Supplementary Methods, Figures, and Tables

Diverse GPCRs exhibit conserved water networks for stabilization and activation

A. J. Venkatakrishnan[1,2,3,4,*], Anthony K. Ma[1,2,3,4,*], Rasmus Fonseca[3], Naomi R. Latorraca[1,2,3,4,5], Brendan Kelly[1,2,3,4], Robin M. Betz[1,2,3,4,5], Chaitanya Asawa[1,2,3,4], Brian K. Kobilka[3,#], Ron O. Dror[1,2,3,4,5#]

[1]Department of Computer Science, Stanford University, Stanford, California 94305, USA.

[2]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, California 94305, USA.

[3]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA.

[4]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

[5]Biophysics Program, Stanford University, Stanford, California 94305, USA.

* indicates equal contribution

[#] To whom correspondence should be addressed. Email: kobilka@stanford.edu; ron.dror@stanford.edu

Materials and Methods

MD simulation system setup

The new simulations were initiated from the crystal structures from the Protein Data Bank (PDB) of inactive DOR (PDB ID: 4N6H), inactive $A_{2A}R$ (PDB ID: 5IU4), active $A_{2A}R$ (PDB ID: 5G53), active M2R (4MQS), and active MOR (5C1M). Prime (Schrödinger, Inc.) was used to model in missing side chains. Hydrogen atoms were added, and protein chain termini were capped with the neutral groups acetyl and methylamide. Residues Asp2x50 and Asp3x49 were protonated in the active-state simulations and deprotonated in the inactive-state simulations. All other titratable residues were left in their dominant protonation state at pH 7.0. The prepared protein structures were aligned on the transmembrane helices using the OPM database (1). For low-resolution structures, *i.e.* structures of active $A_{2A}R$, and M2R, waters were added with Dowser(2), in addition to internal waters that were resolved in the crystal structure. Using Dabble (https://zenodo.org/badge/latestdoi/29268375), the structures were then inserted into a pre-equilibrated palmitoyl-oleoyl-phosphatidylcholine (POPC) bilayer, and solvated with 0.15 M NaCl in explicitly represented water.

MD simulation force field parameters

We used the CHARMM36 parameter set for protein molecules, lipid molecules, and salt ions, and the CHARMM TIP3P model for water. Parameters for the co-crystallized ligands naltrindole (in inactive DOR), NECA (in active $A_{2A}R$), BU72 (in active MOR), and iperoxo (in active M2R) were generated using the CHARMM General Force Field (CGenFF) (3) with the ParamChem server (paramchem.org), version 1.0.0. Parameters for the co-crystallized ligand ZM241385 (in inactive $A_{2A}R$) were obtained from an earlier study (4).

MD simulation protocol

Simulations were performed on GPUs using the CUDA version of PMEMD (Particle Mesh Ewald Molecular Dynamics) in Amber (5). Prepared systems were minimized, then equilibrated as follows: The system was heated using the Langevin thermostat from 0 to 100 K in the NVT ensemble over 12.5 ps with harmonic restraints of 10.0 kcal•mol^{-1}•Å^{-2} on the non-hydrogen atoms of lipid, protein, and ligand, and initial velocities sampled from the Boltzman distribution. The system was then heated to 310 K over 125 ps in the NPT ensemble with semi-isotropic pressure coupling and a pressure of one bar. For all the simulations except active $A_{2A}R$, further equilibration was performed at 310 K with harmonic restraints on the protein and ligand starting at 5.0 kcal•mol^{-1}•Å^{-2} and reduced by 1.0 kcal•mol^{-1}•Å^{-2} in a stepwise fashion every 2 ns, for a total of 10 ns of additional restrained equilibration. For active $A_{2A}R$, further equilibration was performed at 310 K with harmonic restraints on the protein and ligand starting at 5.0 kcal•mol^{-1}•Å^{-2} and reduced by 1.0 kcal•mol^{-1}•Å^{-2} in a stepwise fashion every 2 ns to 1.0 kcal•mol^{-1}•Å^{-2} and then by 0.1 kcal•mol^{-1}•Å^{-2} in a stepwise fashion every 2 ns, for a total of 28 ns of additional restrained equilibration.

