Thermostabilization of the β_1-adrenergic receptor Correlates with Increased Entropy of the Inactive State

Michiel JM Niesen[1], Supriyo Bhattacharya[1], Reinhard Grisshammer[2], Christopher G Tate[3], and Nagarajan Vaidehi[1,*]

[1]Division of Immunology, Beckman Research Institute of the City of Hope, 1500, E. Duarte Road, Duarte, CA-91010, USA.

[2] Membrane Protein Structure Function Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Department of Health and Human Services, Rockville, Maryland 20852, USA

[3]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

[*]Corresponding author; email: NVaidehi@coh.org

TEXT S1

Details of the MD Simulations:

The wt-β_1AR and m23-β_1AR receptors with lipids were minimized by steepest descent energy minimization. MD simulations were performed with Berger lipid parameters [1]. Short-range non-bonded interactions were truncated at 12Å, with the neighbor list updated every 10 fs. Long-range electrostatics was calculated using the smooth particle mesh Ewald (PME) method. Bonds were constrained using the LINCS algorithm [2-3] to allow for a simulation step size of 2 fs. Each system was equilibrated by performing 100 ps of MD at 310 K using a NVT ensemble followed by 5ns of MD under NPT conditions with a pressure of 1bar. The velocity-rescaling thermostat was used for temperature coupling during the equilibration [4] and a Parrinello-Rahman barostat [5] for pressure coupling. The protein and ligand were kept in place during these equilibration steps using position restraints of 100000 kJ/mol/nm^2.

RMSD based clustering

RMSD based conformational clustering was done using the g_cluster command in GROMACS with the gromos clustering algorithm [6]. A RMSD cut-off of 1.5Å was used and snapshots were taken from the trajectories every 100ps. The free energy difference between each cluster was estimated based on their relative population according to equation 1 [7]. The equation will give incorrect absolute free energy values but can be applied to estimate the difference in free energy between two states.

$$FE(x) = -kTln(P(x)) \qquad [1]$$

Here *FE(x)* is the free energy estimate for state *x*, *k* is the Boltzman constant, *T* is the temperature chosen as 310K and, *P(x)* is the probability of finding the protein in state *x*.

Conformational Entropy and Mutual Information Calculation

The use of internal coordinates for conformational entropy calculations has previously been shown to be a more successful strategy than using Cartesian coordinates [8]. We have

neglected the contributions from bond and angle degrees of freedom since they tend to be relatively small. First order conformational entropy was calculated using the Shannon Entropy (equation 2,3) [8-9] for each torsion angle. We used 35 bins in the torsion angle distribution, that choice was based on convergence of the calculated entropy values. A correction for under sampling was applied as has been done previously [10-11].

$$S(x) = -R(\sum_{n=1}^{nrBins}\{P(n)\ln\left(\frac{P(n)}{h(n)}\right)\} + \frac{nrBins-1}{2N}) \quad [2]$$

$$S(x) = R\left(\ln\left(Nh(n)\right) - \frac{1}{N}\sum_{n=1}^{nrBins}\{k(n)\ln\left(k(n)\right)\} - \frac{nrBins-1}{2N}\right) \quad [3]$$

Here R is the gas constant, $P(n)$ the probability of bin number n, $h(n)$ the width of bin n, N the total number of datapoints, and $k(n)$ the number of datapoints in bin n. The correction for variable bin width was only used when looking for a convergence in calculated entropy values dependent on number of bins. Finally the number of bins was set to 35 and the correction for bin width (division by $h(n)$) was removed.

Pair-wise mutual information [8,10-12] between torsion angles was calculated using equation 4. This requires the calculation of the secondary entropy between two degrees of freedom, this entropy can be calculated using equation 3.

$$I(x,y) = S(x) + S(y) - S(x,y) \quad [4]$$

When assessing the difference in Mutual Information (MI) between residue pairs in the mutant and the wild type (as done in Fig. 4 and Fig. S8) the MI was normalized by the entropy of the individual residues (equation 5-6) [11]. This was done in order to determine the ratio of entropy that is lost due to cross correlation, allowing us to better identify high correlation between residues with low individual entropy.

$$I(x,y)^* = \frac{I(x,y)}{S(x)+S(y)} \quad [5]$$

$$I(x,y)^* = 1 - \frac{S(x,y)}{S(x)+S(y)} \quad [6]$$

Normalizing the MI by the entropy of individual residues helps to identify correlation between residues with low individual entropy, and removes the bias towards residue pairs with high individual entropies [11].

Convergence of MD trajectories

To assess the convergence of our MD trajectories, we have calculated the absolute value of the dot product between the PC vectors of pairs of individual simulations (Equation 7) [13]. Each PC vector has 3N degrees of freedom corresponding to the x,y and z coordinates of the displacement for each atom. The PC vectors used were those corresponding to the Cα atoms of the TM helices (Table S1). If the dot product is equal to 1 the PC vectors point in the same direction, if the dot product is 0 the PC vectors are orthogonal. We observe a distribution around 0.8 for the first 4 PCs, indicating that for the TM domain the most dominant principal motions have converged reasonably well (Figure S10).

$$d_{ij} = \left|\sum_{k=1}^{3N} v_k^i * v_k^j\right| \quad [7]$$

Supplemental Discussion

Mutation R68S in m23-β_1AR causes stronger coupling between IC loop 1 and helix 8

The R68S thermostabilizing mutation in m23-β_1AR is at the intracellular end of TM1. In a recent high-resolution structure of carvedilol-bound m23-β_1AR S68$^{1.59}$ was observed to form a hydrogen bond (H-bond) with R355 on helix 8 [14], which was absent in the crystal structure of cyanopindolol-bound m23-β_1AR (PDB ID: 2VT4). The dynamics of the interaction between residues R/S68$^{1.59}$ and R355 H-bond in the MD simulations of wt-β_1AR and m23-β_1AR was quantified. In the wt-β_1AR there was repulsion between R68$^{1.59}$ and R355, while in m23-β_1AR S68$^{1.59}$ and R355 formed a H-bond for part of the simulation. To understand the effect of this H-bond on the structural dynamics of the receptor, we measured the Cα-Cα distance of residues R68$^{1.59}$ to R355 in wt-β_1AR and

