

Stable networks of water-mediated interactions are conserved in activation of diverse GPCRs

A. J. Venkatakrisnan^{1,2,3*}, Anthony K. Ma^{1,2,3*}, Rasmus Fonseca³, Naomi R. Latorraca^{1,2,3,4}, Brendan Kelly^{1,2,3}, Robin M. Betz^{1,2,3,4}, Chaitanya Asawa^{1,2,3}, Brian K. Kobilka³, Ron O. Dror^{1,2,3,4}

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

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

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

⁴Biophysics Program, Stanford University, Stanford, California 94305, USA.

* indicates equal contribution

Abstract

G protein-coupled receptors (GPCRs) have evolved to recognize incredibly diverse extracellular ligands while sharing a common architecture and structurally conserved intracellular signaling partners. It remains unclear how binding of diverse ligands brings about GPCR activation, the common structural change that enables intracellular signaling. Here, we identify highly conserved networks of water-mediated interactions that play a central role in activation. Using atomic-level simulations of diverse GPCRs, we show that most of the water molecules in GPCR crystal structures are highly mobile. Several water molecules near the G protein-coupling interface, however, are stable. These water molecules form two kinds of polar networks that are conserved across diverse GPCRs: (i) a network that is maintained across the inactive and the active states and (ii) a network that rearranges upon activation. Comparative analysis of GPCR crystal structures independently confirms the striking conservation of water-mediated interaction networks. These conserved water-mediated interactions near the G protein-coupling region, along with diverse water-mediated interactions with extracellular ligands, have direct implications for structure-based drug design and GPCR engineering.

Significance

G protein-coupled receptors (GPCRs) represent both the largest class of drug targets and the largest family of human membrane proteins. Recent three-dimensional structures are revealing the presence of many water molecules buried inside GPCRs, but the functional role of these waters has been unclear. Using extensive atomic-level computer simulations, we find that although most of these waters are highly mobile, a few are stable. These stable water molecules mediate state-dependent and state-independent polar networks that are conserved across diverse GPCRs. These findings promise to help guide the rational design of GPCR-targeted drugs.

Introduction

G protein–coupled receptors (GPCRs) are membrane proteins that act as a control panel for our cells: they recognize the presence of extracellular molecules such as hormones or neurotransmitters and stimulate intracellular signaling pathways in response. Humans have over 800 GPCRs, which share a common architecture comprising seven transmembrane (TM) helices (1).

GPCRs transmit signals across the cell membrane by transitioning from an inactive conformational state to an active conformational state. The active conformational state, which is favored by binding of certain ligands to the extracellular side of the GPCR, couples to and stimulates G proteins. Available structures indicate that diverse GPCRs undergo similar conformational changes upon activation (2). However, these GPCRs have evolved to recognize dramatically different ligands: human GPCRs recognize molecules ranging from amines, nucleosides, lipids, and steroids to entire proteins, in most cases with high specificity. This ability to stimulate cellular pathways in response to the presence of specific extracellular ligands makes GPCRs excellent drug targets, and in fact roughly one-third of all drugs act by binding to GPCRs (3). The process by which diverse ligands bind to GPCRs and favor activation is of great interest in molecular biology and drug discovery but remains incompletely understood.

Recent high-resolution crystal structures reveal dozens of water molecules buried in the transmembrane region of GPCRs (4–9). Water has been shown to play a critical role both in the binding of drugs to their targets (10) and in the function of many proteins (11, 12), raising the question of how the water molecules in GPCRs behave and what function they might serve. Indeed, this question was raised even before the high-resolution structures were available (13, 14).

Unfortunately, answering this question from crystal structures alone is difficult, because each crystal structure represents a static snapshot. In reality, water molecules may be highly mobile, and GPCRs themselves are constantly changing conformation (15). Moreover, most of the available GPCR crystal structures were solved at cryogenic temperatures at which water molecules are expected to be much less mobile than at body temperature. Even recovering average positions of water molecules from crystallographic data can be difficult when the corresponding electron density is not highly and unambiguously localized.

Here, we investigate the structural dynamics of water molecules in a diverse set of GPCRs using molecular dynamics simulations, which allow us to follow the trajectory of each individual water molecule at physiological temperature and in a non-crystallographic membrane environment. We compare water dynamics not only between different GPCRs but also between

the active and inactive conformational states of the same GPCR. Our results show that the water molecules observed in GPCR crystal structures are not all equal: a few are stable in their crystallographically observed position, but most are highly mobile, spending only a small fraction of their time near that position.

Remarkably, stable water molecules near the G-protein-binding site form a network of polar interactions (hydrogen bonds) that rearranges upon receptor activation. This network is almost perfectly conserved across the GPCRs we examined, in sharp contrast to water-mediated interactions with ligands, which vary widely from one GPCR to another. This conserved water network appears to play a key role in the GPCR activation mechanism that was preserved as GPCRs evolved to recognize different ligands.

Results

We focus on class A GPCRs, which represent both the largest class of GPCRs and the best characterized class structurally. We chose functionally distinct and evolutionarily distant ligand-activated class A GPCRs for which both inactive-state and active-state structures are available: β_2 adrenergic receptor (β_2 AR (16, 17)), M2 muscarinic receptor (M2R (18, 19)), μ opioid receptor (MOR (20, 21)), and adenosine A_{2A} receptor (A_{2A} R (22, 23)). We also included δ opioid receptor (DOR) as it has a particularly high-resolution inactive-state crystal structure (1.8 Å) (24). Collectively, these GPCRs bind to different types of ligands (amines, peptides, nucleosides) and different G proteins (Gs, Gi). All simulations included an explicitly represented lipid bilayer, water, salt ions, and the co-crystallized ligand (**Table S1**).

Only a minority of crystallographic water molecules are stable

In simulations at physiological temperature, only a few of the waters resolved in high-resolution GPCR crystal structures remain near their crystallographically observed positions (**Fig. 1a** and **Fig. S1**). Our simulations are consistent with crystallographic data in that the same regions of the receptor are populated by waters in both cases (**Fig. S1**), but in simulation, most of the waters swim about, showing little preference for the specific positions occupied by waters in the crystal structures. A few crystallographic water positions, however, are consistently occupied by water molecules in simulation. We refer to these crystallographic water molecules as “stable,” even though an individual water molecule at such a position is generally replaced by a different water molecule at the same position at least every few hundred nanoseconds. In DOR simulations, for example, the water in the “kink” of transmembrane helix 6 (TM6) was highly stable with a water molecule observed within 1 Å of the crystallographic position for 99% of the time. On the other hand, the waters in the receptor core were highly mobile; for some of these, a water molecule was found within 1 Å of the crystallographic position less than 10% of the time.

To further characterize the behavior of water molecules in the structural framework of the

receptor, we investigated the polar interactions (hydrogen bonds) in the receptor formed by the water molecules (**Fig. 1b**). We focused on water-mediated interactions in which a pair of receptor residues is connected by one or two water molecules in the network of hydrogen bonds. We estimated the stability of a given water-mediated interaction based on the frequency of the interaction's presence in simulation (see **Methods**).

