Journal of Structural Biology 212 (2020) 107634

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Protein-water hydrogen-bond networks of G protein-coupled receptors: Graph-based analyses of static structures and molecular dynamics

Éva Bertalan [a], Samo Lešnik [a,b], Urban Bren [b,c], Ana-Nicoleta Bondar [a,*]

[a] Freie Universität Berlin, Department of Physics, Theoretical Molecular Biophysics Group, Arnimallee 14, D-14195 Berlin, Germany
[b] University of Maribor, Faculty of Chemistry and Chemical Engineering, Smetanova ulica 17, SI-2000 Maribor, Slovenia
[c] University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, SI-6000 Koper, Slovenia

ARTICLE INFO

Keywords:
Dynamic hydrogen-bond network
Graphs
Network analysis
Clustering algorithm
GPCR
Opioid receptor

ABSTRACT

Protein and protein-water hydrogen bonds shape the conformational energy landscape of G Protein-Coupled Receptors, GPCRs. As numerous static structures of GPCRs have been solved, the important question arises whether GPCR structures and GPCR conformational dynamics could be described in terms of conserved hydrogen-bond networks, and alterations of these hydrogen-bond networks along the reaction coordinate of the GPCR. To enable efficient analyses of the hydrogen-bond networks of GPCRs we implemented graph-based algorithms, and applied these algorithms to static GPCR structures from structural biology, and from molecular dynamics simulations of two opioid receptors. We find that static GPCR structures tend to have a conserved, core hydrogen-bond network which, when protein and water dynamics are included with simulations, extends to comprise most of the interior of an inactive receptor. In an active receptor, the dynamic protein-water hydrogen-bond network spans the entire receptor, bridging all functional motifs. Such an extensive, dynamic hydrogen-bond network might contribute to the activation mechanism of the GPCR.

1. Introduction

G Protein-Coupled Receptors (GPCRs) are seven-helical membrane proteins that mediate communication between cells and their environment. In response to an extracellular signal, such as binding of a ligand, the GPCR undergoes a conformational change leading to the activated state of the receptor that binds a G protein at the cytoplasmic side. Both GPCRs (Hauser et al., 2017; Lappano and Maggiolini, 2011) and G proteins (Li et al., 2019) are targets for drug design, and knowledge of conformational changes associated with activation of the GPCR could inform and guide the design of new therapeutics. Here, we present a methodology that relies on graph-based algorithms and clustering techniques from machine learning to compute hydrogen (H)-bond networks from data sets of experimental GPCRs structures, and from computer simulations. We use this methodology to identify H-bond interaction networks that could contribute to long-distance propagation of conformational change in GPCRs.

The largest family of GPCRs is class A, the rhodopsin-like family (Erlandson et al., 2018). Class A GPCRs share conserved motifs denoted as molecular switches –amino acid residues of the switches undergo structural rearrangements upon activation of the GPCR (Trzaskowski et al., 2012) (Fig. 1). At the cytoplasmic side, the DRY motif (D3.49, R3.50, Y3.51) is implicated in G-protein binding of activated GPCR (Deupi et al., 2012; Rasmussen et al., 2012). The DRY motif is separated from NPxxY (N7.49, P7.50, and Y7.53) in the inactive receptor; in the activated receptor, Y7.53 reorients and bridges with Y5.58 via water, and Y5.58 interacts with R3.50 of DRY (Deupi et al., 2012; Venkatakrishnan et al., 2019; Yuan et al., 2014). At the center of the receptor, W6.48 of the CWxP motif (C6.47, W6.48, P6.50) relocates upon receptor activation and approaches P5.50 of the PIF motif (P5.50, I3.40, F6.44) (Deupi et al., 2012). The sodium binding pocket, present in many Class A GPCRs (Katritch et al., 2014), is delineated by D2.50, N3.35, S3.39 and it is relatively close to the ligand-binding pocket (Fig. 1B).

A fundamental question is how binding of a ligand at the extracellular side of the GPCR leads to conformational changes at the cytoplasmic side, such that the GPCR samples an activated conformation that can bind the G protein. Protein-water H-bond networks, as observed in static structures and molecular dynamics (MD) simulations, are thought to shape the conformational energy landscape of GPCRs (Blankenship et al., 2015; Jardon-Valadez et al., 2009, 2010; Murakami and Kouyama, 2008; Varma et al., 2019; Venkatakrishnan et al., 2019). There are, at the moment, more than 350 structures for GPCRs in Protein Data Bank,

* Corresponding author.
E-mail address: nbondar@zedat.fu-berlin.de (A.-N. Bondar).

https://doi.org/10.1016/j.jsb.2020.107634
Received 21 June 2020; Received in revised form 6 September 2020; Accepted 24 September 2020
Available online 29 September 2020

PDB (Berman et al., 2000). This large data set of static GPCR structures could be used to identify H-bond networks that characterize GPCRs.

To this aim, we used concepts from graph theory and machine learning and implemented a methodology to catalogue and characterize internal H-bond networks in static GPCR structures. We augment the analyses of static GPCR structures with atomistic MD simulations of two class A GPCRs, the δ-opioid receptor, DOR, and the κ-opioid receptor, KOR, which we chose because their static structures are well characterized (Che et al., 2020, 2018; Claff et al., 2019; Fenalti et al., 2014, 2015; Granier et al., 2012; Huang et al., 2015; Koehl et al., 2018; Manglik et al., 2012; Wu et al., 2012), and because they are of direct relevance for developing safe pain medication (Seth et al., 2018).

We found that many of the static structures of GPCRs have conserved water sites, and that some of these water sites are part of a central cluster of H-bonds that connects groups of the sodium-binding pocket with the CWxP and NPxxY motifs; in structures solved at high resolution, several other smaller H-bond clusters are observed, such that at least one H-bond cluster is present along most of the length of the GPCRs. The central H-bond cluster we identified in static GPCR structures is sampled in MD simulations of the DOR and KOR. In active-like KOR, the central H-bond cluster connects to a local cluster at the cytoplasmic side, such that an extensive network of H-bonds, with a linear projection of ~ 68 Å, extends from the extracellular to the cytoplasmic side of the GPCR via groups of all switch motifs.

2. Methods

Set-A, Set-B, and Set-C of static GPCR crystal structures, and internal waters. We used the database for experimentally solved and predicted GPCR structures, GPCR-EXP. As accessed on 22.11.2019, there were 369 structures from 4 families of GPCRs: there were 326 class A GPCR structures, 22 structures for class B, 9 for class C, and 12 for class F. From this database we first extracted as *Set-A* structures of class A GPCRs, i.e., rhodopsin-like receptors; *Set-A* is representative for the best-studied family of GPCRs, for which structures with good resolution are available (Schemes 1 and 2).

Three-dimensional structures of GPCRs were downloaded from the Orientations of Proteins in Membranes server, OPM (Lomize et al., 2011). OPM gives structures oriented along the membrane normal, and an estimation of the membrane planes. We used this information about the location of the membrane plane to select, for each GPCR, the TM helical region.

N and C termini and loop segments in GPCRs are diverse in terms of structure (Fig. S1D) and amino acid composition (Isberg et al., 2015; Mirzadegan et al., 2003) and are not included in the BW numbering scheme. For simplicity, and since we focus on inter-helical H-bond networks of the TM region of GPCRs, we exclude loops and termini from the structural overlaps of GPCRs. We overlapped the TM helices of each GPCR structure with Chimera (Pettersen et al., 2004) using as a reference the A2A Adenosine receptor structure PDB ID:5NM4 (Weinert et al., 2017). This A2A structure was chosen as a reference because its resolution of 1.7 Å is the highest from the entire dataset.

We consider a water molecule as internal if its oxygen atom is within the membrane region indicated by OPM. We found that the number of internal waters largely depends on the resolution at which the structure was solved – the better the resolution, the higher the number of internal waters (Fig. S1).

Structures solved at a resolution of 2 Å or better tend to have at least

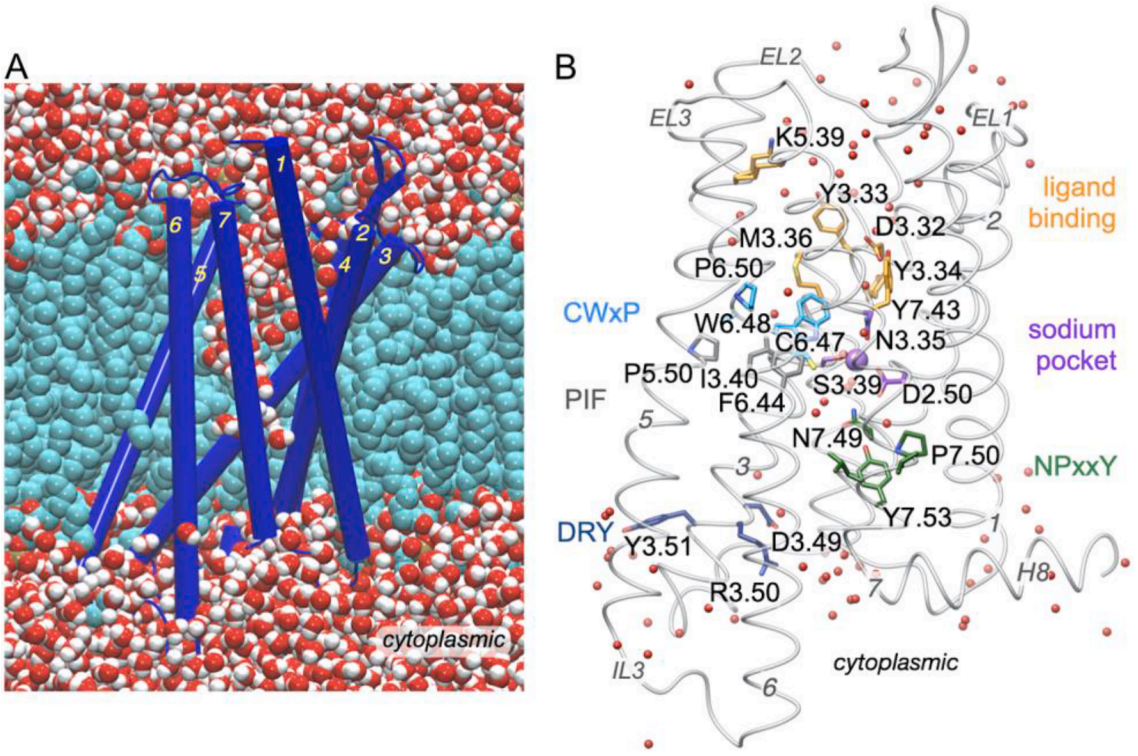


Fig. 1. Architecture and conserved motifs of class A GPCRs. (A) Cut-away view of the DOR embedded in a hydrated lipid membrane, based on a coordinate snapshot from simulations reported here. (B) Conserved motifs of class A GPCRs shown on the DOR structure PDB:4N6H (Fenalti et al., 2014). We use the Ballesteros and Weinstein (BW) scheme (Ballesteros and Weinstein, 1995), which assigns each amino acid residue a number that indicates the transmembrane (TM) helix, and a number that indicates the location relative to most conserved amino acid residue of that TM. Groups of CWxP, PIF, DRY and NPxxY are colored light blue, gray, dark blue and green, respectively; conserved groups of the orthosteric ligand-binding site are colored orange, and groups of the allosteric sodium pocket are colored purple. Red dots indicate water oxygen atoms, and the purple sphere, the sodium ion. Fig. 1A was prepared with Visual Molecular Dynamics, VMD (Humphrey et al., 1996); all other molecular graphics were prepared with UCSF Chimera (Pettersen et al., 2004). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of simulations performed. We report the length of the production runs and the average Cα RMSD values with standard deviations computed for the TM helices. Repeat simulations of the DOR, KOR-inactive, and KOR-active, were performed starting from the same coordinates as the main simulation.

Protein	Length (ns)	RMSD (Å)
DOR	220	1,7 ± 0,2
	200, repeat	1,9 ± 0,2
KOR-inactive	205	1,8 ± 0,1
	200, repeat	1,9 ± 0,2
KOR-active	195	1,6 ± 0,2
	200, repeat	1,9 ± 0,3

30 internal waters, with only a handful of structures solved at this resolution having 15–20 waters (Fig. S1). Most of the structures solved at a resolution of 2.5 Å or higher have at least 10 internal waters (Fig. S1). The average number of internal waters computed for all structures solved at a resolution of at least 2.5 Å is 25. The minimum resolution of a structure that contains 25 internal waters is 2.3 Å. Based on this qualitative assessment, and to be able to compute protein-water H-bond networks, we chose as *Set-A* the 25 GPCR structures that were solved at resolution 2.3 Å or better, and contain at least 25 internal water molecules (Table S1).

The criteria for selecting structures are, however, not unique, and we analyzed separately structures solved at lower resolution and/or with

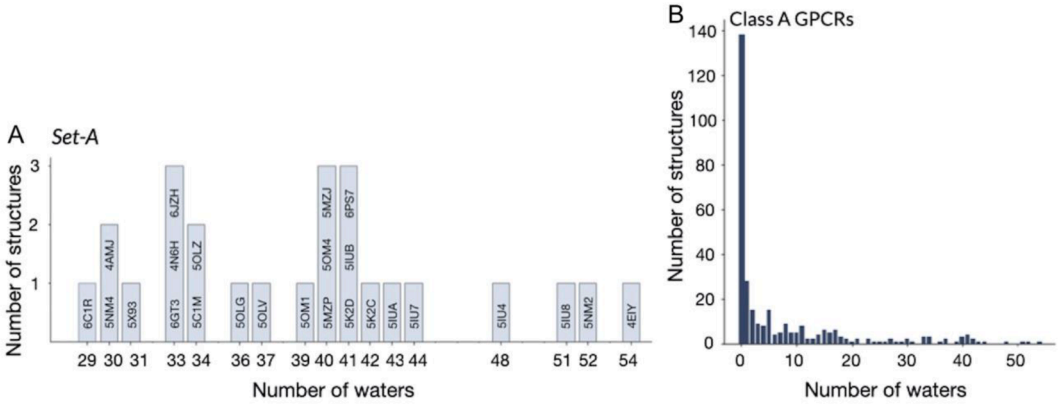


Fig. 2. Internal waters in static GPCR structures. (A) Number of internal waters in GPCR structures from *Set-A* (panel A). Structures are identified by their PDB IDs, which are listed in Table S1 together with the resolution at which structures were solved. (B) Histogram of the number of internal waters from all class A GPCR structures in the GPCR-EXP database, as accessed on 22.11.2019. Additional analyses of the number of internal waters in static GPCR structures are presented in Figure S1.