S68$^{1.59}$ to R355 in m23-β_1AR during the MD simulations (Fig. S7). While wt-β_1AR shows a broad distribution of Cα distances, m23-β_1AR shows a relatively narrow distribution indicating that the H-bond between S68$^{1.59}$ and R355 stabilizes helix 8 and TM1, thus contributing to the increased thermostability of the mutant. We also quantified the propensity for H-bond formation by calculating the distribution of distances between the charged nitrogen atoms of the R355 sidechain and the polar oxygen of the S68$^{1.59}$ sidechain (Fig. S7B). The bimodal distribution of distances observed suggests that S68$^{1.59}$ exists in a dynamic equlibrium between a H-bonded state with R355 having favorable enthalpy (peak at 2.8Å), and a solvent exposed state with a broader range of distances (6-11Å) when the H-bond is weak. The bimodal distribution of R355 conformations is consistent with the different orientations of the side chain observed in the crystal structures of m23 [14]. The stabilization could come from enthalpy and entropy in this case.

Table S1: TM Definitions used for PCA analysis and RMSD based clustering.

TM	First Residue	Last Residue
1	44	68
2	76	104
3	114	144
4	156	178
5	206	232
6	294	313
7	324	341

Table S2: Definition of the Intracellular (IC) and Extracellular (EC) half of the receptor as used for quantifying entropy and mutual information per domain.

Domain	Residues included
Extracellular half	33-40,89-126,168-218,301-334
Intracellular half	41-88,127-167,219-300,335-358

Table S3: Calculated reduction in entropy due to correlated movement (called Mutual Information) of residues. The difference in mutual information between mutant and wild type β_1AR for all residues is shown in column 2 and for residues in TM domain only is shown in column 3. m23-β_1AR has more correlated movement between residues in the EC end of the receptor, while wildtype has more crosstalk in the IC end and between IC and EC. Units are TΔI (kcal/mol/K), the difference in pairwise mutual information.

Domain	For all residues	For TM residues only
EC-EC	0.16	0.07
EC-IC	-0.15	-0.05
IC-IC	-0.57	-0.35

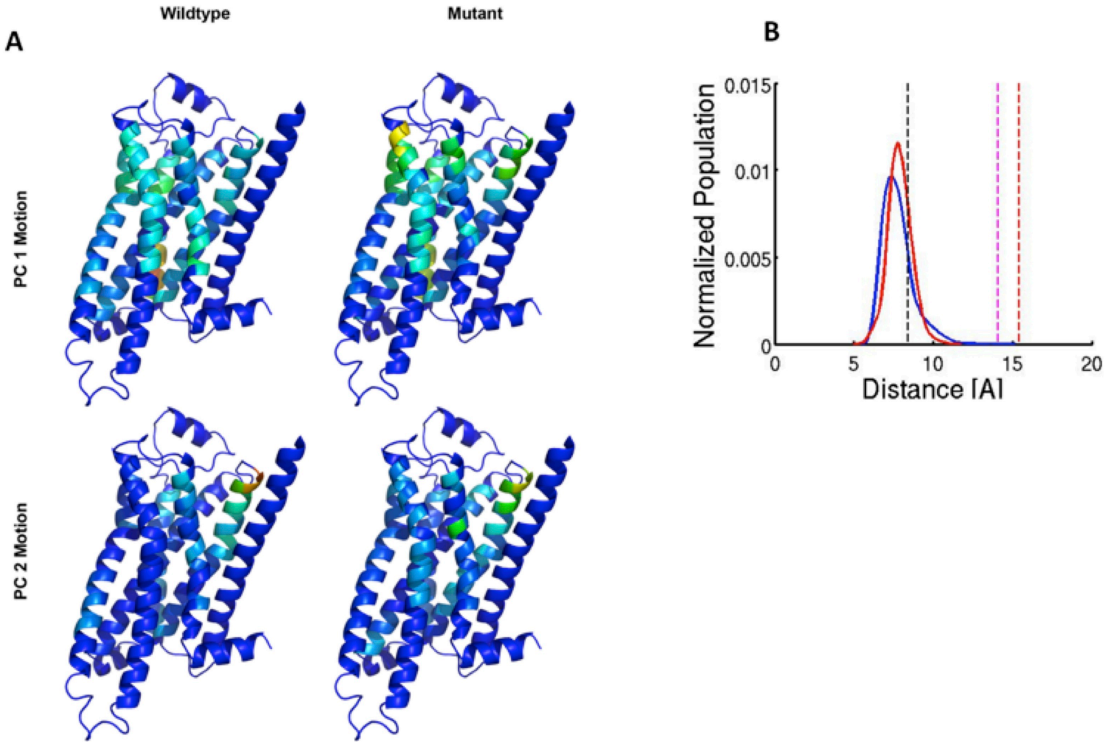
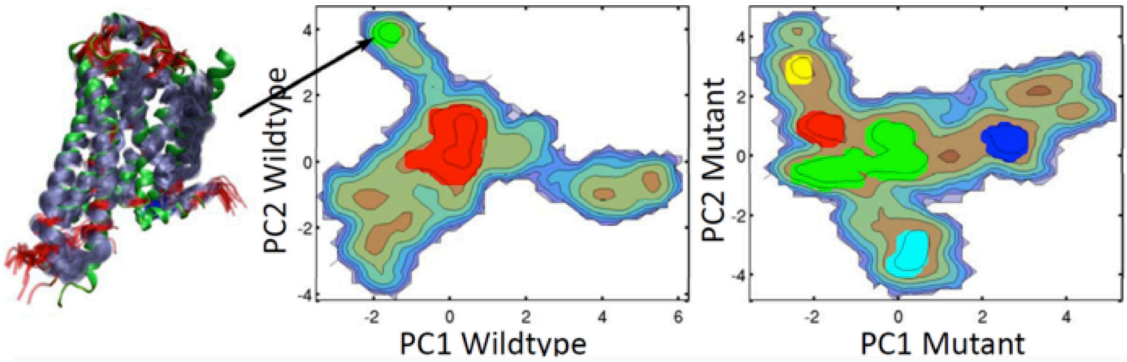


