Supporting information for

"Title"

Exploring the activation process of the β2AR-G_s complex

"Authors"

Chen Bai[1,2]*, Junlin Wang[2], Dibyendu Mondal[1], Yang Du[2], Richard D Ye[2], Arieh Warshel[1]*

[1.]Department of Chemistry, University of Southern California, Los Angeles, CA 90089-1062, U.S.A.

[2.]School of Health and Life Sciences, The Chinese University of Hong Kong, 2001 Longxiang Road, Longgang District, Shenzhen 518172, China

Constructing the assemblies

In this work we used experimental crystal structures together with homology modeling technique to model the G_s^{GDP}, the experimental intermediate state, and the GDP free open-in state used for calculations. The G_s part of G_s^{GDP} structure comes from chain A of XRD structure with PDB ID 6EG8[1]. Since this structure only has the G_s^{GDP} part, to model the β2AR-G_s^{GDP} complex, we obtained the activated β2AR part from X-ray diffraction (XRD) structure with PDB ID: 3SN6[2]. For the intermediate structure, for the interaction and orientation between β2AR and the last 21 residues on α5 we used XRD structure 6E67[1] (the β2AR part of this structure and 3SN6 are almost identical), the rest part of the G protein comes from PDB ID 6EG8. For the GDP free open-in structure, we used XRD structure with PDB ID: 3SN6.

After obtaining the three major endpoint states of the system, we used targeted MD to generate the intermediate conformations between each pair of them. Next, we add membrane particles around the transmembrane bundle and perform relaxation runs on each structure until the energy is converged. We trim the structures into coarse grained (CG) representation and evaluate their CG free energies to obtain the conformational free energy change profile. For each intermediate structure we perform series of docking calculates on GDP + Mg^{2+}. From nucleotide binding pocket to the bulk we get 25 binding free energies at a 1 Å interval. Combining these set of data and the conformational free energy change we could obtain the free energy map that is shown in main text Figure 3. The mutational free energy change is calculated by the differences between the solvation free energy of a specific residue after it was mutated into Alanine. The

details of each method used in this work is described in details below. All calculations are performed by Molaris-XG package[3, 4].

Coarse-Grained (CG) Model

The coarse-grained model used in this work is one of most reliable models for studying complex protein system like membrane proteins. In this CG models the sidechain of a protein residue is represented as a simplified united atom, whereas the main chain atoms are represented explicitly as shown in Figure S2. The simplified united atom is generally placed at the mass center of the sidechain (for polar and nonpolar residues) or at the center of the charged group of a residue (for ionizable residues). In our CG model, the solvent is treated implicitly and to represent a membrane a grid of effective atoms is used. A consistent treatment of the electrostatic free energy is a key factor for the success of this CG model in explaining the energetics of many complicated and large biological systems.[5-7]

The overall CG free energy is given by the expression:

$$\Delta G_{\text{fold}} = \Delta G_{\text{main}} + \Delta G_{\text{side}} + \Delta G_{main/side} = c_1 \Delta G_{\text{side}}^{\text{vdw}} + c_2 \Delta G_{\text{solv}}^{\text{CG}} + c_3 \Delta G_{\text{HB}}^{\text{CG}} + \Delta G_{\text{side}}^{\text{elec}} + \Delta G_{\text{side}}^{\text{polar}} + \Delta G_{\text{side}}^{\text{hyd}} + + \Delta G_{main/side}^{vdw} \quad \text{(S1)}$$

where $\Delta G_{\text{solv}}^{\text{CG}}$ and $\Delta G_{\text{HB}}^{\text{CG}}$ stand for the main-chain solvation and hydrogen-bond energy, and $\Delta G_{\text{side}}^{\text{elec}}$, $\Delta G_{\text{side}}^{\text{polar}}$, $\Delta G_{\text{side}}^{\text{hyd}}$, $\Delta G_{\text{side}}^{\text{vdw}}$ and $\Delta G_{\text{main/side}}^{\text{vdw}}$ are the electrostatic, polar, hydrophobic, van der Waals component of the side-chain energy and the main chain/side chain van der Waals contribution, respectively. The constants c_1, c_2 and c_3 are scaling coefficients, taking values of 0.10, 0.25 and 0.15, respectively in the current implementation.

