Received: 31 August 2023 | Revised: 3 February 2024 | Accepted: 2 March 2024

DOI: 10.1111/bph.16387

Themed Issue: Complexity of GPCR Modulation and Signaling (ERNST)

THEMED ISSUE REVIEW

Graph-based algorithms to dissect long-distance water-mediated H-bond networks for conformational couplings in GPCRs

Éva Bertalan[1] | Matthew Joseph Rodrigues[2] | Gebhard F. X. Schertler[2] | Ana-Nicoleta Bondar[3,4]

[1]Physikzentrum, RWTH-Aachen University, Aachen, Germany

[2]Laboratory of Biomolecular Research, Paul Scherrer Institut, Villigen, Switzerland

[3]Forschungszentrum Jülich, Institute of Computational Biomedicine, Jülich, Germany

[4]Faculty of Physics, University of Bucharest, Măgurele, Romania

Correspondence
Ana-Nicoleta Bondar, Forschungszentrum Jülich, Institute for Computational Biomedicine (IAS-5/INM-9), Wilhelm-Johnen-Straße, Jülich 52428, Germany; University of Bucharest, Faculty of Physics, Atomistilor 405, Măgurele 077125, Romania.
Email: nbondar@fizica.unibuc.ro; a.bondar@fz-juelich.de

Funding information
H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 701647; European Research Council, Grant/Award Number: 951644; DFG, German Research Foundation, Grant/Award Number: 491111487

Abstract

Changes in structure and dynamics elicited by agonist ligand binding at the extracellular side of G protein coupled receptors (GPCRs) must be relayed to the cytoplasmic G protein binding side of the receptors. To decipher the role of water-mediated hydrogen-bond networks in this relay mechanism, we have developed graph-based algorithms and analysis methodologies applicable to datasets of static structures of distinct GPCRs. For a reference dataset of static structures of bovine rhodopsin solved at the same resolution, we show that graph analyses capture the internal protein–water hydrogen-bond network. The extended analyses of static structures of rhodopsins and opioid receptors suggest a relay mechanism whereby inactive receptors have in place much of the internal core hydrogen-bond network required for long-distance relay of structural change, with extensive local H-bond clusters observed in structures solved at high resolution and with internal water molecules.

LINKED ARTICLES: This article is part of a themed issue Complexity of GPCR Modulation and Signaling (ERNST). To view the other articles in this section visit http://onlinelibrary.wiley.com/doi/10.1111/bph.v182.14/issuetoc

KEYWORDS

bovine rhodopsin, C-Graphs, conserved hydrogen-bond networks, G protein coupled receptors, opioid receptors, protein-water hydrogen-bond networks, visual rhodopsins, water-mediated hydrogen-bond networks

1 | INTRODUCTION

G protein coupled receptors (GPCRs) are seven-helical membrane proteins that eukaryotic cells use to mediate communication with the external world. To do this, GPCRs bind an activating ligand, called an agonist, at their extracellular side, they undergo a conformational change and become activated. The GPCR then activates heterotrimeric G protein partners at the cytoplasmic side by catalysing exchange of **guanosine diphosphate (GDP)** for **GTP**. This causes the trimer to dissociate into Gα and Gβγ, which then interact with effector proteins, activating downstream signalling pathways that ultimately lead to a physiological response (Choi et al., 2012). Humans

Abbreviations: BW, Ballesteros-Weinstein; CWxP, Cys-Trp-x-Pro, where x indicates any amino acid residue; DRY, Asp-Arg-Tyr; JSR-1, jumping spider rhodopsin-1; OPM, Orientations of Proteins in Membranes; PCA, Principal Component Analysis; PIF, Pro-Ile-Phe; SACLA, SPring-8 Angstrom Compact free electron Laser; SFX, serial femtosecond crystallography.

have >800 GPCR sequences (Fredriksson et al., 2003), with specialized GPCRs responsible for mediating cellular responses to diverse extracellular stimuli, such as photons, protons and other ions, odorant molecules, hormones, neurotransmitters and proteins (Wacker et al., 2017). GPCRs are important drug targets, with about one third of the drugs targeting GPCRs (Hauser et al., 2017), and GPCRs contributing ~12% of all human protein drug targets (Santos et al., 2017). Knowledge of how GPCRs work is essential, as it can guide the development of new therapeutics.

Experimental and computational data have documented the key role of hydrogen (H)-bonds and of internal water molecules for GPCR function (see, e.g. Lesca et al., 2018; Rosenbaum et al., 2009; Venkatakrishnan et al., 2013, 2019). Class A GPCRs, which we discuss here, share several conserved motifs, also known as molecular switches, which rearrange along the GPCR activation pathway (Fritze et al., 2003; Goncalves et al., 2010; Trzaskowski et al., 2012; Vogel et al., 2008; Yuan et al., 2014; Zhou et al., 2019). The Ballesteros–Weinstein (BW) numbering scheme facilitates comparison of GPCR primary amino acid sequences, using a common numbering scheme (Ballesteros & Weinstein, 1995). In this scheme, the residues of each alpha helix are numbered by the helix number followed by the position of the residue relative to the most conserved residue in the helix, which is numbered