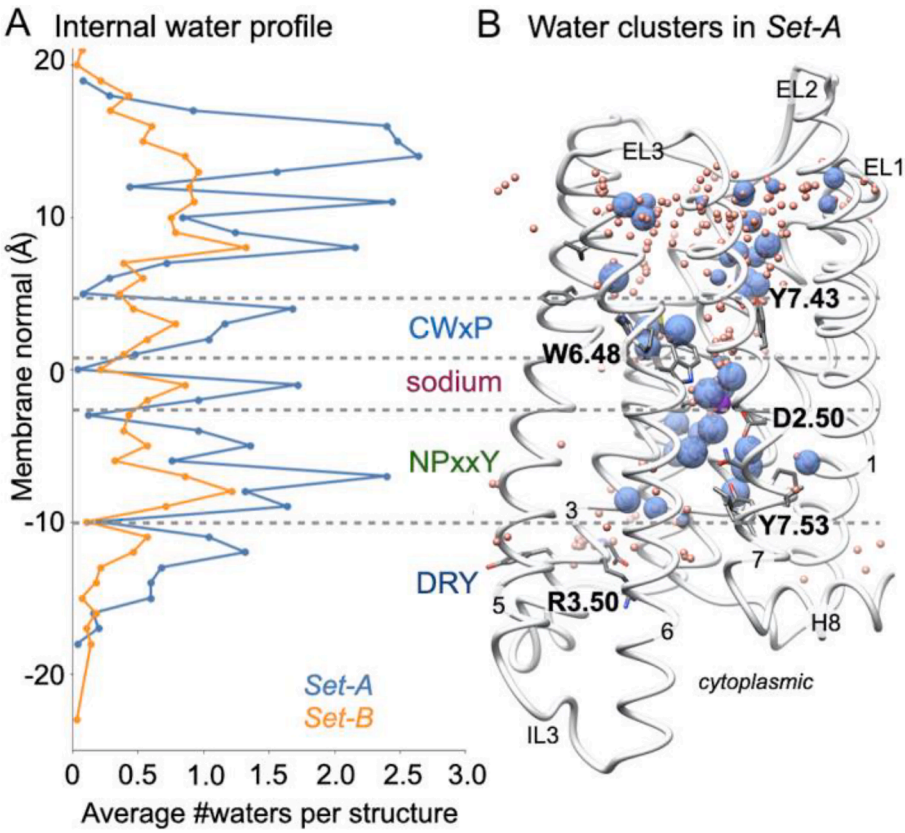


Fig. 3. Location of internal waters in static GPCR structures. Conserved motifs separate regions where waters are likely to be found. (A) Distribution of internal waters along the membrane normal in GPCR structures from *Set-A* vs. *Set-B*. (B) Molecular graphics illustrating internal water clusters in *Set-A*. For protein coordinates we used the DOR structure PDB:4N6H (Fenalti et al., 2014). Tests used for DBSCAN settings, and additional analyses of internal waters in static GPCR structures, are presented in Figures S2-S5.

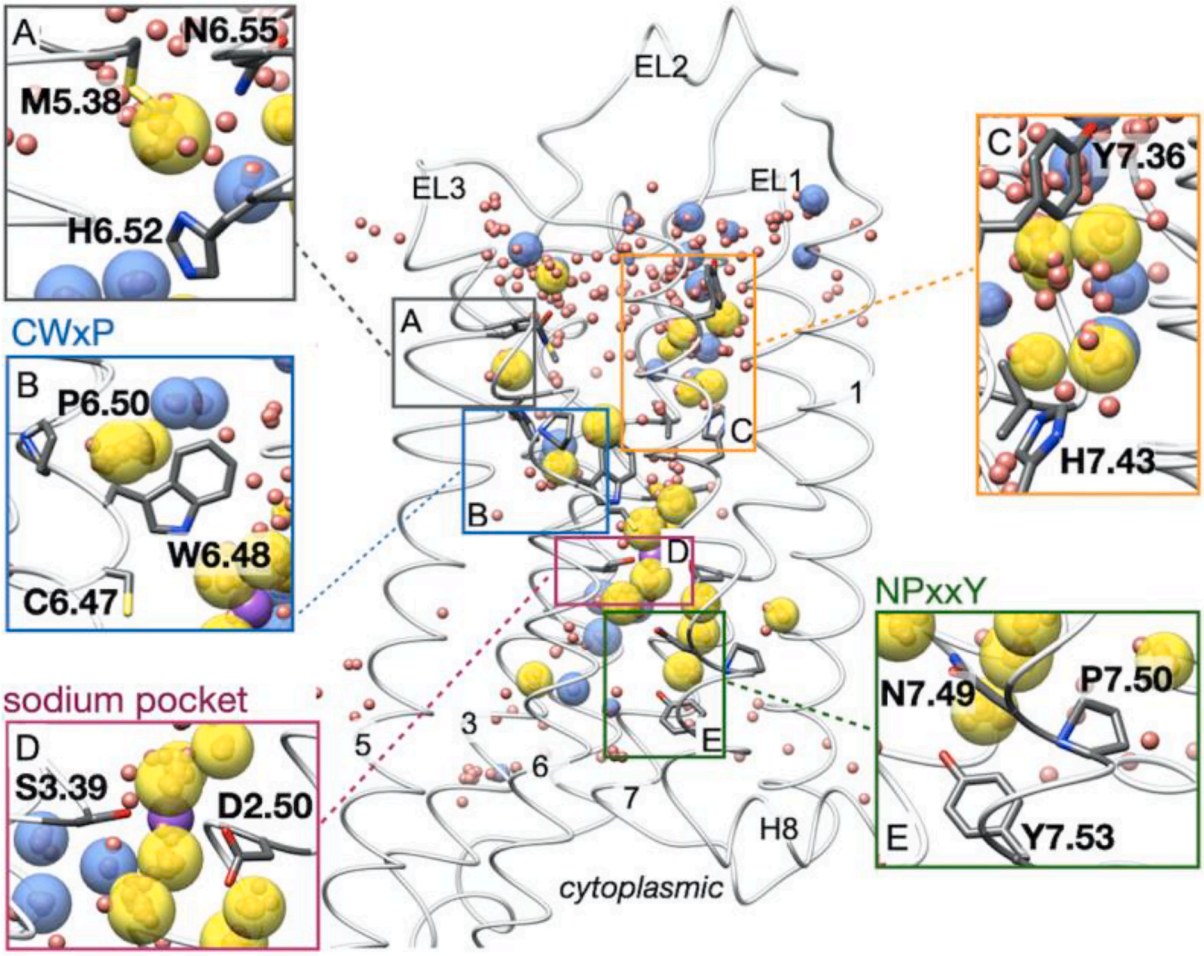


Fig. 4. Conserved water clusters in GPCRs. Spheres indicate water clusters obtained with DBSCAN. Spheres colored yellow indicated conserved water sites present in both *Set-A* and *Set-B*; spheres colored blue are present only in *Set-A*; spheres colored light-red indicate internal waters in the superimposed structures of *Set-A* that are not part of a cluster, and thus are considered outliers. The central panel gives an overview of the distribution of the water clusters and of the outlier waters. (A-E) Close view of conserved water sites near molecular switches. The molecular graphics are based on the A2A structure (PDB ID:5IU4) from ref. (Segala et al., 2016). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

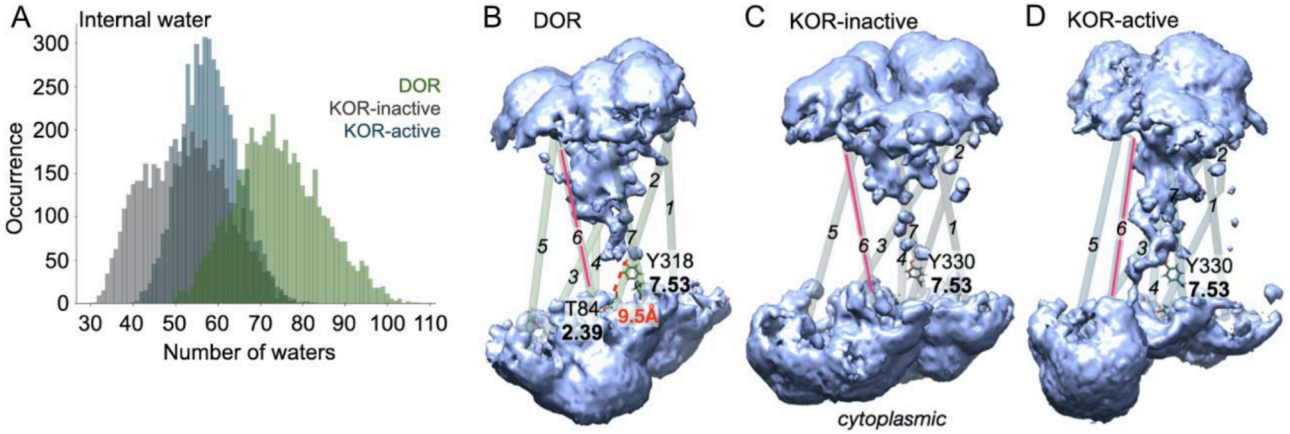


Fig. 5. Internal waters in MD simulations of opioid receptors. Note that, during simulations, DOR and KOR remain structurally stable, with relatively small values of the Cα root mean squared distances (RMSD) for the α-helical region (Figures S6 A-C). (A) Histogram of the number of internal waters during the last 100 ns of MD simulations of KOR-active (blue), KOR-inactive (gray), and DOR (green). (B-D) Volume map of water occupancy in the last 10 ns of each simulation. In DOR (panel B) and KOR-inactive (panel C), the water density is interrupted at Y7.53. Molecular graphics for the volume of water occupancy were generated with Chimera by considering water molecules within 5 Å of the protein. That more waters visit the DOR than the KOR could be due to the DOR having a less tilted and longer TM1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

fewer internal waters. *Set-B* contains 28 structures with resolution 2.3–2.5 Å, and with 10–25 internal water molecules (Table S2). *Set-C* contains 255 structures with resolutions below 2.5 Å; of the *Set-C* structures, 139 lack coordinates for internal water, and the remaining 116 have on the average 3 internal waters.

Clustering algorithm to identify clusters of internal water molecules, and tests for internal waters in Set-A and Set-B of GPCR structures. To determine regions inside GPCR structures where there is a high probability to find waters, we used Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996), which is an unsupervised machine learning algorithm.

For each structure, we used as data points the Cartesian coordinates of the oxygen atoms of the internal waters. We used DBSCAN to group the closely packed data points into clusters, and to mark the low-density regions as outliers. In what follows we summarize the algorithm used to cluster internal waters of GPCRs with DBSCAN; we implemented this algorithm in Python with Scikit-learn (Pedregosa et al., 2011).

Compared to other clustering algorithms such as *k*-Means (Alsabti et al., 1997), DBSCAN has the advantage that it does not require a predetermined number of clusters. Instead, DBSCAN uses two hyperparameters. The first of these hyperparameters is *eps,* defined as the maximum distance between two samples that are considered as neighbors, i.e., members of the same cluster. The second hyperparameter is the minimum number k of data points within a neighborhood to be considered as a cluster. We chose this number as $k = 5$, that is, at least 20% of the structures from the data set must have water at the same position for these waters to be assigned to the same cluster; in the case of *Set-A*, this would mean at least 5 of the 25 structures.

To find the optimal value for *eps* in the case of the internal waters of GPCRs, we calculated the Euclidean distance of the $k = 5$ nearest waters to each water molecule, and plotted the distances in ascending order to determine the density distribution of the data points. The optimal *eps* value is located around the steepest curvature of the graph (Naik and Sawant, 2013), thus between 1 and 2 (Fig. S2A).

To evaluate the quality of the clustering we used the Silhouette score (Rousseuw, 1987), which indicates how well the algorithm assigns data points to clusters: the more compact the cluster is, the better the clustering. The silhouette s(*i*) for each data point *i* is given by the equation

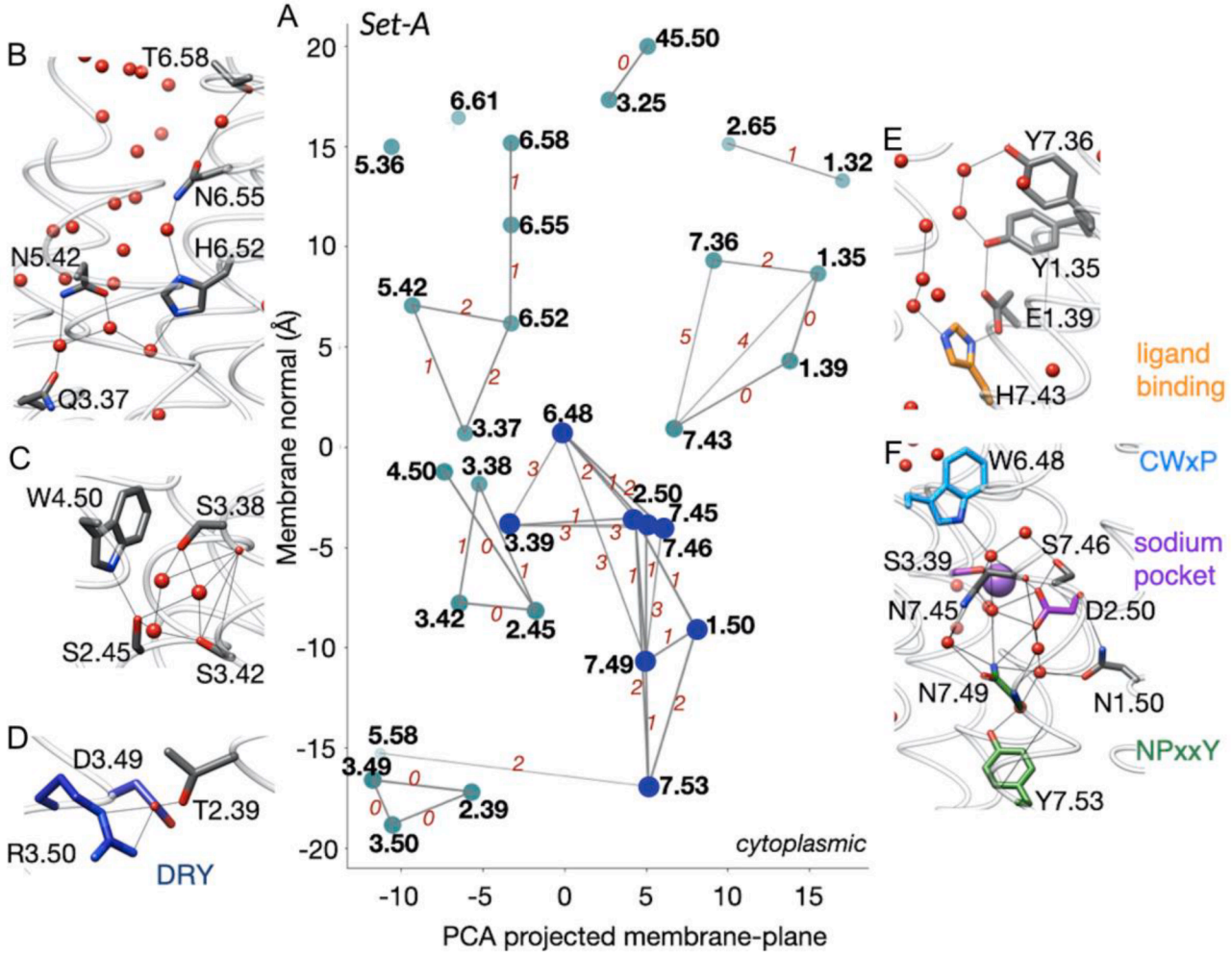


Fig. 6. Conserved H-bond clusters in static GPCR structures. Amino acid residues are labeled according to the BW scheme. (A) Distribution of conserved H-bond clusters identified for *Set-A*. For clarity, we show only H-bonds present in at least 40% of the structures in *Set-A*. Nodes on the graph without any connection are amino acid residues involved in H-bonds with different partners, such that the H-bond frequency is less than 40% within the dataset. The size of the nodes and the thickness of the edges indicate the H-bond frequency; numbers in red give the average number of water molecules that bridge two amino acid residues. C3.25 and C45.50 are disulfide bridged, and are included in the H-bond graph to help illustrate the relative location of H-bond clusters in the protein. (B-F) Molecular representation of conserved local H-bond clusters shown schematically in panel A. Water oxygen atoms are shown as red spheres, and H-bonds as thin gray lines. (F) Molecular representation of the conserved core network of GPCR structures. The molecular graphics are based on the A2A Adenosine receptor, PDB ID:5NM4 (Weinert et al., 2017). Corresponding analyses for *Set-B* structures are presented in Figure S12. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