Figure S1. A. Regions of the receptor that contribute significantly to the principal component 1 (PC1) and PC2. Regions in red show highest movement, with gradation to regions in deep blue that show the least contribution to PC1 and PC2. Colorscale is the same across all figures. **B**. The distance between TM3-TM6 (distance between Cα atoms of R139$^{3.50}$ and L289$^{6.34}$), in blue for wt-b$_1$AR and in red for m23-b$_1$AR. The black dashed line indicates the distance in the inactive state crystal structure of β_1AR, the purple dashed line is in the active state crystal structure of β_2-adrenergic

receptor with the nanobody, and the red dashed line is in the fully active state crystal structure of the β_2-adrenergic receptor with the Gs protein bound.



wt-β_1AR Clusters	RMSD [Å]	PE [kcal/mol]	Pop [%]
C1 (red)	1.4	-2669	88
C2 (green)	2.8	-2531	5

m23-β_1AR Clusters	RMSD [Å]	PE [kcal/mol]	Pop [%]
C1 (red)	1.8	-2534	11
C2 (green)	1.5	-2593	53
C3 (blue)	2.3	-2514	12
C4 (yellow)	2.2	-2452	13
C5 (cyan)	2.3	-2546	6

Figure S2. Properties of conformational clusters by PC coordinates. Lowest potential energy is found in the cluster closest to the crystal structure in both m23-b$_1$AR and wt-b$_1$AR trajectories. Mutant has multiple clusters that are well populated and similar in potential energy. The conformational cluster with just 5% of the population for the wild type shows disorder in TM1 and TM7 that could be aggregation prone.

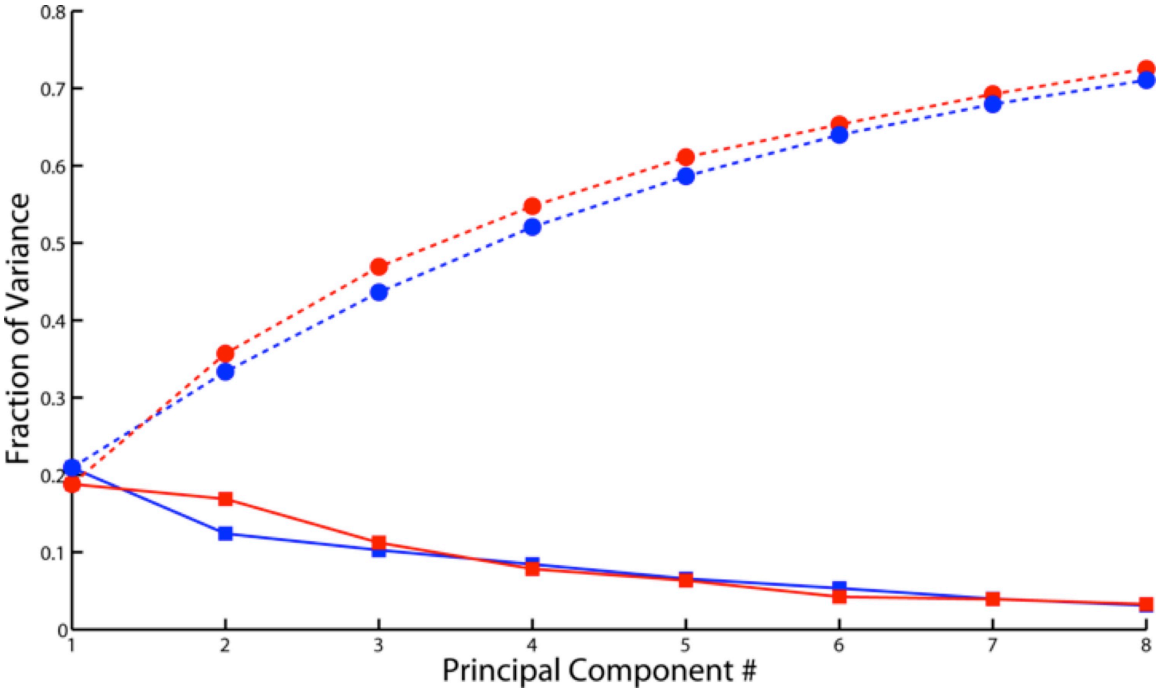


Figure S3. The fraction of variance captured by the PCs for wt-b_1AR (blue) and m23-b_1AR (red). Dashed lines with circles show the cumulative fraction of variance captured by the PCs, solid lines with squares show the fraction of variance captured by the current PC.

