Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

The Molecular Mechanism of $P2Y_1$ Receptor Activation

Shuguang Yuan, H. C. Stephen Chan, Horst Vogel, Slawomir Filipek, Raymond C. Stevens,* and Krzysztof Palczewski**

anie_201605147_sm_miscellaneous_information.pdf

Supporting Information

Methods

Loop filling and refinements. Since the intracellular loop ICL2 for each structure was removed for the insertion of a fusion protein, the loop refinement protocol in Modeller[1] V9.10 was used to reconstruct and refine this region. A total of 20000 loops were generated for each receptor and the conformation with the lowest DOPE (Discrete Optimized Protein Energy) score was chosen for receptor construction. Repaired models were submitted to Rosetta V3.4 for loop refinement with kinematic loop modeling methods[2]. Kinematic closure (KIC) is an analytic calculation inspired by robotics techniques for rapidly determining possible conformations of linked objects subject to constraints. In the Rosetta KIC implementation, 2N - 6 backbone torsions of an N-residue peptide segment (called non-pivot torsions) were set to values drawn randomly from the Ramachandran space of each residue type, and the remaining 6 phi/psi torsions (called pivot torsions) were solved analytically by KIC.

Protein structure preparations. All protein models were prepared in Schrodinger suite software under the OPLS_2005 force field[3]. Hydrogen atoms were added to the repaired crystal structures at physiological pH (7.4) with the PROPKA[4] tool to optimize the hydrogen bond network provided by the Protein Preparation tool in Schrodinger. Constrained energy minimizations were carried out on the full-atomic models, with heavy atom coverage to 0.4 Å.

Ligand structure preparations. All ligand structures were obtained from the PubChem[5] online database. The LigPrep module in Schrodinger 2015 suite software was introduced for geometric optimization by using the OPLS_2005 force field. The ionization state of ligands was calculated with the Epik[6] tool employing Hammett and Taft methods in conjunction with ionization and tautomerization tools[6].

Molecular dynamics simulations. The membrane system was built in Schrodinger Desmond[7] GUI, with the receptor crystal structure pre-aligned in the OPM (Orientations of Proteins in Membranes) database[8]. Pre-equilibrated 110 POPC lipids coupled with 92000 TIP3P water molecules in a box ~ 68 Å x 68 Å x 96 Å were used to build the protein/membrane system. We modeled the protein, lipids, water and ions using the CHARMM36 force field[9]. Ligands were assigned with CHARMM CgenFF forcefield[10]. Ligand geometry was submitted to the GAUSSIAN 09 program[11] for optimization at the Hartree-Fock 6-31G* level when generating forcefield parameters. The system was

gradually heated from 0 K to 310 K followed by a 1ns initial equilibration at constant volume with the temperature set at 310 K. Both the ligand molecule and backbone of the protein were restrained by a force constant of 10 kcal/(mol.Å^2) during this step. Next, an additional 40 ns restrained equilibration was performed at constant pressure and temperature (NPT ensemble; 310 K, 1 bar), and the force constant was trapped off gradually from 10 to 0 kcal/mol. All bond lengths to hydrogen atoms were constrained with M-SHAKE. Van der Waals and short-range electrostatic interactions were cut off at 10 Å. Long-range electrostatic interactions were computed by the Particle Mesh Ewald (PME) summation scheme. All MD simulations were done in Desmond[7]. The MD simulation results were analyzed in Gromacs [12] and VMD [13]. The solvent accessible surface area was calculated in Gromacs.

The interacellular transmembrane helix distances (TM3-TM6, TM6-TM7 and TM7-TM3) were caculated by the mass center distance of each investigated interacellular region. Specifically, residues 145-153 defined TM3, residues 257-265 defined TM6 and residues 318-325 defined TM7. Calculations were done in Gromacs.