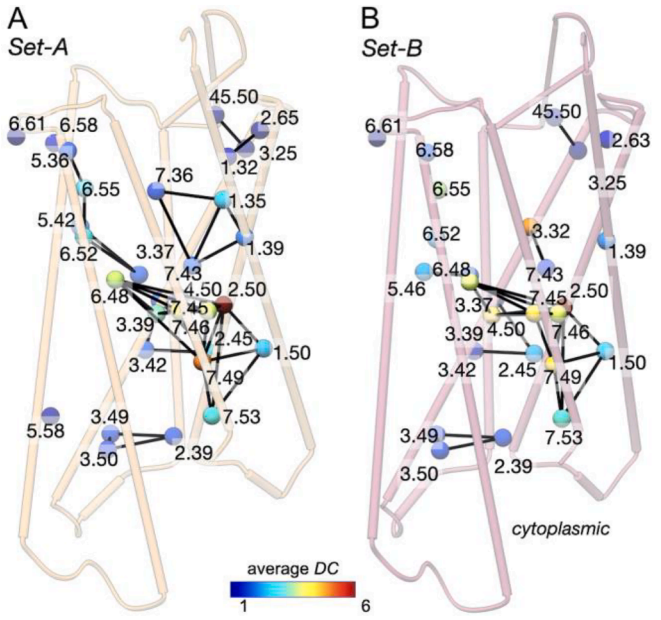


Fig. 7. Centrality measures for conserved H-bond networks of *Set-A* and *Set-B*. Spheres indicate amino acid residues part of the conserved network and are colored according to their average *DC* values (Table S3, S4). Black lines represent the conserved H-bonds. (A-B) Conserved H-bond networks of GPCR *Set-A* (panel A) and *Set-B* (panel B). *BC* values of the H-bond groups are presented in Figure S16.

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \quad (1)$$

where a is the mean distance between point i and all other data points in the same cluster, and b is the minimum mean distance between point i and another, nearest cluster (Rousseuw, 1987). Averaging over the s(i) values gives the Silhouette score, SS:

$$SS = \bar{s}(eps) \quad (2)$$

where $\bar{s}$ is the average silhouette value. The Silhouette Score *SS* depends on the parameters of the clustering algorithm; a high *SS* score indicates compact clusters that are well distinguished from each other (Rousseuw, 1987).

We calculated the number of clusters, the number of outliers, and *SS* values for each value of *eps* ranging from 1 to 2, with step size of *eps* of 0.1 (Fig. S2B). The highest *SS* corresponds to 47 clusters and 169 outlier points in *Set-A* (Fig. S3), and gives the optimal value for *eps* as 1.4.

Water molecules of *Set-B* were assigned to the 47 cluster centers found in *Set-A* with the same *eps* and minimum cluster size value *k* as used for *Set-B*. The combined clustering analyses for internal waters of *Set-A* and *Set-B* were then used to derive a subset of conserved clusters present in both sets (Scheme 3). *Set-C* structures contain few or no internal water molecules, and thus were not subjected to DBSCAN analyses of conserved water sites.

Graphs of H bonds. Graphs are defined as a pair of sets of vertices (nodes) and edges (links between the nodes). In the analyses we report here, *nodes* are amino acid residues that can H-bond, and *edges* are H-bonds between amino acid residues. H-bonds between two protein amino acid residues can be direct, or mediated by H-bonded water molecules. We label the nodes of the graphs according to the BW scheme.

Computations of H-bond networks in GPCR structures. To identify H-bond networks of GPCRs crystal structures we used the TM domains as described above, and an additional 4 Å slab on each side of the membrane that allowed us to include in computations the extracellular loop EL2 and the cytoplasmic DRY motif. To compute H-bond networks from simulation trajectories we used the entire protein.

All analyses of H-bond networks were performed with Bridge (Siemers et al., 2019). Bridge represents the protein H-bond network as a graph whose nodes are protein H-bonding groups, and whose edges are protein–protein H-bonds, or water-mediated H-bonds between two protein groups. With the Connected Components search, Bridge identifies continuous connections between two nodes of the graph of H-bonds (Siemers et al., 2019). In simulation time series, when two end

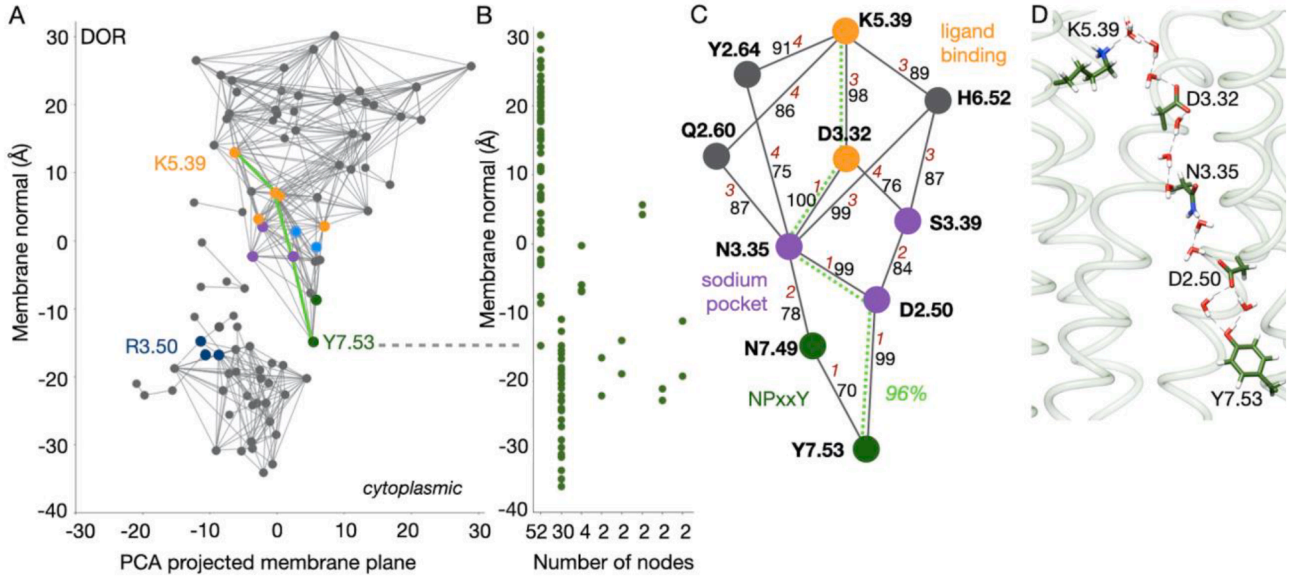


Fig. 8. Water-mediated H-bond networks in DOR simulations. (A) Internal protein-water H-bond network. For clarity, we show H bonds with occupancies ≥ 10% during the last 100 ns of the simulation. Nodes in color other than gray represent amino acid residues that are part of a conserved GPCR motif. (B) Linear length of continuous H-bond networks. (C) Close view of shortest H-bond paths between K5.39 and Y7.53. Numbers in italics red indicate the average number of waters of the bridge, and numbers in gray indicate the occupancy of the bridge. The green dotted line indicates the long-distance H-bond path between K5.39 and Y7.53. (D) Molecular graphics of the long-distance water-mediated path that connects K5.39 and Y7.53. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

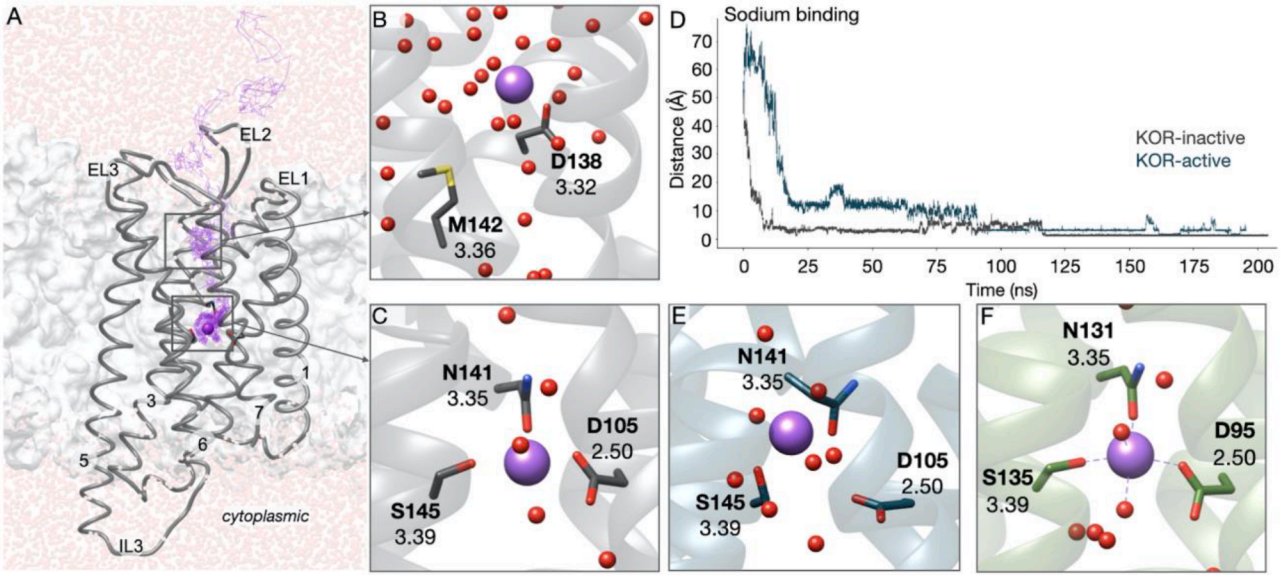


Fig. 9. Sodium ion interactions in opioid GPCR. (A) A sodium ion enters from the bulk and binds near the internal sodium-binding pocket of KOR-inactive. (B-C) Close view of interaction sites of the sodium ion in KOR-inactive. First, within ~ 5–10 ns, the sodium ion binds near D3.32 and M3.36 (panel B), then it moves at the sodium-binding pocket delineated by N3.35, D2.50, S3.39 (panel C). (D) Time series of the distance between the sodium ion and the mid-point of the binding pocket in simulations of KOR-inactive (gray trace) and KOR-active (green trace); in DOR, the sodium-binding site remained empty. Distances are reported in Å. (E) Close view of the groups interacting with the sodium ion at the end of the KOR-active simulations. (F) Close view of groups interacting with the sodium ion in the starting DOR structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

nodes are inter-connected by several segments of protein-water H-bond wires, the overall occupancy of the entire H-bond path between the two end nodes is reported as a Joint Occupancy, *JO* (Siemers et al., 2019).

For searches of H-bond networks we considered H-bonds between protein sidechains and water-mediated H-bonds between these protein groups. For simplicity, we restricted to 5 the number of H-bonded waters.

Linear length of H-bond networks of GPCRs. Initial analyses of the H-bond network of the GPCR structures from our training set indicated extensive networks of H-bonds that can include protein groups from all 7 TM helices of the protein. To derive simple representations of these complex H-bond networks, we projected the H-bond network on the membrane normal, and computed the linear length of the projected network.

We define the linear length of a H-bond path as the distance, reported in Å, between the *z*-coordinates of the Cα atoms of protein groups which form the continuous H-bond chain projected on the membrane normal (Scheme 4).

Visualization of three-dimensional (3D) H-bond graphs in 2D planes. To visualize 3D H-bond graphs in 2D planes, the Cartesian *x*- and *y*-coordinates (i.e., coordinates within the membrane plane) of each atom was projected into a 1D line computed with Principal Component

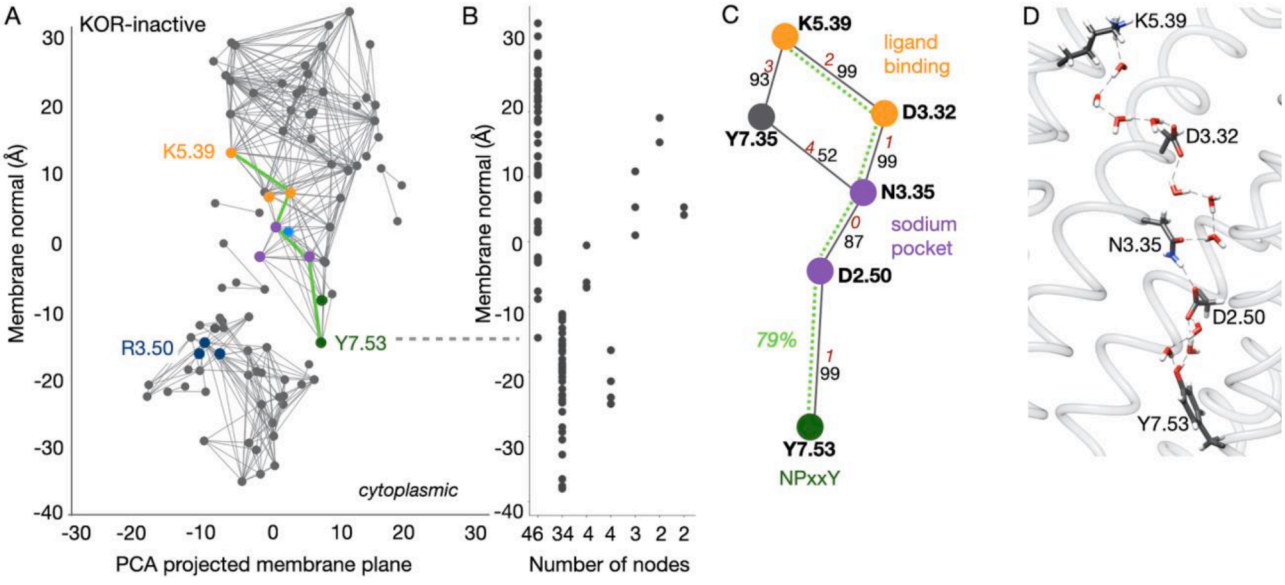


Fig. 10. Water-mediated H-bond network in KOR-inactive simulations. (A) Extended protein-water H-bond networks in the inter-helical region of the receptor. There are two main networks, one extending from the extracellular side to Y7.53, and the other from −10 Å of the middle of the receptor to the cytoplasmic site. (B) Linear length of continuous H-bond networks. (C) A long-distance H-bond path with high *JO* value (79%) connects K5.39 to Y7.53. (D) Molecular graphics of the long-distance path identified in panel C.