Multiple independent simulations were initialized from the final snapshot of the restrained equilibration. These simulations were conducted in the NPT ensemble at 310 K and 1 bar, using a Langevin thermostat and Monte Carlo barostat. Simulations used periodic boundary conditions. A time step of 4.0 fs with hydrogen mass repartitioning (6) was used for inactive DOR, inactive $A_{2A}R$, active M2R, and active $A_{2A}R$, and a time step of 2.5 fs was used for active MOR. The simulation frames of inactive DOR, inactive $A_{2A}R$, active M2R, and active $A_{2A}R$ were written every 200 ps and those of active MOR were written every 100 ps.

The active-state simulations were maintained in the active state using harmonic restraints applied to receptor residues within 5 Å of the co-crystallized nanobody for active M2 receptor and active MOR and within 5 Å of the mini G protein for active $A_{2A}R$ receptor. For active M2 receptor and active $A_{2A}R$ receptor, the harmonic restraints were 5.0 kcal•mol^{-1}•Å^{-2}. For active MOR, the harmonic restraints were 1.0 kcal•mol^{-1}•Å^{-2}. The exclusion of the co-crystallized nanobody or mini G protein allows us to reduce the size of the simulated system and thus achieve longer simulation lengths.

Bond lengths to hydrogen atoms were constrained using SHAKE. Non-bonded interactions were cut off at 9.0 Å, and long-range electrostatic interactions were computed using the particle mesh Ewald (PME) method with an Ewald coefficient β of approximately 0.31 Å and B-spline interpolation of order 4. The FFT grid size was chosen such that the width of a grid cell was approximately 1 Å.

Computation of water-mediated interactions

Hydrogen bonds in the MD simulations were computed using the HBonds plugin in VMD (7) with the following geometric criteria: donor to acceptor distance less than 3.5 Å and donor-hydrogen-acceptor angle greater than 110°. Hydrogen bonds in the high-resolution crystal structures were computed based on the donor to acceptor distance of 3.2 Å only, as crystal structures generally lack hydrogens. For the crystal structure-based analysis, we considered high-resolution structures (of resolution 2.1 Å or lower) of diverse GPCRs in inactive state: Orexin-2 (PDB ID: 5WQC), D4 dopamine (PDB ID: 5WIU), δ-opioid (PDB ID: 4N6H), A_{2A} adenosine (PDB ID: 5IU4) and β_1 adrenergic (PDB ID: 4BVN) receptors. For the active state, crystal structures of the following GPCRs were considered: rhodopsin (PDB ID: 2X72), viral chemokine receptor US28 (PDB ID: 4XT1) and μ-opioid receptor (PDB ID: 5C1M). A water-mediated interaction between a pair of residues is defined when both residues form hydrogen bonds with the same water molecule or a pair of water molecules that in turn are linked by a hydrogen bond. In the ligand-associated analyses, water-mediated interactions are defined to occur when a water molecule or pair of water molecules connect a ligand atom and a residue through hydrogen bonds. We only considered interactions mediated by residues that were at least 4 positions apart in TM1-7 and helix 8 (H8) in order to avoid local short-range interactions. We evaluated the presence of water-mediated interactions between structurally equivalent residues across different GPCR structures. Structural equivalence was assigned using the GPCRdb numbering scheme (8, 9), which is developed based on the Ballesteros Weinstein numbering scheme (10) and corrects for helical bulges and

constrictions. In the GPCRdb numbering scheme, every residue is addressed using two numbers separated by an 'x'. The first number denotes the helix (1-8) and the second number denotes the residue position relative to the most conserved position on that helix, which is assigned the number 50. For example, 3x51 denotes a residue in transmembrane helix 6, one position after the most conserved residue (3x50). A bulge residue is assigned the same number as the preceding residue followed by a 1, *e.g.* 551 for a bulge following position 55.