While the electrostatic, van der Waals and all bonding interactions for all the main chain atoms are calculated using our regular ENZYMIX force field,[4] the side-chain contributions are calculated with special functional forms. The electrostatic contributions of the side-chain atoms are obtained using eq. S2.

$$\Delta G_{\text{side}}^{\text{elec}} = -2.3RT \sum_i Q_i^{\text{MC}}\left(pKa_i^i - pKa_i^{\text{w}}\right) + \Delta G_{QQ}^{(\text{f})} - \Delta G_{QQ}^{(\text{uf})} + \Delta G_Q^{\text{dev}} \qquad \text{(S2)}$$

In eq. S2, pKa_i^{w} is the *pKa* of the *i*th ionizable residue in water and Q_i^{MC} denotes the Monte Carlo (MC) averaged charge of the *i*th ionizable residue. The interaction potentials that occur between the ionized groups in unfolded and the folded states of the protein are denoted by $\Delta G_{QQ}^{(\text{uf})}$ and $\Delta G_{QQ}^{(\text{f})}$, respectively. In order to estimate the *pKa* of the *i*th residue in the protein environment (pKa_i^i) we employ the Monte Carlo Proton Transfer (MCPT) Method. In the MCPT simulation, at each MC step a proton transfer between a random pair of ionizable residues or an ionizable residue and the bulk solvent is attempted. This procedure continues until the electrostatic free energy of the folded protein converges and the charge configuration of the folded protein at that MC step is accepted as protein configuration to calculate the CG free energy. The last term in eq. S2 is a correction term that takes care of the change in the ionization of the side chains upon the unfolding process.

The pKa_i^i of the *i*th residue in eq. S2 is calculated using:

$$pKa_i^i = pKa_i^{\text{w}} - \frac{\text{sgn}\left(Q_i^{\text{ion}}\right)}{2.3RT}\Delta G_{\text{self},i} \qquad \text{(S3)}$$

where $\text{sgn}\left(Q_i^{\text{ion}}\right)$ represents the sign function of the charge of the *i*th residue in its ionized form (and its values are: +1 for LYS, HIS, ARG, and -1 for ASP, GLU). The term $\Delta G_{\text{self},i}$ in eq. S3, denotes the change in self-energy of moving an ionizable residue from water to the protein. This

self-energy term is a key to the reliable representation of the electrostatic effect of a protein system. $\Delta G_{\mathrm{self},i}$ can be expressed as:

$$\Delta G_{\mathrm{self},i} = \sum_i \left[U_p^{\mathrm{self}}\left(N_p^i\right) + U_{np}^{\mathrm{self}}\left(N_{\mathrm{np}}^i\right) + U_{\mathrm{mem}}^{\mathrm{self}}\left(N_{\mathrm{mem}}^i\right) \right] \quad \text{(S4)}$$

Where in eq. S4 the index i runs over all ionizable residues and N_{p}^i, N_{np}^i, and N_{mem}^i are the numbers of polar and nonpolar residues, and the number of membrane grid nodes surrounding the i^{th} residue. U designates an effective potential and the explicit function form of this potential can be found in ref. [8].

Along with the electrostatic the polar and hydrophobic contributions of the side chain are also evaluated using effective potentials with functional form like the ones used to calculate the self-energies of the ionizable residues (see ref. [8]). The details of the functional forms and other CG parameters in the energy functions can be found in ref. [8]. In this study we used a membrane model for which the hydrophobic contributions are scaled down by a factor of ~3.6 and the polar contributions of sidechains are not included. This model is called membrane model 1. This model has been shown to provide a good agreement for folding free energies of several membrane-associated peptides.[8]

Semi-microscopic version of Protein Dipole Langevin Dipole method (PDLD/S-LRA/β) of solvation free energy and Binding Free energy calculation