Only a fraction of the crystallographically observed water-mediated interactions are stable in simulation, because most of the water molecules are highly mobile, forming many transient, alternative polar interactions. For example, of the 42 water-mediated interactions in the DOR crystal structure, only 19 are formed in 60% or more of simulation frames (**Fig. 1b** and **Fig. S2**).

Conserved networks of state-independent and state-dependent water-mediated interactions in GPCR activation

Might water molecules play a role in GPCR activation? To address this question, we first compared water-mediated interactions in simulations of inactive and active GPCR states. We also compared the water-mediated interactions across different GPCRs (β_2 AR, A_{2A} R, MOR, and M2R). For the comparisons, we analyzed interactions between residues in structurally equivalent positions at different receptors. Structurally equivalent residues were identified using GPCRdb numbers (25), which are receptor-independent generic numbers for referencing structurally equivalent positions in GPCRs (see **Methods**).

We found that a surprisingly large set of stable water-mediated interactions (i.e., those formed over 60% of the time in simulation) were shared across *all* receptors examined in a given activation state (**Fig. 2**): six interactions mediated generally by four waters in the inactive state, and 12 interactions mediated generally by five waters in the active state.

These conserved interactions represent 20–30% of the stable water-mediated interactions observed in each receptor individually. In contrast, a majority of the remaining stable interactions are found in just one of A_{2A} R, β_2 AR, M2R and MOR. For example, of 32 stable water-mediated interactions in active-state M2R, 12 are shared across all four active-state receptors, and 14 of the remaining 20 are unique to M2R.

Among the conserved, stable, water-mediated interactions, three interactions are maintained in a state-independent manner, *i.e.* they are present in both the inactive states and the active states of A_{2A} R, β_2 AR, M2R and MOR. While one of these interactions stabilizes a helical kink of TM6, the other two water molecules are proximal to the G protein-coupling region and mediate interactions between residues on TM1 and TM2 and between residues on TM7 (**Fig. 3a**).

Two additional surprising observations link the conserved, stable, water-mediated interactions to

the GPCR activation process. First, the remaining conserved, stable water-mediated interactions are in the lower half of the GPCR, near the G protein-binding interface, where the largest structural rearrangements occur upon activation. Second, although these interactions are conserved across diverse GPCRs in a given activation state, they differ dramatically between the active and inactive states. Every conserved stable water molecule near the G protein-binding interface changes interaction partners upon activation, and a number of water molecules form conserved, stable water-mediated interactions in the active state but not in the inactive state. Water-mediated interactions between TMs 1, 2, and 7 are stably formed in the inactive states (**Fig. 4a**). This network is disrupted upon activation and a new stable network of water-mediated interactions is formed between TM3, TM5, TM6, and TM7 (**Fig. 4a**). Tyr7x53 (GPCRdb number) of the NPxxY motif acts as a switch between the inactive and active water-mediated networks. In the inactive state, Tyr7x53 interacts with a water molecule at the TM1-TM2 interface, and upon activation Tyr7x53 switches conformation and interacts with water molecules at the interface of TMs 3, 5, and 6.

All the side-chains involved in this rewiring of interactions with stable waters upon activation are highly conserved as polar residues across class A GPCRs (>89%) (**Table S2**). Taken together, our observations indicate that stable waters near the G protein-coupling interface form a conserved polar interaction network that rearranges upon receptor activation. The fact that this water-mediated activation switch was preserved as the GPCR family evolved suggests that it is functionally important.

Comparative analysis of crystal structures confirms conservation of state-dependent and state-independent water-mediated interactions

We postulated that the conserved stable water-mediated interactions identified from the MD simulations would correspond to water-mediated interactions that are maintained across crystal structures of different receptors. In order to investigate this, we compared the water-mediated interactions from all available high-resolution crystal structures (see **Methods**): inactive Orexin-2, D4 dopamine, δ -opioid, A_{2A} adenosine and β_1 adrenergic receptors, and active μ -opioid receptor (**Fig. 3b**). We compared the water-mediated interactions across structurally equivalent positions (see **Methods**). A network of seven water-mediated interactions involving three water molecules is maintained in all these crystal structures (**Fig. 3b**). Interestingly, this network of water-mediated interactions is highly similar to the network of conserved, state-independent, stable interactions observed in the MD simulations. All the three water-mediated interactions found to be conserved and stable across the simulations (**Fig. 3a**) are maintained consistently across the crystal structures (**Fig. 3b**). The remaining interactions that are consistently maintained in the crystal structures involve the same three waters as the conserved stable water-mediated interactions in the simulations.

Similarly, we compared the water-mediated interactions across high-resolution crystal structures of inactive states and crystal structures of active states. We found a network of two water-mediated interactions formed by Tyr7x53 to be maintained exclusively in the inactive state across Orexin-2, D4 dopamine, δ -opioid and adenosine A_{2A} receptors (**Fig. 4b**). The only exception where this network is absent is in the case of inactive β_1 adrenergic receptor, and this is explained by the presence of a Tyr7x53Leu thermostabilizing mutation, which would prevent formation of any polar interactions. Interestingly, the conserved inactive state-dependent network identified from crystal structures is identical to the conserved, stable water-mediated interaction network found in the MD simulations of inactive states (**Fig. 4a,b**). Similarly, we also found a network of three water-mediated interactions formed by Tyr7x53 to be maintained exclusively in the crystal structures of active state rhodopsin, chemokine receptor US28, and μ -opioid receptor (**Fig. 4b**). All the three water-mediated interactions identified from the crystal structures of active states are present in the network of conserved, stable water-mediated interactions found in the MD simulations of the active states (**Fig. 4a,b**).

Taken together, the comparative analysis of crystal structures supports the presence of conserved state-dependent and state-independent water-mediated interactions in GPCRs.

Stable water molecules in the ligand-binding pocket vary between different GPCRs

A number of the water molecules in high-resolution crystal structures of GPCRs form direct interactions with the ligand, and previous studies have pointed out the importance of water in determining ligand affinities and binding rates (26). We thus also examined the occupancy of waters in the binding pocket and the stability of water-mediated interactions between ligands and receptor residues in simulation.

Most waters in the binding pocket are swimming about. Certain waters and water-mediated interactions, however, are stable. For example, in simulations of inactive-state DOR, two waters in the binding pocket are highly stable near their crystallographic positions (**Fig. 5a**). These two stable waters mediate interactions between a hydroxyl group on the ligand and polar residues on TMs 5 and 6.

No stable water-mediated interactions between the ligand and the receptor appear to be conserved across diverse GPCRs. Our simulations showed no such interactions conserved across all the receptors examined, and only very limited structurally equivalent interactions between subsets of these receptors (**Fig. 5b**). This variability in water-mediated interactions across ligand-binding pockets suggests that, unlike the G protein-coupling region, the binding pocket has been adapted to exploit different water-mediated interactions in different GPCRs as the GPCR family evolved.