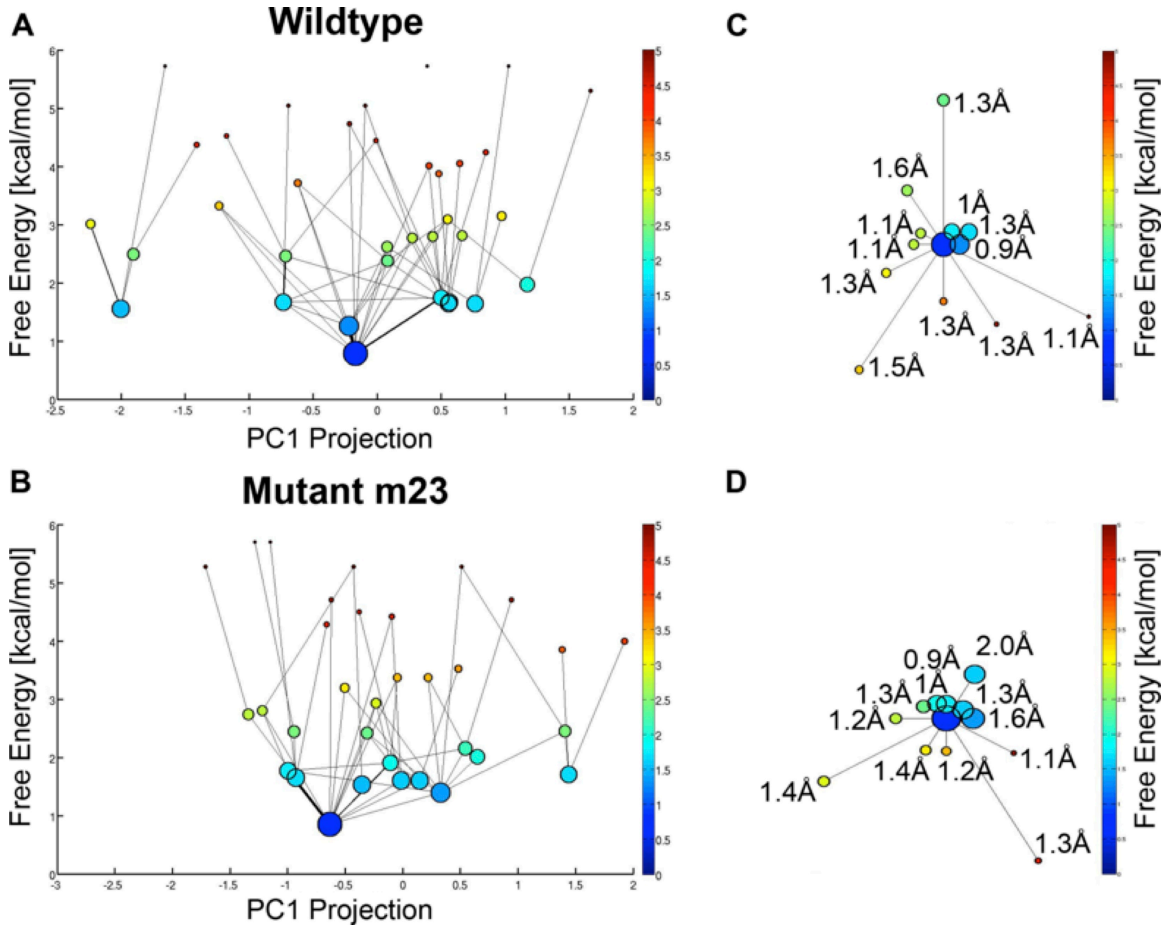


Figure S4. Conformational clusters calculated by RMSD for wt-b$_1$AR and m23-β$_1$AR. The clusters are colored by their free energies (kcal/mol). The size of the clusters indicates the population of the cluster. The thickness of the lines joining the clusters show the number of transitions observed between the clusters during the MD simulations. **A-B** Show the clusters plotted versus their free energy (Equation 1) on the y-axis and the combined PC coordinate (same coordinate used in **A** and **B**) on the x-axis. **C-D** Show only the most populated cluster (center) and the clusters with direct transitions to and from this cluster (surrounding the center cluster). Distance from the center cluster is inversely related to number of transitions, clusters that are closer to the center have more transitions to the main cluster.

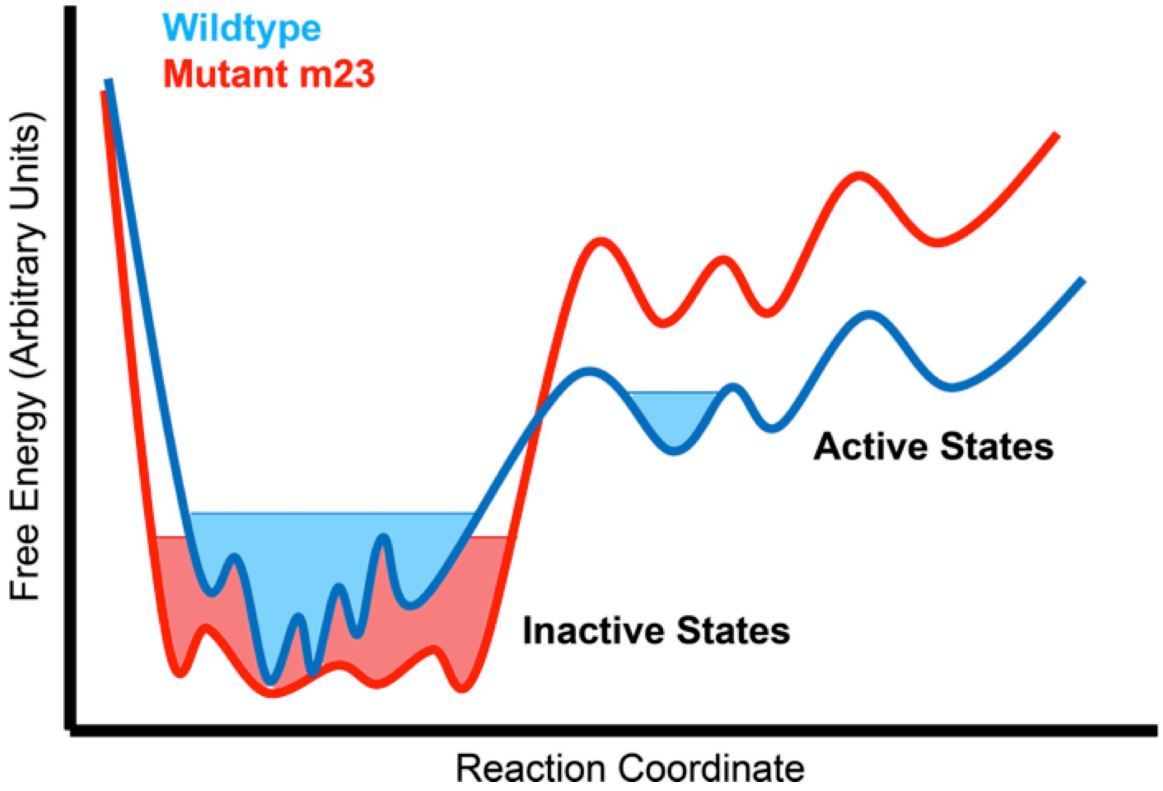


Figure S5. Schematic representation of the receptor population distribution in wt-b_1AR (blue) and m23-b_1AR (red). Note that in m23-b_1AR the population is shifted towards the inactive state of the receptor.