Scaled Protein Dipole Langevin Dipole (PDLD/S) is a method to calculate the electrostatic energies of a system in a semi-microscopic level.[9] In general, microscopic models produce accurate representation of a simulation system, but with large compensating contributions in

energy calculations, it is hard to obtain accurate results from these models. On the other hand, microscopic models produce accurate (may not be precise) results by implicitly assuming the compensating contributions by using large dielectric constant. We wanted to keep the benefits of both models (microscopic and macroscopic) and therefore a scaling factor is used with microscopic PDLD calculations to implicitly consider the compensating contributions. In PDLD/S model the solvent molecules are represented by a grid of Langevin dipoles (LD). The LD representation of solvent considers average polarization of the solvent, whereas to use averaged protein configurations multiple snapshots of protein configurations are taken from long molecular dynamics simulation. That leads to a consistent PDLD/S calculation. The averaging in PDLD/S is performed using linear response approximation (LRA). Furthermore, we utilize the linear interaction method (LIE) to approximates the non-electric binding contribution by a scaled vdW term. Thus, the overall method is called PDLD/S-LRA/β. In this work, we have used the PDLD/S-LRA/β to calculate the solvation free energies of different residues in their protein environment. The details of the method are discussed in many of our previous works.[4, 10] We used the POLARIS[4] module of Molaris-XG software [11] to calculate the solvation free energy of the C-terminal residues of the G-proteins.

The following equation represents the free energy change

$$\Delta G_{bind}^{LRA/\alpha} = \frac{1}{2}\left(\langle U_{elec,l}^{p}\rangle_{l} - \langle U_{elec,l}^{w}\rangle_{l}\right) + \frac{1}{2}\left(\langle U_{elec,l}^{p}\rangle_{l'} - \langle U_{elec,l}^{w}\rangle_{l'}\right) + \beta\left(\langle U_{vdW,l}^{p}\rangle_{l} - \langle U_{vdW,l}^{w}\rangle_{l}\right) \quad \text{(S5)}$$

where $\langle U_{elec,l}^{p}\rangle_{l}$ is the electrostatic contribution for the interaction between the ligand and its surroundings, p and w designate protein and water, respectively, and ℓ and ℓ′ designate the

ligand in its actual charged form and the "non-polar" ligand, where all of the residual charges are set to 0. Previously, this method was called LRA/α, but since the new element is the β term relative to our original LRA treatment, we prefer the name LRA/β[12].

"Targeted Molecular Dynamics" for generating intermediate configurations between two end-states of a simulation system

The targeted molecular dynamics (TMD) simulation is used to generate intermediate conformations of a protein between two given end point conformations of the same protein. In TMD, a mapping relaxation simulation is performed where the system is forced to remain close to a constraint coordinate $r_{0,m}$ at each mapping frame (m). The constraint coordinate is denoted by:

$$r_{0,m} = r_{m-1}^{\text{avg}} + \lambda\left(r_{\text{trgt}} - r_{m}^{\text{avg}}\right) \tag{S7}$$

In eq. S7, r_{m-1}^{avg} is the average co-ordinate of the protein system from previous mapping window (m-1) and r_{trgt} denotes the co-ordinate of the target protein structure. The term λ in eq. S7 is the mapping parameter whose value varies between 0 to 1. In the case of the initial frame, r_{m-1}^{avg} is the co-ordinate of initial structure of the protein. In order to constrain the co-ordinate of a system, a simple quadratic energy function (eq. S8) is used in TMD. If the system is constrained close to a reference coordinate $r_{0,m}$ at a relaxation mapping frame m, then the energy function takes a form,

$$E_{\text{TMD}} = k\left(r - r_{0,m}\right)^2 \tag{S8}$$

In eq. S8, *r* denotes the instantaneous coordinate of the system at the relaxation mapping frame *m*. Whereas, *k* is the applied force constant. In this work, 151 frames of constraint relaxation runs were performed using k = 0.5 kcal/(mol×Ang^2) to drive the system from r_{init} to r_{trgt}.

Plasmid construction

(1) pcDNA3.1 β2AR construction: The β2AR gene was amplified by Polymerase Chain Reaction (PCR) from pFastBac1 β2AR using β2AR Forward and β2AR Reverse primer (Genewiz, Suzhou, China) (Table S1). The fragment of β2 AR was digested by restriction enzymes Hind III (NEB, Beverly, MA, USA) and XhoI (NEB, Beverly, MA, USA) and then cloned into pcDNA3.1 vector.

(2) pcDNA3.1 G_s WT (G_s wild type), Gs D381A, Gs M386A construction: For construction, pFastBac1 G_s WT was served as template. Fragments of G_s WT and Gs mutants were amplified by using following three primer pairs: G_s WT Forward, G_s WT Reverse; G_s WT Forward, G_s D381A Reverse; G_s WT Forward, G_s M386A Reverse (Genewiz, Suzhou, China) (Table S1). Then the three fragments were digested by Hind III and XhoI and cloned into pcDNA3.1 vector.