Discussion

The human genome includes hundreds of class A GPCRs, and these receptors have evolved to recognize extremely diverse extracellular ligands—ranging from nucleosides and amines to peptides and even entire proteins—often with very high specificity. On the other hand, all of these GPCRs couple to the same set of intracellular G proteins, which share very similar structures. One might thus imagine that GPCRs would have evolved a common activation mechanism. Indeed, previous studies have identified conserved networks of amino acid contacts and ionic interactions that appear to facilitate such a mechanism (2, 27).

Our results suggest that certain water molecules also play an important, conserved role in GPCR activation. In particular, we discovered a network of highly stable water molecules near the G protein-binding site that is conserved almost perfectly across diverse class A GPCRs and that rearranges in a conserved manner upon GPCR activation. Waters also play a critical role in mediating ligand binding in GPCRs, but the water-mediated interactions in the ligand-binding pocket vary widely among class A GPCRs, likely reflecting the fact that the binding pockets of these different GPCRs have evolved to recognize different ligands.

Our results complement those of previous experimental and computational studies. Recent high-resolution crystal structures have revealed the presence of dozens of waters within GPCRs. Analyses of earlier, lower-resolution crystal structures have suggested the presence of certain conserved structural water molecules (13, 14). Radiolytic footprinting has provided insights into tightly bound waters and dynamic waters in rhodopsin (28). Simulation-based studies have shown the presence of activation associated internal water pathways in GPCRs (28, 29) and indicated a role for waters in ligand binding (30–32).

A few caveats are in order. First, the force fields used for MD simulation are not perfect, although simulations have proven useful in elucidating water dynamics that are difficult to characterize experimentally (33, 34). Second, the strengths of hydrogen bonds in proteins can vary depending on factors such as the local structural environment (35), but our current analysis does not take this into account. Third, we lack a direct method to determine the functional importance of a water molecule in a particular position. We thus use stability and conservation across receptors as rough proxies for importance. Fourth, our analysis of stable water molecules and water-mediated interactions does not cover all the ways in which water may contribute to GPCR function; for example, displacement of certain water molecules from the binding pocket upon ligand binding is a determinant of ligand-binding affinity (36). Finally, our analysis has focused on class A GPCRs as they are the most extensively structurally characterized class of GPCRs. Future work will be necessary to understand the role of water

molecules in the activation of other GPCR classes.

Our findings have several implications for GPCR drug discovery and GPCR engineering. First, understanding the behavior of water molecules in the binding pocket can aid structure-based drug discovery (26). Our results show that crystallographically resolved waters cannot all be treated as equal. One needs to characterize the dynamics of water molecules (e.g. through simulations) to identify the stable waters, which can then be exploited for drug design (37).

Second, the conserved stable waters and water-mediated interactions identified in this study can aid in structural modeling. Homology models of GPCRs are widely used for drug discovery, both when no crystal structure of the target is available and when the available structures are not in the desired activation state. The conserved water molecules identified in this study for active-state and inactive-state GPCRs can be incorporated into homology models and used to refine those models. Likewise, these water molecules can also be incorporated into low-resolution experimental GPCR structures, such as cryo-EM structures (38), in which few waters are resolved.

Third, mutating GPCR residues to stabilize either the active state or the inactive state has proven useful in GPCR crystallography as well as in functional investigations. The distinct water-mediated interaction networks associated with different activation states provide new avenues for biochemically perturbing the conformational landscape of GPCRs. In particular, residues interacting with the conserved, stable waters identified in this study are excellent candidate sites for introducing mutations to shift the conformational ensemble towards the active or inactive state.

More generally, our results reveal a “hidden” structural feature of GPCRs, a network of stable water-mediated interactions that is not evident from crystal structures alone. Our study also highlights the untapped potential of comparing structural dynamics across a drug target family to identify shared and distinct molecular features.

Material and Methods

MD simulation trajectories

The MD simulation trajectories of inactive β_2 AR, M2R and MOR and active β_2 AR were obtained from previously published studies (**Table S1**). New MD simulations were performed for inactive DOR and A_{2A} R and active A_{2A} R, M2R, and MOR (**Table S1**), as described below. The simulations were all performed using CHARMM force fields (**Table S1**). We verified that the simulations initiated from the inactive-state and active-state structures were maintained in their respective states.

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 (39). For low-resolution structures, *i.e.* structures of active A_{2A}R, and M2R, waters were added with Dowser(40), 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 iperexo (in active M2R) were generated using the CHARMM General Force Field (CGenFF) (41) 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 (42).

MD simulation protocol

Simulations were performed on GPUs using the CUDA version of PMEMD (Particle Mesh Ewald Molecular Dynamics) in Amber (43). 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 (44) 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.

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 (45) 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 (25), which is developed based on the Ballesteros Weinstein numbering scheme (46) 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 (**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 to 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 computing water density, a grid with 1 Å³ 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 GPCRDdb (25) 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.

Visualisation

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 (45) and PyMOL.

References

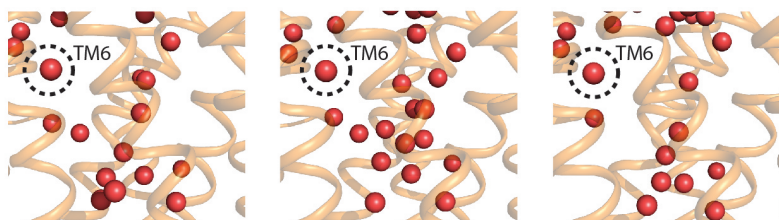
1. Fredriksson R, Lagerström MC, Lundin L-G, Schiöth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* 63(6):1256–1272.

2. Venkatakrisnan AJ, et al. (2016) Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* 536(7617):484–487.
3. Santos R, et al. (2017) A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* 16(1):19–34.
4. Huang W, et al. (2015) Structural insights into μ -opioid receptor activation. *Nature* 524(7565):315–321.
5. Segala E, et al. (2016) Controlling the Dissociation of Ligands from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength. *J Med Chem* 59(13):6470–6479.
6. Fenalti G, et al. (2014) Molecular control of δ -opioid receptor signalling. *Nature* 506(7487):191–196.
7. Suno R, et al. (2018) Crystal Structures of Human Orexin 2 Receptor Bound to the Subtype-Selective Antagonist EMPA. *Structure* 26(1):7–19.e5.
8. Wang S, et al. (2017) Dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science* 358(6361):381–386.
9. Miller-Gallacher JL, et al. (2014) The 2.1 Å resolution structure of cyanopindolol-bound β 1-adrenoceptor identifies an intramembrane Na⁺ ion that stabilises the ligand-free receptor. *PLoS One* 9(3):e92727.
10. Breiten B, et al. (2013) Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J Am Chem Soc* 135(41):15579–15584.
11. Levy Y, Onuchic JN (2006) Water mediation in protein folding and molecular recognition. *Annu Rev Biophys Biomol Struct* 35:389–415.
12. Bellissent-Funel M-C, et al. (2016) Water Determines the Structure and Dynamics of Proteins. *Chem Rev* 116(13):7673–7697.
13. Pardo L, Deupi X, Dölker N, López-Rodríguez ML, Campillo M (2007) The role of internal water molecules in the structure and function of the rhodopsin family of G protein-coupled receptors. *Chembiochem* 8(1):19–24.
14. Angel TE, Chance MR, Palczewski K (2009) Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc Natl Acad Sci U S A* 106(21):8555–8560.
15. Latorraca NR, Venkatakrisnan AJ, Dror RO (2017) GPCR Dynamics: Structures in Motion. *Chem Rev* 117(1):139–155.
16. Rasmussen SGF, et al. (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450(7168):383–387.
17. Rasmussen SGF, et al. (2011) Crystal structure of the β 2 adrenergic receptor-Gs protein