Computing frequencies of water-mediated interactions

Water-mediated interactions were computed for every frame of the MD simulations. The stability of a water-mediated interaction between a pair of residues was defined as the fraction of frames in which either a direct or extended water-mediated interaction is formed. The stability of water-mediated interactions was computed for various receptors in both the inactive and active states (**SI Appendix, Supplementary Table S1**).

Water occupancy and water density analysis

To compute the occupancy of water molecules proximal to each crystallographic water position, we identified amino acid residues within 5 Å of the crystallographic water's position and aligned all simulation frames on these residues. We then calculated the fraction of simulation frames in which a water molecule is within 1 Å of the crystallographic water position.

To compute water density, a grid with 1 $Å^3$ voxels was superimposed upon a GPCR structure. We then computed the fraction of simulation frames in which an oxygen atom of a water molecule is contained within each voxel. The water density map was output as a ".dx" file and loaded into VMD with an 'isosurface' representation to render a water density map.

Sequence analysis

We obtained the human class A non-olfactory GPCR sequences from GPCRdb (11) and computed the percentage conservation of each amino acid at every position in TMs 1-7 and helix 8. The percentage of polar amino acids at each position was computed by adding the percentages of the following amino acids: Asn, Asp, Glu, Gln, His, Ser, Thr, and Tyr (**SI Appendix**, **Supplementary Table S2)**.

Visualization

Water-mediated interactions between residues were visualized using flareplots (https://gpcrviz.github.io/flareplot/). Structural visualization of protein structure and water-mediated interaction networks was performed using VMD (7) and PyMOL.