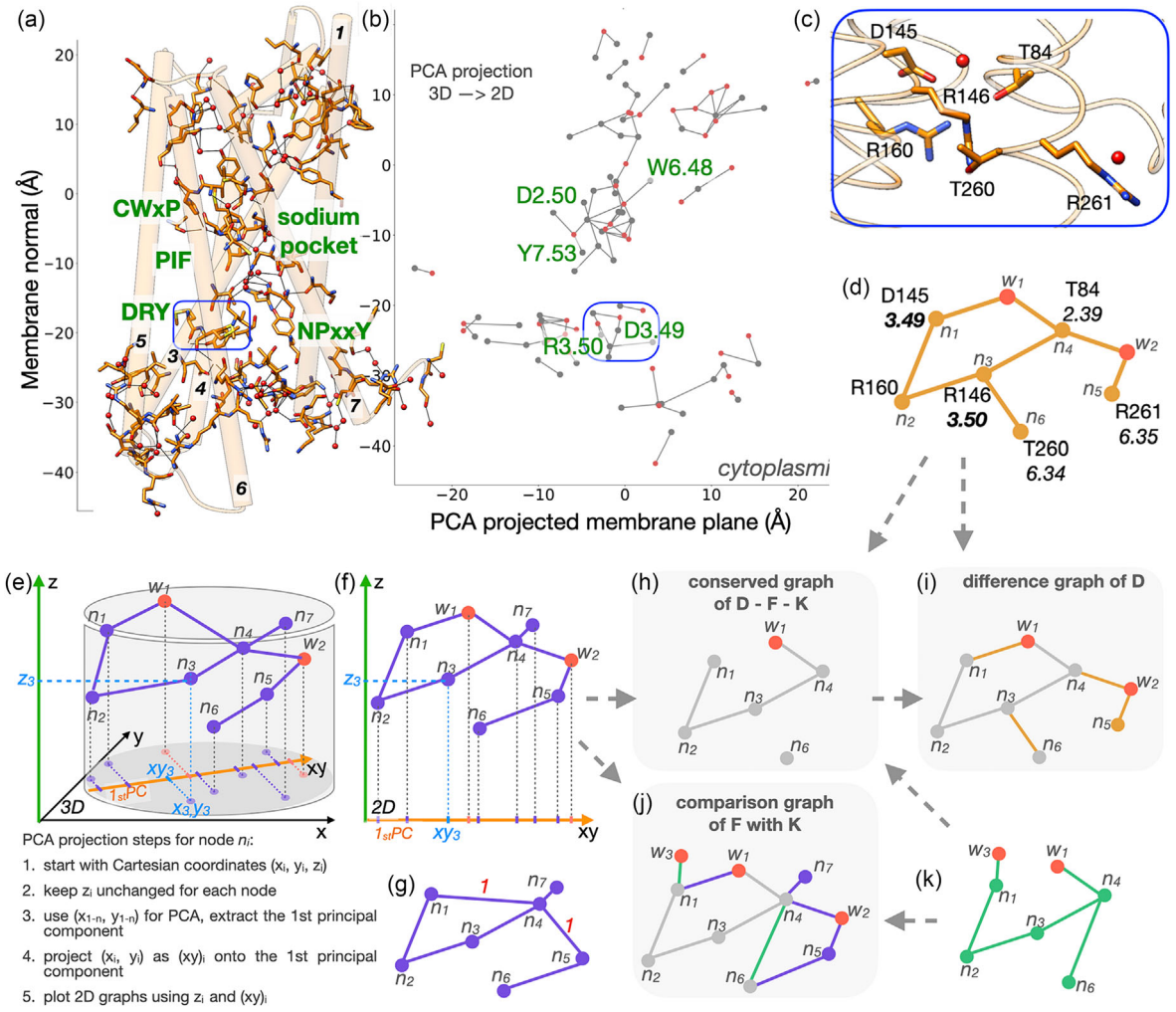

FIGURE 1 C-graphs computes H-bond graphs of GPCRs and projects them onto the membrane normal. (a) The inactive δ receptor structure (pdb 4n6h) is used as a reference in this illustration. Conserved functional motifs of class A GPCRs are labelled. (b) The sidechain–water H-bond graph of the reference structure from panel a is represented with the Cα coordinates of the H-bonding sidechains projected onto the membrane normal (vertical axis), and on the PCA projected membrane plane. Grey nodes are sidechains part of the H-bond network and red nodes, water molecules. (c) Local H-bond cluster of DRY-R146$^{3.50}$. (d) Schematic representation of the local H-bond cluster from panel c; nodes of are labelled n_1–n_6 for protein sidechains and w_1, w_2, for water molecules. (e,f) Schematic representation of a local cluster in the membrane receptor with cylindrical symmetry (panel e) and its PCA projection as summarized in panel e and illustrated in panel f; in both panels e and f, nodes representing water molecules are shown explicitly as red dots. (g) The water wire representation of the H-bond cluster from panel f, now with bridging water molecules represented implicitly by the numbers '1' along the corresponding edges. (h–j) Conserved (panel h), difference H-bond graph (panel i) and comparison (panel j), computed, with a conservation threshold of 100%, for the H-bond clusters illustrated in panels d, f and k. The conserved H-bond graph consists of nodes n_1–n_6 and water w_1. The difference graph consists of nodes and edges which are absent from the networks of shown in panels f or k. The comparisons graph shows with grey the nodes and edges present in both f and k networks, and colour codes with purple the nodes and edges only present in f and with green the ones only present in k.

50. For example, D3.49 refers to an aspartate residue located in helix 3, directly N-terminal of the most conserved residue. These BW numbers are commonly used to refer to the conserved motifs.

The DRY motif, composed of D3.49, R3.50 and Y3.51 (Figure 1), is at the cytoplasmic side of the receptor, where it participates in the binding of the G protein (Rasmussen et al., 2011). Located relatively close to the core of the receptor, Y7.53 of the NPxxY motif (N7.49, P7.50, Y7.53) rearranges during GPCR activation so as to sample distinct interactions with a wire of internal water molecules (Yuan et al., 2014) and bridges via water to Y5.58, and from Y5.58 to DRY-R3.50 (Deupi et al., 2012; Venkatakrishnan et al., 2019). The CWxP (C6.47, W6.48, P6.50) and PIF (P5.50, I3.40, F6.44) motifs are located approximately at the core of the receptor (Figure 1). Rearrangement of W6.48 during GPCR activation associates with changes in internal water H-bonding (Deupi et al., 2012; Standfuss et al., 2011). Many class A GPCRs bind sodium at a site delineated by D2.50, S3.39, other polar sidechains and water molecules (Katritch et al., 2014). Interactions and dynamics at the sodium-binding site are thought to couple to agonist binding and protein conformational dynamics (Katritch et al., 2014). Water molecules play key roles along the activation of GPCRs, as they help stabilize discrete conformational states, participate in ligand binding and modulate receptor function (Lesca et al., 2018).

The discrete structural events summarized above are milestones along the activation path of GPCRs. Whether and how GPCRs use water-mediated H-bond networks to relay changes in structure and dynamics from the ligand-binding site to the G-protein binding site remain fundamental open questions.

The goal of this review is to discuss the usefulness of recently developed graph-based algorithms for analyses of water-mediated H-bond networks of GPCRs. As model systems for GPCRs, we will primarily focus on bovine **rhodopsin**, which we will compare with other visual receptors, **opioid receptors** and the **adenosine A_{2A} receptor**. We will largely restrict ourselves to discussing H-bond networks extracted from datasets of static structures solved with X-ray crystallography.

2 | WATER-MEDIATED H-BOND NETWORKS IN CRYSTAL STRUCTURES OF BOVINE RHODOPSIN

Much about water in GPCRs was revealed, at least initially, from work on bovine rhodopsin (BHR), the receptor whose importance as a model system for GPCRs is underlined by the fact that class A GPCRs are also known as 'rhodopsin-like' (Deupi, 2014). In the dark, rhodopsin covalently binds its endogenous ligand, **retinal**, in the 11-cis inverse agonist configuration. Upon absorption of light in the visible wavelength range, the retinal photo-isomerizes to **all-trans retinal**, which is an agonist configuration. This mechanism of activation, whereby the ligand is already bound to the inactive receptor, and the receptor can be simply activated by light, makes rhodopsin an ideal system to study the general principles of GPCR activation.

A role for water molecules in mediating ligand–receptor interactions was initially postulated based on spectroscopy data of rhodopsin (Eilers et al., 1999; Nagata et al., 1997; Verhoeven et al., 2001). Discrete water molecules and water-mediated H-bond networks were then observed in crystal structures of rhodopsin—the first structure, solved at 2.8-Å resolution, had an interhelical water molecule located within H-bond distance from $D83^{2.50}$ and the carbonyl group of $G120^{3.35}$, and close to $N302^{7.49}$, such that it could mediate contacts between helices 2, 3 and 7 of the receptor (Palczewski et al., 2000). The subsequently refined structure indicated additional water molecules could occupy internal cavities, including two cavities near the retinal ligand, and the suggestion was made that protein motions could allow water molecules to visit these sites (Teller et al., 2001). Seven discrete water molecules near conserved residues of class A GPCRs were then consistently found in both rhodopsin polypeptide chains of the crystal structure solved at 2.6-Å resolution—including a cluster of three water molecules part of a local H-bond cluster with $N302^{7.49}$ of the NPxxY motif and $D83^{2.50}$, and two water molecules near the retinal—one water bridging E181 and S186 via H-bonds, and the second water molecule, near $E113^{3.28}$ (Okada et al., 2001); the two water molecules near the retinal were subsequently confirmed by the higher, 2.2-Å resolution structure (Okada et al., 2004). At about the same time, the refined structure of dark-state rhodopsin revealed ordered water molecules that mediate an H-bond network extending from $W265^{6.48}$ in the vicinity of the retinal to NPxxY and what became known as 'the complex counterion' of the retinal Schiff base, the interaction between an ordered water molecule and the negatively charged $E113^{3.28}$ counterion (Li et al., 2004).

Subsequently, three-dimensional structures were solved for rhodopsin in the early intermediates of the photocycle, batho and lumi (Nakamichi et al., 2007; Nakamichi & Okada, 2006) and in active (meta) conformations (Choe et al., 2011; Standfuss et al., 2011). Structures of the two meta conformations of rhodopsin, solved at resolutions of 2.9- to 3.0-Å resolution, indicated significantly fewer internal water molecules than the 2.2-Å resolution dark state structure (16–19 vs. 37 internal waters), which can make it challenging to directly compare their water-mediated H-bond networks (Bertalan et al., 2021).

Most recently, a set of five structures were solved for rhodopsin at 1.8-Å resolution at two Free Electron Laser facilities: two structures for the dark state solved with serial femtosecond crystallography (SFX) at SACLA (SPring-8 Angstrom Compact free electron Laser) and the SwissFEL, rhodopsin at 1 and 10 ps after illumination solved with TR-SFX (time-resolved SFX) at the SwissFEL (free-electron laser at the Paul Scherrer Institute [PSI]) and the structure at 100 ps after illumination, solved with TR-SFX at SACLA (Gruhl et al., 2023). This dataset of structures provides an unprecedented opportunity to evaluate the rearrangements of water-mediated H-bond networks upon illumination. Because the structures were solved at the same resolution and because for each intermediate after illumination there is a reference dark state structure solved at the same facility, it is reasonable to anticipate that there are minimal variations in H-bonding caused by lack of waters in lower resolution structures (Bertalan & Bondar, 2023) or from the structure solving protocol.

3 | GRAPH-BASED APPROACHES TO EVALUATE THE WATER-MEDIATED H-BOND NETWORKS OF GPCRS

We have briefly summarized above some of the key findings, from structural biology, on the role of internal water molecules in rhodopsin function. More generally, internal water-mediated H-bonds and H-bond networks are central for the allosteric coupling mechanisms used by GPCRs to ensure long-distance communication between the ligand- and the G-protein binding sites (Venkatakrishnan et al., 2019). There is currently a wealth of data from structural biology on structures of GPCRs, including at resolutions high enough to observe water molecules: according to GPCRdb (accessed on July 16th, 2023), there are currently 1177 static structures of GPCRs that could be used to evaluate intramolecular interactions that may shape GPCR conformational dynamics.