- complex. *Nature* 477(7366):549–555.
18. Kruse AC, et al. (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* 504(7478):101–106.
 19. Haga K, et al. (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482(7386):547–551.
 20. Manglik A, et al. (2012) Crystal structure of the μ -opioid receptor bound to a morphinan antagonist. *Nature* 485(7398):321–326.
 21. Huang W, et al. (2015) Structural insights into μ -opioid receptor activation. *Nature* 524(7565):315–321.
 22. Carpenter B, Nehmé R, Warne T, Leslie AGW, Tate CG (2016) Structure of the adenosine A(2A) receptor bound to an engineered G protein. *Nature* 536(7614):104–107.
 23. Segala E, et al. (2016) Controlling the Dissociation of Ligands from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength. *J Med Chem* 59(13):6470–6479.
 24. Fenalti G, et al. (2014) Molecular control of δ -opioid receptor signalling. *Nature* 506(7487):191–196.
 25. Isberg V, et al. (2016) GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res* 44(D1):D356–64.
 26. Jazayeri A, Dias JM, Marshall FH (2015) From G Protein-coupled Receptor Structure Resolution to Rational Drug Design. *J Biol Chem* 290(32):19489–19495.
 27. Isom DG, Dohlman HG (2015) Buried ionizable networks are an ancient hallmark of G protein-coupled receptor activation. *Proc Natl Acad Sci U S A* 112(18):5702–5707.
 28. Angel TE, Gupta S, Jastrzebska B, Palczewski K, Chance MR (2009) Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc Natl Acad Sci U S A* 106(34):14367–14372.
 29. Yuan S, Filipek S, Palczewski K, Vogel H (2014) Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat Commun* 5:4733.
 30. Yuan S, et al. (2015) The Mechanism of Ligand-Induced Activation or Inhibition of μ - and κ -Opioid Receptors. *Angew Chem Int Ed Engl* 127(26):7670–7673.
 31. Dror RO, et al. (2011) Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci U S A* 108(32):13118–13123.
 32. Bortolato A, Tehan BG, Bodnarchuk MS, Essex JW, Mason JS (2013) Water network perturbation in ligand binding: adenosine A(2A) antagonists as a case study. *J Chem Inf Model* 53(7):1700–1713.

33. Tajkhorshid E, et al. (2002) Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* 296(5567):525–530.
34. de Groot BL, Grubmüller H (2001) Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* 294(5550):2353–2357.
35. Pace CN, Nick Pace C (2009) Energetics of protein hydrogen bonds. *Nat Struct Mol Biol* 16(7):681–682.
36. Snyder PW, et al. (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc Natl Acad Sci U S A* 108(44):17889–17894.
37. Manglik A, et al. (2016) Structure-based discovery of opioid analgesics with reduced side effects. *Nature* 537(7619):185–190.
38. Zhang Y, et al. (2017) Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature* 546(7657):248–253.
39. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. *Bioinformatics* 22(5):623–625.
40. Zhang L, Hermans J (1996) Hydrophilicity of cavities in proteins. *Proteins: Structure, Function, and Genetics* 24(4):433–438.
41. 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.
42. Guo D, et al. (2016) Molecular Basis of Ligand Dissociation from the Adenosine A2A Receptor. *Mol Pharmacol* 89(5):485–491.
43. 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.
44. 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.
45. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38.
46. 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.

- a. Mobility of water molecule in the transmembrane region of a GPCR
(e.g. δ opioid receptor)
Positions of waters in different conformations sampled in simulation



- b. Water-mediated interactions in crystal structure
42 water-mediated interactions
- Water-mediated interactions stable in simulation
19 stable water-mediated interactions

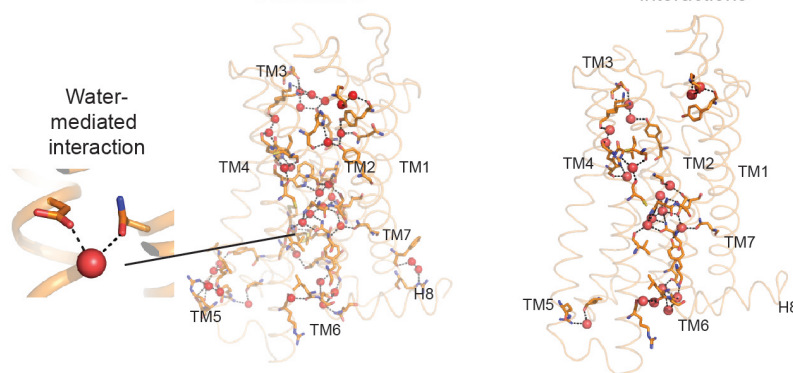


Figure 1

Fig. 1. Stability of water molecules and water-mediated interactions in GPCRs. (a) Positions of water molecules (red spheres) in one region of a GPCR (the δ opioid receptor [DOR]) in three simulation snapshots. Receptor transmembrane helices are shown as orange ribbons. Most waters are mobile, but the water at the TM6 kink (circled) is not. (b) Left: water-mediated interactions present in the crystal structure of DOR. Right: water-mediated interactions formed over 60% of the time in simulation. Hydrogen bonds forming water-mediated interactions are shown as black dashed lines, and the residues involved in these interactions are shown with orange sticks.