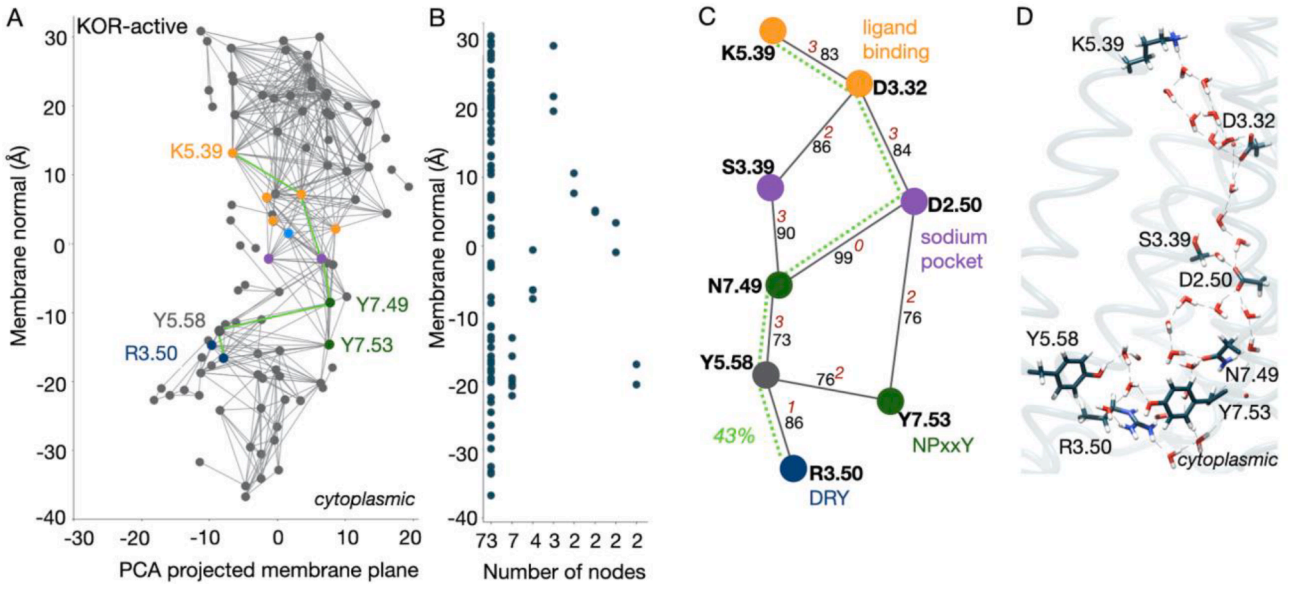


Fig. 11. Dynamic H-bond network of KOR-active. (A) In KOR-active simulations, a H-bond network extends from the cytoplasmic to the extracellular sides of the receptor. (B) Linear length of continuous H-bond networks. (C) A continuous H-bond path with $JO = 43\%$ interconnects the DRY motif, the ligand-binding pocket, the sodium pocket, and the NPxxY motif. (D) Molecular graphics illustrating the long-distance path identified in panel C.

Analysis, PCA. This line is defined by the orthonormal eigenvector with the largest variance, i.e., largest eigenvalue, of the diagonalized covariance matrix of the atomic coordinates relative to their average values. The *z*-coordinate of each atom was kept along the membrane normal such that, in all 2D-projections of the H-bond graphs, the distance along the membrane normal (i.e., the *z*-axis of the Cartesian coordinate system) between two protein groups gives their relative location. All calculations of the 2D planes were performed in Python with Scikit-learn (Pedregosa et al., 2011).

Shortest H-bond paths, Betweenness Centrality and Degree Centrality. The shortest path between two nodes is the path with the least number of edges (Scheme 5). The Betweenness Centrality (*BC*) (Freeman, 1977) of a node n is defined as

$$BC(n) = \sum_{i \neq n \neq j} \frac{\sigma_{ij}(n)}{\sigma_{ij}} \tag{3}$$

where σ_{ij} is the total number of shortest pathways between nodes i and j, and $\sigma_{ij}(n)$ is the number of shortest paths that include that node. The Degree Centrality (*DC*) is the number of connections a node has. In a H-bond graph, the *DC* value of a specific amino acid residue gives the number of H-bonds with its neighbors (Scheme 5).

Conserved nodes and edges of a H-bond graph. We used Bridge to compute the H-bond network of each GPCR structure from *Set-A*, *Set-B*,

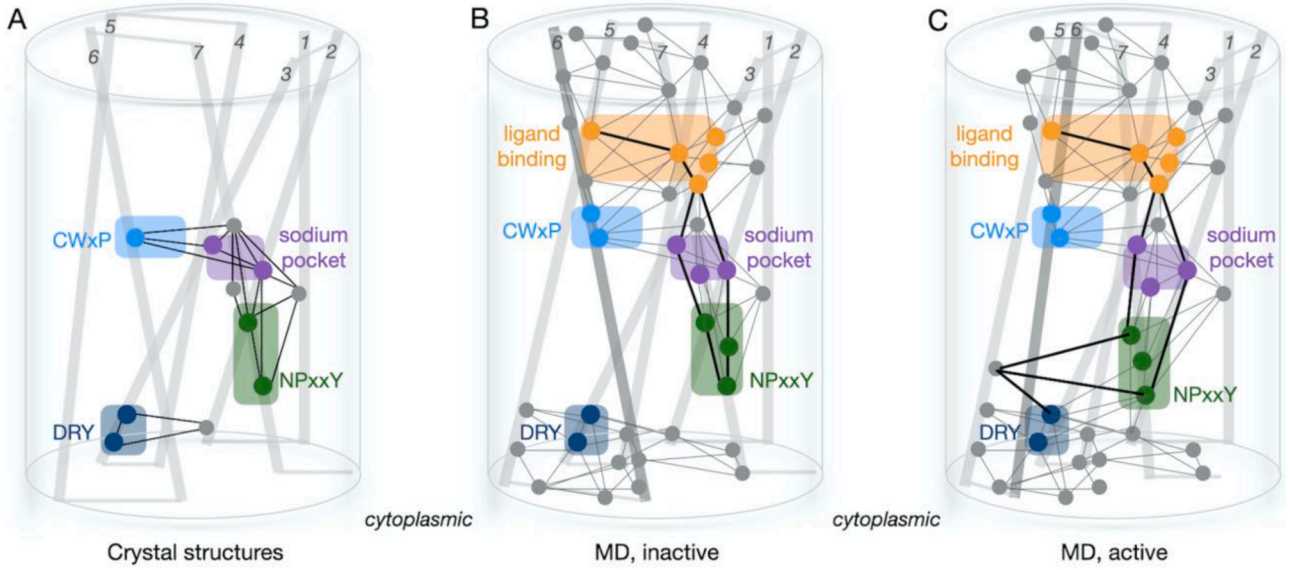
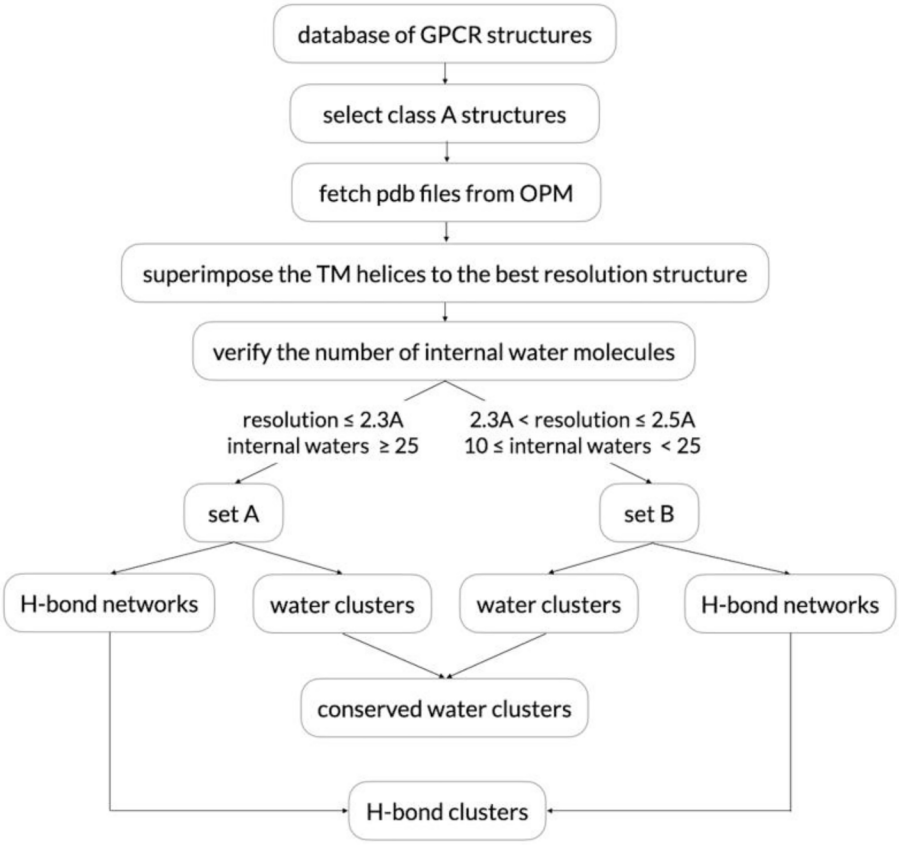


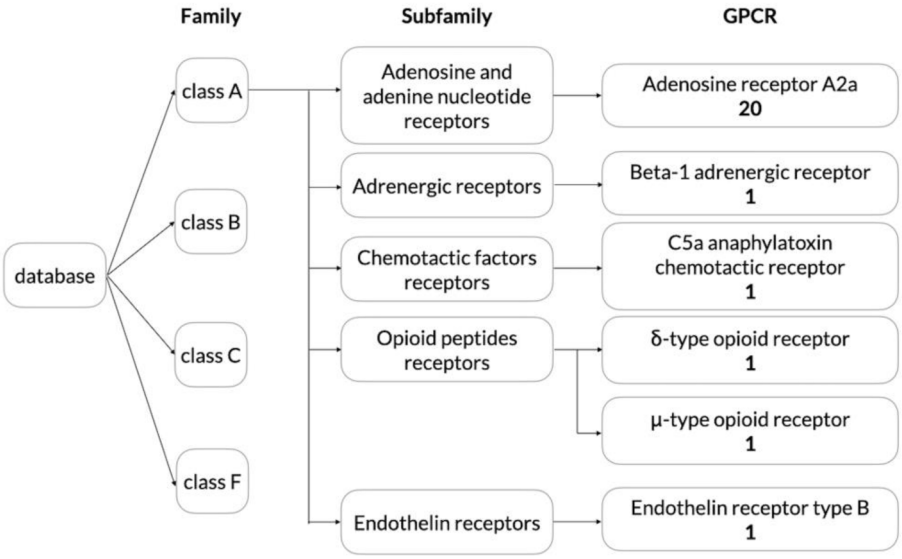
Fig. 12. Schematic representation of main observations. (A) Analysis of 53 GPCR structures from *Set-A* and *Set-B* revealed a conserved core H-bond cluster interconnecting the CWxP and NPxxY motifs and the sodium-binding pocket, and a local H-bond cluster of the DRY motif. (B) Simulations of the DOR and KOR-inactive indicate a H-bond network connects the ligand-binding pocket at the extracellular side to the NPxxY motif at the cytoplasmic half; this network includes all groups of the core H-bond network identified in static GPCR structures. The extended core network and the DRY motif network remain separated throughout these simulations. (C) Simulations of KOR-active indicate an extensive water-mediated H-bond network whereby the extended core network and the DRY motif network have connected and span the interior of the GPCR from the extracellular to the cytoplasmic side.



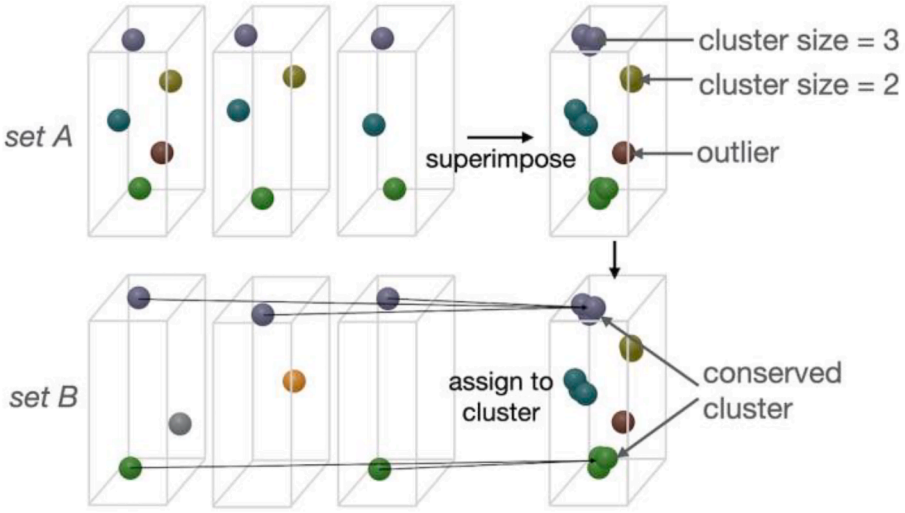
Scheme 1. Work flow used to establish *Set-A* and *Set-B* of GPCR structures to analyze water and H-bond clusters. We selected and assigned the structures to one of the two sets according to the resolution at which the structures were solved, and to the number of internal waters. All other static structures of class A GPCRs from the database were included in *Set-C*.

and *Set-C*. This led to an ensemble of H-bond graphs for each set, which we used to compute the frequency of occurrence of H-bond groups and of H-bonds between two specific protein groups.