The challenge is how to efficiently analyse and compare structures—of the same receptor solved in different experimental conditions or of structures of distinct receptors, for example in inactive versus active conformations, and how to display these comparisons among H-bond networks. To tackle this challenge, we have recently developed the graph-based algorithm C (onserved)-Graphs (C-Graphs), which allowed us to directly compare and dissect H-bond networks of GPCRs with different amino acid residue sequences (Bertalan et al., 2020, 2021; Bertalan & Bondar, 2023) and of microbial rhodopsins (Bertalan & Bondar, 2023; Morizumi et al., 2023). The principles of H-bond graph computations and C-Graphs computations for GPCRs are illustrated in Figure 1 and summarized below.

Briefly, an *H-bond graph* consists of nodes, which are the H-bonding groups, and edges, which are direct H-bonds or water-mediated bridges between these H-bonding groups. H-bond graphs computed from static protein structures have as nodes (H-bonding groups) protein groups, ligand molecules, or water molecules solved with the static protein structure. A *local H-bond cluster* consists of a subset of H-bonding groups interconnected to each other via H-bonds. We can use Connected Component Analyses (Cormen et al., 2009) to extract local clusters from the H-bond graphs in the graph-based algorithm Bridge (Siemers et al., 2019).

We compute H-bonds using standard geometric criteria. For static protein structures that lack coordinates for H atoms, as is the case of all GPCR structures discussed here, we use as *H-bond criterion* a distance ≤3.5 Å between the H-bond donor and acceptor hetero-atoms. An H-bond angle criterion, typically within 60°, may be added to structures that include coordinates for H atoms. The *length of a water bridge* between two protein sidechains is given by the number of H-bonding water molecules in the bridge. Salt bridges, such as between Asp/Glu and Arg/Lys side chains, are computed and reported together with the other H-bonds.

H-bond graphs for each of the individual structures are computed with the graph-based algorithm Bridge/Bridge2 (Siemers et al., 2019; Siemers & Bondar, 2021). These computations are very efficient: tests using single cores showed that Bridge2 needed <1 s to compute 596 H-bonds between sidechains of **covid-19 spike protein S** and 3.1 s for its 12368 water-mediated H-bonds (Siemers & Bondar, 2021); compared to MDAnalysis (Gowers et al., 2016; Michaud-Agrawal et al., 2011), Bridge was ∼216 times faster in a test computation of the water-mediated H-bond network of the C1C2 channelrhodopsin chimera (Siemers et al., 2019).

To compare H-bond graphs computed for distinct GPCRs, we align their structures in the membrane, either by using Chimera (Pettersen et al., 2004) or taking prealigned structures from the database Orientations of Proteins in Membranes, OPM (Lomize et al., 2011). We then represent nodes of the H-bond graph by the Cα coordinates of the amino acid residues projected onto the membrane normal (vertical, *z* axis of the Cartesian coordinate system), and onto the first principal component computed with Principal Component Analysis (PCA) for the membrane (*x*-*y*) plane (Bertalan et al., 2020, 2021). PCA finds the principal components (eigenvectors) of the data that capture the most variance and then it projects the data points onto a lower-dimensional space defined by the chosen eigenvectors. More specifically, within C-Graphs the Cartesian *x*-*y* coordinates of the Cα atoms of the protein residues part of the H-bond network and of the network's water oxygen atoms, are projected onto a line given by the orthonormal eigenvector with the largest eigenvalue of the diagonalized covariance matrix of the atomic coordinates. The advantage of using PCA to project the membrane plane is that we can present a 2D point (atomic coordinates along the *x*, *y* axes) as a single point, which is then plotted along the horizontal axis of the H-bond graphs (Figure 1e,f).

In Figure 1b, we illustrate an H-bond graph calculation for the structure of the inactive **δ receptor** bound to **naltrindole** (Fenalti et al., 2014). For this H-bond graph and for all other H-bond graphs we present here, we used the standard distance criterion of 3.5 Å between the donor and acceptor hetero-atoms. The protein was aligned along the membrane normal (Figure 1a) and a molecular graphics prepared using Chimera (Pettersen et al., 2004). To illustrate how water molecules participate to the internal H-bond network of the receptor, we show them explicitly as graph nodes and colour them differently from nodes representing protein groups. Typically, we colour nodes representing water molecules red, and nodes representing protein sidechains, grey. In the H-bond graph shown as a 2D plot (Figure 1b), the vertical axis gives the projections along the membrane normal (*z* axis) of the coordinates of the Cα atoms of the amino acid residues (or of the water oxygen atoms); the horizontal axis represents the PCA projections of the coordinates of these atoms in the membrane plane (*x*-*y* plane). This graph representation shows that the conserved NPxxY-Y7.53, located at about −15 Å along the membrane normal is part of a large H-bond cluster that includes D2.50 and CWxP-W6.48, and which extends ∼25 Å towards the extracellular side (Figure 1b). About 10 Å to the cytoplasmic side of Y7.53, DRY-D3.49 and R3.50 are in a small local H-bond cluster that lacks connections to the Y7.53 cluster (Figure 1b), that is, there are no H-bonds (edges) between the core NPxxY-Y7.53 and DRY-D3.49 H-bond clusters (Figure 1b).

To facilitate analyses of protein–water H-bond networks in datasets of static structures of GPCRs, we developed automated comparisons of H-bond graphs computed from datasets of static protein

structures. We initially relied on the matrices of common nodes and common edges, these are square matrices whose elements are nodes and edges common to structures P_i and P_j of the dataset (Bertalan et al., 2020). The diagonal matrix elements of these matrices gave the number of nodes and respectively, of edges in the graphs of the individual structures P_i, whereas the off-diagonal elements, gave the nodes and edges conserved between pairs of structures (Bertalan et al., 2020). For computational efficiency, our subsequent implementation of conserved H-bond graph computations within C-Graphs stores the conserved nodes and edges as lists and no longer uses matrices (Bertalan et al., 2021). That C-Graphs is computational efficient is illustrated by the finding that, for a dataset of 16 static structures of the A_{2A} receptor, each with ~58 nodes and ~61 edges in the H-bond graph, it took 16 s to complete a C-Graphs calculation (Bertalan et al., 2021).

Depending on the specific aspects of interest, H-bond networks computed from a dataset of static GPCR structures may be dissected with the help of (i) conserved, (ii) difference and (iii) comparison graphs (Bertalan et al., 2021). These types of graphs are discussed below and illustrated in Figure 1c–k for three small H-bond clusters. The H-bond cluster shown in Figure 1d is used as a reference; its nodes and edges correspond to the local H-bond cluster of DRY-D3.49 identified in Figure 1b and shown in Figure 1c as a molecular representation; this local H-bond cluster consists of six protein nodes labelled n_1 to n_6, two water molecules labelled w_1 and w_2, and eight edges (H-bonds). We further make use of two hypothetic H-bond clusters, shown in Figure 1f,k, to illustrate types of H-bond graph analyses for datasets of structures. To facilitate inspection of the H-bond graph computations detailed below, the protein nodes and the edges are coloured differently in each of these three structures. For clarity, we only use red to indicate water nodes.

i. A *conserved H-bond graph* (Figure 1h) computed for a set of static structures consists of the H-bonds and H-bonding groups that are present in all structures according to a preset conservation threshold. The conserved H-bond graph illustrated in Figure 1h consists of the nodes and edges present in all three H-bond clusters used here for the illustration (Figure 1d,f,k); that is, it corresponds to a 100% conservation threshold. This H-bond graph consists of protein nodes n_1–n_4 and n_6, water w_1, and four edges, with n_6 shown without an incident edge because its H-bond partners are different in each of the three clusters subjected to the computation.
ii. The *difference graph* of a given structure of the dataset shows the H-bonds and H-bonding groups present in that structure, but not in the other structures of the dataset. In Figure 1i, the difference H-bond graph of the reference structure (shown in Figure 1d), indicates with grey the nodes and edges present in the conserved graph, and with orange, the protein nodes and the edges present in the reference structure, but absent in one of the other two structures of the dataset (as shown in Figure 1d,k). This graph shows that the reference structure is characterized by four unique H-bonds, of which two are mediated by water w_2.
iii. The *comparison graph* between two structures colour-codes nodes and edges that are present in both structures, versus in either of the structures (Figure 1j). The comparison graph computed for the structures shown in Figure 1f,k colours grey protein nodes n_1–n_4 and n_6, as these are present in both structures, and with the corresponding colour code those nodes and edges that are present only in either of the structure (compare Figure 1j with Figure 1f,k). Thus, for two static protein structures, the conserved and comparison graph provide the complete information to compare the H-bond networks; for datasets of at least three structures, the difference graph is needed to evaluate H-bonds and H-bonding groups that distinguish one particular structure from the other structures of the dataset.