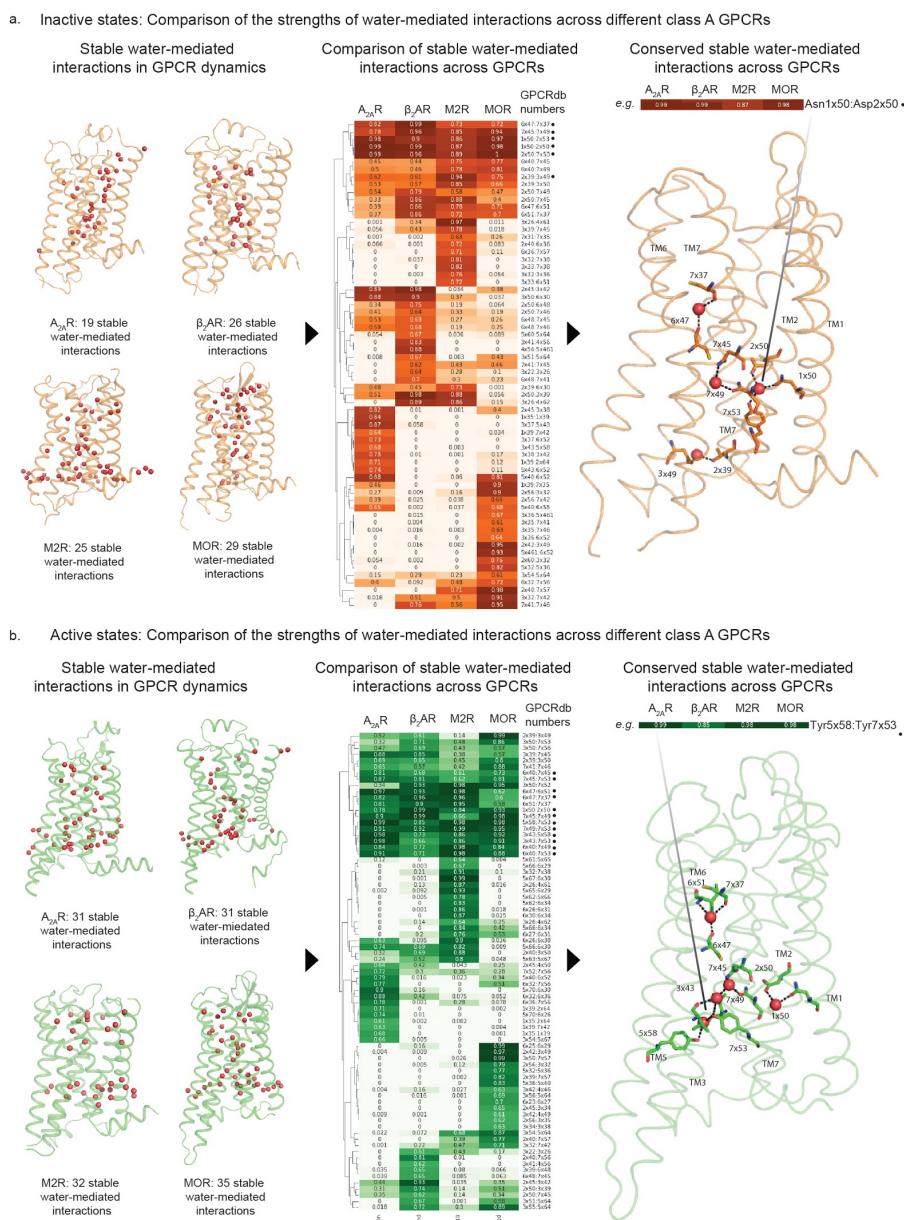
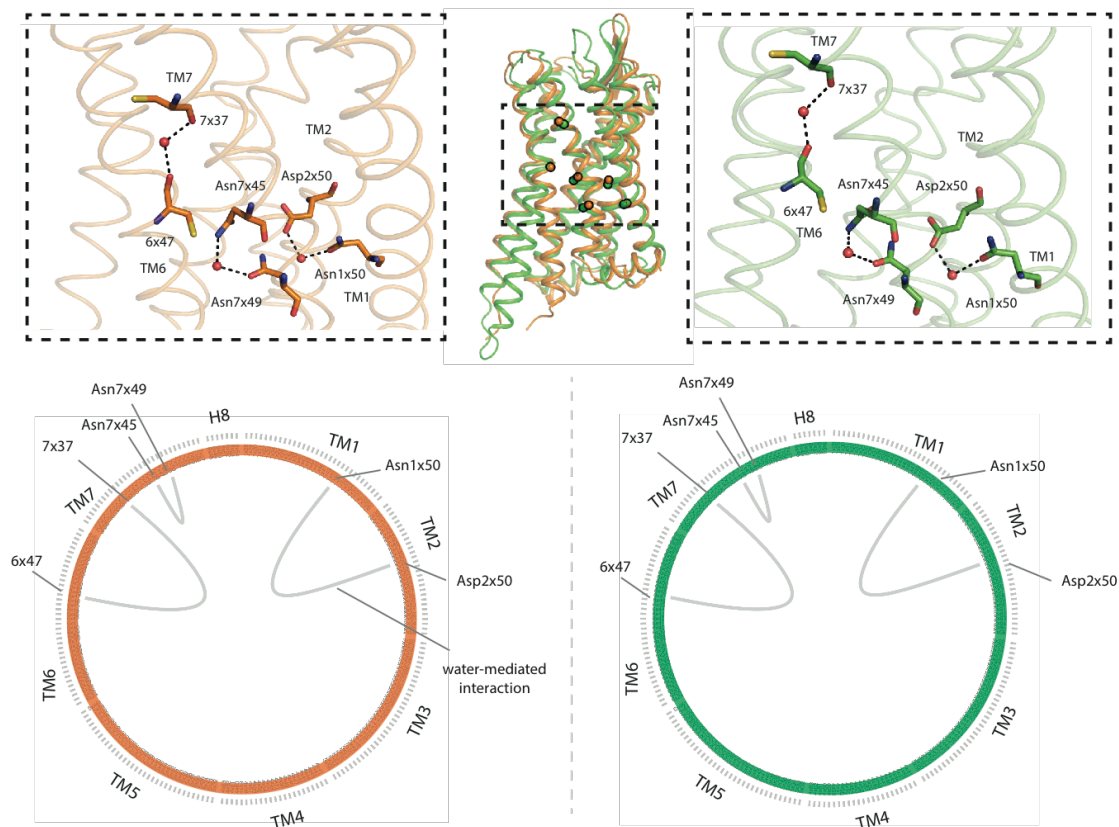


Figure 2

Fig. 2. Conserved water-mediated interactions in inactive and active states of GPCRs. Comparison of the stability of structurally equivalent water-mediated interactions across different GPCRs in (a) inactive states and (b) active states. Four functionally distinct and evolutionarily distant class A GPCRs were compared: *A_{2A}R*, *β₂AR*, M2R, MOR. The stability of the water-

mediated interactions is estimated based on the frequency of formation in simulations. (a-b) Left: Frames from simulations of four functionally distinct GPCRs. Receptors are shown as ribbons and water molecules are shown as spheres. A schematic slider is shown to indicate that the each image represents a simulation frame. Middle: Heatmap comparing the stability of water-mediated interactions. 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 (marked with black dot). Right: Water-mediated interactions that are stable (frequency > 60%) across all four receptors mapped onto a GPCR structure. Hydrogen bonds forming water-mediated interactions are shown as black dashed lines, with the residues involved shown as sticks. Inactive state is shown in orange and active state is shown in green.

a. Water-mediated interactions stable in MD simulations and common to the inactive and the active states of GPCRs



b. Conserved water-mediated interactions across the high-resolution structures of class A GPCRs