Lipid order calculation. The lipid order calculation was done by MEMBPLUGIN, a plugin in VMD. The hydrophobic tail structure of lipids belonging to the bilayer is intimately linked to certain structural properties such as bilayer condensation, membrane fluidity and membrane thickness. One way to evaluate hydrophobic tail arrangement is to assess the order of these hydrophobic tails in terms of the lipid order parameter, -S_{CH}. This parameter quantifies the order of phospholipid hydrocarbon tails by averaging (ensemble and time) the angle, θ, for each C-H bond with respect to the bilayer normal, independently for every CH_2 group in each lipid acyl tail over a given lipid molecule in the membrane. The -S_{CH} is calculated for each methyl group by the following equation:

$$-S_{CH} = -\left\langle \frac{3cos^2\theta - 1}{2} \right\rangle$$

The angular brackets indicate ensemble and time averaging. Lipid order parameters can be measured experimentally in form of deuterium order parameters S_{CD} by NMR spectroscopy if lipid hydrocarbon chains are selectively deuterated[14]. To calculate average S_{CH}, the fatty acid carbons of each lipid species in the atom selection are classified by an index that corresponds to the atom position on the fatty acid chain. The

average S_{CH} is computed from the S_{CH} of each one of the hydrogen atoms bonded to a specific carbon. The sum of average S_{CH} of all carbons with the same index for all the simulation frames is then computed (without distinction between different lipid species, fatty acid chain stereochemical numbering or frame). This sum is divided by the (number_of_atoms_with_this_index * num_of_frames) resulting in the S_{CH} shown in the plot. Fully detailed computations can be found in the output files. Since most lipids in the simulated system contacted the surface of the receptor and each lipid diffused very frequently in and out during the simulation, we included all lipids during the calculations.

Average water density calculation. Water density was calculated in Volmap plugin in VMD [13]. The protein backbone of each frame was superimposed prior to the calculations. Volmap creates a map of the weighted atomic density at each grid point. This is done by replacing each atom in the selection with a normalized Gaussian distribution of width (standard deviation) equal to its atomic radius. The Gaussian distribution for each atom is then weighted by using an optional weight read from one of the atoms' numerical properties, and defaults to a weight of one. The various Gaussians are then additively distributed on a grid. The meaning of final map will depend of the weights of mass. The average water density was calculated based on the final 0.1 μs frames of each long time scale MD simulation. Final output results were visualized in VMD.

Interaction fingerprint calculations. The interaction fingerprint between protein and ligand was done with IChem[15]. We first extracted 500 snapshots from the final 0.1 μs MD simulation, and then submitted them to IChem. IChem can convert protein−ligand coordinates into a fingerprint (TIFP) of 210 integers registering the corresponding molecular interaction pattern. TIFP fingerprints have been calculated for ca. 1000 protein−ligand complexes, enabling a broad comparison of relationships between interaction pattern similarities and ligand or binding site paired similarities. The ligands of $P2Y_1R$ in this work displayed only two different types of interactions: an ionic interaction and a hydrophobic contact. We defined the calculation zone within 6 Å of the ligand mass center and kept the default parameters of IChem in our calculations.

Residue communication network calculations[16]**.** Correlated atomic fluctuations of a particular receptor state were characterized with Bio3DT as reported elsewhere[16-18]. The network nodes represent residues, which are connected through edges weighted by their

constituent atomic correlation values. Community analysis and node centrality with Bio3D and suboptimal path calculation with WISP software [19] were performed on each network to characterize its properties and identify residues involved in the dynamic coupling of distal sites. Parameters for the suboptimal path analysis included input source and sink nodes, as well as the total number of paths to be calculated. The last parameter was set to 500 paths, which was found to yield converged results in all cases [16].

Table S1. Simulated systems of $P2Y_1R$.

GPCR-ligand	Ligand type	Continuous water pathway	State	Simulation time (μs)
$P2Y_1R$*-BPTU[1]	antagonist	no	inactive	2x2
$P2Y_1R$*-ADP[1]	agonist	yes	active	3x2
$P2Y_1R$-MRS2500[2]	antagonsit	no	inactive	2x2
$P2Y_1R$-ADP[2]	agonist	yes	active	3x2

[1]Based on the crystal structure of $P2Y_1R$ with bound BPTU (PDB: 4XNV). BPTU: 1-(2-(2-(tert-butyl)phenoxy)pyridin-3-yl)-3-(4-(trifluoromethoxy)phenyl)urea.