The H-bond clusters used to illustrate the graph computations above are small enough such that nodes representing water molecules that are part of the clusters can be shown explicitly. This, however, is often not the case when analysing protein–water H-bond networks of an entire GPCR, as structures solved at high resolution may contain numerous internal water molecules and protein–water H-bonds (Figure 2a–d). To simplify the display of the conserved, difference and comparison H-bond graphs computed for GPCR structures, we employ the water wire representation illustrated in Figure 1g, whereby we only show protein nodes, edges (H-bond connections) between two protein nodes can be either direct H-bonds or water-mediated bridges between the nodes, with up to three water molecules in each bridge (i.e. water molecules are not shown explicitly as graph nodes). The number—or, for a set of structures, the average number—of water molecules in each bridge can be indicated for each edge, for example, in Figure 1g, the '1's along the edges between n_1–n_4 and n_4–n_5 indicate that those connections are water-mediated bridges, each with one water molecule per bridge.

Direct comparisons of the H-bond graphs of GPCRs with different amino acid residue sequences are possible because nodes (amino acid residues) of the H-bond graphs can be assigned a unique numbering scheme: the Ballesteros-Weinstein (BW) scheme (Ballesteros & Weinstein, 1995). Pre-aligned GPCR sequences with amino acid residues renumbered according to the generic numbering based on the BW scheme are downloaded from GPCRdb (Kooistra et al., 2021). GPCR structures for which the sequence conservation value meets the criterion set by the user are selected for the analysis. As Bridge/Bridge2 and C-Graphs read coordinate files in standard Protein Data Bank (PDB) format, in order to compute the conserved H-bond graphs for GPCRs with distinct amino acid residue sequences, the C-Graphs Graphical User Interface renumbers the residues according to their GPCRdb generic BW numbers (Bertalan et al., 2021).

When a conserved H-bond graph is computed for, say, two GPCRs with different amino acid residue sequences, Seq-1 and Seq-2, the location of the graph nodes corresponds to the reference structure, let us say, Seq-1. For the difference or comparison graph of Seq-2 relative to the conserved graph of Seq-1 and Seq-2, common nodes are taken from the reference structure Seq-1; unique nodes present only in Seq-2 are taken from this structure.

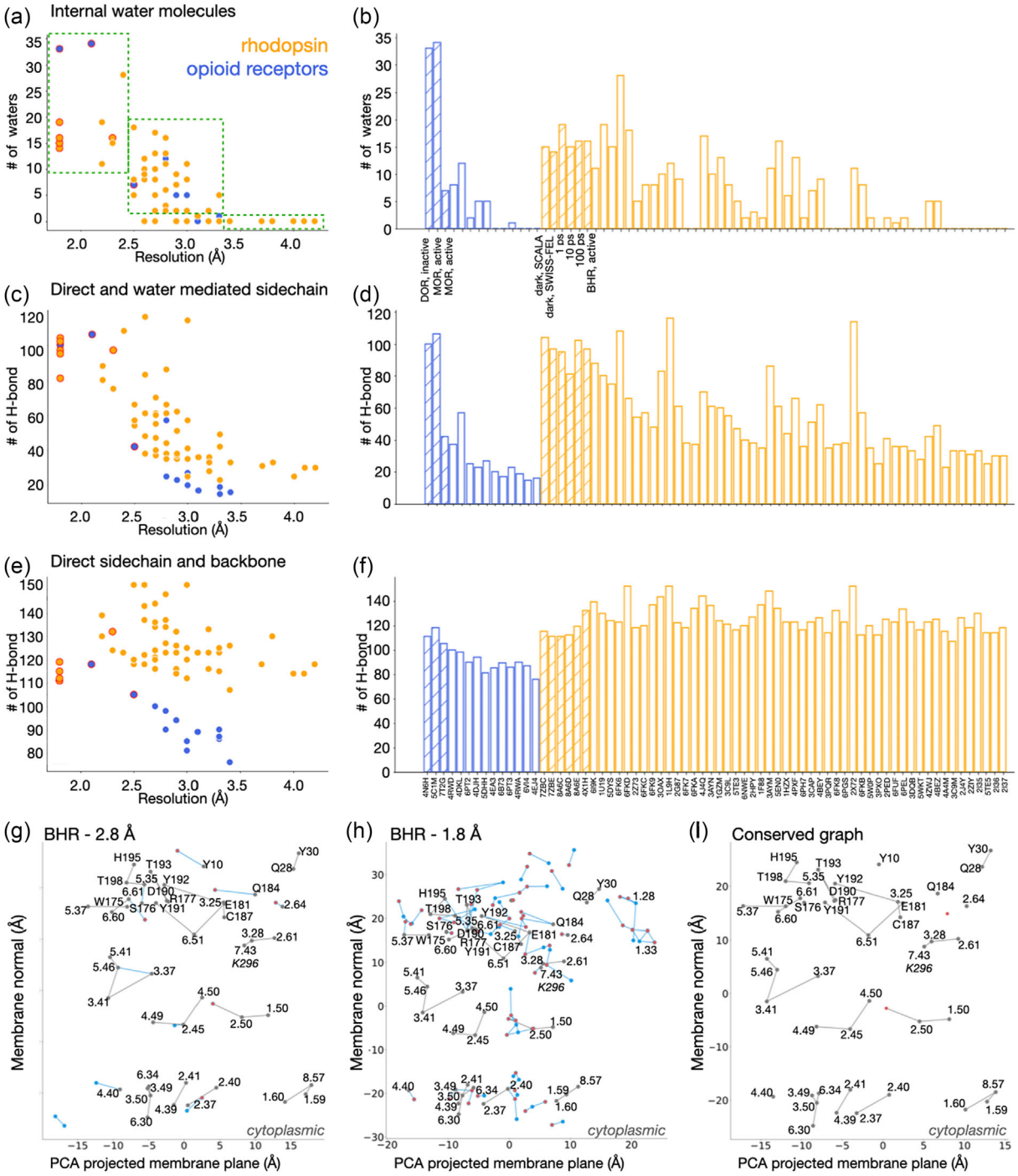

FIGURE 2 The H-bond network of a GPCR depends on the resolution at which its structure was solved. The data points and bars from panels a–f are based on 57 static structures of visual rhodopsins and 14 static structures of opioid receptors. (a,b) The number of internal water molecules as a function of the resolution at which the structure was solved (panel a) and in each structure of the dataset (panel b). In panel b, OR and rhodopsin structures explicitly discussed in the text are ordered according to their being inactive versus active and the bars are striped; the remaining structures are ordered according to the resolution. Note that 17 rhodopsin structures lack internal water molecules. (c,d) The total number of direct and water-mediated sidechain H-bonds as a function of the resolution (panel c) and for each structure of the dataset (panel d). (e,f) The total number of direct H-bonds between sidechains and between sidechains and backbone, shown as a function of the resolution (panel e) and for each structure of the dataset (panel f). (g–i) H-bond graphs of rhodopsin (BHR) structures solved at 2.8-Å resolution (pdb 1f88, panel g) compared to 1.8-Å resolution (pdb 7zbc, panel h) and their conserved graph (panel i). The graphs were computed for sidechain–sidechain and sidechain–water H-bonds; the 1.8-Å resolution structure was used as a reference for the conserved graph computation.

Below, we illustrate the usefulness of H-bond graph computations to dissect GPCR H-bond networks for a dataset of 12 static structures of GPCRs: six structures of rhodopsin, of which five are high-resolution structures of the dark (inactive) state and of intermediates 1, 10 and 100 ps after illumination (Gruhl et al., 2023), and one structure of active rhodopsin (Blankenship et al., 2015); one structure of dark-state squid rhodopsin, SQR (Murakami & Kouyama, 2008), one of inactive jumping spider rhodopsin-1 (JSR-1) bound to **9-cis retinal** (Varma et al., 2019); three receptor (opiate receptor) structures—the naltrindole-bound inactive δ receptor (Fenalti et al., 2014), the active-like **μ receptor** bound to the morphinan agonist **BU72** and an antibody (Huang et al., 2015), and active μ receptor bound to the agonist mitragyne pseudoindoxyl and to **Gi1** (Qu et al., 2022). Finally, our dataset also includes one structure of the inactive A_{2A} receptor. We note that the nomenclature we use for the opioid receptors, the A_{2A} receptor and the naltrindole ligand conforms to *BJP*'s Concise Guide to Pharmacology (Alexander et al., 2021). For the three visual rhodopsin pigments we include in this work, we use our own abbreviations to indicate the organism.

