Supporting Information

Activation of Corticotropin-Releasing Factor 1 Receptor: Insights from Molecular Dynamics Simulations

Rajesh Singh, Navjeet Ahalawat, and Rajesh K. Murarka*

SI Text

Active state ensemble

Active state conformations obtained from aMD trajectories were not enough for calculation of mutual information based correlations. To overcome the sampling effect, we carried out normal MD simulations starting from 50 randomly selected active state conformations. Each simulation was carried out for 10 ns (a total of 500 ns) and coordinates were saved every 2 ps. The active state considered in this work represents conformations that were satisfying the active state criteria defined by two distances: cytoplasmic TM2-TM6 distance (21.5-23.5 Å) and His155-Glu209 distance (10-13 Å), respectively.

Calculation of Mutual Information:

The configurational space of a molecule can be described in a standard Cartesian coordinate system or in an internal coordinate system of bond lengths, bond angles, and torsional angles. The use of internal coordinates for conformational entropy calculations has previously been shown to be more successful strategy than using Cartesian coordinates[1]. In this work we have considered only φ, ψ, and heavy-atoms χ torsional angles and neglected the contribution from bond and angle degrees of freedom since they tend to be relatively small[2-4]. Conformational entropies were calculated using the Shannon Entropy[5] (equation 1 & 2) for each torsion angle (and torsion angle pairs). We used 35 bins with adaptive partitioning to calculate the torsion angle distributions. A correction for under sampling was applied similar to previous works[3, 6, 7].

$$H(A) = -R\left(\sum_{i=1}^{M_A} p(a_i)\log p(a_i) + \frac{M_A-1}{2N}\right) \quad (1)$$

Here R is the gas constant, $p(a_i)$ the probability of i^{th} bin, N the total number of data points, and M_A denotes the number of discrete histogram bins with nonzero probability. Similarly, entropy for torsional angle pair can be defined as

$$H(A,B) = -R\left(\sum_{i=1}^{M_A}\sum_{j=1}^{M_B} p(a_i,b_j)\log p(a_i,b_j) + \frac{M_{AB}-1}{2N}\right) \quad (2)$$

Pair-wise mutual information (MI) was calculated using equation 3.

$$I(A,B) = H(A) + H(B) - H(A,B) \quad (3)$$

The residue based MI was calculated as the average MI between all torsion pairs involving the two residues.

Calculation of allosteric pathways

To calculate allosteric pathways we adapted mutual information (MI) based method. This method uses MI derived from torsional degrees of freedom to calculate allosteric pathways between each pair of residues that showed above-average MI ($MI > MI_{avg}$) and were >10 Å apart in the receptor structure[3]. We first constructed an undirected inter-residue contact network, where the residues represent nodes and the inter-residue contacts represent the edges of the network. Two residues are in contact if the average distance between Cα atoms of the residue pair were within 10 Å of one another. The edge weights were calculated as $MI_{max} - MI_{ab}$, where MI_{max} is the maximum MI among all residue pairs in the receptor and MI_{ab} is the MI between the two residues connected by that edge[3]. Weights of all edges with $MI < MI_{avg}$ were set to zero to avoid selection of pathways with weak MI. The allosteric pathways were calculated using the shortest-path algorithm by Dijkstra[8], as implemented in the NetworkX[9] package of python.

Clustering of allosteric pipelines

Top 500 allosteric pathways based on MI of terminal residues were selected for further analysis. Clustering of pathways was done based on their mutual proximity in the receptor structure. To define the

proximity of two pathways, we used "overlap" parameter defined by Bhattacharya and co-workers[3], which is the fraction of nodes of both pathways that are within cutoff distance (10Å) of one another. We clustered the pathways using hierarchical clustering method implemented in SciPy package of python. To determine the optimal number of clusters, we calculated (a) intraoverlap: average overlap of pathways within each clusters, and (b) interoverlap: average overlap across clusters. The optimal number of clusters was chosen as the one that maximized intraoverlap, while minimizing interoverlap. The cluster separation efficiency (SE) between two clusters i and j is given by equation 4. The optimum number of clusters can be obtained by maximizing the average SE over all clusters as discussed by Bhattacharya and co-workers[3],

$$\mathrm{SE(i,j)} = \frac{\mathrm{Intraoverlap(i)+Intraoverlap(j)}}{2} - \mathrm{Interoverlap(i,j)} \qquad (4)$$

Volume calculation of the G protein-binding site

Center atoms were identified at the G protein-binding site in each structure and an "inclusion" region is defined in a text-based POVME input file. Since both receptors, CRF1R structures (PDB 4K5Y) and the GPCR-Gs complex (PDB 3SN6), belong to different classes of GPCRs different residues near to the binding site were identified to best cover the inclusion region at the binding pocket. Grid spacing of 1.0 Å was used to create a field of equidistant points and then the volume is measured using POVME2.[10,11]

References

(1) Killian, B. J.; Kravitz, J. Y.; Somani, S.; Dasgupta, P.; Pang, Y. P.; Gilson, M. K. *J.Mol. Biol.* **2009**, *389*, 315.

(2) Niesen, M. J.; Bhattacharya, S.; Grisshammer, R.; Tate, C. G.; Vaidehi, N. *J. Phys. Chem. B* **2013**, *117*, 7283.