[2]Based on the crystal structure of $P2Y_1R$ with bound MRS2500 (PDB: 4XNW). MRS2500: (1′R,2′S,4′S,5′S)-4-(2-Iodo-6-methylaminopurin-9-yl)-1-[(phosphato)methyl]-2(phosphato)bicycle[3.1.0]-hexane.

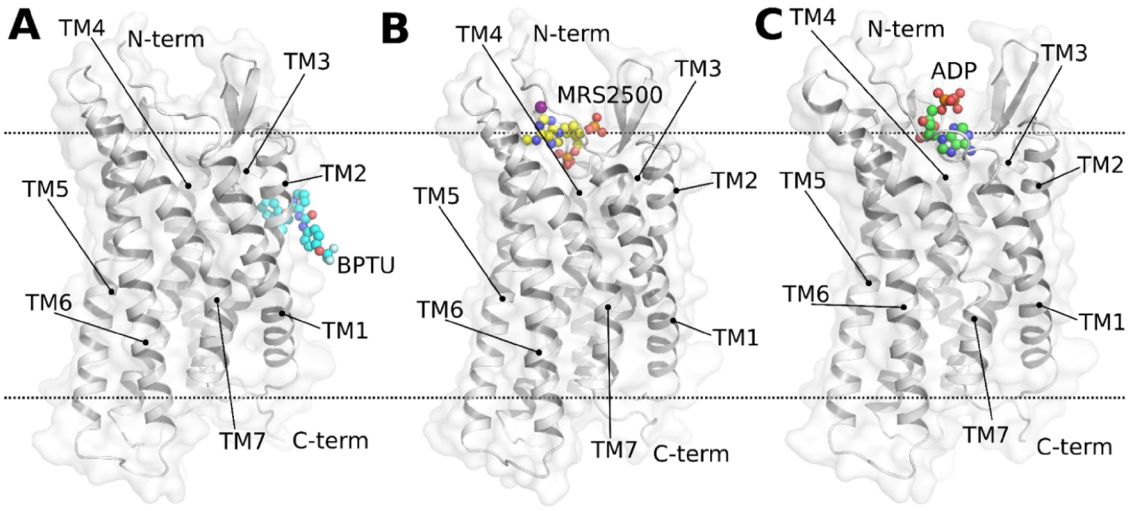

Figure S1. Locations of $P2Y_1R$ ligands at the end of MD simulations. The receptor structure (grey cartoon) in each panel originates from the same initial crystal structure (pdb: 4XNW). (**A**) $P2Y_1R$ bound with antagonist BPTU. (**B**) $P2Y_1R$ bound with antagonist MRS2500. (**C**) $P2Y_1R$ bound with agonist ADP.