Except for the structure of Gi1-bound active μ-opioid receptor, which was solved using cryo-Electron Microscopy (Qu et al., 2022), our dataset includes only static GPCR structures solved with X-ray crystallography. With this dataset, we thus study visual receptors mostly in inactive conformations—as their structures were solved at high resolution—but also compare H-bond networks of inactive versus active GPCRs in two distinct class A GPCRs, visual receptors and opioid receptors.

4 | H-BOND GRAPHS DEPEND ON THE RESOLUTION AT WHICH THE STRUCTURES WERE SOLVED

As discussed above for rhodopsin structures, the resolution at which the structure was solved impacts the overall picture of the internal H-bonding of the receptor, particularly with respect to the internal water molecules. A systematic assessment of the relationship between resolution and the overall picture of H-bonding is presented in Figure 2a–f for visual rhodopsins and for opioid receptors. There are currently 57 structures of visual rhodopsins –rhodopsin, SQR and JSR-1—solved at resolutions from 4.2 to 1.8 Å. We counted, for each structure of this dataset, the number of internal water molecules, which we defined as the water molecules located within the membrane planes indicated by OPM. Thus, in the case of rhodopsin, we count as internal water molecules all waters whose oxygen atom is located from approximately −20 to +20 Å along the membrane normal in a projection as illustrated in Figure 1a. Of the 57 visual rhodopsin structures, 17 lack internal water molecules (Figure 2b).

The plot of the number of internal water molecules as a function of the resolution (Figure 2a) shows that rhodopsin structures solved at a resolution better than 2.5 Å have 10–25 internal waters; most of the structures solved at about 2.5- to 3.5-Å resolution have five to 19 internal waters, though some have only one to four waters; structures solved at 3.5–4.0 Å have one to two internal water molecules (Figure 2a). That is, the general trend is a decrease in the number of internal water molecules as the resolution worsens. By contrast, the resolution impacts much less the direct H-bonds between protein sidechains, and between sidechain and backbone (Figure 2e). Taken together, the separate analyses for the internal water molecules (Figure 2a) versus direct protein H-bonds (Figure 2e) indicate that the decrease, with decreasing resolution, in the number of protein–water H-bonds (Figure 2c), is largely due to the missing water molecules in structures solved at low resolution.

To illustrate the effect of the missing water molecules on the overall picture of the protein–water H-bond network, in Figure 2g,h, we compare the H-bond graphs computed for two dark-state rhodopsin structures solved at different resolution, 1.8 Å (Gruhl et al., 2023) versus 2.8 Å (Palczewski et al., 2000). The high-resolution structure reports coordinates for 91 water molecules, of which 15 are internal; by contrast, the lower resolution structure has in total 16 water molecules, of which three are internal. We wondered how the difference of 12 internal water molecules could reflect in the overall picture of the H-bond network of rhodopsin.

Figure 2g–i illustrates the difference and conserved H-bond graphs computed for these two rhodopsin structures by considering the direct sidechain–sidechain H-bonds and internal water-mediated bridges between sidechains. The 1.8-Å resolution structure has 104 H-bonds in the sidechain–water H-bond graph and 304 H-bonds in the sidechain–backbone–water graph. The total numbers of H-bonds in the 2.8-Å structure are less than half: 38 H-bonds in the sidechain–water graph and 148 H-bonds in the sidechain–backbone–water graph. The conserved H-bond graph shows that most of the local H-bond clusters of the 2.8 Å structure are also present in the 1.8-Å structure (Figure 2i): only 10 H-bonds from the graph at lower resolution structure are absent from the higher resolution structure, five of which are water-sidechain H-bonds and five sidechain–sidechain H-bonds (Figure 2g). Most of the H-bonds present only at lower resolution are at the cytoplasmic or extracellular sides of the receptor, where there may be artefactual conformational differences induced by differences in crystal packing.

The difference H-bond graph of the high-resolution structure (Figure 2h) shows water-mediated H-bonding at the core of the receptor at the conserved residue 2.50. At the extracellular side, there is a local cluster of four protein sidechains and six water molecules (see the cluster of $E33^{1.28}$ and $S38^{1.33}$); each water molecule of this cluster has 1–3 H-bonds. Some water molecules with 3 H-bonds are also observed in a local H-bond cluster of the δ receptor structure used for illustration in Figure 1b (see the cluster at residue $K214^{5.39}$). Although water H-bonding may be overestimated when computing H-bonds only based on distance without an H-bond angle criterion, we note that neutron scattering data indicate an average of 3.6 H-bonds per ambient water molecule (Soper et al., 1997), several water molecules each within 2.7–3.1 Å of three potential H-bond partners were found in the electron density map of aquaporin Aqy1 solved at 0.88-Å resolution (Eriksson et al., 2013) and discrete water molecules with three

or more H-bonds were identified at ligand-binding sites of high-resolution protein structures (Poornima & Dean, 1995).

The discussion above underlines the major contribution of discrete internal water molecules to the protein–water H-bond networks of GPCRs. Because the H-bond network of a GPCR is likely underestimated by a lower resolution structure that lacks internal water molecules, one possible approach would be to add water molecules at conserved sites identified in structures solved at high resolution (Bertalan et al., 2021). Dedicated algorithms to place water molecules in protein structures already exist—most notably, the web server HomolWat, which places in GPCR structures homologous water molecules found in higher-resolution structures (Mayol et al., 2020) and Dowser++ (Morozenko & Stuchebrukhov, 2016). Though an assessment of the accuracy of water placement in, for example, rhodopsin, is beyond the scope of this work, we note that in our preliminary test HomolWat placed about 80 water molecules in the chain A of the 2.8-Å rhodopsin structure.

A caveat of our analysis above for the visual rhodopsin structures is that the dataset includes, for rhodopsin and SQR, structures of both the dark states and of later photocycle intermediates. Thus, part of the variation in the number of internal water molecules and/or protein H-bonds might arise from differences in the protein structure and the associated change in the number and location of internal water molecules. Moreover, the observation that the number of internal water molecules tends to decrease with the resolution, whereas the number of protein H-bonds has a significantly less pronounced resolution dependence, may not be easily extrapolated from visual rhodopsins to other GPCRs. As we illustrate in Figure 2a,c,e, structures of opioid receptors tend to show, with decreasing resolution, a similar decrease in the number of water molecules and protein H-bonds. The structures we used below for detailed discussion were solved at a resolution of 2.5 Å or better; we stress however, that this criterion is not unique. A more detailed analysis of the number of internal water molecules in a combined dataset of 53 GPCR structures solved at resolutions of 2.5 Å or better, and with at least 10 internal water molecules was presented by us recently (Bertalan et al., 2020).

5 | THE TIME-RESOLVED SNAPSHOTS OF WATER-MEDIATED H-BOND NETWORKS IN RHODOPSIN

Very recently, ultrafast time-resolved crystallography at room temperature identified the structural events changes that occur along early steps of the activation path of rhodopsin—a data set of three rhodopsin structures at distinct time points after illumination, 1, 10 and 100 ps, and the two respective dark-state structures solved at SACLA and SwissFEL (Gruhl et al., 2023). The high resolution at which these structures were solved allowed for numerous water molecules to be observed: there are 87–91 waters in the two dark-state structures; the three structures solved at 1, 10 and 100 ps after illumination report coordinates for 88, 55 and 75 water molecules. We present in Figure 3 H-bond graphs computed for these structures.

The H-bond graph of the SACLA dark-state structure (Figure 3a) has no fewer than 300 H-bonds. Most of these H-bonds involve water, with the 32 sidechain–sidechain and 83 sidechain–backbone H-bonds contributing ~11% and 28% of the total number of H-bonds; the remaining H-bonds of the graph involve water.

The comparison H-bond graph indicates that just a handful of H-bonds distinguish the SACLA from the SwissFEL structure (Figure 3c). At the cytoplasmic side, residue 1.59 is within H-bond distance from 1.60 and 8.57 only in the SACLA structure, although the electron density at this position is relatively poorly defined, suggesting that the sidechain of residue 1.59 is flexible and may not be unambiguously modelled in a single conformation. In extracellular loop 2, R177 and H195 are within H-bond distance in the SACLA H-bond graph, but not in the graph computed for the SwissFEL structure, as the latter the contact distance between the residues increases from 3.42 to 3.51 Å. That the R177–H195 H-bond is absent from the SwissFEL structure is thus caused by its falling just outside the 3.50-Å cut-off for the H-bond criterion used to compute the H-bond graphs. Another difference between the H-bond graphs of the SACLA and SwissFEL dark-state rhodopsin structures is that only the latter has an H-bond between T4 and N15 (Figure 3c).