(3) Bhattacharya, S.; Vaidehi, N. *Biophys. J.* **2014**, *107*, 422.

(4) McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. *J. Chem. Theory Comput.* **2009**, *5*, 2486.

(5) Shannon, C. E. *1963. M.D. computing : computers in medical practice* **1997**, *14*, 306.

(6) Steuer, R.; Kurths, J.; Daub, C. O.; Weise, J.; Selbig, J. *Bioinformatics* **2002**, *18*, S231.

(7) Pandini, A.; Fornili, A.; Fraternali, F.; Kleinjung, J. *Faseb J.* **2012**, *26*, 868.

(8) Dijkstra, E. W. *Numer. Math.* **1959**, *1*, 269.

(9) Hagberg, A. A.; Schult, D. A.; Swart, P. J. *Proceedings of the 7th Python in Science Conference (SciPy2008)* **2008**, 11.

(10) Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A., POVME: an algorithm for measuring binding-pocket volumes. *J. Mol. Graph. Model.* **2011,** 29, (5), 773-776.

(11) Durrant, J. D.; Votapka, L.; Sørensen, J.; Amaro, R. E., POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **2014,** 10, 5047-5056.

(12) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.

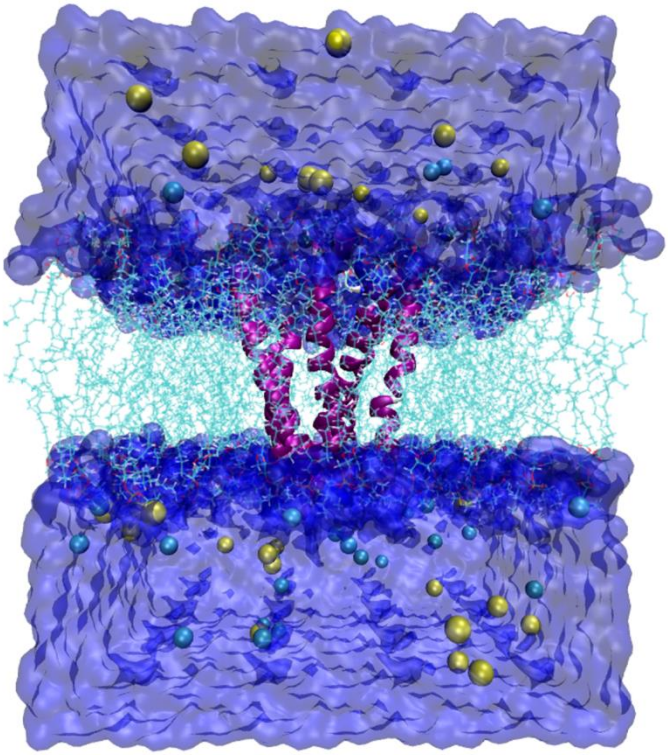


Figure S1. Schematic representation of CRF1R inserted into the POPC bilayer and solvated in an aqueous medium neutralized by 0.15 M NaCl.

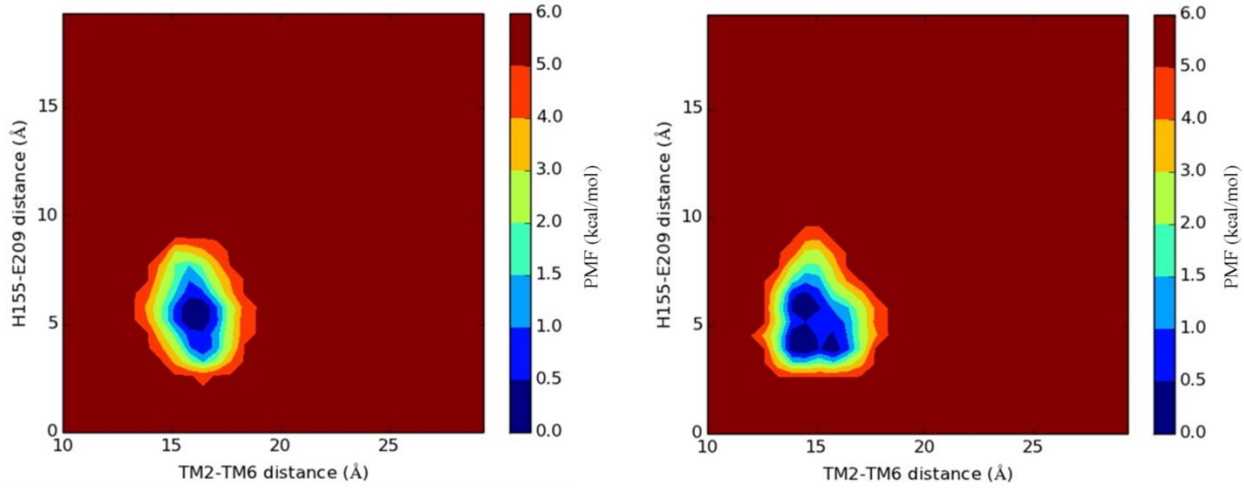
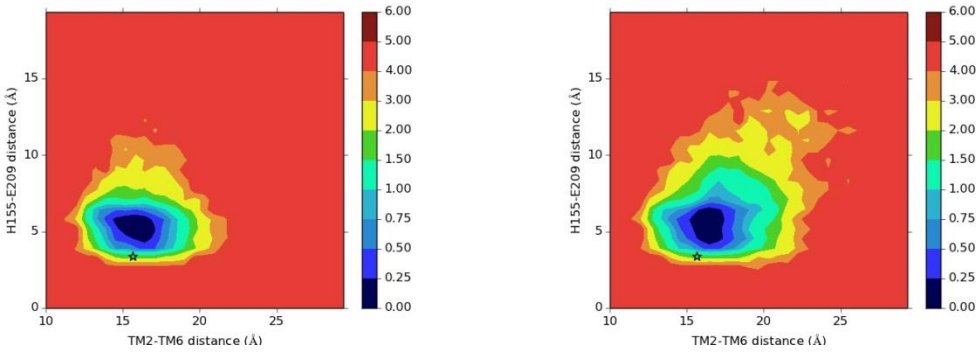
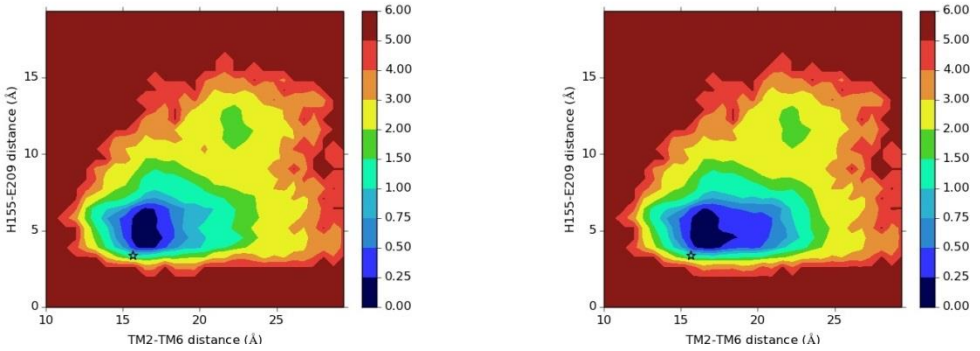