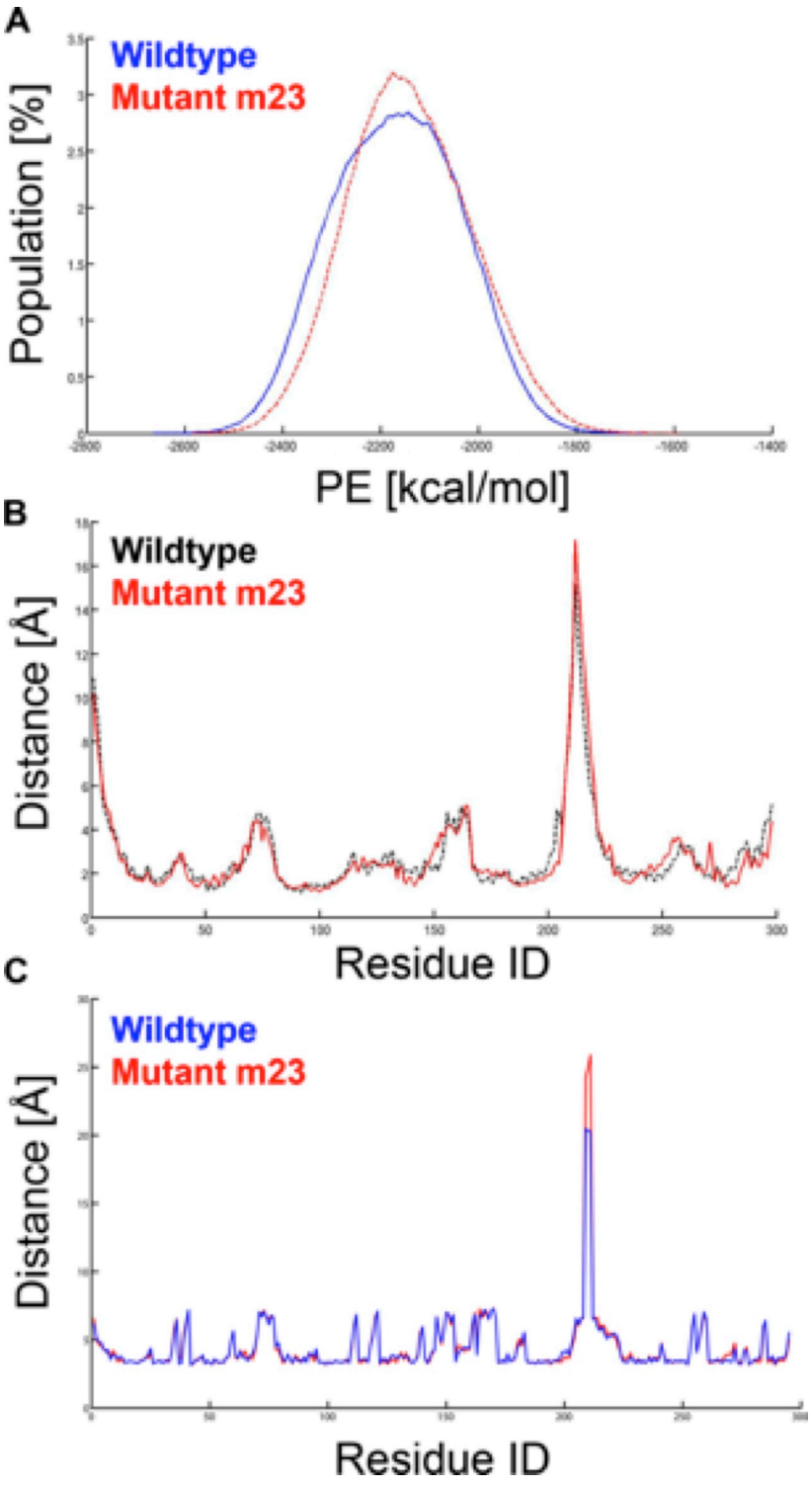


Figure S6. A. Distribution of protein potential energy over total MD trajectories. The mean value for the wild type distribution (-2167±130 kcal/mol, blue solid line) and the mutant distribution (-2140±126 kcal/mol, red dotted line) shows no difference. The mutant distribution has a higher peak, indicating that the receptor samples

more conformations with a similar potential energy. **B.** The difference in Cα root mean square deviation of each residue from the average structure from MD simulations for the m23-β_1AR (red) and wt-β_1AR (black). **C.** i-i+4 distances measured over the MD trajectories for m23-b_1AR (red) and wt-b_1AR (blue). The little difference between mutant and wild type in these two plots indicate that the average structures, and the secondary structures from the MD simulations do not differ significantly.