These minor differences between the two dark-state structures are due to subtle variations in sidechain positions in flexible regions of the protein and small changes in inter-residue distances close to the H-bond distance threshold. They are irrelevant for the internal protein–water H-bond network and unlikely to be biologically significant. We conclude that, within the limitations of the H-bond analyses based on a distance criterion, the H-bond networks of the two dark state rhodopsin structures are essentially the same (Figure 3c).

Relative to the SwissFEL dark-state structure, both the 1- and 10-ps structures have one additional H-bond at the extracellular side. This H-bond (R177–H195) is present in the SACLA dark state model and is therefore unlikely to be light induced. At the cytoplasmic side, compared to the SwissFEL dark state, the 1-ps structure has three additional H-bonds (residue pairs 1.60–1.59, 8.57–1.59 and 8.49–K67); relative to the dark-state the 10-ps structure has one additional H-bond at the cytoplasmic side, between residues 6.32 and 6.36 (Figure 3d,e).

The structure at 100 ps after illumination has about the same total number of H-bonds as the SACLA dark state (Figure 3a,b,f), but it has one less H-bond connection at the core of the receptor, seven at the extracellular side and one at the cytoplasmic side (Figure 3f). Compared to the SACLA dark-state, the 100-ps structure has three more H-bond connections at the extracellular side and one H-bond at the cytoplasmic side (Figure 3f).

Taken together, four H-bond connections are present in the 100 ps structure but absent in the corresponding dark state (K16–S14, S14–T4, T193–S176, N310$^{8.47}$–E249$^{6.32}$). A few H-bond connections are present in both dark-state structures but absent from the three snapshots after illumination (N8–R177, N8–S22, R177–C110$^{3.25}$, R177–N111$^{3.26}$, N111$^{3.26}$–N8 and C110$^{3.25}$–N8). The H-bond connection between 6.48 and 7.49 is missing from the H-bond graph of the structure at 100 ps after illumination.

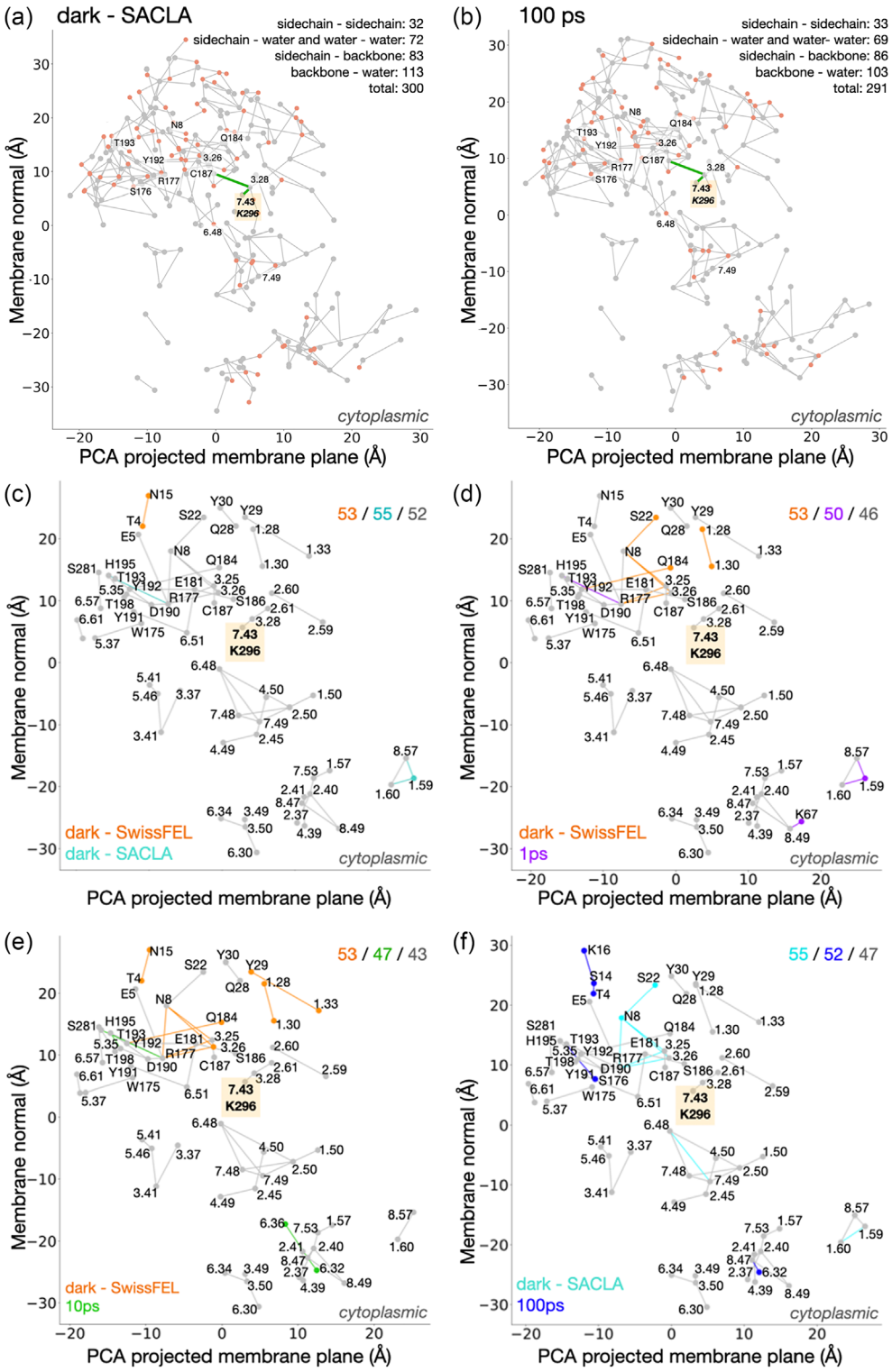

FIGURE 3 H-bond graphs of early structural changes in rhodopsin (BHR). For clarity, nodes representing H-bonding water molecules are shown explicitly as red dots only in panels a and b. The retinal molecule is covalently bound to K296$^{7.43}$ via a protonated Schiff base. (a,b) H-bond graphs for the SACLA dark state (pdb 7zbc, panel a) and the structure 100 ps after illumination (pdb 8a6e, panel b), including sidechain–backbone H-bonds. There are 300 H-bonds in panel a and 291 H-bonds in panel b. The numbers in the upper right corners of the two plots show the numbers of various H-bonds included in the computation. (c) Comparison H-bond graph of the two dark-state structures solved at SACLA versus SwissFEL (pdb 7zbe). (d–f) Comparison H-bond graphs of the BHR structures 1 ps (pdb 8a6c), 10 ps (pdb 8a6d) and 100 ps after illumination and their corresponding dark state solved at the same facility. For panels c–f, numbers in the right upper corner indicate H-bond connections (directly between protein sidechains and water bridges with up to three H-bonded waters between sidechains), colour-coded for the structures; numbers in grey indicate how many of these H-bond connections are conserved between the two structures.

The H-bond graphs indicate that differences in H-bond connections (absent/present in an intermediate state relative to the dark state) can be detected relatively far away from the source of structural change, which is the photo-isomerized retinal. These H-bond differences could be due to uncertainties in the precise orientation of groups close to the extracellular bulk, as bulk-exposed groups tend to have larger B factors. Many of the detected H-bond differences are also attributable to disruption of water bridges between amino acid side chains; these water molecules are more plentiful further from the retinal-binding site and close the bulk solvent interfaces. We further note that details such as shortening or lengthening of selected H-bonds in the vicinity of the retinal Schiff base, as revealed by the inspection of the high-resolution electron density maps of the dark-state versus intermediates after illumination (Gruhl et al., 2023), are not captured by the graphs in Figure 3c–f. We currently explore approaches to map such H-bond distance changes