Figure S2. Two-dimensional potential of mean force profiles obtained from nMD simulations of the antagonist-bound (left) and the apo trajectory (right) (750 ns each; Table 1) for distances between TM2 and TM6 intracellular region vs. His155$^{2.50}$ and Glu209$^{3.50}$.



100 ns 200 ns



300 ns 400 ns

Figure S3. Two-dimensional potential of mean force (reweighted) profiles obtained from the aMD trajectory of apo receptor (Sim VI), at 100, 200, 300 and 400 ns, for distances between TM2 and TM6 intracellular region vs. $His155^{2.50}$ and $Glu209^{3.50}$. The star represents the position of the antagonist-bound inactive crystal structure (PDB 4K5Y).

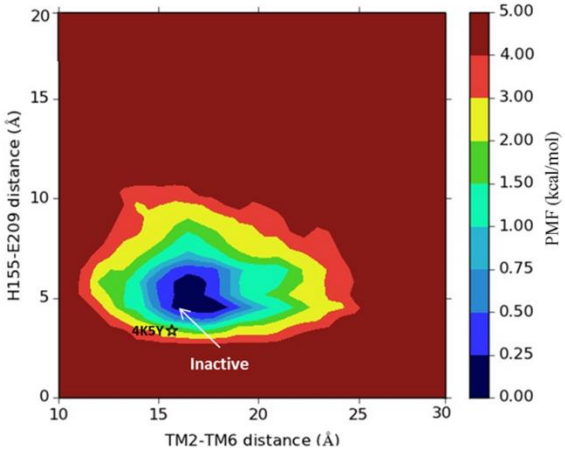


Figure S4. Two-dimensional potential of mean force (reweighted) profile obtained from aMD trajectory of the antagonist-bound receptor (500 ns - cumulative total of five trajectories) for distances between TM2 and TM6 intracellular region vs. His155$^{2.50}$ and Glu209$^{3.50}$. The star represents the position of the inactive crystal structure (PDB 4K5Y).