(3) pcDNA3.1 G_s C379A construction: pcDNA3.1 G_s WT was served as template. Two fragments were obtained by using following two primer pairs: G_s WT Forward, Gs C379A Reverse; Gs C379A Forward, pcDNA330R (table S1). The full length of Gs C379A fragment was obtained by overlap PCR using above mentioned two fragments. Then the Gs C379A fragment was digested by Hind III and XhoI and cloned into pcDNA3.1 vector.

Sequencing results show that the sequence of plasmid is right.

cAMP production in different cell lines

HeLa or HEK 293 cells were co-transfected with pcDNA3.1 β2AR and pcDNA3.1 G_s WT (Gs wild type) by using Lipofectamine™ 3000 (Invitrogen, Carlsbad, CA, USA). After 24h incubation, 5μL (total 20000 cells) cell solution and 5μL different concentrations of ISO were added to 384-well plate (OptiPlate-384, #6007290, PerkinElmer, Waltham, MA, USA) and incubated for 40min at 37℃. The cAMP levels were detected according the cAMP kit protocol (LANCE cAMP ultra assay kit ,Cat # TRF0263;PerkinElmer, Inc, Waltham, MA，USA). Briefly, 5μL 4X Eu-cAMP tracer working solution and 5μL 4X Ulight-anti-cAMP working solution were added into the cell solution in 384-well plate and incubated for 60 min at 37℃. TR-FRET ratio(665nm/615nm) was detected by Envision Multilabel Plate Reader (PerkinElmer, Inc, Waltham, MA，USA). The results were shown in Fig. S3.

The results show that there was no significant difference in different concentrations ISO treatment in HEK 293 cells. In other word, ISO has weak ability to induce cAMP production in β2 AR and Gs WT co-transfected HEK 293 cells. However, ISO can induce cAMP production in HeLa cells (The Y axis is TR-FRET ratio 665nm/615nm, cAMP production increased as the ratio of 665nm/615nm decreased). Therefore, HeLa cells were used for cAMP assay

Effect of different cell densities on cAMP production

HeLa cells were co-transfected with pcDNA3.1 β2 AR and pcDNA3.1 G_s wild type (G_s WT) or G_s mutants（D381A, M386A and C379A） by using Lipofectamine™ 3000. After 24h incubation, 5μL (total 3000 cells or 6000 cells) cell solution and 5μL different concentrations of ISO were added to 384-well plate and incubated for 40min at 37℃. The cAMP levels were

detected according the cAMP kit protocol. TR-FRET ratio (665nm/615nm) was detected by Envision Multilabel Plate Reader. The results were shown in Fig. S4.

The results show that the TR-FRET ratio (665nm/615nm) in 3000 cell group is higher than 6000 cell group. Therefore, the lower cell densities (3000cell/well) were used for follow-up experiments.

References

1. Liu, X.; Xu, X.; Hilger, D.; Aschauer, P.; Tiemann, J. K. S.; Du, Y.; Liu, H.; Hirata, K.; Sun, X.; Guixà-González, R.; Mathiesen, J. M.; Hildebrand, P. W.; Kobilka, B. K., Structural Insights into the Process of GPCR-G Protein Complex Formation. *Cell* **2019,** *177* (5), 1243-1251.e12.
2. Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. A.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K., Crystal structure of the β2 adrenergic receptor–Gs protein complex. *Nature* **2011,** *477* (7366), 549-555.
3. Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A., Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems. *Annual Review of Physical Chemistry* **2011,** *62* (1), 41-64.
4. Lee, F. S.; Chu, Z. T.; Warshel, A., Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. *Journal of Computational Chemistry* **1993,** *14* (2), 161-185.
5. Rychkova, A.; Warshel, A., Exploring the nature of the translocon-assisted protein insertion. *P Natl Acad Sci USA* **2013,** *110* (2), 495-500.
6. Bai, C.; Warshel, A., Revisiting the protomotive vectorial motion of F-0-ATPase. *P Natl Acad Sci USA* **2019,** *116* (39), 19484-19489.
7. Alhadeff, R.; Vorobyov, I.; Yoon, H. W.; Warshel, A., Exploring the free-energy landscape of GPCR activation. *P Natl Acad Sci USA* **2018,** *115* (41), 10327-10332.
8. Vorobyov, I.; Kim, I.; Chu, Z. T.; Warshel, A., Refining the treatment of membrane proteins by coarse-grained models. *Proteins* **2016,** *84* (1), 92-117.
9. Russell, S. T.; Warshel, A., Calculations of Electrostatic Energies in Proteins - the Energetics of Ionized Groups in Bovine Pancreatic Trypsin-Inhibitor. *J Mol Biol* **1985,** *185* (2), 389-404.
10. Schutz, C. N.; Warshel, A., What are the dielectric "constants" of proteins and how to validate electrostatic models? *Abstr Pap Am Chem S* **2002,** *223*, C75-C75.
11. Warshel, A.; Chu, Z.; Villa, J.; Strajbl, M.; Schutz, C.; Shurki, A.; Vicatos, S.; Plotnikov, N.; Schopf, P., Molaris-XG, v 9.15. *University of Southern California: Los Angeles* **2012**.
12. Singh, N.; Warshel, A., Absolute binding free energy calculations: on the accuracy of computational scoring of protein-ligand interactions. *Proteins* **2010,** *78* (7), 1705-23.