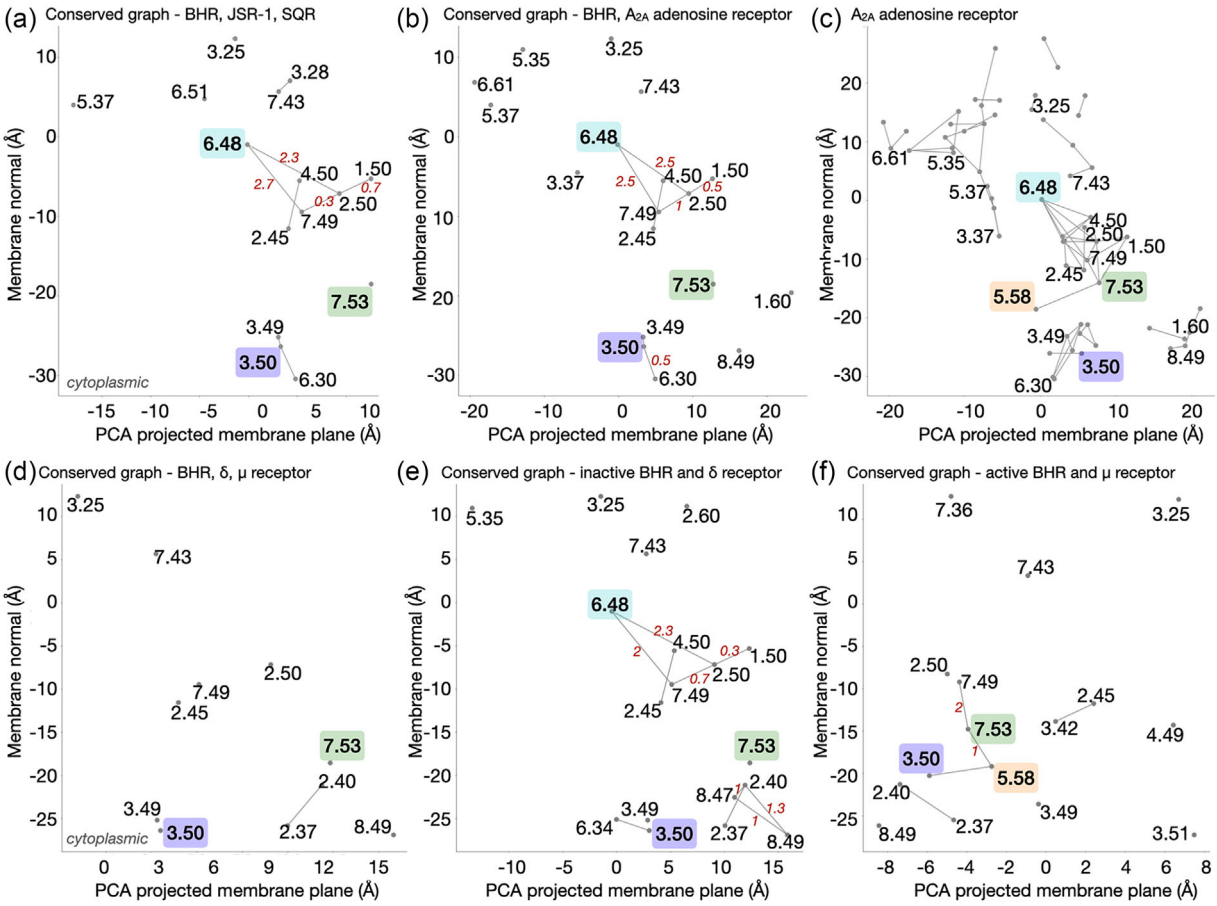

FIGURE 4 Illustration of conserved H-bond graphs computed for distinct class A GPCRs. Numbers in red italic fonts along the edges indicate the average number of water molecules in H-bond bridges between protein sidechains. The pdb entries for the dark-state rhodopsin (BHR) are listed in the legend of Figure 3. (a) Conserved H-bond graph computed for dark-state SACLA BHR, dark-state squid rhodopsin (SQR) (pdb 2z73) and iso-JSR1 (pdb 6i9k). (b) Conserved H-bond graph computed for dark-state SACLA BHR and the high-resolution structure of inactive A_{2A} receptor (pdb 5iu4). (c) The protein–water H-bond graph of the A_{2A} adenosine receptor; for clarity, we do not show water molecules explicitly as nodes, and label only residues that are part of the conserved H-bond network with BHR shown in panel e. (d) Conserved graph computed for the dark-state BHR structures solved at SACLA versus SwissFEL, active BHR (pdb 4x1h), inactive δ receptor (pdb 4n6h) and two active μ receptor structures, pdb 5c1m, solved at 2.1-Å resolution and pdb 7t2g, solved at 2.5-Å resolution. (e) Conserved graph for the dark state BHR structures (SACLA and SwissFEL) and the inactive δ receptor structure (pdb 4n6h, 1.8-Å resolution). (f) Conserved graph computed one active BHR (pdb 4x1h) and two active μ receptor structures. Conserved graph computations shown in panels a, b, d and e, used the SACLA dark-state BHR structure; for the graph shown in panel f, the reference structure was the active μ receptor structure 5c1m.

onto H-bond graphs while also avoiding over-crowding the H-bond graphs.

How each bond that breaks or forms as rhodopsin changes conformation contributes to the free energy landscape of the receptor is unclear. Observations on the backbone H-bond stabilities of outer membrane protein (OmpA) were interpreted to suggest that, for a large membrane protein whose hydration changes during its reaction cycle, conformational dynamics would be shaped by changes in the solvation of the sidechains, and less by changes in the strengths of H-bonding (Lessen et al., 2020). This might indeed be the case for rhodopsin, because osmotic pressure measurements indicated that formation of an active meta-rhodopsin intermediate couples with an influx of about 80–100 bulk water molecules that enter the interior of the receptor (Fried et al., 2022).

6 | H-BOND NETWORKS OF INACTIVE VERSUS ACTIVE GPCRS

The H-bond graphs discussed above for the five high-resolution structures rhodopsin indicate extensive H-bonding of the receptor (Figure 3). The retinal Schiff base (residue K296, 7.43), is part of a local H-bond cluster with residues 2.61 and 3.28 in the sidechain–water H-bond networks of all structures of the dataset (Figure 3); when the backbone groups are included in the H-bond graph computation, the Schiff base connects, via 3.28, to the extracellular H-bond cluster that includes the conserved C187 (see the thick green lines in Figure 3a,b), but it remains disconnected from the H-bond clusters at the cytoplasmic side of the Schiff base—for example, from the core H-bond cluster of 6.48.