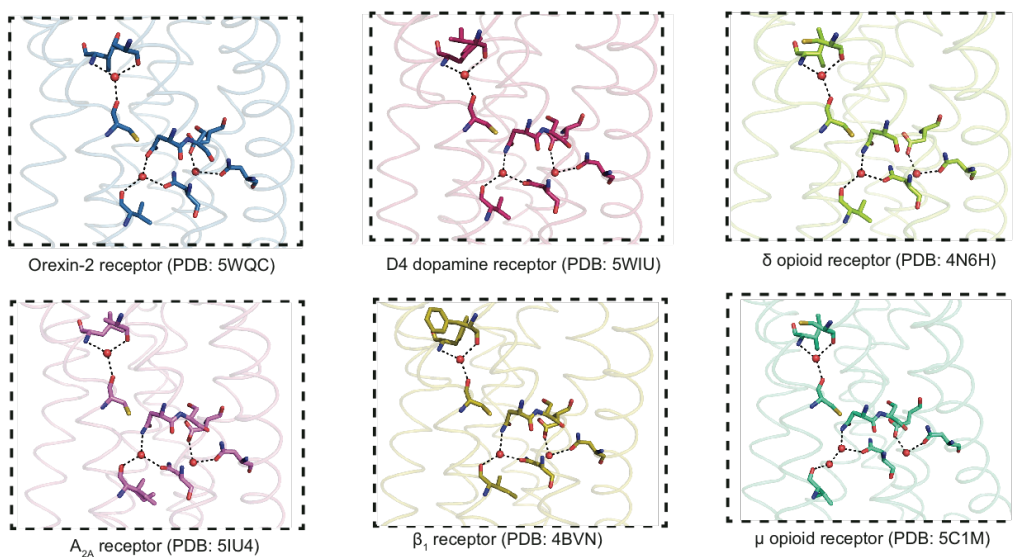
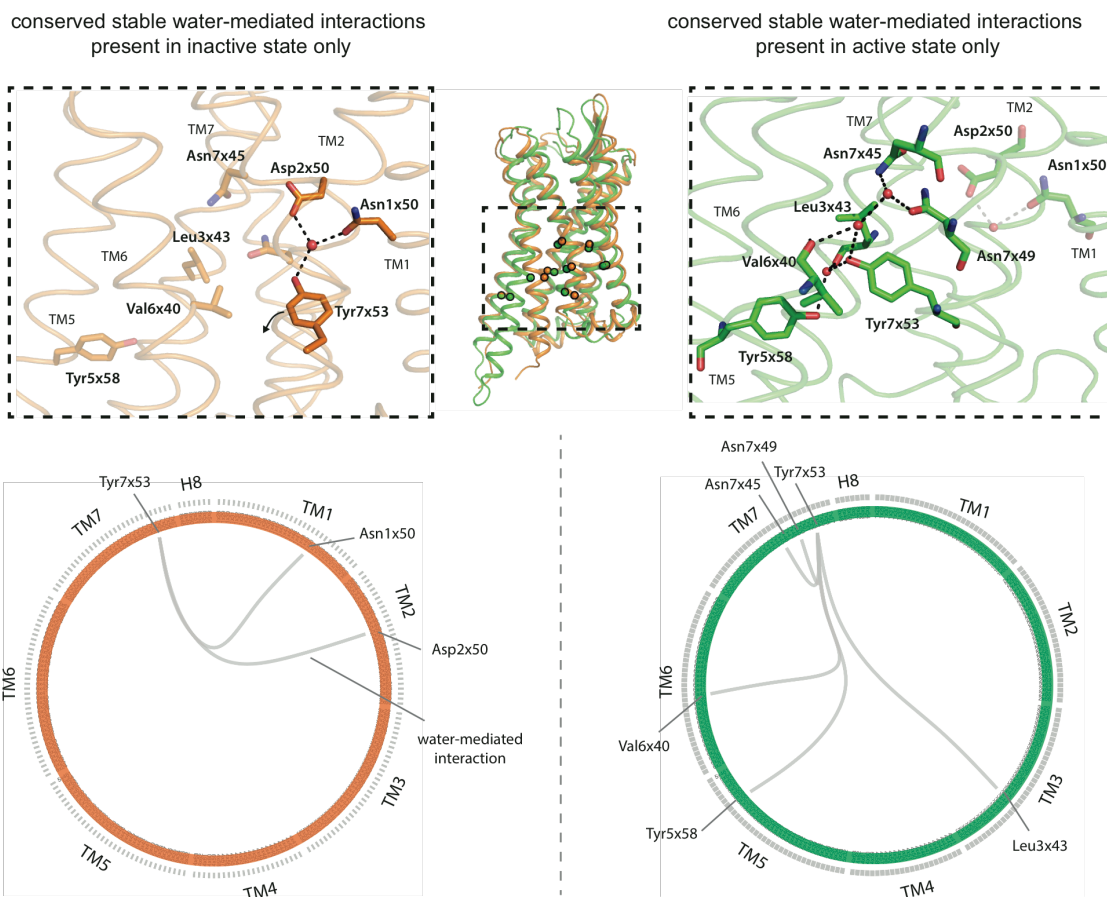


Figure 3

Fig. 3. State-independent water-mediated interaction network: conserved and stable water-mediated interactions maintained across the inactive and the active states of diverse GPCRs. (a) Stable water-mediated interactions (frequency > 60%) present in both the inactive (orange) and the active (green) states of diverse GPCRs. Top: the conserved stable water-mediated interactions mapped onto a structure. Residues are shown as sticks, water molecules are shown as spheres, and hydrogen bonds forming water-mediated interactions are shown as black dashed lines. Bottom: The conserved stable water-mediated interactions are shown using 'flareplots'. In flareplots, the amino acid residues in the transmembrane helices and helix 8 are shown as points on the circle. The water-mediated interactions between the residues are shown as chords connecting the points on the circle. (b) Water-mediated interactions common across the high-resolution structures (resolution of 2.1 Å or lower) of diverse GPCRs: inactive 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, and active μ -opioid receptor (PDB ID: 5C1M). Residues are shown as sticks, water molecules are shown as spheres, and hydrogen bonds forming water-mediated interactions are shown as black dashed lines.

a. Stable water-mediated interactions in MD simulations of the inactive and the active states of class A GPCRs



b. Conserved water-mediated interactions in the inactive and the active state crystal structures of class A GPCRs