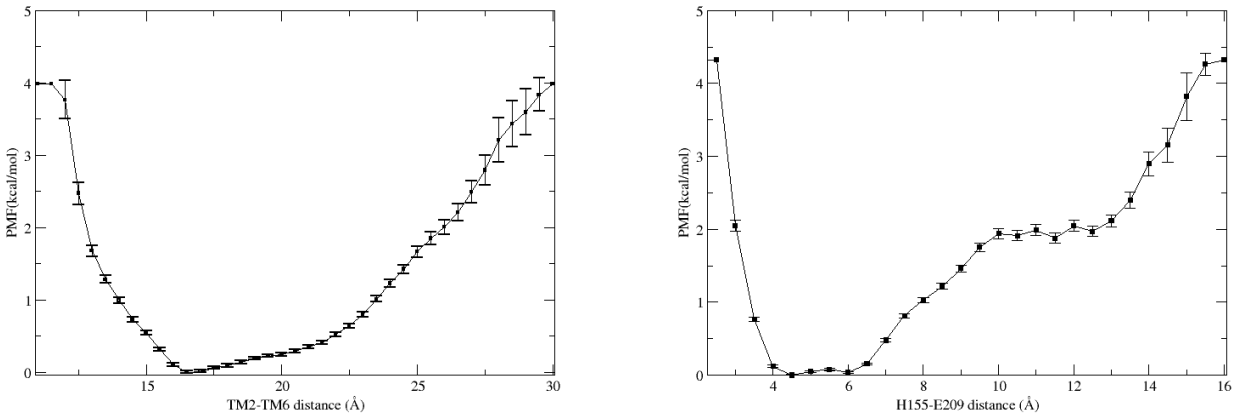


Figure S5. One-dimensional potential of mean force profiles with error bars (standard errors), calculated from the aMD simulations of the apo trajectory (Sim VI), for intracellular distances between TM2 and TM6 (left) and between His155$^{2.50}$ and Glu209$^{3.50}$ (right).

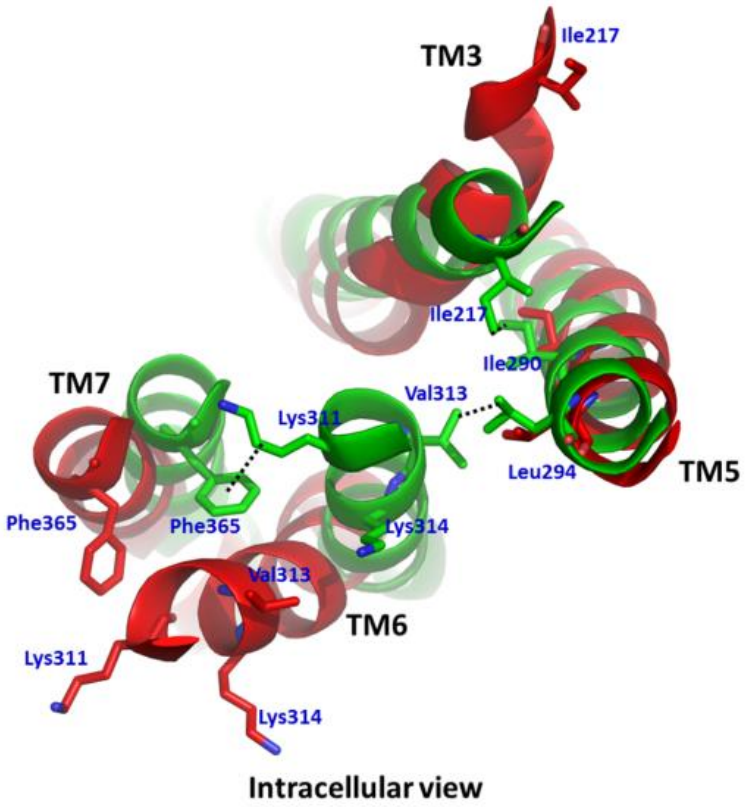


Figure S6. Superimposed structures of inactive state (green; PDB 4K5Y) and active state (red) of CRF1 receptor. Black lines show the interactions towards the cytoplasmic side of TM3, TM5, TM6 and TM7.

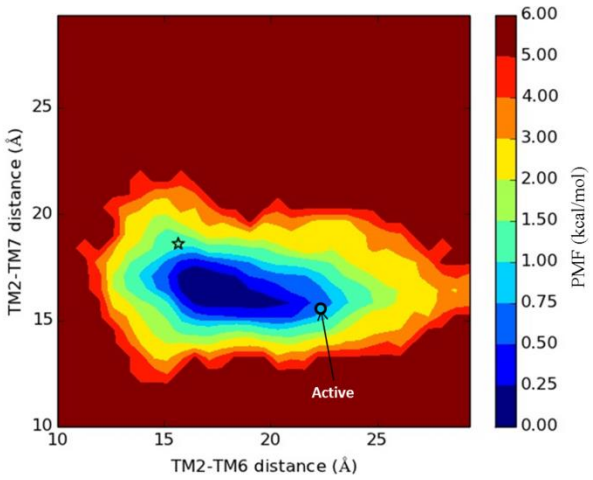


Figure S7. Two-dimensional potential of mean force (reweighted) profile obtained from aMD trajectory (Sim VI) of the apo receptor for distances between TM2 and TM6 intracellular region vs. TM2 and TM7 extracellular region. Star and circle represent positions of the inactive crystal structure (PDB 4K5Y) and the putative active state, respectively.