The occurrence of a node (H-bond group) is given by the number of structures in which that group is involved in H-bonding, as illustrated in Scheme 6. The occurrence of a H-bond between two protein groups gives the number of structures in which the same two protein groups are H-bonded (Scheme 6). We consider that nodes or edges are conserved when their occurrence is $\geq 40\%$. With the conserved nodes and edges we assembled conserved H-bond graphs for *Set-A*, *Set-B*, and *Set-C*.



Scheme 2. Composition of *Set-A* of GPCR structures. *Set-A* includes 25 class A GPCR structures solved at resolutions of 2.3 Å or higher. Each structure contains at least 25 internal waters. Database refers to the initial set of structures obtained from GPCR-EXP, from which we took the annotation of the GPCR proteins (see also Table S1).

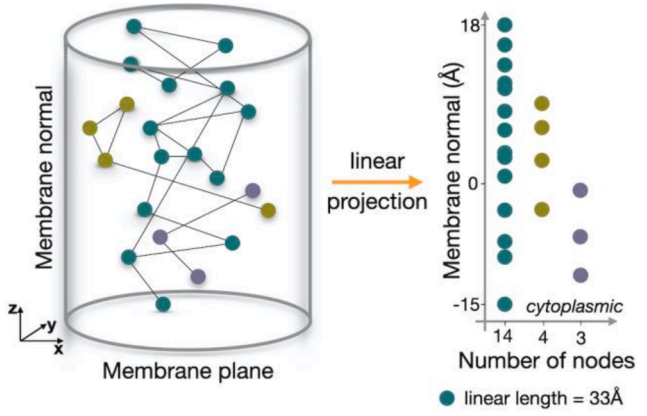


Scheme 3. Algorithm used to identify conserved water clusters in GPCR structures. The three blocks delineated by thin gray lines represent three protein structures, and the small spheres indicate water oxygen atoms. To identify water clusters in *Set-A*, GPCR structures were superimposed onto the A2A structure used as a reference. The ensemble of the superimposed structures was analyzed using DBSCAN (Ester et al., 1996) to cluster waters according to spatial proximity. The size of each cluster is given by the number of waters in that cluster. We set the minimum cluster size to $k = 5$; waters present in fewer structures are tagged as outliers. For *Set-B*, following structure superimposition with the reference A2A structure (Weinert et al., 2017), waters are assigned to the clusters identified for *Set-A*.

Similarity between graphs of H-bonds. To quantify similarities between H-bond networks we computed matrices of common edges, *CE*, and common nodes, *CN*.

For a pair of GPCR structures, *CN* is given by the number of amino acid residues that are nodes of H-bond graphs in both structures, and *CE* is given by the number of H-bonds present in both structures. For example, if in two structures S_i and S_j amino acid residues 1.50 and 7.53 are H-bonded to each other, groups 1.50 and 7.53 are counted towards the *CE* value of S_i and S_j, and the H-bond between 1.50 and 7.53 is counted toward *CN*. More generally, we consider that the H-bond graphs computed for structures S_i and S_j have in common E_{ij} edges and N_{ij} nodes, and compute *CE* and *CN* matrices according to the equations

$$CE = \begin{pmatrix} E_{ii} & \cdots & E_{j,i} \\ \vdots & \ddots & \vdots \\ E_{i,j} & \cdots & E_{j,j} \end{pmatrix} \quad (3)$$



Scheme 4. Schematic representation of the algorithm used to compute the linear length of H-bond networks. The small spheres represent Cα atoms of amino acid residues part of a H-bond network, and lines between the small spheres indicate direct or water-mediated H-bonding. In this scheme we depict three networks, with small spheres colored according to the network they belong to. The *z*-coordinates of Cα atoms of amino acid residues involved in continuous H-bond networks are projected onto the membrane plane. The linear length of a network is given by the distance in Å between the z-coordinates of the Cα atoms at the cytoplasmic vs. extracellular extremities of the network.

$$CN = \begin{pmatrix} N_{ii} & \cdots & N_{j,i} \\ \vdots & \ddots & \vdots \\ N_{i,j} & \cdots & N_{j,j} \end{pmatrix} \quad (4)$$

with $E_{i,j} = E_{j,i}$ and $N_{j,i} = N_{i,j}$. The diagonal matrix elements E_{ii} and N_{ii} give the number of edges and, respectively, the number of nodes of the H-bond graph computed for each structure S_i (Scheme S1B, S1D), whereas the off-diagonal matrix elements give the number of nodes and edges that are conserved between structures S_i and S_j. We computed matrices *CE* and *CN* separately for *Set-A* and *Set-B*.

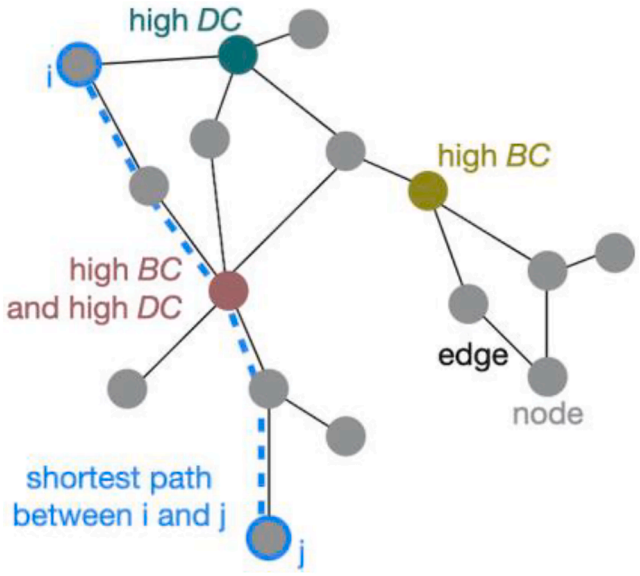
Protein structures used for MD simulations, and simulation setup. To probe the dynamics of internal protein-water H-bond networks of GPCRs, we performed MD simulations of the DOR and the KOR (Table 1). For the starting coordinates we used *i)* the KOR in an inactive conformation, PDB ID:4DJH (Wu et al., 2012); *ii)* the KOR in an active-like conformation, PDB ID:6B73 (Che et al., 2018); *iii)* the DOR in an inactive conformation, PDB ID:4N6H (Fenalti et al., 2014). For simplicity, and to distinguish between the simulations, in what follows we will refer to the three MD simulations as KOR-active, KOR-inactive, and DOR, respectively.

The DOR structure includes 286 amino acid residues, from S37 to D322. The KOR-inactive (Wu et al., 2012) and KOR-active structures (Che et al., 2018) have 12 and 5 missing internal amino acid residues, respectively. Missing internal groups and missing loops were built with Modeller 9.21 (Eswar et al., 2006). Using Modeller, we generated 50 models for each receptor structure, with 10 loop refinements; models were evaluated with the DOPE and MOLPDF energy function, and the loop with the lowest MOLPDF score was chosen in each case. The resulting KOR structures have 280 amino acid residues, from S55 to D334.

All titratable amino acid residues were considered in standard protonation states, i.e., Asp/Glu are negatively charged, Arg/Lys, positively charged, and His groups are singly protonated on the Nδ atom. Due to uncertainties regarding the accuracy of force-field parameters for ligands bound to the ORs, ligands were removed from all structures.

As the KOR structures lack sodium ion at the allosteric binding pocket, to facilitate direct comparison between the three simulations, all three MD simulations were started without a bound sodium ion.

The orientation of the proteins in the membrane was obtained from OPM (Lomize et al., 2011). We used CHARMM-GUI (Wu et al., 2014) to place the proteins in a hydrated bilayer of 1-palmytoyl-2-leoyl-*sn*-glycero-3-phosphatidylcholine (POPC), with ions added for charge neutrality. The complete simulation systems contain 138.105 atoms for DOR, and 159.999 and 146.911 atoms for KOR-inactive and KOR-active,



Scheme 5. Shortest H-bond paths and centrality measures of a H-bond network. The path colored blue connects nodes *i* and *j* via the least number of nodes, and it is thus the shortest distance path between these two nodes. A node has high *DC* value when many direct edges connect it to other nearby edges, and high *BC* value when it is part of many short-distance paths that interconnect two other nodes.

respectively.

Protocol for MD simulations. We used CHARMM36m for proteins, ions, and lipids (Brooks et al., 1983; Feller and MacKerell Jr., 2000; Huang et al., 2016; Klauda et al., 2010; MacKerell Jr. et al., 2004, 1998), and the TIP3P water model (Jorgensen et al., 1983). All MD simulations were performed using NAMD (Kalé et al., 1999; Phillips et al., 2005). Coulomb interactions were computed with smooth Particle Mesh Ewald summation (Darden et al., 1993; Essmann et al., 1995). For short-range real space interactions we used a switching function from 10 Å with cutoff at 12 Å. Following geometry optimization and heating to 300 K, simulations continued with equilibration with velocity rescaling according to the standard CHARMM-GUI protocol, and production runs without any restraints. Heating and equilibration were performed using the *NVT* ensemble (constant number of particles *N*, constant volume *V*, and constant temperature *T*) with an integration step of 1 fs. Production runs were performed in the *NPT* ensemble (constant pressure $P = 1$ bar), using a Langevin dynamics scheme (Feller et al., 1995; Martyna et al., 1994) with damping coefficient of 5 ps^{-1} and an integration step of 2 fs. Coordinates were saved every 10 ps. Unless specified otherwise, average values were computed from the last 100 ns of each simulation.

Computations of dynamic water-mediated H-bond paths from MD simulations. To estimate the number of water molecules in the inter-helical region of the DOR and KOR during MD simulations, we monitored waters whose oxygen atom was simultaneously within 5 Å of any protein heavy atom, and within 5 Å of the plane of the lipid phosphate groups. This 5 Å distance was proposed to exclude fluctuations in the number of protein-bound waters that can exchange with the bulk during the simulations (Kučerka et al., 2008). Waters in the inter-helical region of the DOR and KOR were included in computations of protein-water H-bond paths using Bridge (Siemers et al., 2019).

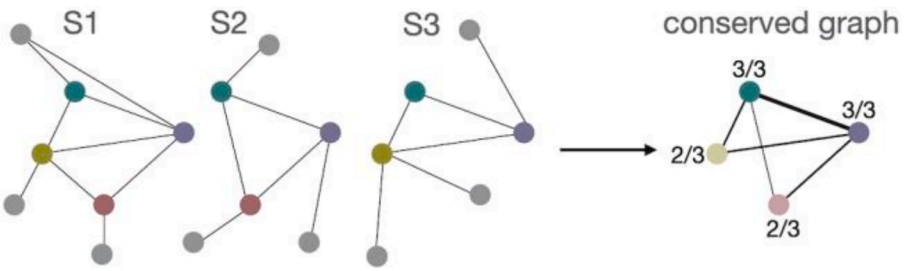
3. Results and discussion

We developed a methodology to identify and analyze conserved water sites and conserved protein-water H-bond networks in ensembles of static GPCR structures from structural biology, and in MD simulation trajectories. We established three sets of GPCR structures that are distinguished by the resolution at which the structures were solved, and by the number of internal waters. We used static GPCR structures to compute, using the methodology presented here, conserved H-bond networks, and to dissect the role of internal waters in mediating long-distance couplings between different segments of the GPCRs. MD simulations on opioid GPCRs indicate waters can help establish a remarkable, dynamic network of H-bonds that spans the receptor.

Internal waters in GPCR structures and MD simulations. Activation is thought to associate with formation of a continuous water-mediated H-bond network in the TM region (Venkatakrishnan et al., 2019; Yuan et al., 2014). Static structures of *Set-A* have 29–54 internal waters (Fig. 2A), but many of the *Set-B* and *Set-C* structures have few, or no internal waters (Fig. 2B, S1). Absence of water coordinates, or a small number of internal waters, in a static protein structure, could be due to the limited resolution at which a structure is solved, to water molecules being disordered (Gottschalk et al., 2001; Levitt and Park, 1993; Quillin et al., 2006), and/or a temperature-dependent probability for a water molecule to occupy a protein cavity (Bondar et al., 2008; Gottschalk et al., 2001). Indeed, GPCR structures solved at higher resolution tend to contain more internal waters, the largest number of internal waters, 54, being found for a structure solved at a resolution of 1.8 Å (Liu et al., 2012) (Figs. 2, S1, Table S1).

As a first step towards evaluating internal protein-water H-bond networks of GPCRs, we sought to identify sites where internal waters are more likely to be found. For *Set-A*, we found with DBSCAN 47 water clusters (see blue spheres on Fig. 3B), with an average of 17 waters in one cluster (Fig. S4). Most of these clusters are found in 5 regions of the protein, separated from each other by molecular switches (Fig. 3B). *Set-B* structures contain 22 of the 47 water clusters identified for *Set-A* (yellow spheres in Fig. 4).

Water clusters common to *Set-A* and *Set-B* tend to be located near molecular switches and conserved groups (Fig. 4B-E). A prominent water cluster delineated by amino acid residues 6.52, 6.55 and 5.38, contains waters present in 24 structures of *Set-A*, and 8 structures of *Set-B*. This water cluster appears to be a common feature of adenosine receptors, which constitute most of the *Set-A* structures we studied; in the



Scheme 6. Conserved H-bond graphs in structures of GPCRs. For each H-bond graph computed for structures S_i, the algorithm identifies amino acid residues that are common. The conserved H-bond graph reported contains four nodes (amino acid residues), two of which are present in 3 of the 3 structures (3/3), and two are present only in 2 structures (2/3).

A2A structure we used for the molecular representation in Fig. 4 (Segala et al., 2016), the water cluster is delineated by H6.52, N6.55, and M5.38. The large clusters also tend to contain waters with small B factors (Fig. S5), i.e., waters with reduced mobility, which could explain their being solved even in *Set-B* structures with more modest resolution.

The structure of the DOR receptor used to initiate the simulations (Fenalti et al., 2014) has 33 internal waters (Fig. 2A). The other two structures used for simulations, of the KOR-inactive (Wu et al., 2012) and KOR-active (Che et al., 2018) structures have 5 and 0 internal waters, respectively. Many more water molecules visit transiently the inter-helical region of the DOR and KOR. In each simulation there is a relatively broad distribution of the number of internal waters, with peaks centered at around 50–75 waters (Fig. 5A, S6), suggesting the TM domain of the ORs can be visited, at least transiently, by numerous water molecules. That most waters solved in the DOR structure are mobile in simulations was also observed by (Venkatakrishnan et al., 2019). Likewise, MD simulations on A2A indicated waters in the TM region are highly dynamic, though the average number of waters remained constant (Leitner et al., 2020).

The number of internal waters in simulations is compatible with the upper limit of the number of waters observed in static structures (Fig. 2A). The average number of internal waters found from simulations is also close to the number of conserved water sites identified with DBSCAN for Set-A (Fig. 3), suggesting the settings we used for DBSCAN computations are reasonable.