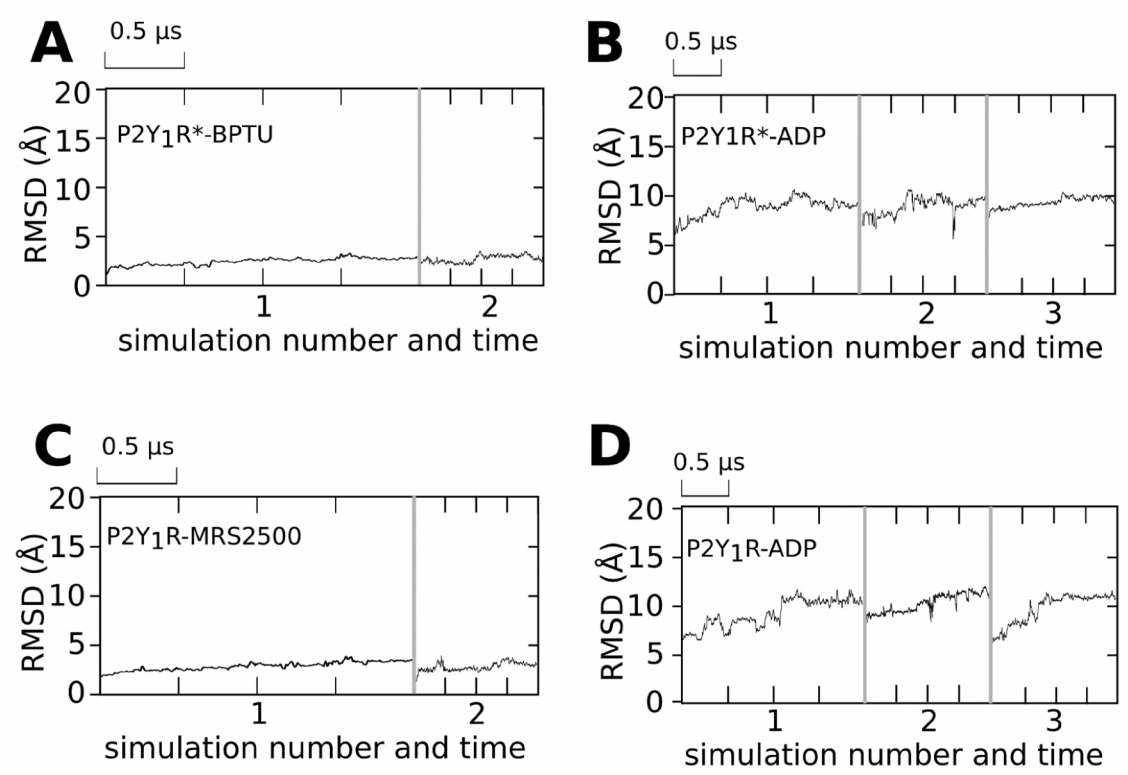

Figure S2. The root-mean-square deviation (RMSD) of the atomic position of the ligands for (**A**) $P2Y_1R$*-BPTU, (**B**) $P2Y_1R$*-ADP, (**C**) $P2Y_1R$-MRS2500, and (**D**) $P2Y_1R$-ADP.