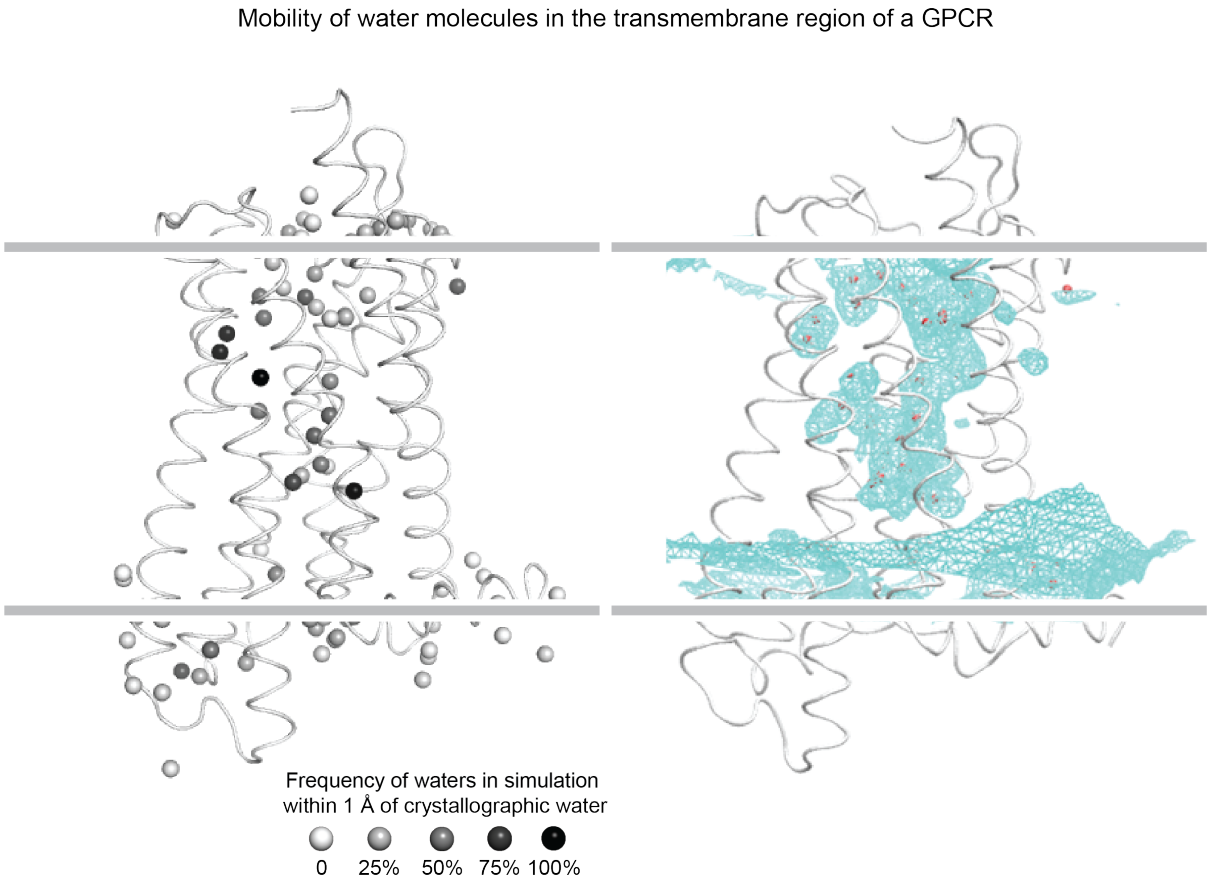
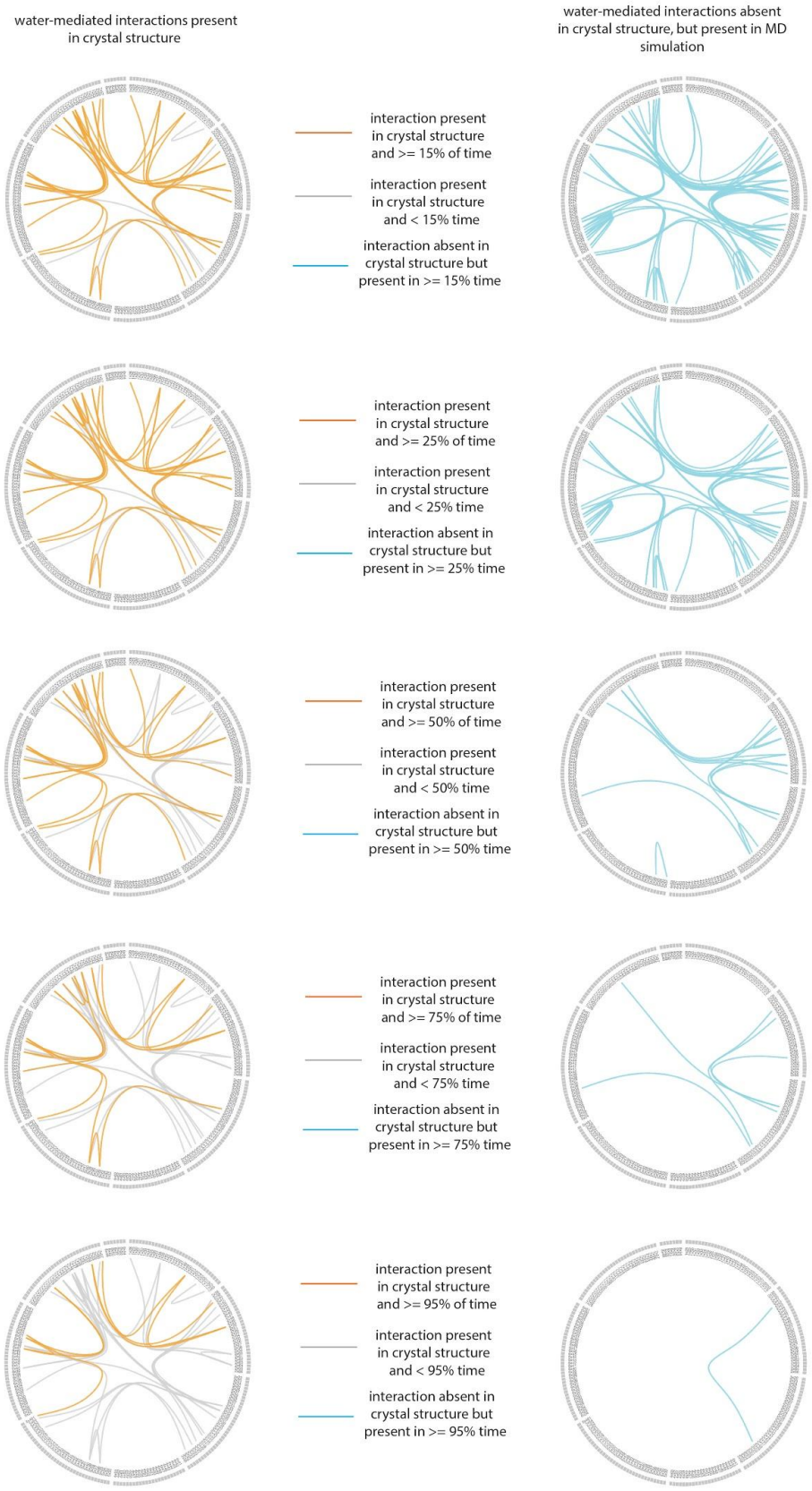