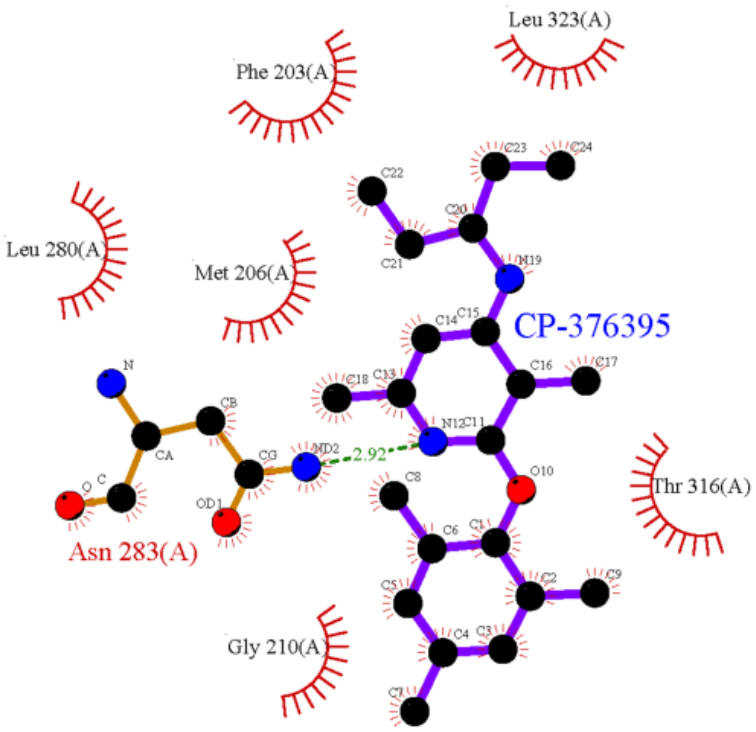


Figure S8. Interactions of antagonist CP-376395 with residues of CRF1R (PDB 4K5Y), generated using LigPlot (https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT).

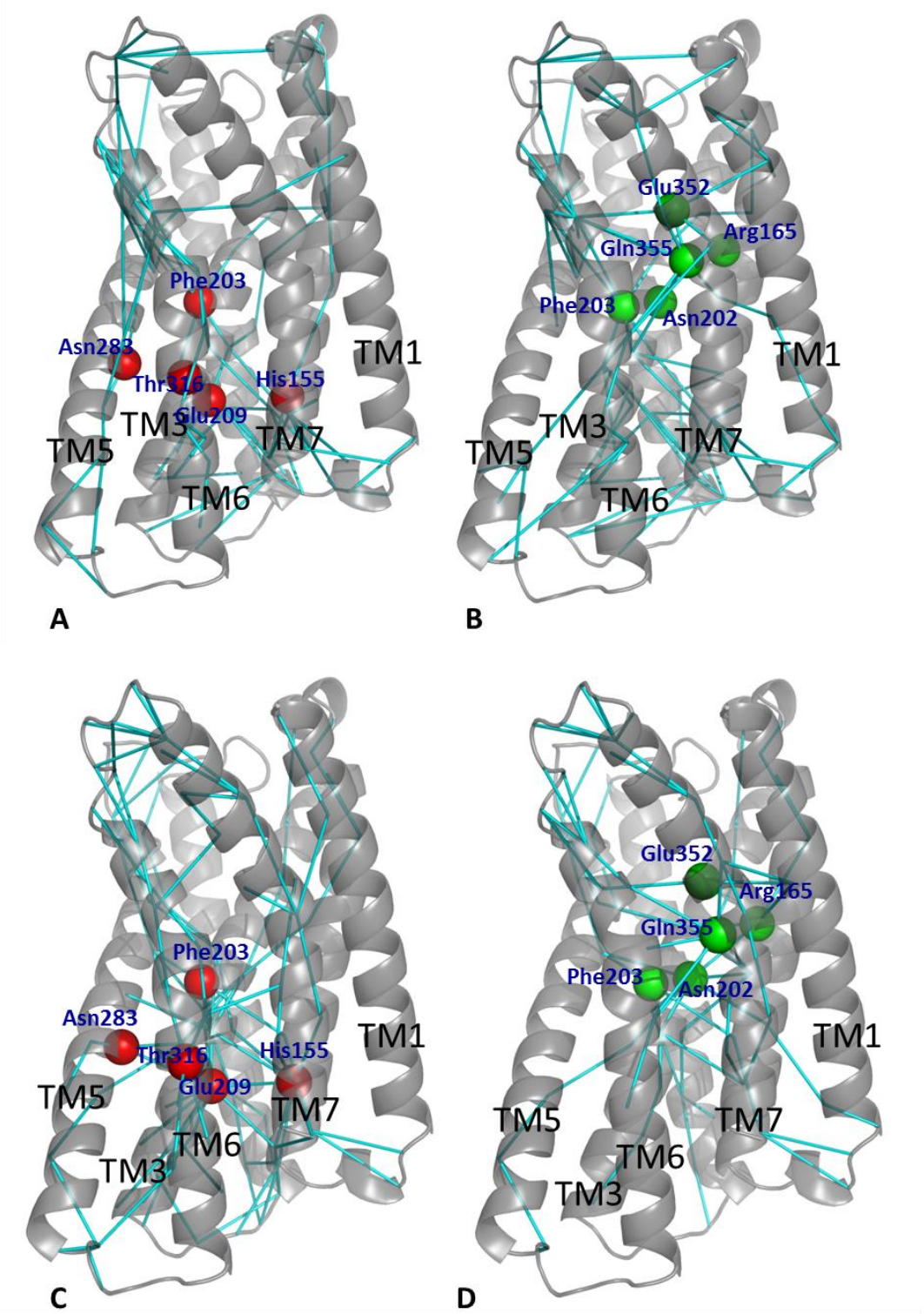


Figure S9. Locations of activating (red balls) and inactivating mutations (green balls), and the number of active state pathways (A & B) and the inactive state pathways (C & D) mediated through these residues, portrayed on the crystal structure (PDB 4K5Y).