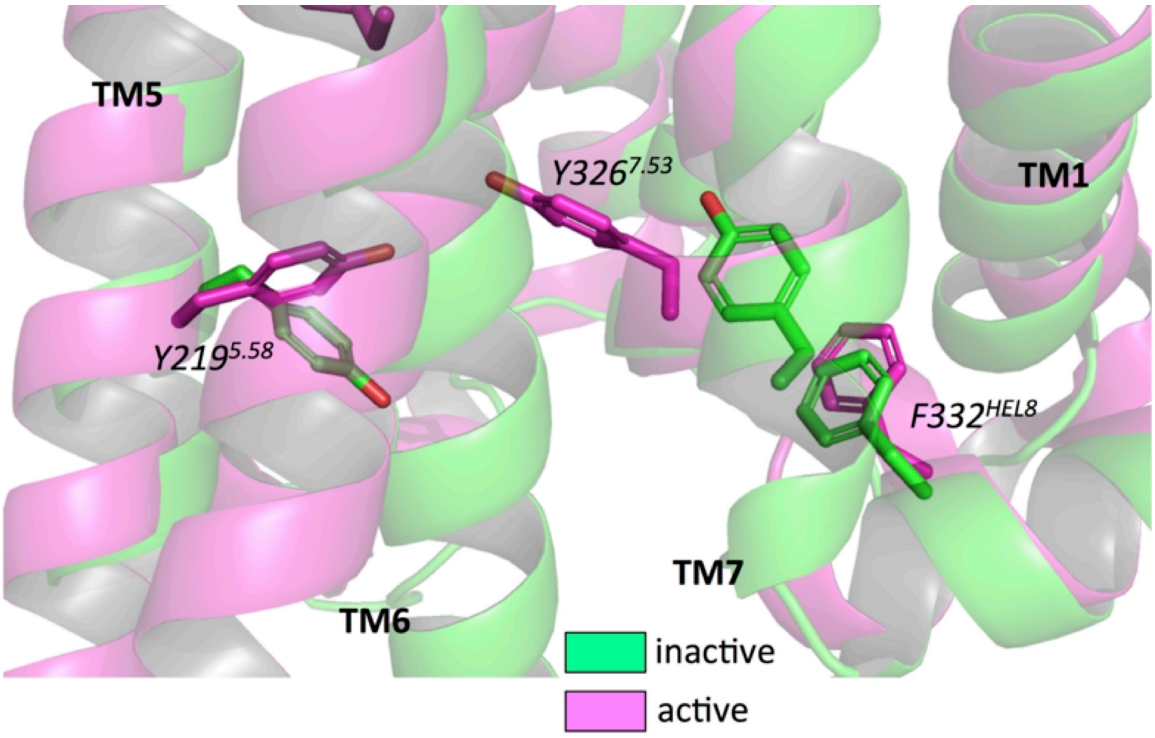


Figure S7. Difference in orientation for $Y219^{5.58}$ and $Y326^{7.53}$ between inactive (pdb: 2RH1) and active state (pdb: 3SN6) of β_2AR. These residues are conserved between b_1AR and b_2AR.