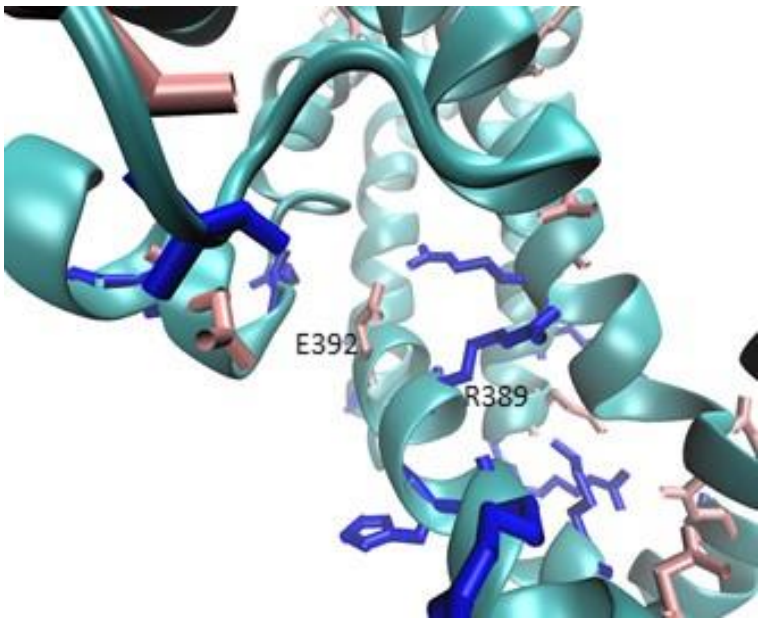

Fig. S1. Positive charged (blue) and negative charged (pink) residues around E392 and R389 on α5.