Fig. S1. Occupancy of water molecules in GPCR simulation. (a) The occupancy of water molecules in simulations of DOR at crystallographic water positions. The occupancy of each position is shown on a white (0% occupancy) to black (100% occupancy) spectrum. Occupancy of waters in simulation is calculated as the fraction of simulation time where waters are present within 1 Å of each crystallographic water position. (b) The density of water molecules in the transmembrane region. The mesh encloses regions where waters in simulation are present at least 1% of simulation frames in each grid cell of dimension 1 Å x 1 Å x 1 Å. Crystallographic water positions are shown by red spheres. These almost always fall within the mesh-enclosed regions, indicating that despite the low occupancy of waters in simulation at most crystallographic water positions, waters tend to occupy the same regions of the receptor in crystal structures and simulations.

Varying stability of water-mediated interactions in GPCR simulation (*e.g.* inactive state DOR)



Supplementary Figure S2

Fig. S2. Stability of water-mediated interactions present in crystal structures and simulation. Water-mediated interactions present in MD simulation of a GPCR were stratified into three sets: interactions that are present in crystal structure and in simulations above a frequency threshold (orange), interactions that are present in crystal structure and in simulations below a frequency threshold (gray), interactions that are absent in crystal structure but are observed in simulation above a frequency threshold (gray). The water-mediated interactions were stratified at different frequency thresholds: 15%, 25%, 50%, 75%, and 95%.