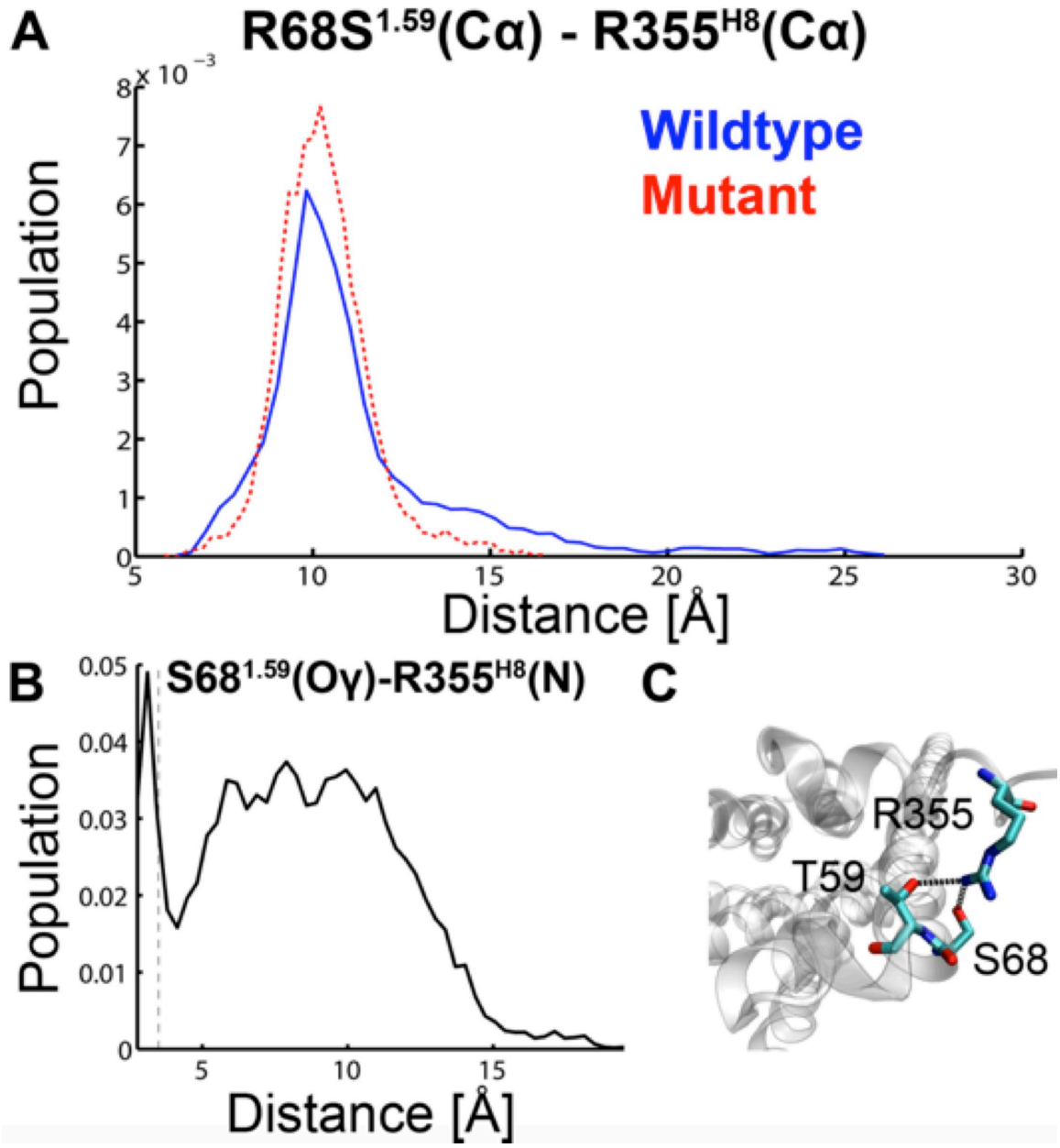


Figure S8. The R68S thermostabilizing mutation causes increased interaction between ICL1 and helix 8. The distances between atoms in residues 68 and 355 were measured from ~500,000 structures generated within a 1.05ms MD simulation performed for both wt-β_1AR and m23-β_1AR **A.** The distance between ICL1 and helix 8 is more stable in m23-β_1AR as measured by the distance between the Cα atoms of residue 68 and R355. **B.** Population of the distance distribution between S68 and R355 shows a hydrogen bond is formed (most populated distance

is between 2.9 and 3.3 Å). **C.** snapshot of the hydrogen bond (dashed black line) between S68 and R355 extracted from the simulations (C, cyan; O, red, N, blue).

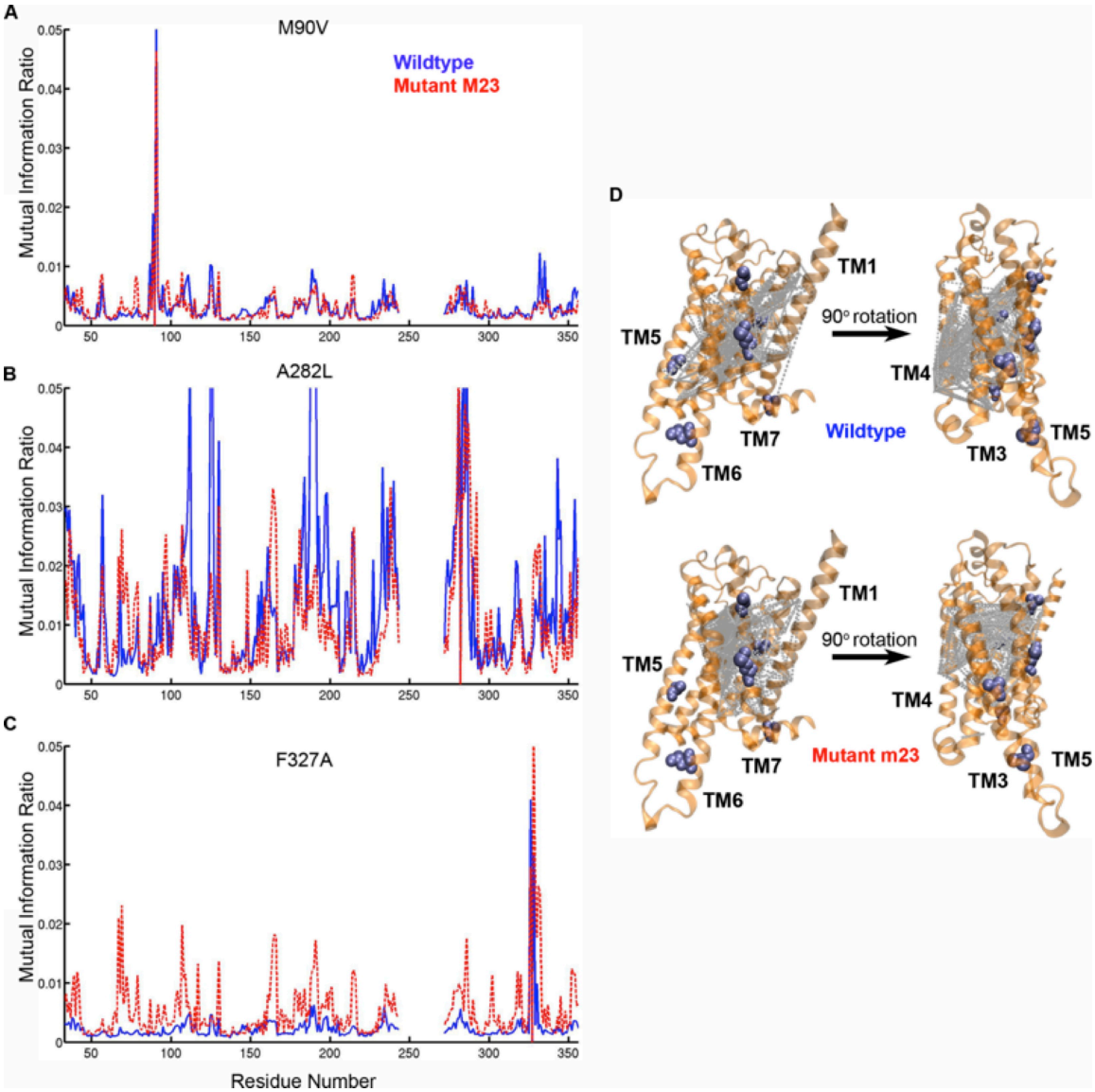


Figure S9. Mutual Information of all residues in the receptor with mutated residues **A.** M90V, **B.** A282L and **C.** F327A. No significant differences are found for M90V. The data for A282L reflected the high flexibility of this region in the simulations due to the lack of structure of ICL3 observed in the β_1AR structure. **D.** Residue pairs with increased mutual information in the wt-b_1AR and m23-b_1AR respectively are indicated with grey dashed lines, mutated residues are shown as blue spheres. In

wt-b_1AR there is increased correlation in the IC half of the receptor, while for m23-b_1AR most of the correlation is in the EC half.