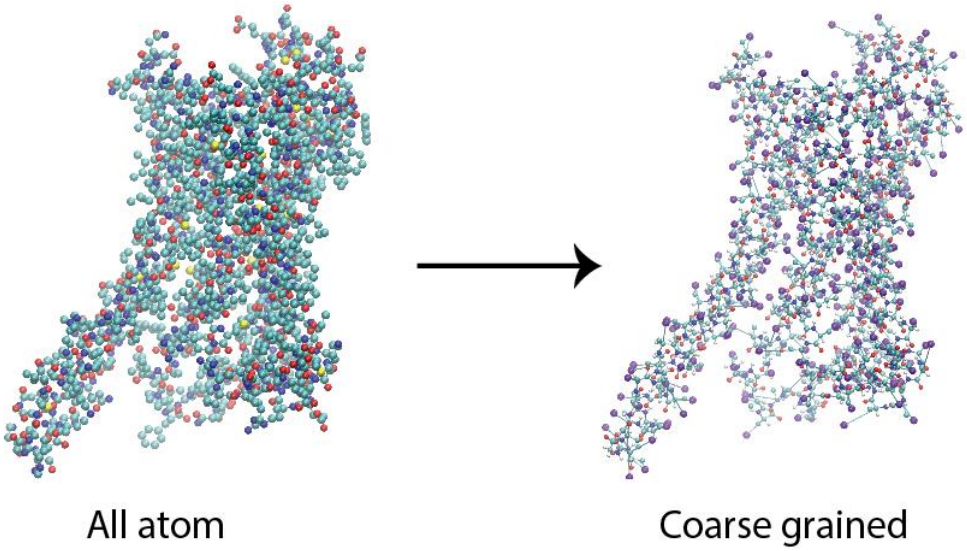

Fig. S2. A visual presentation of the CG model used in this study. The β2AR of PDB 3SN6 is shown in an all-atom CPK representation in the left and in a CG CPK representation on the right. In the CG model, the CB represent an entire sidechain of a residue (shown in violet) in the CG model (see text for details). The atoms are colored cyan, red, blue and yellow for carbon (and CB in all-atom presentation), oxygen, nitrogen and sulfur, respectively.

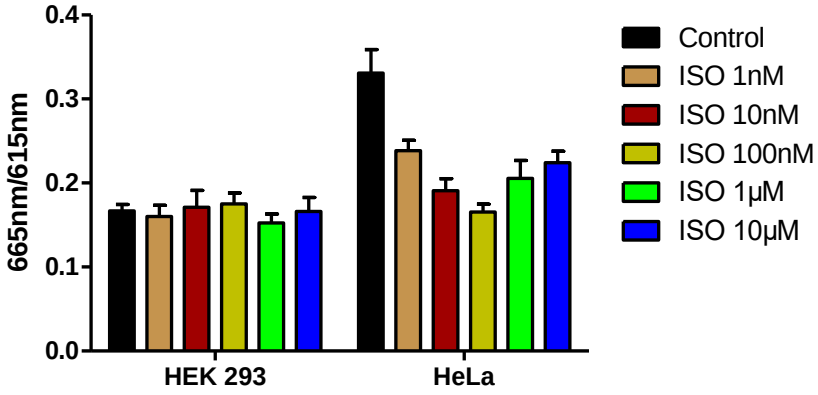

Fig.S3 cAMP production in different cell lines. HeLa or HEK 293 cells were co-transfected with β2 AR and Gs WT and treated with different concentrations of ISO. TR-FRET ratio (665nm/615nm) was detected by Envision Multilabel Plate Reader.

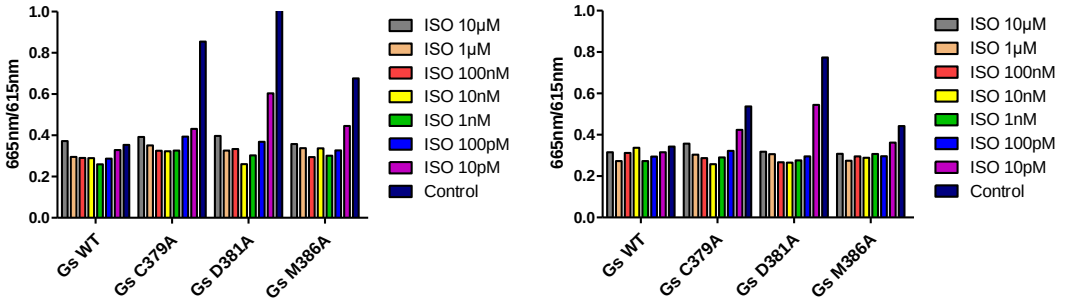

Fig. S4. Effect of different cell densities on cAMP production. HeLa cells (cell densities of the left figure:3000 cells/well, cell densities of right figure: 6000 cells/well) were co-transfected with β2 AR and either Gs WT or one of the Gs mutants, then treated with different concentrations of ISO, TR-FRET ratio (665nm/615nm) was detected by Envision Multilabel Plate Reader.