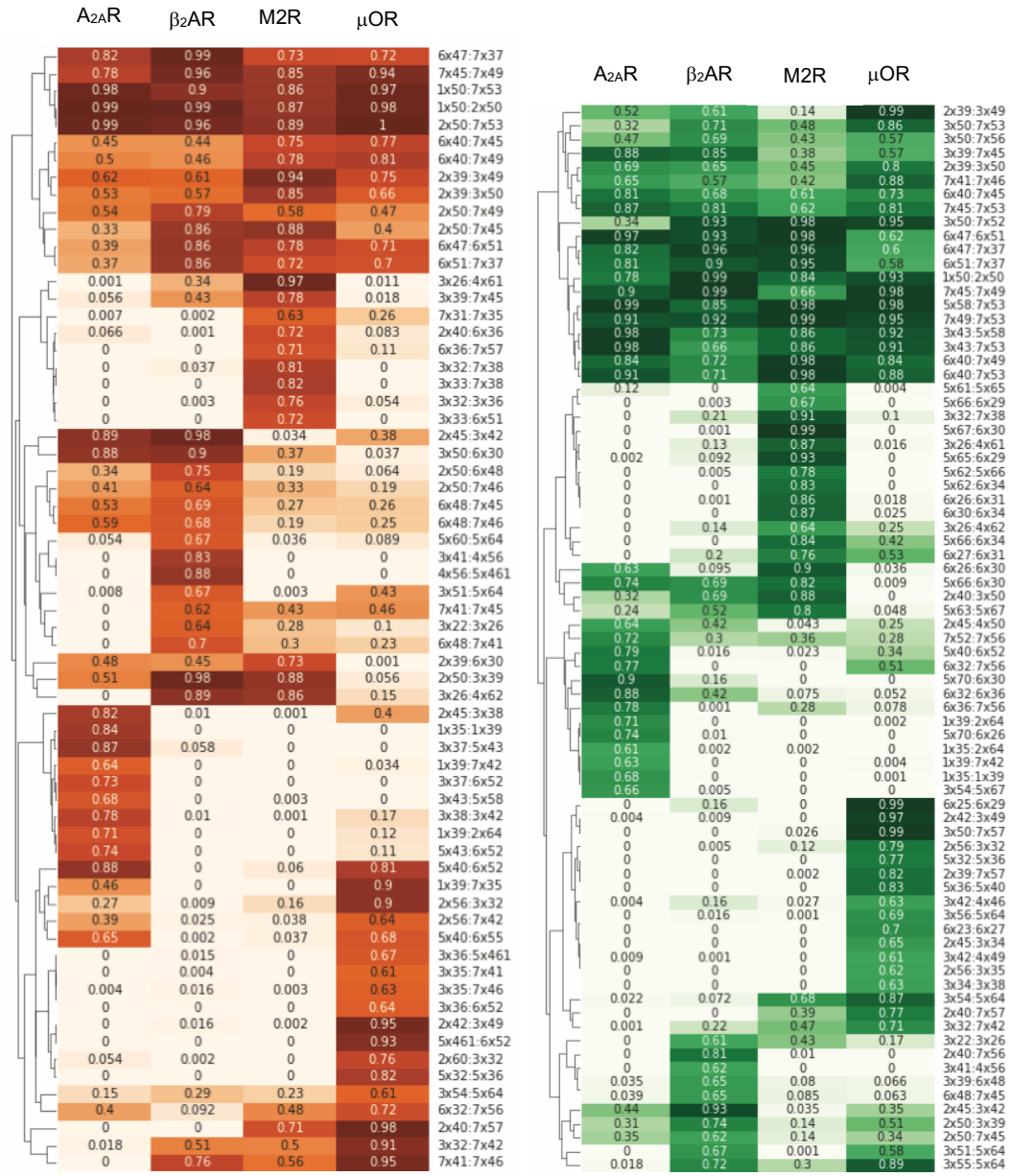


Fig. S3. Heatmap comparing the stability of water-mediated interactions in inactive states (left panel) and active states (active panel). Columns indicate receptors and rows indicate pairs of structurally equivalent positions, shown using GPCRdb numbers. The cells indicate the measure of stability of water-mediated interactions. Stability is shown as percentage values and using corresponding colors. Lighter shades indicate water-mediated interactions that have low stability and darker shades indicate water-mediated interactions that have high stability. Rows with dark cells in all columns indicate interactions that have highly stability in all the receptors being compared.