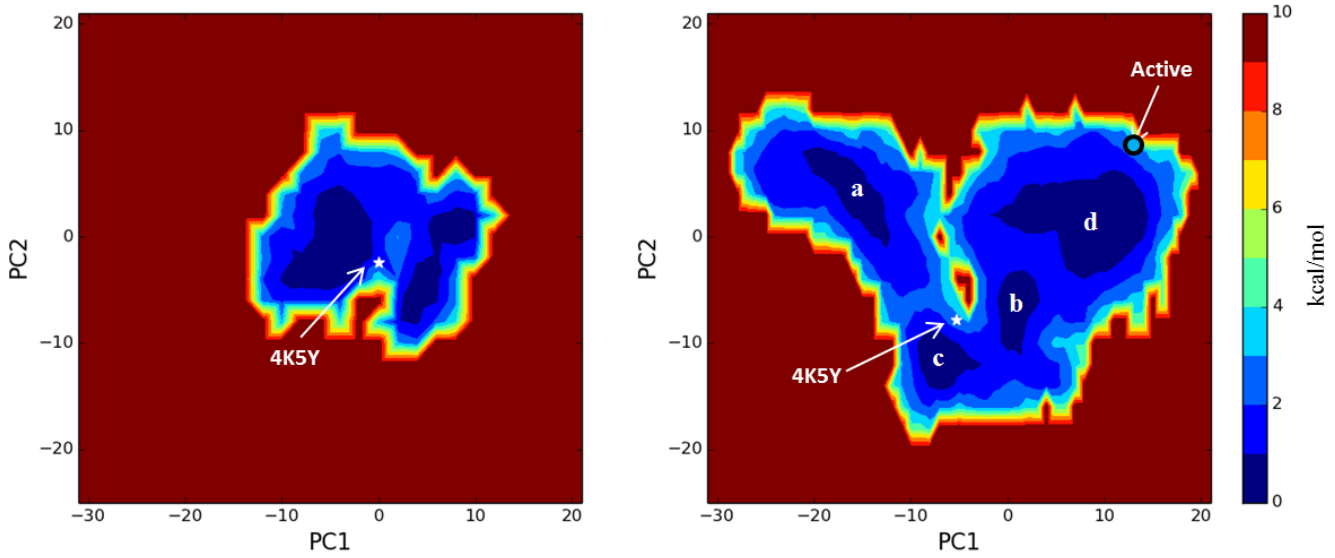


Figure S10. Free energy landscape (reweighted) along the first two principal components obtained from aMD simulations of the antagonist-bound (left) and the apo (right) trajectories. Star and circle represent positions of the inactive crystal structure (PDB 4K5Y) and a representative structure of the putative active state, respectively.