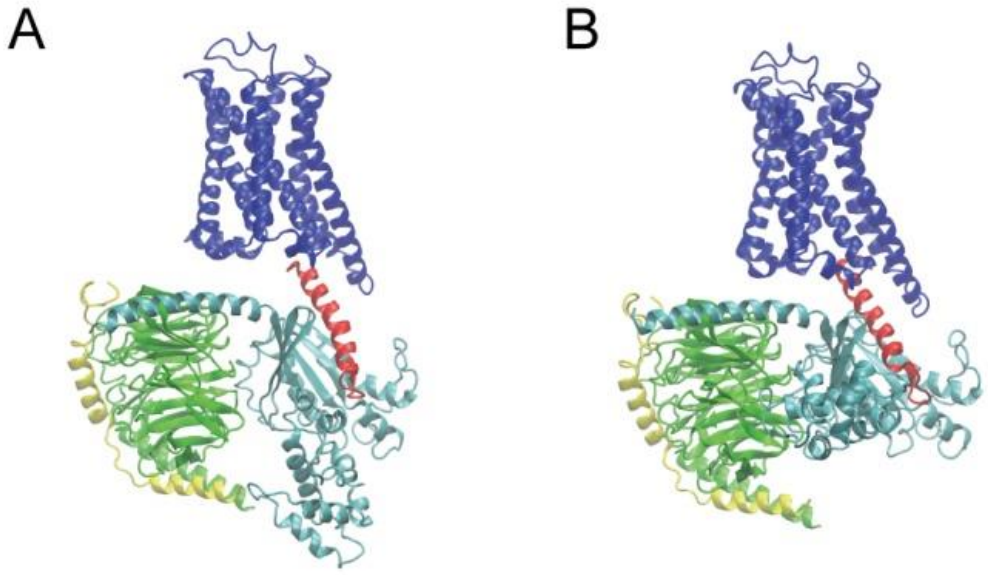

Fig. S5. The two structures that correspond to point A and B on the free energy surface of Fig 3. Color code is same as in Fig 2. Note the large conformation change at the α domain (cyan).

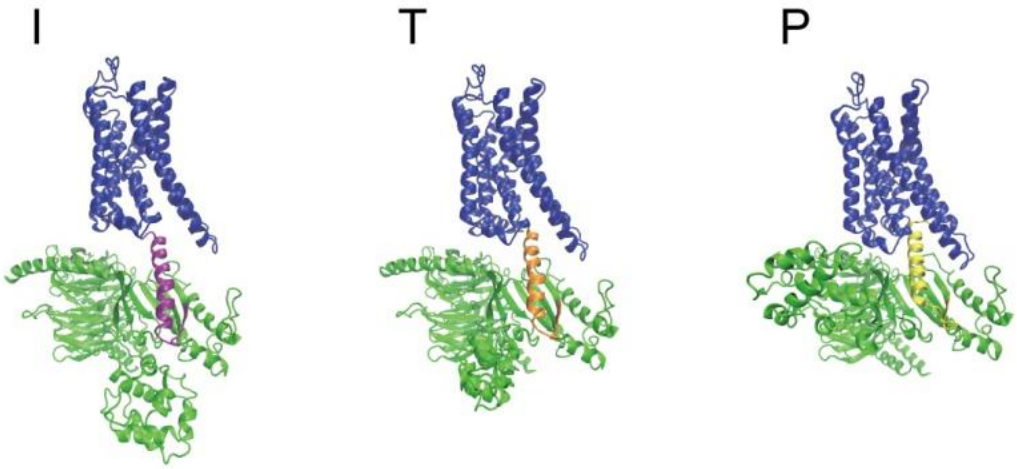

Fig. S6. Structures of the I, T, and P state of Fig 2. α 5 is colored in purple, orange, and yellow in the three states, respectively. Note the intrusion of α 5, the conformational change of Gs protein, and the shortening of the distance between β 2AR receptor and Gs during the conversion process.

Table S1：Primer information

Primer name	Primer sequence
β_2 AR Forward	CCCAAGCTTGCCACCATGGGGCAACCCGGGAA
β_2 AR Reverse	CCGCTCGAGTTACAGCAGTGAGTCATTTGTACT
Gs WT Forward	CCCAAGCTTGCCACCATGGGCTGCCTCGGGAAC
Gs WT Reverse	CCGCTCGAGTTAGAGCAGCTCGTACTGACGA
Gs C379A Forward	AACGACGCCCGTGACATCATTCAGCGCATG
Gs C379A Reverse	GATGTCACGGGCGTCGTTGAACACACGGCG
Gs D381A Reverse	CCGCTCGAGTTAGAGCAGCTCGTACTGACGAAGGTGCAT GCGCTGAATGATCGCACG
Gs M386A Reverse	CCGCTCGAGTTAGAGCAGCTCGTACTGACGAAGGTGCG CGCG
pcDNA330R	CCACACCCGCCGCGCTTAAT