A

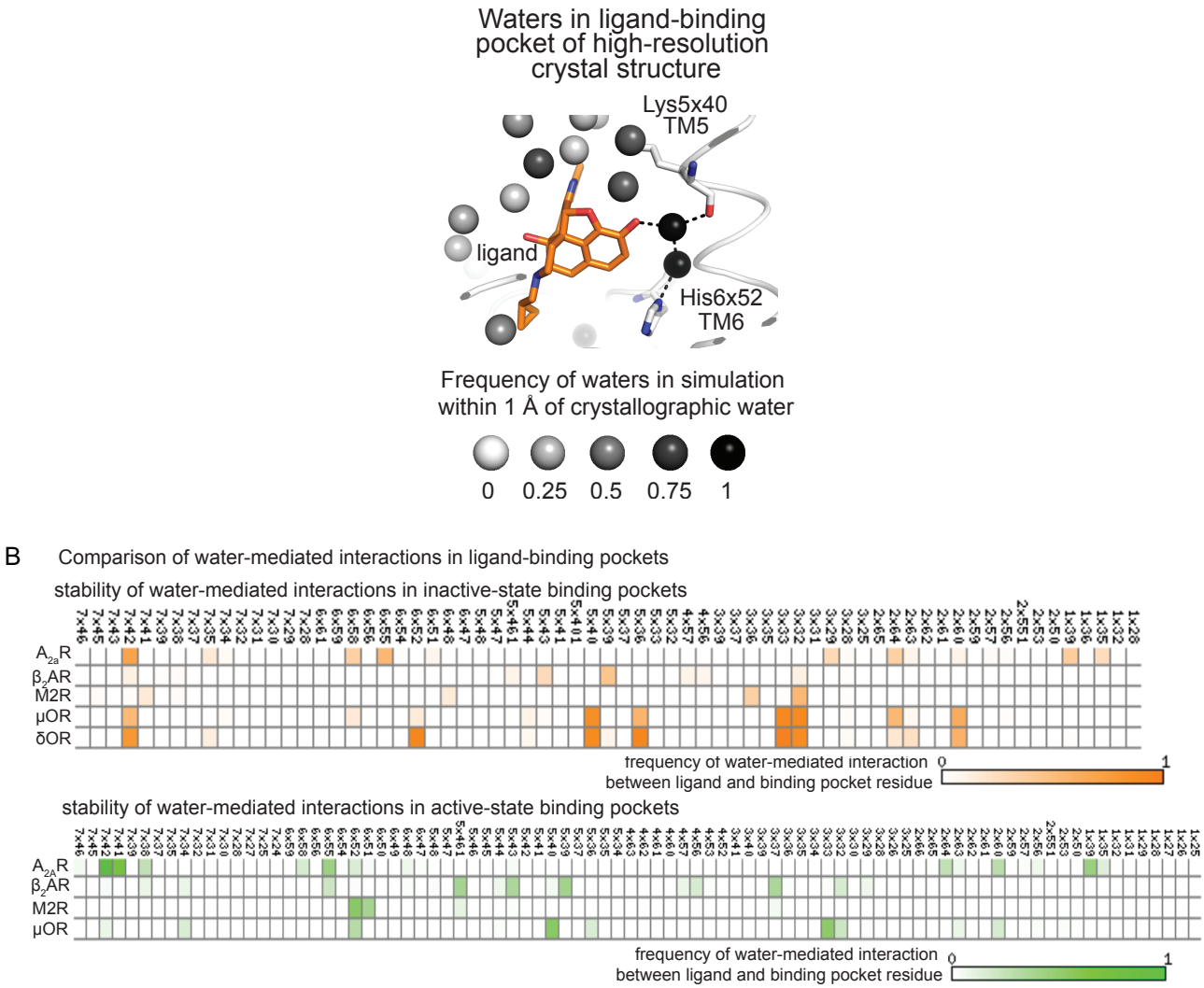


Fig. S4. Diversity of stable water-mediated interactions in ligand-binding pockets of GPCRs. (*A*) Water molecules in the ligand-binding pocket of a high-resolution GPCR structure (DOR). The water occupancy in simulation of each crystallographic water position is shown on a white (0% occupancy) to black (100% occupancy) spectrum. Water occupancy in simulation is calculated as the fraction of simulation frames in which a water molecule is present within 1 Å of each crystallographic water positions. Two highly stable waters and the water-mediated interactions they form in the binding pocket are highlighted. The hydrogen bonds formed by these highly stable water molecules connecting the ligand and the binding pocket are shown as black dashed lines. (*B*) Heatmaps comparing the stability of water-mediated interactions in the ligand-binding pocket of five class A GPCRs: $A_{2A}R$, β_2AR, M2R, MOR, and DOR. In the heatmaps, the rows indicate receptors and the columns indicate structurally equivalent binding pocket positions. The frequency of water-mediated interactions between the ligand and the binding pocket residues in MD simulations is mapped to a spectrum of white to orange (for inactive state) and white to green (for active state). Lighter shades indicate water-mediated interactions that have low stability and darker shades indicate water-mediated interactions that have high stability.