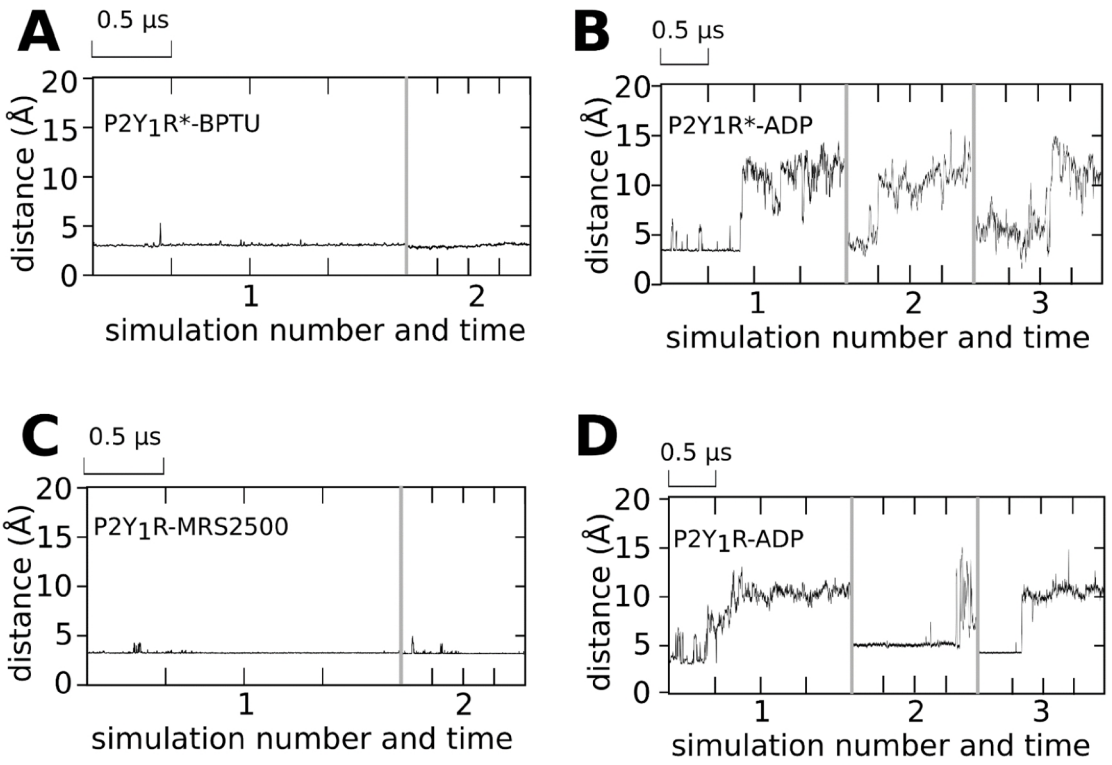

Figure S3. Ion lock distances between $K46^{1.46}$ and $R195^{ECL2}$ of (**A**) $P2Y_1R$*-BPTU, (**B**) $P2Y_1R$*-ADP, (**C**) $P2Y_1R$-MRS2500, and (**D**) $P2Y_1R$-ADP.

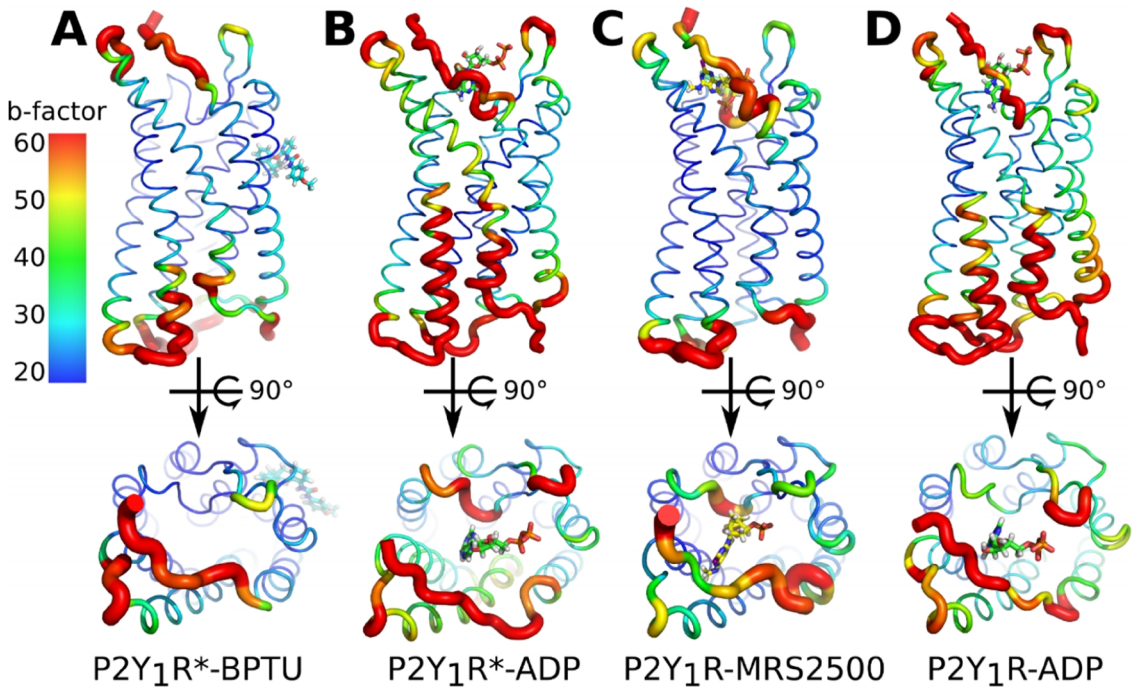

Figure S4. The b-factors (Å^2) of P2Y$_1$R. (**A**) P2Y$_1$R*-BPTU. (**B**) P2Y$_1$R*-ADP. (**C**) P2Y$_1$R-MRS2500. (**D**) P2Y$_1$R-ADP.