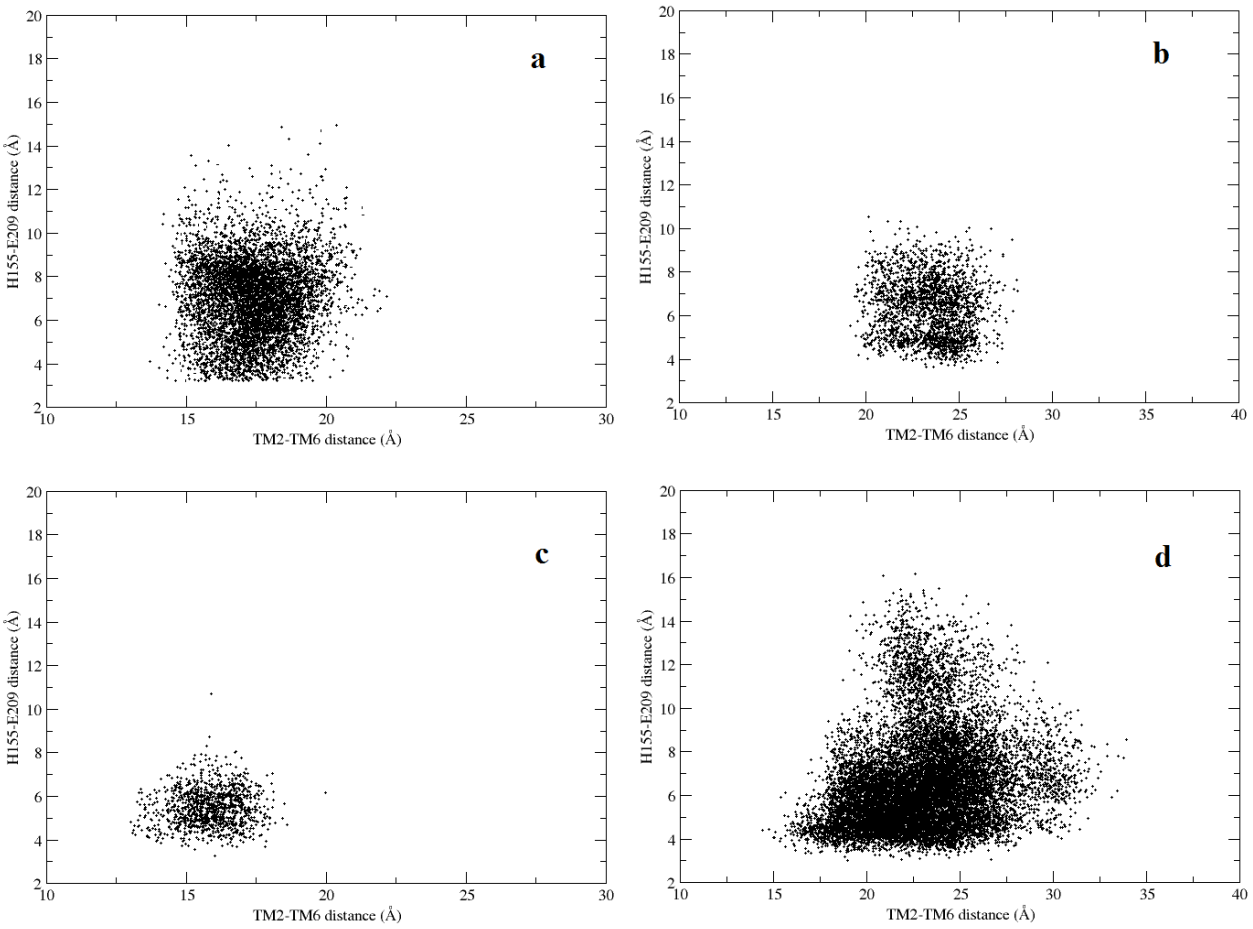


Figure S11. Four clusters identified from the principal component analysis of the apo trajectory.

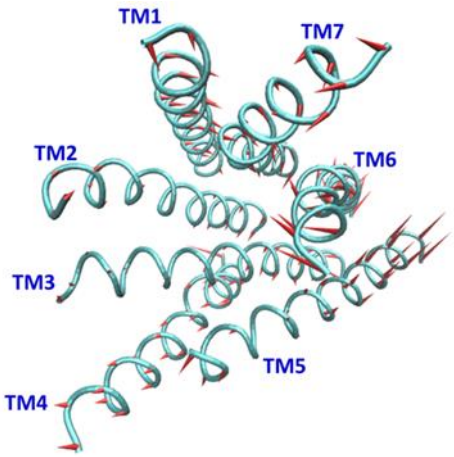


Figure S12. Motions (porcupine plots) along the first principal component (PC1) calculated from aMD simulations of the apo trajectory (extracellular view).

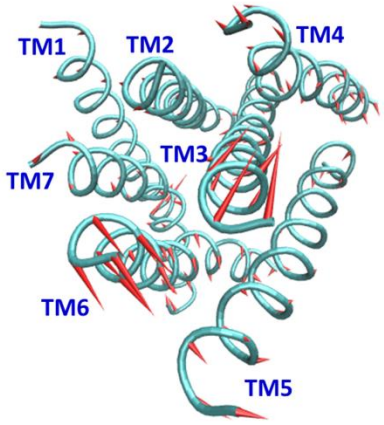


Figure S13. Motions (porcupine plots) along the second principal component (PC2) calculated from aMD simulations of the apo trajectory (intracellular view).

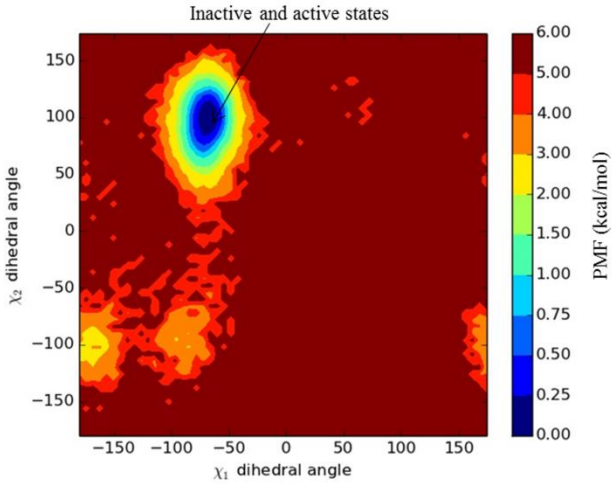


Figure S14. Two-dimensional potential of mean force (reweighted) profile obtained from aMD trajectory (Sim VI) of the apo receptor showing the spread of χ_1 and χ_2 dihedral angles of Trp236$^{4.50}$. Inactive and active conformations of the receptor overlap in the same region centered at χ_1 and χ_2 dihedral angles of -72° and 100°, respectively.

