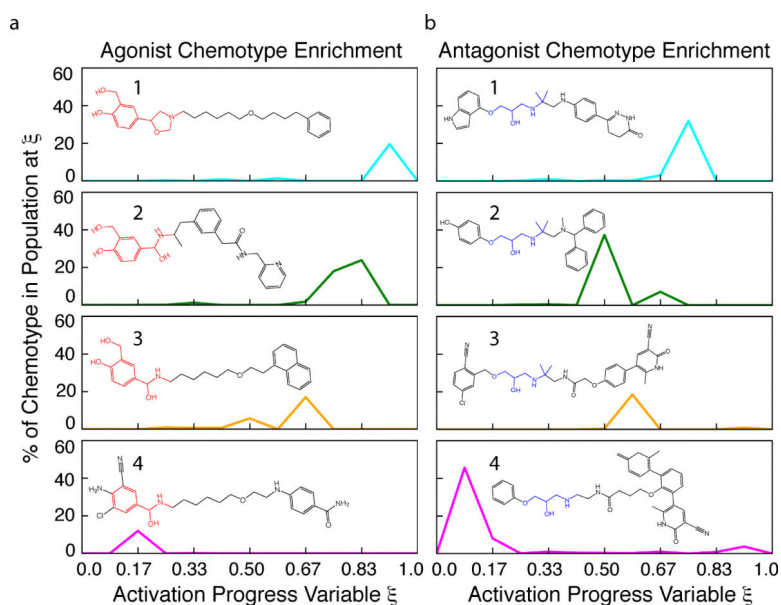
Figure 4. Examples of GPCR Ligand Chemotypes Enriched at MSM States Along Activation Pathways

MSM states from high flux activation pathways were assigned a progress score ξ based on structural metrics in Figure 1 and range from the inactive crystal structure (2RH1) at $\xi=0$ to the active crystal structure (3P0G) at $\xi=1$. Top ranking compounds from a retrospective virtual screen of known β_2AR ligands at each MSM state, and both crystal structures, were clustered according to their 3D shape and chemistry overlap. Four examples of chemotype clusters enriched by select MSM states along the activation pathways are shown, with the percentage of a chemotype represented in the total ligands enriched at a given ξ. (a) Example agonist chemotypes are catecholamine derivatives (**1**, **2** and **3**) and ethanolamine (**4**) derivatives. b) Example antagonist chemotypes (**5, 6, 7** and **8**) share a 2-hydroxy propyl amino core and include a carbostyril substituted pyridazinone (**5**), benzhydryl-amine (**6**), and pyridone nitriles (**7, 8**). The complete distributions along ξ for all agonist and antagonist chemotype clusters are shown in Supplementary Figures S21 and S22. These results show that docking to intermediates identified by MSM Transition Path Theory analysis enriches more diverse chemotypes that could be missed by screens of only a few structures.