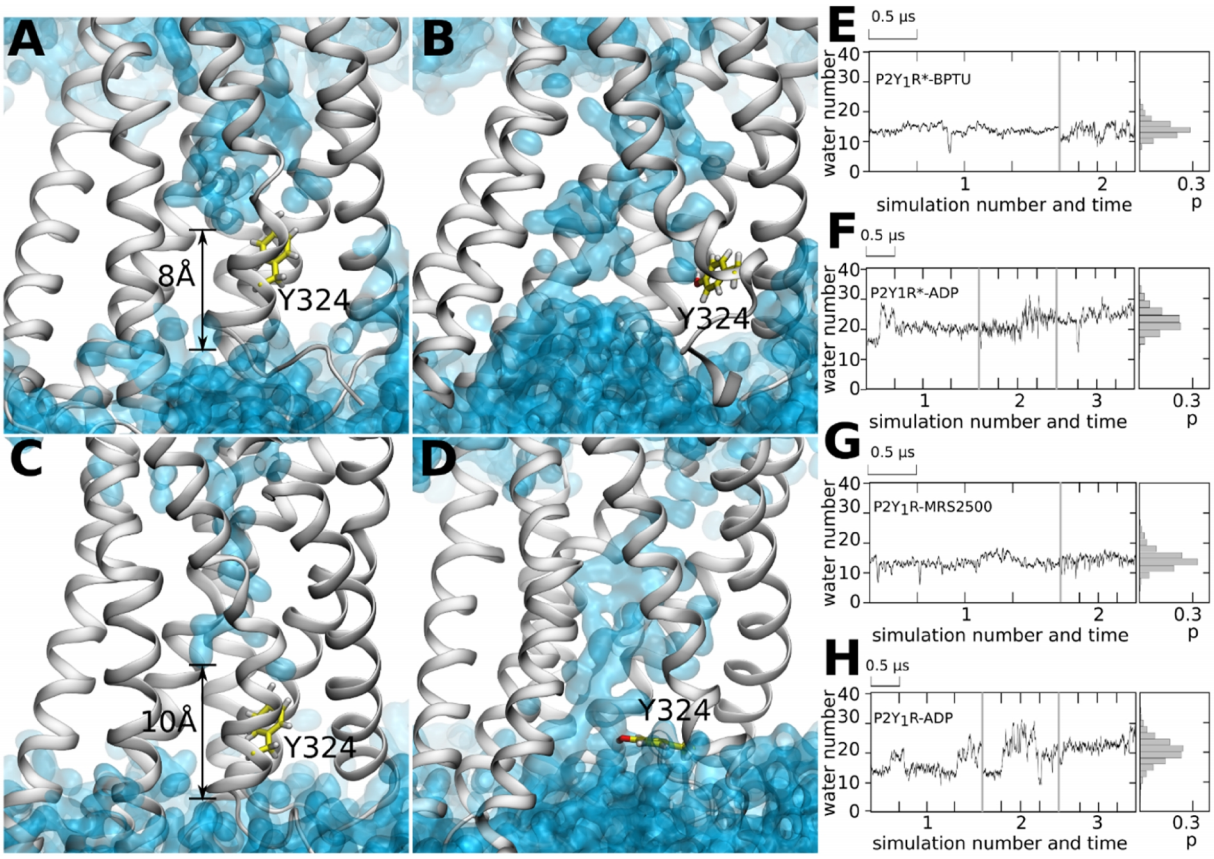

Figure S5. The number of water molecules next to Y324$^{7.53}$. (**A-D**) Water density during the final 0.1 µs MD simulations of P2Y$_1$R*-BPTU (**A**), P2Y$_1$R*-ADP (**B**), P2Y$_1$R-MRS2500 (**C**), and P2Y$_1$R-ADP (**D**). (**E-H**) Numbers of water molecules within 5 Å of Y324$^{7.53}$ in P2Y$_1$R*-BPTU (**E**), P2Y$_1$R*-ADP (**F**), P2Y$_1$R-MRS2500 (**G**), and P2Y$_1$R-ADP (**H**).