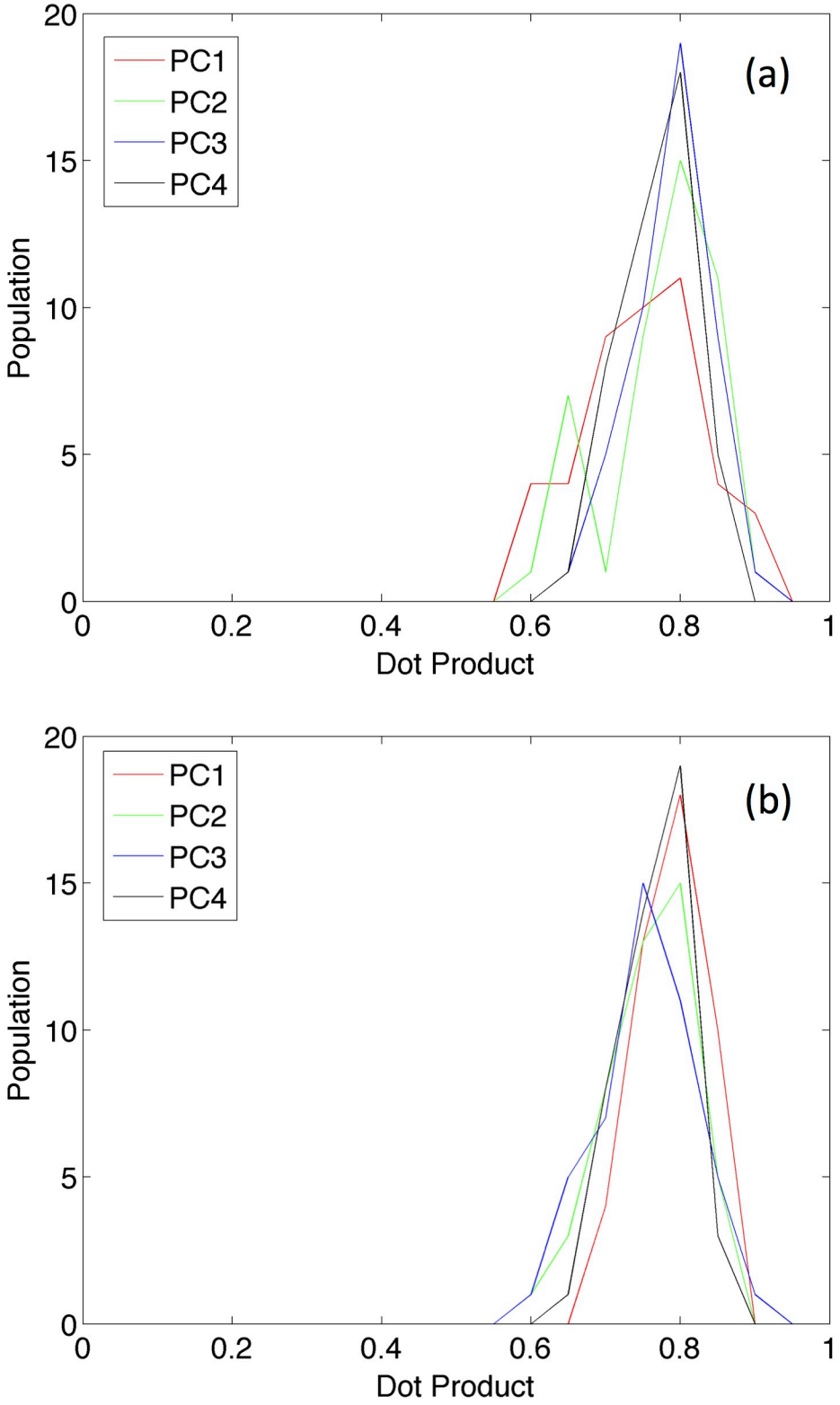


Figure S10. Population distributions of dot products of first four eigenvectors between pairs of individual trajectories of (a) WT and (b) m23 β_1AR. Absolute

values of dot products between normalized eigenvectors are plotted. Only the Cα atoms of TM domains are used. Values closer to 1 indicate that the vectors are in similar direction, while values close to 0 indicate that the vectors are near orthogonal.

References

(1) Berger, O.; Edholm, O.; Jahnig, F. *Biophysical Journal* **1997**, *72*, 2002-2013.

(2) Hess, B. *Journal of Chemical Theory and Computation* **2008**, *4*, 116-122.

(3) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J Comput Chem* **1997**, *18*, 1463-1472.

(4) Nose, S. *J Chem Phys* **1984**, *81*, 511-519.

(5) Parrinello, M.; Rahman, A. *Phys Rev Lett* **1980**, *45*, 1196-1199.

(6) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angewandte Chemie International Edition* **1999**, *38*, 236-240.

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

(8) Killian, B. J.; Kravitz, J. Y.; Somani, S.; Dasgupta, P.; Pang, Y. P.; Gilson, M. K. *Journal of molecular biology* **2009**, *389*, 315-335.

(9) Shannon, C. E. *M D Comput* **1997**, *14*, 306-317.

(10) Steuer, R.; Kurths, J.; Daub, C. O.; Weise, J.; Selbig, J. *Bioinformatics* **2002**, *18*, S231-S240.

(11) Pandini, A.; Fornili, A.; Fraternali, F.; Kleinjung, J. *Faseb J* **2012**, *26*, 868-881.

(12) McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. *Journal of Chemical Theory and Computation* **2009**, *5*, 2486-2502.

(13) Grossfield, A.; Feller, S. E.; Pitman, M. C. *Proteins* **2007**, *67*, 31-40.

(14) Warne, T.; Edwards, P. C.; Leslie, A. G. W.; Tate, C. G. *Structure* **2012**, *20*, 841-849.