GPCR	PDB ID	State	Ligand	Publication	Intracellular interface	Simulation length	Force field
B2AR	2RH1	Inactive	Carazolol	Nygaard et al., Cell 2013 (Condition: 12) DOI: 10.1016/j.cell.2013.01.008	-	10 us	CHARMM27 protein force field with modification to Asp, Glu, and Arg side chains and CHARMM36 lipid force field
DOR	4N6H	Inactive	Naltrindole	-	-	1.7 us, 1.3 us, 2us	CHARMM36
M2R	3UON	Inactive	QNB	Kruse et al., Nature 2012 (Condition: A) DOI: 10.1038/nature10867	-	16.4 us	CHARMM27 protein force field with modification to Asp, Glu, and Arg side chains and CHARMM36 lipid force field
MOR	4DKL	Inactive	Beta-FNA	Huang et al., Nature 2015 DOI: 10.1038/nature14886	-	1.2 us, 1.4 us, 0.6 us	CHARMM36
A2AR	5IU4	Inactive	ZM241385	-	-	1.5 us, 1.5 us, 1.6 us, 1.3 us, 1.5 us	CHARMM36
B2AR	3SN6	Active	BI-167107	Dror et al., Science 2015 (Conditions: 22, 23, 24, 65) DOI: 10.1126/science.aaa5264	Gs protein bound	50 us, 10us, 10 us, 5 us	CHARMM27 protein force field with modification to Asp, Glu, and Arg side chains and CHARMM36 lipid force field
M2R	4MQS	Active	Iperoxo	-	Nanobody interface restrained	1.4 us, 1.6 us, 1.60 us	CHARMM36
MOR	5C1M	Active	BU72	-	Nanobody interface restrained	3.1 us, 2.8 us, 2.9 us	CHARMM36
A2AR	5G53	Active	NECA	-	Mini-Gs protein interface restrained	1.2 us, 1.2 us, 1.1 us, 1.0 us, 1.1 us	CHARMM36

Supplementary Table S1. List of GPCR simulations analyzed in this study

GPCRDB number	Polar amino acids	Asn	Asp	Glu	Gln	His	Ser	Thr	Tyr
1x50	100%	100%	0%	0%	0%	0%	0%	0%	0%
2x50	98%	1%	96%	1%	0%	0%	0%	0%	0%
5x58	89%	7%	0%	0%	1%	1%	1%	2%	77%
7x45	90%	69%	0%	0%	1%	8%	12%	0%	0%
7x49	98%	78%	17%	0%	0%	1%	1%	1%	0%
7x53	92%	0%	1%	0%	0%	0%	0%	0%	91%

Supplementary Table S2. Sequence conservation across human class A GPCRs (non-olfactory) of positions forming conserved stable water-mediated networks exclusively in inactive state and active state of GPCRs

References

1. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. *Bioinformatics* 22(5):623–625.

2. Zhang L, Hermans J (1996) Hydrophilicity of cavities in proteins. *Proteins: Structure, Function, and Genetics* 24(4):433–438.

3. Vanommeslaeghe K, et al. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 31(4):671–690.

4. Guo D, et al. (2016) Molecular Basis of Ligand Dissociation from the Adenosine A2A Receptor. *Mol Pharmacol* 89(5):485–491.

5. Salomon-Ferrer R, Case DA, Walker RC (2012) An overview of the Amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci* 3(2):198–210.

6. Hopkins CW, Le Grand S, Walker RC, Roitberg AE (2015) Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* 11(4):1864–1874.

7. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38.

8. Isberg V, et al. (2015) Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol Sci* 36(1):22–31.

9. Pándy-Szekeres G, et al. (2018) GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res* 46(D1):D440–D446.

10. Ballesteros JA, Weinstein H (1995) [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods in Neurosciences*, pp 366–428.

11. Isberg V, et al. (2016) GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res* 44(D1):D356–64.