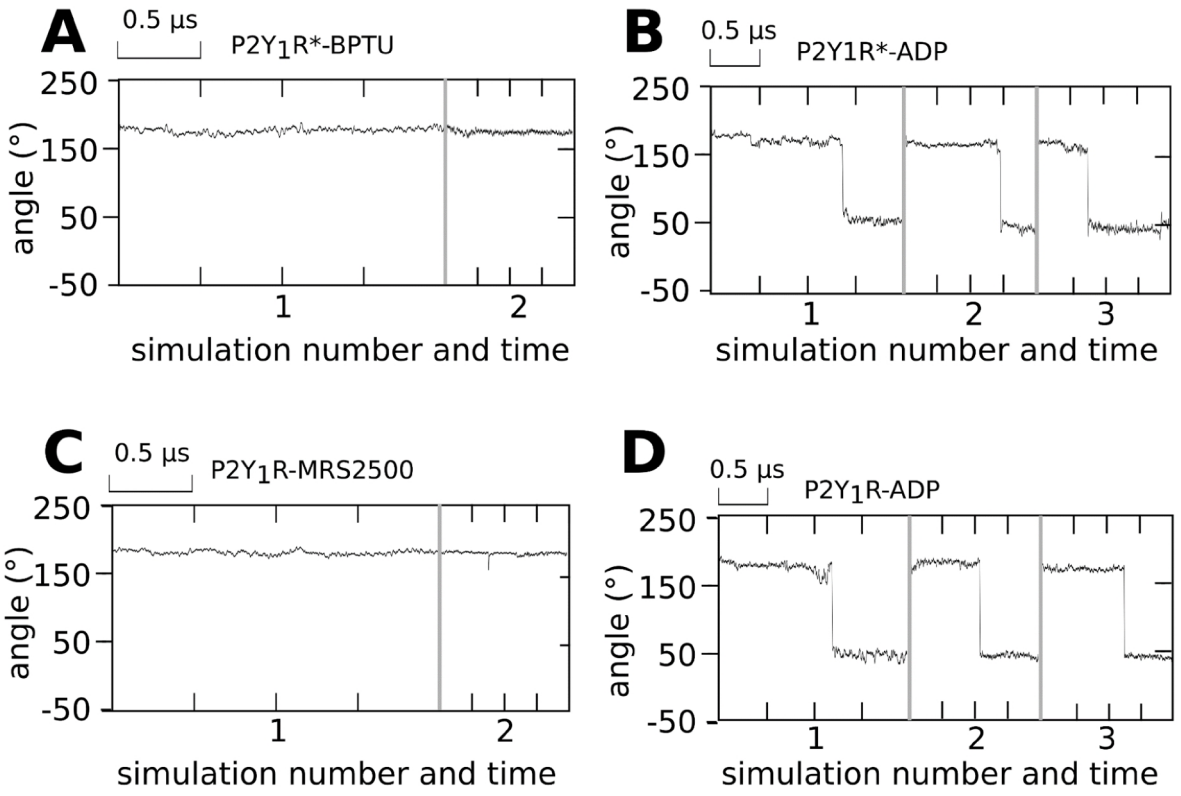

Figure S6. The X_1 angles of Y324$^{7.53}$. (**A)** $P2Y_1R$*-BPTU, (**B**) $P2Y_1R$*-ADP, (**C**) $P2Y_1R$-MRS2500, and (**D**) $P2Y_1R$-ADP.