Common H-bonds inform on similarities between GPCR structures. Protein structures are typically compared in terms of RMSD values. For proteins such as GPCRs, for which internal H-bond networks are thought essential for the conformational dynamics of the protein, the number H-bonds that are common between two GPCRs (i.e., H-bonding between the same amino acid residues within the BW scheme) could provide information about internal rearrangements of protein sidechains during, e.g., ligand binding or activation of the receptor. We calculated, for each structure of *Set-A* and of *Set-B*, the number of H-bonds between protein sidechains, and the number of common H-bonds and H-bonding groups shared by any pair of structures of *Set-A* or *Set-B*.

The 20 adenosine receptor structures included in *Set-A* share many of their H-bonds (Fig. S7) and H-bonding groups (Fig. S8). For example, a structure of A2A receptor with a loop replaced by apocytochrome, PDB 4EIY (Liu et al., 2012), has 62 H-bonds, of which 46 are in common with the 51 H-bonds of PDB ID:5IU4 (Segala et al., 2016), which reports a thermostabilized A2A structure solved in complex with an antagonist ligand, and 36 H-bonds are in common with the A2A structure PDB ID:5NM4 (Weinert et al., 2017) solved at room temperature (Fig. S7). Each of these three A2A structures share with the 52 H-bonds of the DOR only 18–24 H-bonds, that is, the internal H-bond networks of the three A2A structures are much closer to each other than either A2A structure is to the DOR. Even more extreme examples are the C5a anaphylatoxin chemotactic receptor 1, C5AR1 (Liu et al., 2018), which shares with A2A structures only 7–14 of its 41 protein H-bonds (Fig. S7), and the endothelin receptor ETB bound to an antagonist analog (Shihoya et al., 2017), which shares with A2A structures 12–19 of its 69 H-bonds (Fig. S7).

Due to the diversity and lower resolution of *Set-B* structures, H-bond graphs of *Set-B* structures have lower similarity to each other than pairs of structures of *Set-A*: Many pairs of GPCR structures of *Set-B* have in common fewer than 10 H-bonds (Fig. S9). Exceptions are the 4 A2A structures of *Set-B*, which share with each other 32–37 H-bonds (Fig. S9, Table S7), and the D(4) dopamine receptor DRD4 (Wang et al., 2017), which has 51% of its H-bonds in common with an A2A structure (Weinert et al., 2017) (Fig. S10, Table S7).

The relatively small number of common H-bonds between pairs of *Set-B* structures associates with most of the *Set-B* structures having fewer protein–protein and water-mediated H-bonds than *Set-A* structures: Most of *Set-B* structures have ≤ 35 H-bonds (Figs. S9, S10), as compared to ≥ 40 H-bonds in the case of *Set-A* structures (Fig. S7). The largest number of H-bonds for *Set-B* structures, 60, found for the orexin receptor OX, is explained by the 1.96 Å resolution of the structure (Suno et al., 2018).

Taken together, the results on common H-bonds between pairs of GPCR structures suggest that H-bond similarity matrices could be used to refine comparisons of structures of a GPCR solved in different experimental conditions, or of two different GPCR structures solved at similar resolutions. Information on protein–protein H-bonds common among GPCR structures could also be used to improve homology models of GPCRs. GPCR structures solved at modest resolution underestimate the internal H-bond network of the protein.

The core Y7.53 network of H-bonds identified from static GPCR structures. Internal waters located at conserved sites, where they interact with conserved protein groups (Fig. 5), could be considered as structural waters, and we wondered whether these waters help bridge different segments of the protein.

We found that all TM helices can have water-mediated inter-helical connections (Fig. S11). In both *Set-A* and *Set-B*, a conserved H-bond network of eight amino acid residues extends over a linear distance of ~ 17 Å from W6.48 of the CWxP motif to Y7.53 of NPxxY, and via D2.50 of the sodium-binding pocket (Fig. 6A, 6F, S12). In this network, which we denote as *the core Y7.53 network* of the GPCRs, most of the water-mediated bridges are short, with up to 3 waters in a bridge (Fig. 6A, 6F).

In addition to the core Y7.53 network, in *Set-A* we identified several conserved local H-bond clusters. At the cytoplasmic side, a local cluster extends from T6.58 ~ 15 Å deep into the protein to Q3.37, i.e., to a group that is adjacent in the sequence to S3.38 of the core cluster (Fig. 6A, 6B). Another local H-bond cluster reaches from Y7.36 to H7.43 at the ligand-binding site (Fig. 6A, 6E). At the cytoplasmic side, the local H-bond cluster of R3.50 of the DRY motif is close, along the membrane normal, to the core H-bond cluster (Fig. 6A, 6D).

The R3.50 H-bond cluster is present in *Set-B*, as are the H-bonds between residues 2.45 3.42, and 4.50; a number of H-bonds have smaller occurrence, *i.e.*, they are present in some, but not in the majority, of structures (Fig. S12). Overall, the conserved H-bond network of *Set-B* is somewhat sparse (Fig. S12A, Table S10), without the local clusters of *Set-A* (Fig. 6A). For *Set-C* we found even more limited conservation of H-bonding (Table S10), with a few isolated conserved H-bonds (Fig. S12B).

The core Y7.53 H-bond network found for *Set-A* (Fig. 6A) and *Set-B* (Fig. S12A) includes groups essential for function. Experiments on agonist binding to KOR mutants Y7.53L, G7.42K and W6.48K, indicated this pocket is involved in conformational coupling (Che et al., 2018). In the angiotensin II type 1 receptor AT1R, Y7.53 appears important for coupling to the agonist, and D2.50 (D74) is part of a H-bond network that rearranges when 7.46 (N295) rotates upon receptor activation (Wingler et al., 2020). In A2A, the D2.50N mutation abolishes G-protein signaling (White et al., 2018), being thought that D2.50 is at the heart of an allosteric center (Eddy et al., 2018). An Asn at position 7.46, instead of Ser as in most Class A GPCRs, is thought to help AT1R prevent sodium binding and shape conformational dynamics (Suomivuouri et al., 2020). Close to the core Y7.53 network, H-bonding between Y7.43 and D3.32 is thought important for the activation of the KOR (Che et al., 2018).

The overall picture that emerges from analyses of conserved clusters in *Set-A* structures is that the inter-helical region of the receptor hosts a central H-bond cluster that includes groups essential for function, and which neighbors four other smaller, local H-bond clusters. As some of the clusters contain groups proximal in space or sequence to groups from another cluster, the question arises as to whether some or all of these conserved clusters could inter-connect during protein dynamics, leading to a continuous H-bond network that could span the receptor upon activation. *Set-B* structures solved at lower resolution have fewer sidechain-sidechain H-bonds, and thus underestimate the internal H-bond connectivity of GPCRs.

Conserved H-bond networks in static structures of GPCRs in different conformations. Most of the GPCR structures solved at high resolution represent inactive GPCRs. As a consequence, of the 25 structures of our

Set-A, 24 were proposed to represent inactive conformations, and only a MOR structure (Huang et al., 2015) was proposed for an active conformation (Table S1). *Set-A* is thus largely representative of inactive GPCRs. Of the 28 structures from *Set-B*, 19 were proposed for inactive GPCR conformations, 6, intermediate, and 3, for active conformations (Table S2). To test whether the internal H-bond networks could depend on the intermediate conformation proposed for the structure, we separated *Set-A* into *Set-A*$_{inactive}$ (24 structures) and *Set-A*$_{active}$ (one structure). Likewise, we separated *Set-B* structures into *Set-B*$_{inactive}$, *Set-B*$_{intermediate}$, and *Set-B*$_{active}$, and calculated conserved H-bond networks separately for each subset.

As anticipated given they constitute the vast majority of *Set-A*, inactive GPCR structures of *Set-A*$_{inactive}$ have a conserved H-bond network (Fig. S13A) that is largely the same we obtained for the complete *Set-A* (Fig. 6). Both *Set-A*$_{inactive}$ and *Set-A*$_{active}$ contain the core Y7.53 H-bond network (Fig. S13). An important distinction between the conserved H-bond networks of *Set-A*$_{inactive}$ vs. *Set-A*$_{active}$ is that only the latter contains H-bonds between 3.50, 5.58, and 7.53 (Fig. S13). That is, a connection between the central Y7.53 cluster and the DRY local H-bond cluster is absent in the conserved H-bond network of the inactive *Set-A* structures, but it is present in the active structure.

Set-B$_{inactive}$ contains the complete Y7.53 H-bond network (Fig. S14A); overall, the H-bond graph of *Set-B*$_{inactive}$ is largely similar to that of the complete *Set-B* (Fig. S12A). The conserved H-bond networks of *Set-B*$_{intermediate}$ (Fig. S14B) and *Set-B*$_{active}$ (Fig. S14C) contain most of the core Y7.53 network and tend to have in the extracellular half more conserved H-bonds than obtained for *Set-B*$_{inactive}$ (Fig. S14A). Differences in H-bond networks of *Set-B*$_{inactive}$ vs. *Set-B*$_{active}$ can associate with different centrality values of key groups (Tables S8, S9): though both sets give H-bond graphs with high centralities for residues 2.50, 7.49, and 7.53, relative centralities of these groups differ, with highest *DC* value for 2.50 in *Set-B*$_{inactive}$, and for 7.49 in *Set-B*$_{active}$.

In *Set-B*$_{active}$ the Y7.53 network connects to R3.50 of the DRY motif (Fig. S14C), but it contains 3 fewer H-bonding groups than *Set-A*$_{active}$. Differences between the core H-bond networks of *Set-A*$_{active}$ and *Set-B*$_{active}$ could be due to differences in resolution and number of internal waters, or could reflect differences in the amino acid composition: *Set-A*$_{active}$ contains only a MOR, whereas *Set-B*$_{active}$ contains only visual rhodopsin structures –which are an exception among class A GPCRs, in that they lack specific sodium binding (Katritch et al., 2014).

We obtained qualitatively similar results when we combined *Set-A* and *Set-B*, and then separated this combined set of 53 structures according to their annotation of inactive, intermediate, and active conformations (Tables S1, S2). Briefly, the combined set of inactive conformations includes the core network and local H-bond clusters (Fig. S15A). The core network is reduced in the intermediate conformations of the combined set, and connected to residue 3.50 in the active conformations (Figs. S15B,C).

Overall, however, since the number of GPCR structures in the active or inactive conformation is relatively small, and since most of these structures have lower resolution and/or fewer internal waters, their calculated H-bond networks might be less accurate.

D2.50 is a central hub of the internal H-bond network of GPCRs. Centrality measures enable ranking of H-bonding groups according to their importance for connectivity within a H-bond graph (Lazaratos et al., 2020), and to identify protein groups that could be important for protein function (Amitai et al., 2004; Harris et al., 2020; Karathanou and Bondar, 2019). Overall, groups of the core Y7.53 H-bond cluster have highest centrality values (Figs. 7, S16). D2.50 has the highest *DC* value of all protein groups (Fig. 7, Table S8), and high *BC* (Fig. S16, Table S9). Thus, D2.50 appears to be a central hub for the internal H-bond communication network of GPCRs.

A role for D2.50 as central hub, as suggested here, is compatible with the important functional role of D2.50 indicated by the experiments discussed above. D2.50 is also thought to change protonation during the functioning of the GPCR (Vickery et al., 2018; Zhang et al., 2015), a reaction which we suggest could involve S7.46: Similar pairs of hydroxyl group and carboxylate groups within H-bond distance are present in membrane transporters at sites where protonation changes couple to changes in protein conformational dynamics (Bondar and Smith, 2017; Bondar and Lemieux, 2019; del Val et al., 2012, 2014; Siemers et al., 2019).

The extended core H-bond network of the DOR in hydrated lipid membranes. The dynamic H-bond network of the DOR computed from simulations at room temperature indicates a large H-bond cluster that starts at the extracellular region of the protein and extends to Y7.53, interconnecting the ligand-binding site to the sodium-binding pocket and the NPxxY motif (Fig. 8A-C). This H-bond network including all 8 groups that constitute the core Y7.53 network in static structures, and other protein sidechains and internal waters (Fig. 6A, 8A, 8C, S17D). We thus denote this H-bond network as the extended core network of the GPCR. The linear length of the extended core H-bond network is ~ 45 Å (Fig. 8B).

In the extended core Y7.53 network, most of the water-mediated bridges between protein groups are relatively short, with 1–3 waters, such that the *JO* of the shortest distance H-bond path between K5.39 at the ligand-binding site and Y7.53 is very high, ~79–96% (Fig. 8C-D), indicating these two groups are connected almost all the time. In the extended core network, D2.50 retains relatively high *BC* and *DC* values, but the highest centrality values tend to be observed for groups closer to the extracellular side, where numerous waters can visit the protein and establish transient H-bond paths between protein sidechains (Figs. S18, S19).

As in static GPCR structures, the DOR simulations indicate a local H-bond at R3.50 of the DRY motif (Fig. 8A-B, S17D). In MD simulations this local network is, however, significantly more extended, as water molecules can bridge transiently protein groups to give rise to a cluster with a linear length of ~ 20 Å (Fig. 8A, 8B, S17D). Albeit the extended Y7.53 core network and the R3.50 network include groups with about the same location along the membrane normal, these two clusters remain separated, without common H-bonds (Fig. 8A-C).

A qualitatively similar H-bond network was obtained for the repeat DOR simulation (Figures 17A-C), though we note the slightly shorter H-bond path between K5.39 and Y7.53, which passes via D2.50 and D3.32 in the repeat DOR simulation (Fig. S17C), as compared to D2.50, D3.32, and N3.35 in the main DOR simulation (Fig. 8C).

The core H-bond network we find for the DOR is also compatible with that identified in previous DOR simulations (Venkatakrishnan et al., 2019). Both studies indicate a H-bond network that includes residues 1.50, 2.50, 7.45, and 7.49, which are part of the core H-bond network identified in *Set-A* and *Set-B*. The somewhat more extensive internal H-bond network we present here for the DOR is likely due to differences in criteria for water-mediated bridges between amino acid residues of the network. By using Bridge (Siemers et al., 2019), we could calculate efficiently dynamic water-mediated bridges, and found additional interactions mediated by 3–4 waters (Fig. 8C).

KOR-active is spanned by an extensive H-bond network that is interrupted in KOR-inactive. In both simulations of the KOR, a sodium ion from the bulk enters the protein early during the simulations (Fig. 9A), it interacts first with D138 (D3.32) and water (Fig. 9B), and it then moves to the sodium-binding pocket (Fig. 9C). Interactions between the sodium ion and the coordination shell can still fluctuate, such that at the end of the KOR-active simulation, the sodium ion interacts with S3.39 and N3.35, but not with D2.50 (Fig. 9E). Fast sodium binding, within 50 ns, was also observed in the repeat KOR-inactive and KOR-active simulations (Fig. S6E).