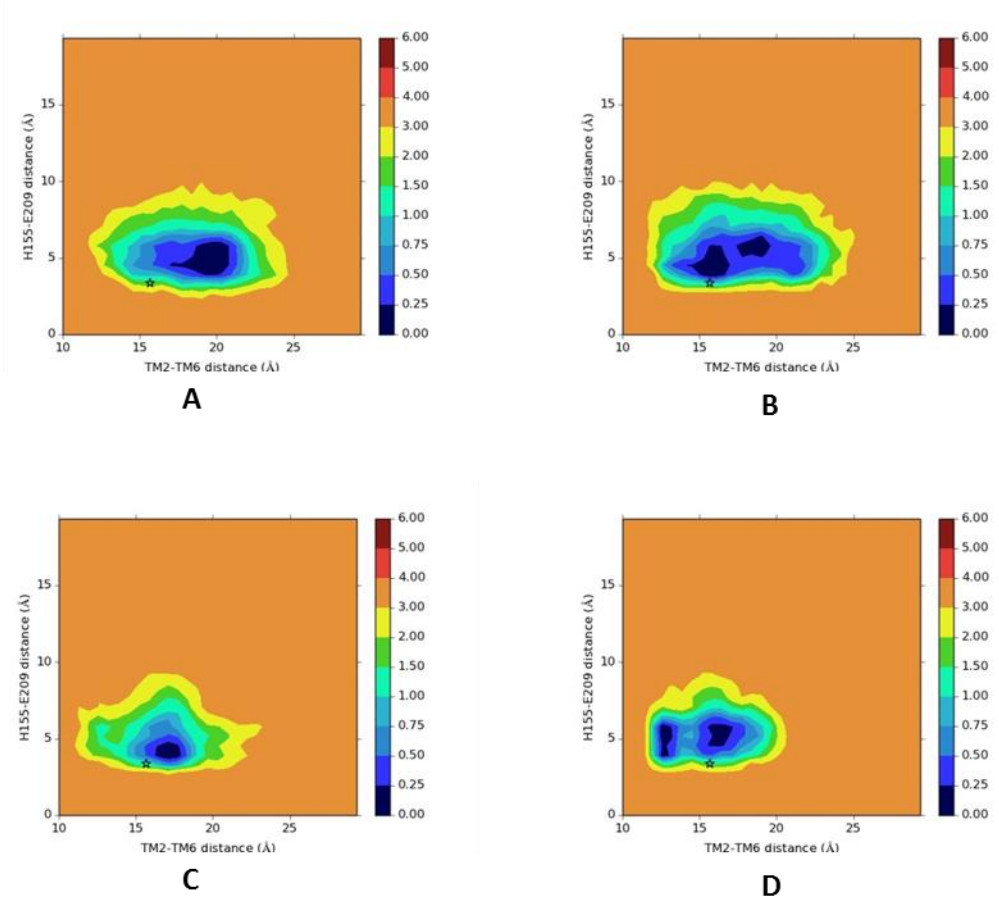


Figure S15. Two-dimensional potential of mean force (reweighted) profiles obtained from the 250 ns of aMD trajectories of the apo receptor: (**A**) Sim II, (**B**) Sim IV, (**C**) Sim VIII, (**D**) Sim X, for distances between TM2 and TM6 intracellular region vs. His155$^{2.50}$ and Glu209$^{3.50}$. The star represents the position of the antagonist-bound inactive crystal structure (PDB 4K5Y).

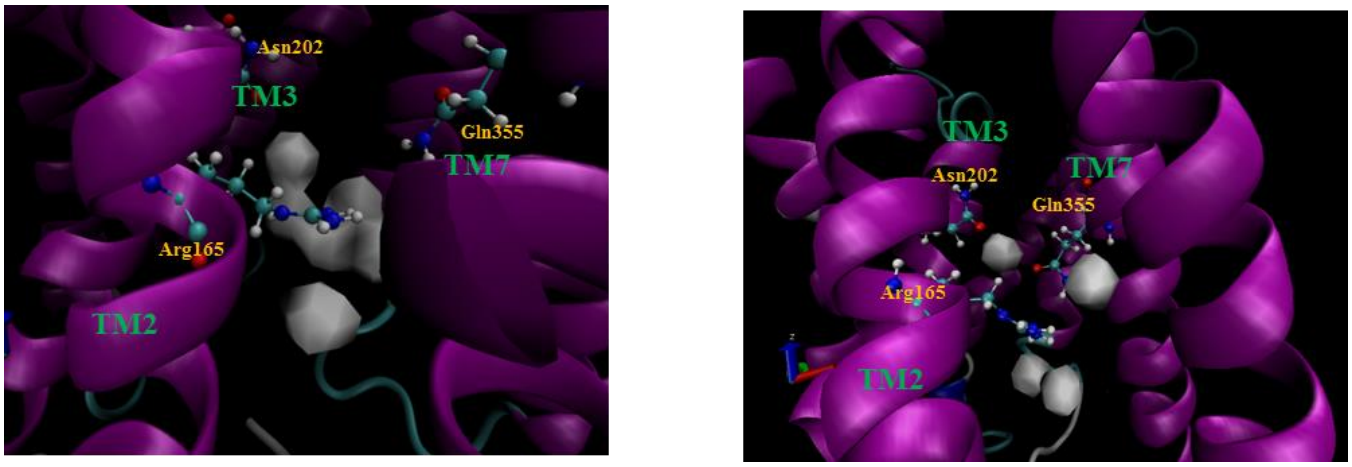


Figure S16. Water density around the conserved residues Arg165$^{2.60}$, Asn202$^{3.43}$ and Gln355$^{7.49}$ in the inactive (left) and active (right) states calculated using Volmap tool in VMD[12].

Table S1. Number of pathways in each allosteric pipeline in the inactive and active conformations of CRF1R, sorted by the pathway population.

Cluster No.	Number of pathways in each cluster	
	Inactive	**Active**
1	125	119
2	105	107
3	59	51
4	55	46
5	48	38
6	41	29
7	25	26
8	18	22
9	16	18
10	8	14
11	-	11
12	-	10
13	-	9