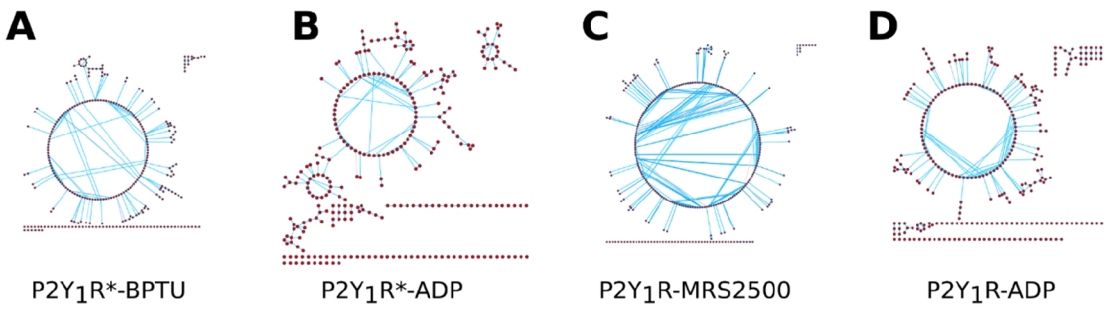

Figure S7. Residue interaction networks of the $P2Y_1R$. (**A**) $P2Y_1R$*-BPTU, (**B**) $P2Y_1R$*-ADP, (**C**) $P2Y_1R$-MRS2500, and (**D**) $P2Y_1R$-ADP. Each residue is represented by a red dot. Line connections indicate contacts between residues (for details, see Methods section). The larger circles in the antagonist-bound sytems indicate multiple residue interactions. The smaller circles with more scatttered dots imply that interactions between side chains inside the receptor are disrupted by helix movements (Figure 3) with subsequent water influx (Figure S4).

References

[1] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, A. Sali, *Curr Protoc Protein Sci* **2007**, *Chapter 2*, Unit 2 9.

[2] D. J. Mandell, E. A. Coutsias, T. Kortemme, *Nat Methods* **2009**, *6*, 551.

[3] D. Shivakumar, J. Williams, Y. J. Wu, W. Damm, J. Shelley, W. Sherman, *Journal of Chemical Theory and Computation* **2010**, *6*, 1509.

[4] C. R. Sondergaard, M. H. M. Olsson, M. Rostkowski, J. H. Jensen, *J Chem Theory Comput* **2011**, *7*, 2284.

[5] Y. L. Wang, J. W. Xiao, T. O. Suzek, J. Zhang, J. Y. Wang, Z. G. Zhou, L. Y. Han, K. Karapetyan, S. Dracheva, B. A. Shoemaker, E. Bolton, A. Gindulyte, S. H. Bryant, *Nucleic Acids Res* **2012**, *40*, D400.

[6] J. R. Greenwood, D. Calkins, A. P. Sullivan, J. C. Shelley, *J Comput Aided Mol Des* **2010**, *24*, 591.

[7] R. C. Chow E, Bowers KJ, Dror RO, Hughes DH, *D. E. Shaw Research Technical Report DESRES/TR* **2008**, *1*.

[8] A. L. Lomize, I. D. Pogozheva, H. I. Mosberg, *J Chem Inf Model* **2011**, *51*, 918.

[9] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, R. W. Pastor, *J Phys Chem B* **2010**, *114*, 7830.

[10] K. Vanommeslaeghe, E. P. Raman, A. D. MacKerell, Jr., *J Chem Inf Model* **2012**, *52*, 3155.

[11] M. J. Frisch, et al , *Gaussian 09, Revision A.1*, Gaussian, Inc., Wallingford CT, **2009**.

[12] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, *Bioinformatics* **2013**, *29*, 845.

[13] W. Humphrey, A. Dalke, K. Schulten, *J Mol Graph Model* **1996**, *14*, 33.

[14] J. Seelig, A. Seelig, *Q Rev Biophys* **1980**, *13*, 19.

[15] G. Marcou, D. Rognan, *J Chem Inf Model* **2007**, *47*, 195.

[16] G. Scarabelli, B. J. Grant, *Biophys J* **2014**, *107*, 2204.

[17] L. Skjaerven, X. Q. Yao, G. Scarabelli, B. J. Grant, *BMC Bioinformatics* **2014**, *15*, 399.

[18] B. J. Grant, A. P. Rodrigues, K. M. ElSawy, J. A. McCammon, L. S. Caves, *Bioinformatics* **2006**, *22*, 2695.

[19] A. T. Van Wart, J. Durrant, L. Votapka, R. E. Amaro, *J Chem Theory Comput* **2014**, *10*, 511.