The result here show that sodium bound at the KOR sodium-binding pocket interacts with both protein and water (Fig. 9C, 9E) is compatible with the crystal structure of sodium-bound DOR (Fig. 9F), and with microsecond-long simulations indicating interactions between D3.32 and a sodium ion entering from the extracellular bulk, and faster binding of sodium to KOR than DOR (Shang et al., 2014). The reason why

sodium entered rapidly the protein in KOR simulations, but not in the two DOR simulations, is unclear. These are distinct protein structures, and details of local protein and water dynamics might shape the energetics of sodium binding. We also cannot exclude that a sodium ion might bind to the DOR in a longer simulation.

The internal H-bond network of KOR-inactive (Figures 10, S20A) is qualitatively similar to that of the DOR (Figures 8, S17D), in that the extended core Y7.53 cluster connects Y7.53 to the cytoplasmic side via the sodium-binding pocket and via the ligand-binding pocket (Fig. 10). Details differ at the sodium-binding pocket, as absence of a sodium ion in DOR associates with water-mediated H-bonding between D2.50, S3.39, and N3.35 (Fig. 8C), whereas in the sodium-bound KOR-inactive D2.50 makes a direct H-bond with N3.35 (Fig. 10C, 10D). This suggests that sodium binding to GPCRs associates with rearrangements of the local H-bond network of D2.50.

As observed for the DOR, in KOR-inactive the local H-bond network of R3.50 includes numerous protein-water H-bonds, yet it remains disconnected from the extended Y7.53 core network (Fig. 10A-B). The repeat KOR-inactive and KOR-active simulations confirm these findings (Fig. S20): As in the KOR-inactive simulation (Fig. 10A), in the repeat KOR-inactive the core Y7.53 H-bond network remains unconnected to the R3.50 network of the DRY motif (Fig. S20A). Likewise, as observed in KOR-active (Fig. 11A), in the repeat KOR-active simulation the core and the DRY motif networks are part of a continuous H-bond network that spans the interior of the GPCR (Fig. S20B). In the repeat simulations on KOR-inactive and the DOR, N7.49 can be part of a path that connects K5.39 and Y7.53 (Figs. S17C, S21C).

A remarkable difference is observed in KOR-active as compared to KOR-inactive and DOR simulations: In KOR-active, the Y7.53 core network connects to the local H-bond network of R3.50, such that the entire TM domain is spanned by a protein-water H-bond network (Figures 11, S20B, S21D). A long-distance protein-water H-bond path with relatively high *JO* value of 43% reaches from K3.59 of the ligand-binding pocket to R3.50 of the DRY motif via Y7.53 and D2.50 (Fig. 11B, 11C).

A continuous protein-water H-bond network spanning the TM region of the KOR-active is compatible with previous simulations on A2A (Lee et al., 2016), though details of H-bond networks sampled with different simulations could depend on the GPCR, ligand being present, and force-field parameters used. To avoid uncertainties regarding the accuracy of the CHARMM force-field description of opioid ligands bound, here we studied the motions of the KOR-active in the absence of a ligand. This caveat is alleviated by the fact that the KOR (Sirohi and Walker, 2015), MOR (Sadée et al., 2005; Wang et al., 2001) and DOR (Costa and Herz, 1989) have basal activity, i.e., spontaneous activity in the absence of an agonist.

The extracellular region of the receptor and, to some extent, the cytoplasmic region, contain a dense network of protein-water H-bonds (Fig. 8A, 10A 11A), such that protein groups located close to the extracellular and cytoplasmic sides have high centrality values when dynamic water bridges are included in the graphs (Figs. S18, S19). When only the most persistent interactions are included in the graphs, D2.50 and D3.32 of the core H-bond network have relatively high centrality values (Figs. S18, S19), as observed in the static structures (Fig. 7). The relative importance of protein groups in centrality computations depends on water molecules present in the network.

4. Conclusions

We implemented graph-based algorithms to identify and characterize protein and protein-water H-bond networks in GPCR structures from structural biology and numerical simulations. The algorithms we presented here include a machine-learning clustering approach to identify conserved water sites in GPCR structures, computations of conserved H-bond clusters and of the linear length of these clusters, similarity measures for H-bonds common to pairs of GPCR structures, and centrality measures to identify groups central for the local and long-distance communication within H-bond networks of GPCRs. For the analyses of H-bond clusters in GPCRs we used three sets of experimental structures: *Set-A*, consisting of 25 GPCR structures solved at resolution of 2.3 Å or better and with at least 25 internal waters, *Set-B*, with 28 structures solved at a resolution between 2.3 Å and 2.5 Å and with at least 10 internal waters, and *Set-C*, with the remaining 255 structures solved at lower resolution and/or with fewer internal waters. To evaluate how dynamic fluctuations in a fluid, hydrated lipid environment could impact internal protein-water H-bond networks, we augmented the analyses of the static GPCR structures by six independent atomistic simulations of two opioid receptors.

Using the algorithms presented here, we studied conserved water sites, i.e., sites where discrete water molecules tend to be present in static GPCR structures. All static GPCR structures included in *Set-A* and *Set-B* have a large, conserved core cluster of eight protein sidechains and water molecules (Fig. 12A), and a local H-bond cluster at the DRY motif. Structures solved at higher resolution have additional, smaller H-bond clusters in the extracellular half of the receptor.

The conserved core cluster is likely to be particularly important for the conformational dynamics of GPCRs, because it bridges the NPxxY motif to the sodium-binding pocket and to the CWxP motif (Fig. 12A). Although the NPxxY and DRY motifs are located relatively close along the membrane normal, in static GPCR structures the conserved core and DRY networks remain separated (Fig. 12A).

During simulations of the DOR and KOR-inactive, protein and water dynamics allow the core H-bond network to expand towards the extracellular side, to the ligand-binding site, such that a continuous protein-water H-bond network can reach from the extracellular ligand binding site deep into the protein, at the NPxxY motif (Fig. 8B, 10B, 12B). In KOR-active simulations the core H-bond network extends even further, to the DRY motif, such that a dense network of protein-water H-bonds contributed by waters and protein groups of all seven TM helices spans the receptor from the extracellular to the cytoplasmic side (Fig. 12C).

The extended H-bond network we identified in KOR-active simulations is compatible with the activation path connecting Y7.53 to the agonist binding site found for class A GPCRs (Zhou et al., 2019). Details of the continuous H-bond network sampled in the absence of bound ligand and G protein might, however, change when a ligand or the G protein binds: Groups of the internal protein H-bond network could interact directly with the ligand –such as D3.32 in the MOR (Huang et al., 2015), or with the G protein –such as R3.50 (Venkatakrishnan et al., 2013). In the future, accurate force-field representation of GPCR ligands could allow us to probe how the dynamic, internal protein H-bond network participates in and respond to ligand binding. Extending such simulations to different types of GPCRs may enable more general predictions about the response of internal H-bond networks to ligand binding.

H-bond networks we identified here are necessarily representative only to the structures included in analyses. As most of the GPCR structures solved at high resolution belong to A2A receptors, the extent to which H-bond networks of A2A receptors inform on structural signatures of the much wider GPCR family, remains to be seen. H-bond networks identified from static protein structures also depend on coordinates solved for water molecules. The accuracy of these water coordinates depends on the resolution at which the protein is solved (Quillin et al., 2006), coordinates for low occupancy or disordered waters might be missing (Gottschalk et al., 2001; Levitt and Park, 1993), and the probability that an internal protein cavity is occupied by water depends on the temperature (Bondar et al., 2008; Gottschalk et al., 2001). With larger data sets of GPCR structures solved at high resolution, the methodologies presented here could enable analyses of the conformational dynamics of GPCRs based on H-bond networks.

CRediT authorship contribution statement

Éva Bertalan: Investigation, Methodology, Validation, Visualization, Writing - original draft. **Samo Lešnik:** Investigation, Visualization, Writing - original draft. **Urban Bren:** Writing - original draft. **Ana-Nicoleta Bondar:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Research was supported in part by the Excellence Initiative of the German Federal and State Governments provided via the Freie Universität Berlin, by the German Research Foundation (DFG) Collaborative Research Center SFB 1078 Project C4 (to A-NB), by an allocation of computing time from the North-German Supercomputing Alliance, HLRN, and by computing resources of the Department of Physics of the Freie Universität Berlin. A-N.B. and S.L. thank the Center for Research Strategy of the FU for financial support of a research stay of S.L. at the FU. U.B. and S.L. acknowledge financial support from the Slovenian Ministry of Science and Education through project grant C3330-19-952021, and from the Slovenian Research Agency programme grants P2-0046 and J1–2471. A-NB acknowledges valuable discussions with Prof. Gebhard Schertler. EB thanks Malte Siemers, Michalis Lazaratos and Oskar Klaja for valuable discussions. We acknowledge the use of the GPCRdb database (http://www.gpcrdb.org).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2020.107634.

References

Alsabti, K., Ranka, S., Singh, V., 1997. An efficient k-means clustering algorithm. Electrical Engineering and Computer Science 43, 1–6.

Amitai, G., Shemesh, A., E., S., Shklar, M., Netanely, D., Venger, I., Pietrovski, S., 2004. Network analysis of protein structures identifies functional residues. J. Mol. Biol. 344, 1135-1146.

Ballesteros, J.A., Weinstein, H., 1995. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G Protein-Coupled Receptors. Methods in Neurosciences 25, 366–428.

Berman, H.M., Westbrook, J., Feng, G., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acid Res. 28, 235–242.

Blankenship, E., Vahedi-Faridi, A., Lodowski, D.T., 2015. The high-resolution structure of activated opsin reveals a conserved solvent network in the transmembrane region essential for activation. Structure 23, 2358–2364.

Bondar, A.-N., Smith, J.C., 2017. Protonation-state coupled conformational dynamics in reaction mechanisms of channel and pump rhodopsins. Photochem. Photobiol. 93, 1336–1344.

Bondar, A.-N., Lemieux, H.J., 2019. Reactions at membrane interfaces. Chem. Rev. 119, 6162–6183.

Bondar, A.-N., Baudry, J., Suhai, S., Fischer, S., Smith, J.C., 2008. Key role of active-site water molecules in bacteriorhodopsin proton-transfer reactions. J. Phys. Chem. B 112, 14729–14741.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem 4, 187–217.

Che, T., English, J., Krumm, B.E., Kim, K., Pardon, E., Olsen, R.H.J., Wang, S., Zhang, S., Diberto, J.F., Sciaky, N., Carroll, F.I., Steyaert, J., Wacker, D., Roth, B.L., 2020. Nanobody-enabled monitoring of kappa opioid receptor states. Nature Comm. 11, 1145.

Che, T., Majumdar, S., Zaidi, S.A., Ondachi, P., McCorvy, J.D., Wang, S., Mosier, P.D., Uprety, R., Vardy, E., Krumm, B.E., Han, G.W., Lee, M.-Y., Pardon, E., Steyaert, J., Huang, X.-P., Strachan, R.T., Tribo, A.R., Pasternak, G.W., Carroll, F.I., Stevens, R.C., Cherezov, V., Katritch, V., Wacker, D., Roth, B.L., 2018. Structure of nanobody-stabilized active state of the kappa opioid receptor. Cell 172, 55–67.

Claff, T., Yu, J., Blais, V., Patel, N., Martin, C., Wu, L., Han, G.W., Holleran, B.J., van der Poorten, O., White, K.L., Hanson, M.A., Sarret, P., Gendron, L., Cherezov, V., Katrich, V., Ballet, S., Liu, Z.-J., Müller, C.E., Stevens, R.C., 2019. Elucidating the active δ-opioid receptor crystal structure with peptide and small-molecule agonists. Science. Advances 5, eaax9115.

Costa, T., Herz, A., 1989. Antagonists with negative intrinsic activity at δ opioid receptors coupled to GTP-binding proteins. Proc. Natl. Acad. Sci. USA 86, 7321–7325.

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: an N x log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092.

del Val, C., White, S.H., Bondar, A.-N., 2012. Ser/Thr motifs in transmembrane proteins: conservation patterns and effects on local protein structure and dynamics. J. Membr. Biol. 245, 717–730.

del Val, C., Bondar, L., Bondar, A.-N., 2014. Coupling between inter-helical hydrogen bonding and water dynamics in a proton transporter. J. Struct. Biol. 186, 95–111.

Deupi, X., Edwards, P., Singhal, A., Nickle, B., Oprian, D., Schertler, G., Standfuss, J., 2012. Stabilized G protein binding site in the structure of constitutively active metharhodopsin-II. Proc. Natl. Acad. Sci. 109, 119–124.

Eddy, M.T., Lee, M.-Y., Gao, Y.-G., White, K.L., Didenko, T., Horst, R., Audet, M., Stanczak, P., McClary, K.M., Han, G.W., Jacobson, K.A., Stevens, R.C., 2018. Allosteric coupling of drug binding and intracellular signaling in the A_{2A} adenosine receptor. Cell 172, 68–80.

Erlandson, S.C., McMahon, C., Kruse, A.C., 2018. Structural basis for G Protein-Coupled Receptor signaling. Annual Reviews of Biophysics 47, 1–18.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G., 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103, 8577–8593.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD-96 Proceedings 96, 226–231.

Eswar, N., Marti-Renom, M.A., Webb, B., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., Sali, A., 2006. Comparative structure modeling with MODELLER. Current Protocols in Bioinformatics 15, 1–30.

Feller, S.E., MacKerell Jr., A.D., 2000. An improved empirical potential energy function for molecular simulations of phospholipids. J. Phys. Chem. B 104, 7510–7515.

Feller, S.E., Zhang, Y., Pastor, R.W., Brooks, B., 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. 103, 4613–4621.

Fenalti, G., Giguere, P.M., Katritch, V., Huang, X.-P., Thomson, A.A., Cherezov, V., Roth, B.L., Stevens, R.C., 2014. Molecular control of δ-opioid receptor signalling. Nature 506, 191–196.

Fenalti, G., Zatsepin, N.A., Betti, C., Giguere, P.M., Han, G.W., Ishchenko, A., Liu, W., Guillemyn, K., Zhang, J., James, D., Wang, D., Weierstall, U., Spence, J.C.H., Boutet, S., Messerchnmidt, M., Williams, G.J., Gati, C., Yefanov, O.M., White, T.A., Oberthuer, D., Metz, M., Yoon, C.H., Barty, A., Chapman, H.N., Basu, S., Coe, J., Conrad, C.E., Fromme, R., Fromme, P., Tourwé, D., Schiller, P.W., Roth, B.L., Ballet, S., Katritch, V., Stevens, R.C., Cherezov, V., 2015. Structural basis for fifunctional peptide recognition at human δ-opioid receptor. Nature Struct. Mol. Biol. 22, 265–268.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35–41.

Gottschalk, M., Dencher, N.A., Halle, B., 2001. Microsecond exchange of internal water molecules in bacteriorhodopsin. J. Mol. Biol. 311, 605–621.