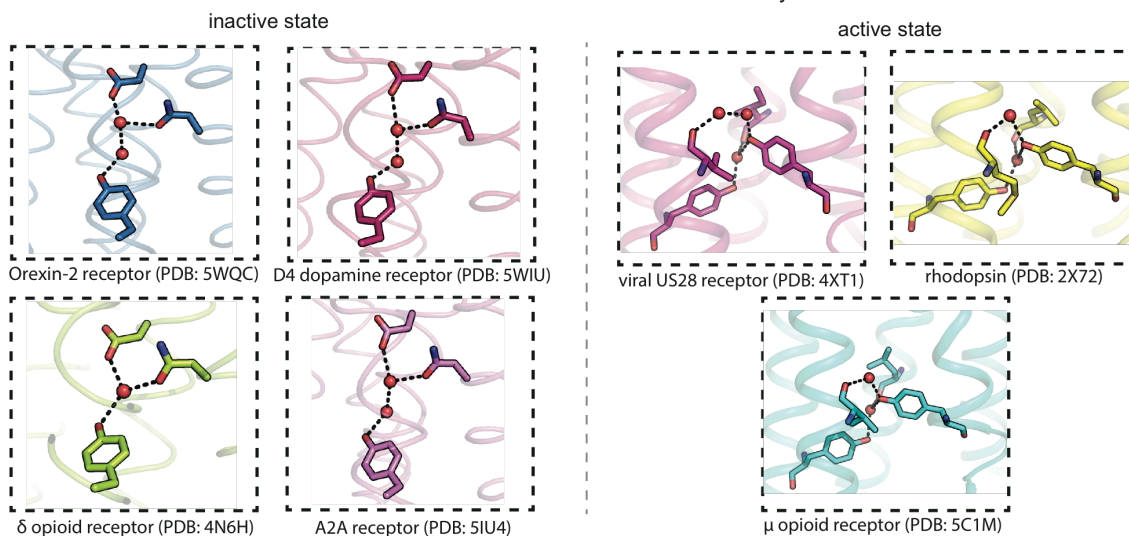


Figure 4

Fig. 4. State-dependent water-mediated interaction network: conserved and stable water-mediated interactions maintained exclusively in the inactive states or the active states of diverse GPCRs. (a) Stable water-mediated interactions (frequency > 60%) present exclusively in all the inactive (orange) or all the active (green) states of diverse GPCRs. Top: the conserved stable water-mediated interactions of inactive state and active states mapped onto structure. Residues are shown as sticks, water molecules are shown as spheres, and hydrogen bonds forming water-mediated interactions are shown as black dashed lines. Bottom: The conserved stable water-mediated interactions are shown using 'flareplots'. (b) Water-mediated interactions present exclusively in the crystal structures GPCRs in inactive state or active state. The high-resolution inactive state structures are of the following GPCRs: Orexin-2 (PDB ID: 5WQC), D4 dopamine (PDB ID: 5WIU), δ -opioid (PDB ID: 4N6H) and A_{2A} adenosine (PDB ID: 5IU4) receptors. The active state crystal structures of the following GPCRs are considered: rhodopsin (PDB ID: 2X72), viral chemokine receptor US28 (PDB ID: 4XT1) and μ -opioid receptor (PDB ID: 5C1M). Residues are shown as sticks, water molecules are shown as spheres, and hydrogen bonds forming water-mediated interactions are shown as black dashed lines.

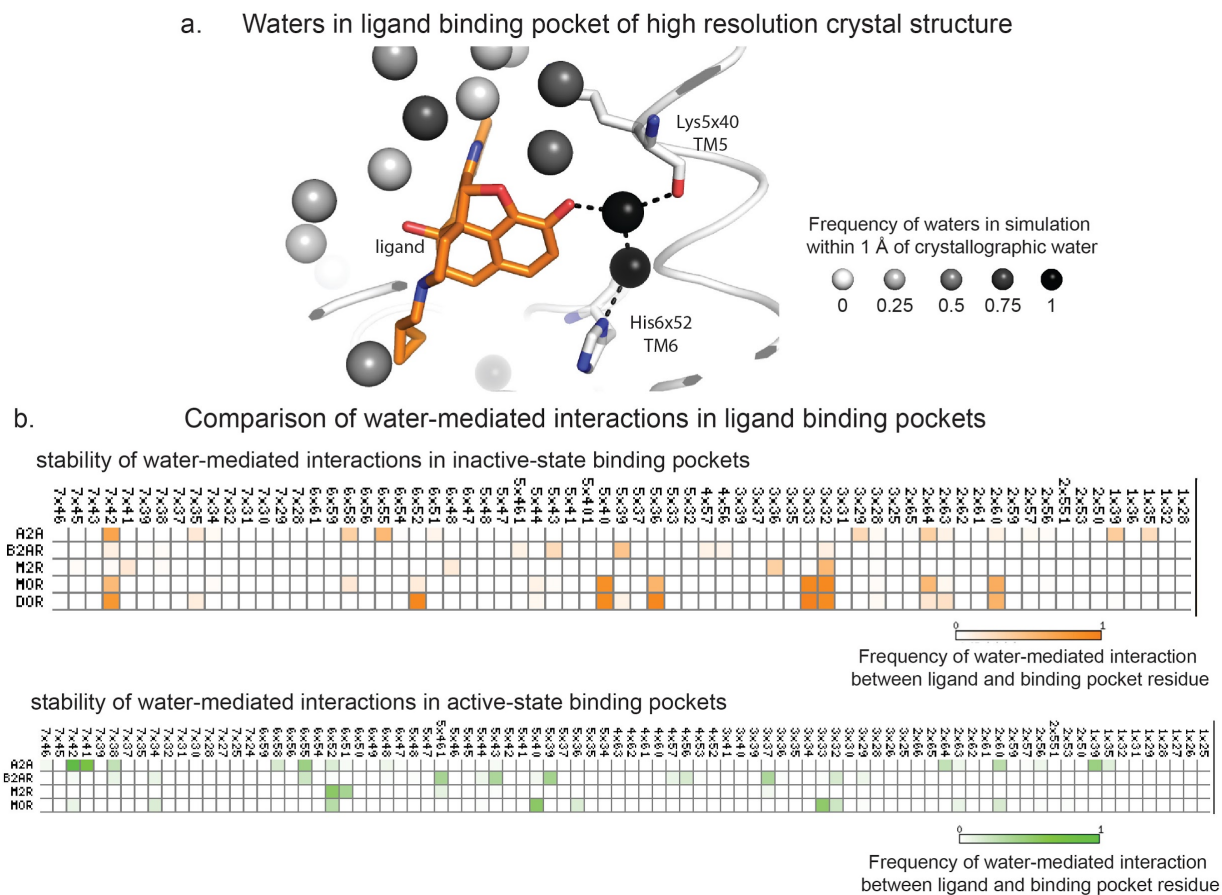
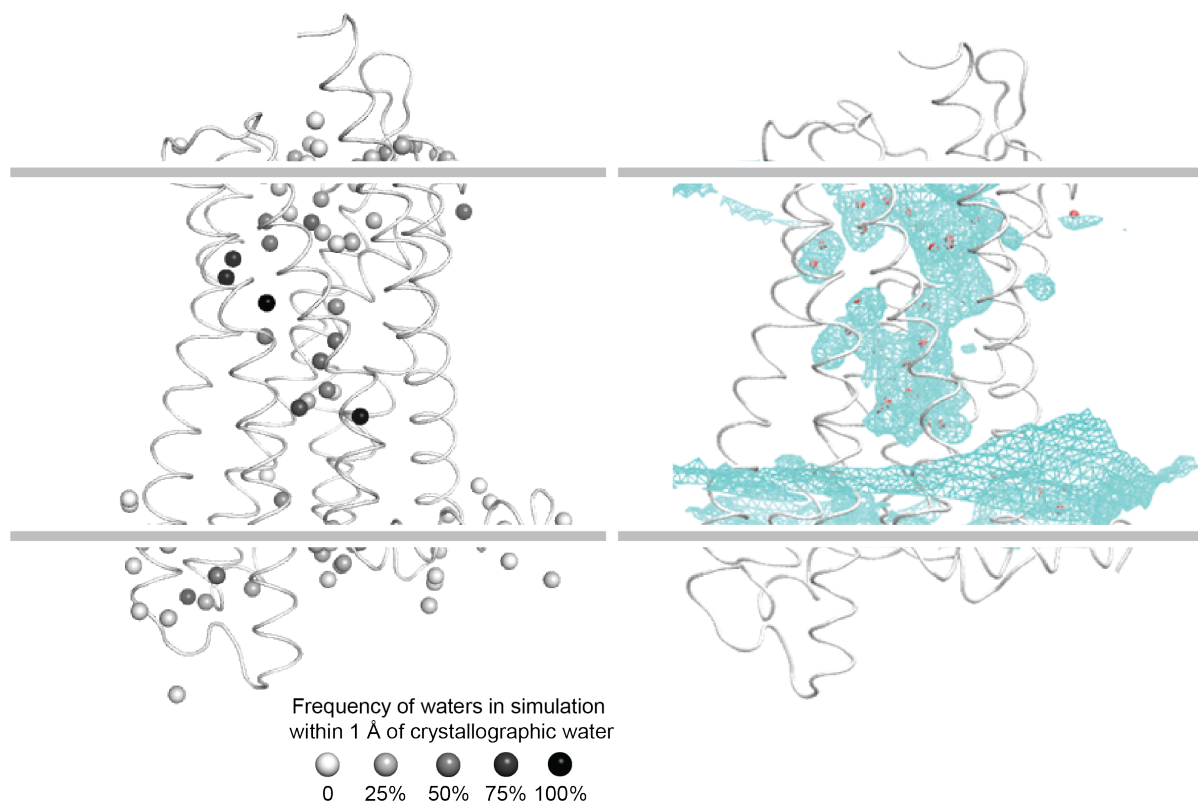