Granier, S., Manglik, A., Kruse, A.C., Kobilka, T.S., Thian, F.S., Weis, W.I., Kobilka, B.K., 2012. Structure of the δ-opioid receptor bound to natrindole. Nature 485, 400–403.

Harris, A., Lazaratos, M., Siemers, M., Watt, E., Hoang, A., Tomida, S., Schubert, L., Saita, M., Heberle, J., Furutani, Y., Kandori, H., Bondar, A.-N., Brown, L.S., 2020. Mechanism of inward proton transport in an Antartic microbial rhodopsin. J. Phys. Chem. B 124, 4851–4872.

Hauser, A.S., Attwood, M.M., Rask-Andersen, M., Schiöth, H.B., Gloriam, D.E., 2017. Trends in GPCR drug discovery: new agents, targets and indications. Nat. Rev. Drug Discovery 16, 829–842.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H., MacKerell Jr., A.D., 2016. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods 40, 71–73.

Huang, W., Manglik, A., Venkatakrishnan, A.J., Laeremans, T., Feinberg, E.N., Sanborn, A.L., Kato, H.E., Livingston, K.E., Thorsen, T.S., Kling, R.C., Granier, S., Gmeiner, P., Husbands, S.M., Trynor, J.R., Weis, W.I., Steyaert, J., Dror, R.O., Kobilka, B.K., 2015. Structural insights into μ-opioid receptor activation. Nature 524, 315–321.

Humphrey, W., Dalke, W., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38.

Isberg, V., de Graaf, C., Bortolato, A., Cherezov, V., Katrich, V., Marshall, F.H., Mordalski, S., Pin, J.-P., Stevens, R.C., Vriend, G., Gloriam, D.E., 2015. Generic GPCR residue numbers - aligning topology maps minding the gaps. Trends Pharmacol. Sci. 36, 22–31.

Jardon-Valadez, E., Bondar, A.-N., Tobias, D.J., 2009. Dynamics of the internal water molecules in squid rhodopsin. Biophys. J. 96, 2572–2576.

Jardon-Valadez, E., Bondar, A.-N., Tobias, D.J., 2010. Coupling of retinal, protein, and water dynamics in squid rhodopsin. Biophys. J. 99, 2200–2207.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935.

Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., Schulten, K., 1999. NAMD2: greater scalability for parallel molecular dynamics. J. Comput. Phys. 151, 283–312.

Karathanou, K., Bondar, A.-N., 2019. Using graphs of dynamic hydrogen-bond networks to dissect conformational coupling in a protein motor. J. Chem. Inf. Model. 59, 1882–1896.
Katritch, V., Fenalti, G., Abola, E.E., Roth, B.L., Cherezov, V., Stevens, R.C., 2014. Allosteric sodium in class A GPCR signaling. Trends Biochem. Sci. 39, 233–244.
Klauda, J.B., Venable, R.M., Freites, J.A., O'Connor, J.W., Tobias, D.J., Mondragon-Ramirez, C., Votrobyov, I., MacKerell Jr., A.D., Pastor, R.W., 2010. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. J. Phys. Chem. B 114, 7830–7843.
Koehl, A., Hu, H., Maeda, S., Zhang, Y., Qu, Q., Paggi, J.M., Latorraca, N., Hilger, D., Dawson, R., Matile, H., Schertler, G.F.X., Granier, S., Weis, W.I., Dror, R.O., Manglik, A., Skiniotis, G., Kobilka, B.K., 2018. Structure of the μ-opioid receptor-G_i protein complex. Nature 558, 547–551.
Kučerka, N., Nagle, J.F., Sachs, J.N., Feller, S.E., Pencer, J., Jackson, A., Katsaras, J., 2008. Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. Biophys. J. 95, 2356–2367.
Lappano, R., Maggiolini, M., 2011. G protein-coupled receptors: novel targets for drug discovery in cancer. Nat. Rev. Drug Discovery 10, 47–60.
Lazaratos, M., Karathanou, K., Bondar, A.-N., 2020. Graphs of dynamic H-bond networks: from model proteins to protein complexes in cell signaling. Curr. Opin. Struct. Biol. 64, 79–87.
Lee, Y., Kim, S., Choi, S., Hyeon, C., 2016. Ultraslow water-medated transmembrane interactions regulate the activation of A_{2A} adenosine receptor. Biophys. J. 111, 1180–1191.
Leitner, D.M., Hyeon, C., Reid, K.M., 2020. Water-mediated biomolecular dynamics and allostery. J. Chem. Phys. 152, 240901.
Levitt, M., Park, B.H., 1993. Water: now you see it, now you don't. Structure 223–226.
Li, J., Ge, Y., Huang, J.-X., Strømgaard, K., Zhang, X., Xiong, X.-F., 2019. Heterotrimeric G protein as therapeutic targets in drug discovery. J. Med. Chem. 63, 5013–5030.
Liu, H., Kim, H.R., Deepak, R.N.V.K., Wang, L., Chung, K.Y., Fan, H., Wei, Z., Zhang, C., 2018. Orthosteric and allosteric activation of the C5a receptor antagonists. Nature Struct. Mol. Biol. 25, 472–481.
Liu, W., Chun, E., Thomson, A.A., Chubukov, P., Xu, F., Katritch, V., Han, G.W., Roth, C. B., Heitman, L.H., IJzerman, A.P., Cherezov, V., Stevens, R.C., 2012. Structural basis for allosteric regulation of GPCRs by sodium ions. Science 337, 232-236.
Lomize, M., Pogozheva, I.D., Joo, H., Mosberg, H.I., Lomize, A.L., 2011. OPM database and PPM server: resources for positioning of proteins in membranes. Nucleic Acid Res. 40, D370–D376.
MacKerell Jr., A.D., Feig, M., Brooks, C.L.I., 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J. Comput. Chem 25, 1400–1415.
MacKerell Jr., A.D., Bashford, D., Bellot, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W. E.I., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M., 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 3586–3616.
Manglik, A., Kruse, A.C., Kobilka, T.S., Thian, F.S., Mathiesen, J.M., Sunahara, R.K., Pardo, L., Weis, W.I., Kobilka, B.K., Granier, S., 2012. Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. Nature 485, 321–326.
Martyna, G.J., Tobias, D.J., Klein, M.L., 1994. Constant-pressure molecular-dynamics algorithms. J. Chem. Phys. 101, 4177–4189.
Mirzadegan, T., Benkö, G., Filipek, S., Palczewski, K., 2003. Sequence analyses of G-Protein-Coupled Receptors: Similarities to rhodopsin. Biochemistry 42, 2759–2767.
Murakami, M., Kouyama, T., 2008. Crystal structure of squid rhodopsin. Nature 453, 363–367.
Naik, M., Sawant, K., 2013. AutoEpsDBSCAN: DBSCAN with Eps automatic for large dataset. International Journal on Advanced Computer Theory and Engineering 2, 11–16.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., M., B., Prettenhofer, P., Weiss, R., Duborg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Phyton. Journal of Machine Learning Research 12, 2825-2830.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - a vizualization system for exploratory research and analysis. J. Comput. Chem 25, 1605–1612.
Phillips, J.C., Braun, B., Wang, W., Gumbart, J., Takjkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., Schulten, K., 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem 26, 1781–1802.
Quillin, M.L., Wingfield, P.T., Matthews, B.W., 2006. Determination of solvent content in cavities in IL-1β using experimentally phased electron density. Proc. Natl. Acad. Sci. 103, 19749–19753.
Rasmussen, S.G.F., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, A.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., 2012. Crystal structure of the β2Adrenergic receptor-Gs protein complex. Nature 477, 549–555.
Rousseuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.
Sadée, W., Wang, D., Bilsky, E.J., 2005. Basal opioid receptor activity, neutral antagonists, and therapeutic opportunities. Life Sci. 76, 1427–1437.
Segala, E., Guo, D., Cheng, R.K.Y., Bortolato, A., Dflorian, F., Doré, A.S., Errey, J.C., Heitman, L.H., IJzerman, A.P., Marshall, F.H., Cooke, R.M., 2016. Controlling the dissociation of ligands from the Adenosine A2A receptor through modulation of salt bridge strength. J. Med. Chem. 59, 6470-6479.
Seth, P., Rudd, R.A., Noonan, R.K., Haegerich, T.M., 2018. Quantifying the epidemic of prescription opioid overdose deaths. AJPH Surveilance 108, 500–502.
Shang, Y., LeRouzic, V., Schneider, S., Bisignano, P., Pasternak, G.W., Filizola, M., 2014. Mechanistic insights into the allosteric modulation of opioid receptors by sodium ions. Biochemistry 53, 5140–5149.
Shihoya, W., Nishizawa, T., Yamshita, K., Inoue, A., Hirata, K., Kadji, F.M.N., Okuta, A., Tani, K., Aoki, J., Fukjiyoshi, Y., Doi, T., Nureki, O., 2017. X-ray structures of the endothelin ET_B receptor bound to clinical antagonist bosentan and its analog. Nature Struct. Mol. Biol. 9, 758–764.
Siemers, M., Lazaratos, M., Karathanou, K., Guerra, F., Brown, L.S., Bondar, A.-N., 2019. Bridge: A graph-based algorithm to analyze dynamic H-bond networks in membrane proteins. J. Chem. Theory Comput. 15, 6781–6798.
Sirohi, S., Walker, B.M., 2015. Maturational alterations in constitutive activity of medial prefrontal cortex kappa-opioid receptors in Wistar rats. J. Neurochemistry 135, 659–665.
Suno, R., Kimura, K.T., Nakane, T., Yamashita, K., Wang, J., Fujiwara, T., Yamanaka, Y., Im, D., Horita, S., Tsujimoto, H., Tawaramoto, M.S., Hirokawa, T., Nango, E., Tono, K., Kameshima, T., Hatsui, T., Joti, Y., Yabashi, M., Shimamoto, K., Yamamoto, M., Rosenbaum, D.M., Iwata, S., Shimamura, T., Kobayashi, T., 2018. Crystal structures of human orexin 2 receptor bound to the subtype-selective antagonist EMPA. Structure 26, 7–19.
Suomivuouri, C.-M., Lattoraca, N.R., Wingler, L.M., Eismann, S., King, M.C., Kleinhenz, A.L.W., Skiba, M.A., Staus, D.P., Kruse, A.C., Lefkowitz, R.J., Dror, R.O., 2020. Molecular mechanism of biased signaling in a prototypical G protein-coupled receptor. Science 367, 881–887.
Trzaskowski, B., Latek, D., Yuan, S., Ghoshdastider, U., Debinski, A., Filipek, S., 2012. Action of molecular switches in GPCRs - Theoretical and experimental studies. Curr. Med. Chem. 19, 1090–1109.
Varma, A., Mutt, E., Mühle, J., Panneels, V., Terakita, A., Deupi, X., Nogly, P., Schertler, G.X., Lesca, E., 2019. Crystal structure of jumping spider rhodopsin-1 as a light sensitive GPCR. Proc. Natl. Acad. Sci. 116, 14547–14556.
Venkatakrishnan, A.J., Deupi, X., Lebon, G., Tate, C.G., Schertler, G.F., Babu, M.M., 2013. Molecular signatures of G-protein-coupled receptors. Nature 494, 185–194.
Venkatakrishnan, A.J., Ma, A.K., Fonseca, R., Latorraca, N.R., Kelly, B., Betz, R.M., Asawa, C., Kobilka, B.K., Dror, R.O., 2019. Diverse GPCRs exhibit conserved water networks for stabilization and activation. Proc. Natl. Acad. Sci. 116, 3288–3293.
Vickery, O.N., Carvalheda, C.A., Zaidi, S.A., Pisliakov, A.V., Katritch, V., Zachariae, U., 2018. Intracellular transfer of Na^+ in an active-state G-Protein-Coupled Receptor. Structure 26, 171–180.
Wang, D., Raehal, K.M., Bilsky, E.J., Sadée, W., 2001. Inverse agonists and neutral antagonists at μ opioid receptor (MOR): possible role of basal receptor signaling in narcotic dependence. J. Neurochemistry 77, 1590–1600.
Wang, S., Wacker, D., Levit, A., Che, T., Betz, R.M., McCorvy, J.D., Venkatakrishnan, A. J., Huang, X.-P., Dror, R.O., Shoichet, B.K., Roth, B.L., 2017. D_4 dopamine receptor high-resolution structures enable the discovery of selective agonists. Science 358, 381–386.
Weinert, T., Olieric, N., Cheng, R., Brünle, S., James, D., Ozerov, D., Gashi, D., Vera, L., Marsh, M., Jaeger, K., Dworkowski, F., Panepucci, E., Basu, S., Skopintsev, P., Doré, A.S., Geng, T., Cooke, R.M., Liang, M., Prota, A.E., Paeels, V., Nogly, P., Ermler, U., Schertler, G., Hennig, M., Steinmetz, M.O., Wang, M., Standfuss, J., 2017. Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. Nature. Comm. 8.
White, K.L., Eddy, M.T., Gao, Z.-G., Han, G.W., Lian, T., Deary, A., Patel, N., Jacobson, K. A., Katritch, V., Stevens, R.C., 2018. Structural connection between activation microswitch and allosteric sodium site in GPCR signaling. Structure 26, 259–269.
Wingler, L.M., Skiba, M.A., McMahon, C., Staus, D.P., Kleinhenz, A.L.W., Suomivuouri, C.-M., Lattoraca, N.R., Dror, R.O., Lefkowitz, R.J., Kruse, A.C., 2020. Angiotensin and biased analogs induce structurally distinct active conformations within a GPCR. Science 367, 888–892.
Wu, E.L., Cheng, X., Jo, S., Rui, H., Song, K.C., Dávila-Contreras, E.M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R.M., Klauda, J.B., Im, W., 2014. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J. Comput. Chem 35, 1997–2004.
Wu, H.-L., Wacker, D., Mileni, M., Katritch, V., Han, G.W., Vardy, E., Liu, W., Thomson, A.A., Huang, X.-P., Carroll, F.I., Mascarella, S.W., Westkaemper, R.B., Mosier, P.D., Roth, B.L., Cherezov, V., Stevens, R.C., 2012. Structure of the human k-opioid receptor in complex with JDTic. Nature 485, 327–332.
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. Nature Comm. 5, 4733.
Zhang, X.C., Cao, C., Zhou, Y., Zhao, Y., 2015. Proton transfer-mediated GPCR activation. Protein. Cell 6, 12–17.
Zhou, Q., Yang, D., Wu, M., Guo, Y., Guo, W., Zhong, L., Cai, X., Dai, A., Jang, W., Shakhanovich, E.I., Liu, Z.-J., Stevens, R.C., Lambert, N.A., Babu, M.M., Wang, M.-W., Zhao, S., 2019. Common activation mechanisms of class A GPCRs. eLife 8, e50279.