Figure 5

Fig. 5. 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_{2A}AR, 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.

Mobility of water molecules in the transmembrane region of a GPCR



Supplementary Figure S1

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)

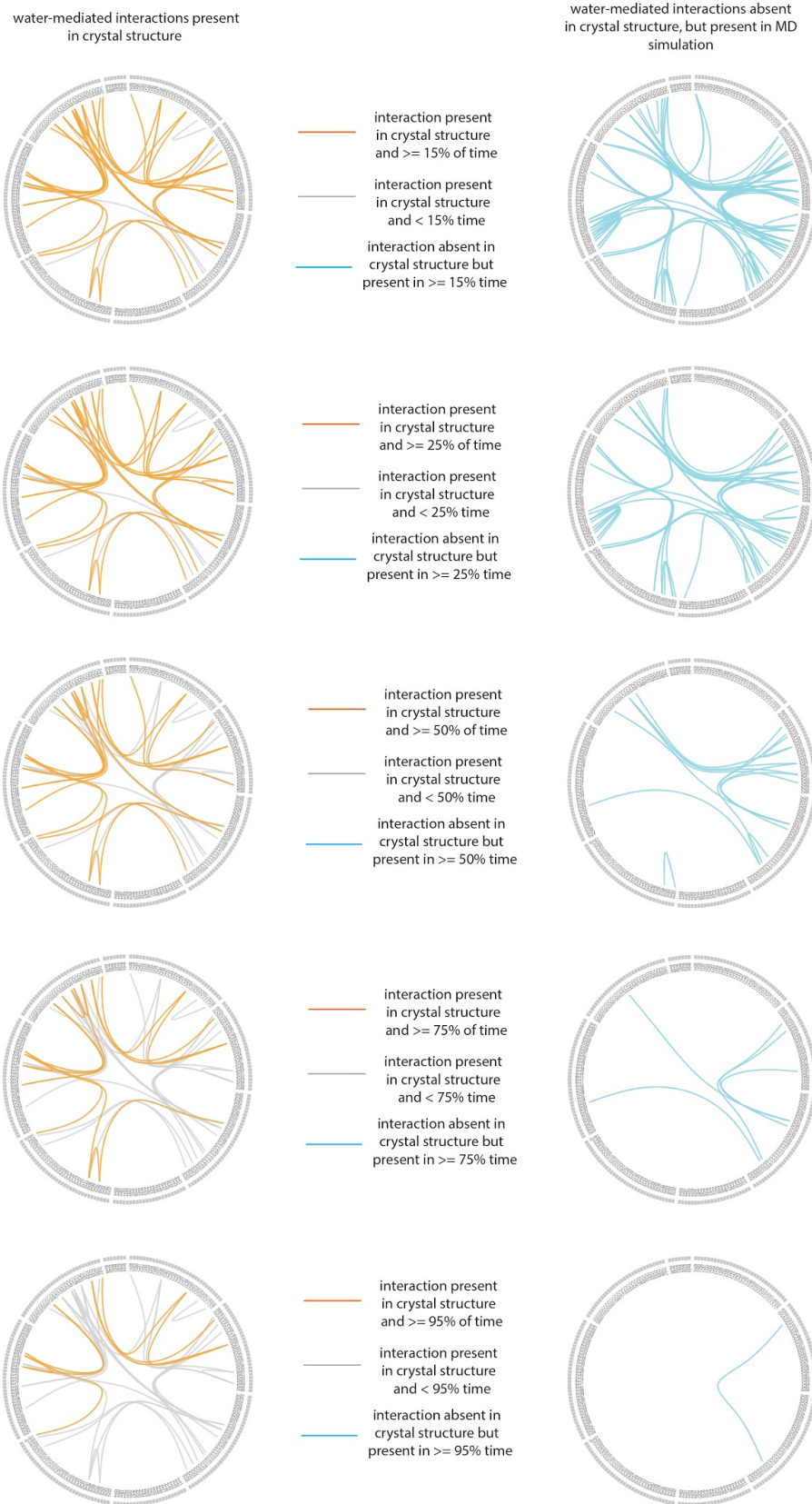


Fig. S2. Stability of water-mediated interactions present in crystal structures and simulation. Water-mediated interactions present in MD simulation of a GPCR were stratified into three sets: interactions that are present in crystal structure and in simulations above a frequency threshold (orange), interactions 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%.

GPCR	PDB ID	State	Ligand	Publication	Intracellular interface	Simulation length	Force field
B2AR	2RH1	Inactive	Carazolol	Nygaard et al., Cell 2013 (Condition: 12)	-	10 us	CHARMM27 protein force field with modification to Asp, Glu, and Arg side chains and CHARMM36 lipid force field
DOR	4N6H	Inactive	Naltrindole	unpublished	-	1.7 us, 1.3 us, 2us	CHARMM36
M2R	3UON	Inactive	QNB	Kruse et al., Nature 2012 (Condition: A)	-	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	-	1.2 us, 1.4 us, 0.6 us	CHARMM36
A2AR	5IU4	Inactive	ZM241385	unpublished	-	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)	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	unpublished	Nanobody interface restrained	1.4 us, 1.6 us, 1.60 us	CHARMM36
MOR	5C1M	Active	BU72	unpublished	Nanobody interface restrained	3.1 us, 2.8 us, 2.9 us	CHARMM36
A2AR	5G53	Active	NECA	unpublished	Mini-Gs protein interface restrained	1.2 us, 1.2 us, 1.1 us, 1.0 us, 1.1 us	CHARMM36

SupplementaryTable 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