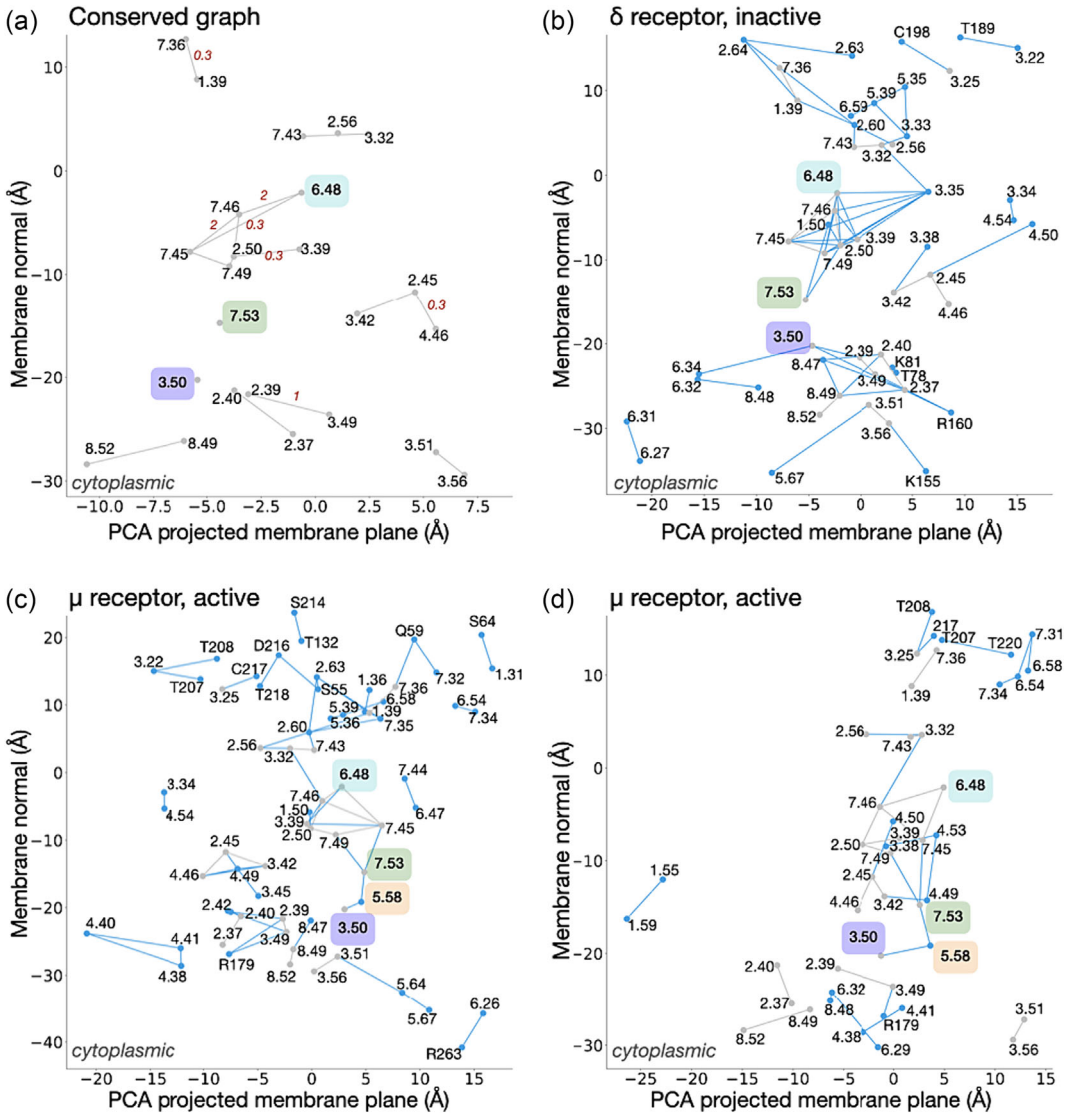

FIGURE 5 H-bond graphs of inactive versus active opioid receptors (ORs). (a) Conserved H-bond graph of the active μ and inactive δ receptor structures used for this illustration. Italic numbers in red along the graph edges (H-bond connections) indicate the average length of the water bridge between H-bonding sidechains. The inactive δ receptor structure was used as a reference to plot the conserved H-bond graph. (b,c) H-bond graphs computed for the inactive δ receptor structure (1.8 Å resolution, pdb 4n6h) versus active μ receptor structure (2.07 Å resolution, pdb 51cm). (d) H-bond graph of the active μ opioid receptor (2.50 Å resolution, pdb 7t2g). Structures of inactive μ (4dkl, 2.8-Å resolution) and κ receptors (4djh, 2.9-Å resolution) were excluded due to their resolution being insufficient to reliably compute water-mediated H-bond networks.

A core H-bond cluster of CWxP-6.48 is also present, for example, in dark-state SQR, inactive 9-cis JSR-1, and inactive A_{2A} receptor and δ receptor (Figures 4 and 5b). In all these structures, the core cluster of CWxP-6.48 includes 2.50 and a few other sidechains and water molecules, with an overall cluster length along the membrane normal of ∼15 Å (Figure 4).

The core CWxP-W6.48 H-bond cluster is important to consider, as all participating sidechains are known to be important for function and some sidechains re-orient during receptor activation: CWxP-W6.48 reorients during receptor activation (Deupi et al., 2012); two interhelical clusters of tightly packed hydrophobic and hydrophilic groups include residue 2.45, with one of these clusters also including residue 1.50 and the other cluster, residue 4.50 (Sanchez-Reyes et al., 2017); 2.50 (Asp in opioid receptors) is part of the sodium-binding pocket present in many of the class A GPCRs (Katritch et al., 2014); Y7.49 is part of the NPxxY motif.

Both structures of active μ receptor have the core 6.48 H-bond cluster and in both structures this cluster connects to DRY-3.50 and Y5.58 (Figure 5c,d). Direct H-bonding (3.50–5.58) and water-mediated bridges between residues Y7.53–Y5.58, and Y5.58–3.50, and 7.53–7.49 are illustrated by the H-bond graphs of the active μ receptor (Figure 5c,d) and active rhodopsin (Figure 4f).

H-bond rearrangements at 7.53 upon receptor activation are captured by the conserved graph shown in Figure 5a. This graph, which was computed for the inactive δ receptor and the two active μ receptor structures, indicate 7.53 as a node without any incident edge (H-

bond), that is, 7.53 is H-bonded in all three structures, but it has distinct H-bond partners in at least two of the structures.

The conserved graph computed for six structures of inactive (dark-state) and active rhodopsin, together with inactive versus active opioid receptors (Figure 4d), has 11 H-bonding groups. These are sidechains that H-bond in all six structures, but have distinct H-bond partners; an H-bond connection between residues 2.37 and 2.40 is conserved in all six structures. A functional role of the H-bond between 2.37 and 2.40 is not clear, but both of these residues were suggested to be involved in an exit pathway for the sodium ion in the activated receptor (Vickery et al., 2018).

7 | CONCLUSIONS

Water-mediated H-bond networks are thought essential for the allosteric conformational coupling mechanism that GPCRs use to translate extracellular agonist binding into structural change leading to activation and G-protein binding at the other side of the membrane. Graph-based computations that enable efficient computations of all H-bonds present in structures in entire datasets allow us to dissect H-bond fingerprints that characterize distinct GPCRs and the structures sampled along their activation. The information provided by structural biology does however, depend upon the resolution at which GPCR structures are solved. Within the limitations of our data set, the relationship between resolution, number of internal water molecules and the overall picture of the internal H-bond network, appear to depend somewhat on the GPCR (Figure 2), which might reflect different protocols used to solve structures experimentally.

A limitation of the H-bond graph analyses originates in the geometric criterion typically used to identify H-bonds. Because H-bonds are represented provided the distance H-bond criterion is met, without details of H-bond length changes, as long as the donor and acceptor hetero-atoms remain within H-bond distance, any strengthening or weakening of the H-bond will not be captured by the H-bond graph computation. For example, the conserved graph for rhodopsin contains the H-bond between the retinal Schiff base and E113, as indeed these two groups remain H-bonded (Figure 3), but the graph lacks direct information about changes of this H-bonding. The distance between the retinal Schiff base and the two carboxylic oxygen atoms of E113 change by 0.2–0.3 Å in the 1-ps structure as compared to the dark state rhodopsin (Gruhl et al., 2023). Further developments of the graph computations to capture such subtle rearrangements of the H-bond networks are currently pursued by our laboratories.

The high-resolution structures of rhodopsin discussed here indicate that water plays an essential role in mediating H-bond networks of GPCRs. Taken together, more than half of the H-bonds of the H-bond graphs shown in Figure 3a,b involve water molecules. Because the great majority of existing GPCR structures are solved at resolutions insufficient for accurate information on their internal water molecules, the important question that arises is whether GPCR structures that have sufficiently high resolution and internal water content may provide clues about H-bond networks of GPCRs in general. To evaluate water-mediated H-bond networks of such structures, one possible approach would be to use specialized tools to add water molecules. Water molecules and water-mediated H-bond networks also tend to be rather dynamic, and atomic-level molecular dynamics simulations have identified, for different GPCRs, water-mediated networks that help interconnect functionally important protein sites (see, e.g, Bertalan et al., 2020; Kapur et al., 2023; Lesnik et al., 2023; Venkatakrishnan et al., 2019). Molecular dynamics simulations may also be performed to evaluate which H-bond networks are stable and which are susceptible to change upon ligand binding or mutation. Such changes could then be identified with C-Graphs, which automatizes the comparison of H-bond networks sampled by GPCRs in independent atomic-level simulations; in the case of JSR-1, for example, C-Graphs identified rearrangements in the same region of the H-bond network for two distinct mutations near the retinal Schiff base, which could be interpreted to suggest the presence of a mechanism for conformational coupling (Bertalan et al., 2021).

Conserved H-bond graph computations indicate that structures of inactive visual rhodopsins, as well as of inactive A_{2A} and δ receptors, share a conserved core cluster that includes residues CWxP-6.48 and D2.50 (Figure 4). Whereas in the inactive structures the core CWxP-6.48 and D2.50 cluster lacks H-bond connections to 3.50, water-mediated H-bond connections between 7.53 and 3.50 via 5.58 are found in the graphs of active rhodopsin and μ receptor (Figure 4f). The finding that there are more H-bonds and H-bonding groups in the conserved graph of inactive, than active rhodopsin versus μ receptor (Figure 4b,c) could be interpreted to suggest sequence-specific H-bond changes upon receptor activation. However, to conclude on the common features of the H-bond networks of active GPCRs, a larger dataset of high-resolution structures with reliable water coordinates would be needed.

The H-bond graph computations presented in this work illustrate the applicability of the C-Graphs H-bond graph computations to directly evaluate the distribution of the entire protein–water H-bond network in GPCRs and to evaluate, within the limitations of the dataset of structures and of the H-bond criteria used, structural rearrangements in active versus inactive GPCRs, including of GPCRs with different amino acid residue sequences. Future developments we plan include developing intuitive representations, within the H-bond graphs, of details of H-bond changes and long-distance correlations between such changes upon GPCR activation.

Our discussion here was focused on the applicability of C-Graphs to GPCRs, for which we have developed the algorithm and its graphical user interface (Bertalan et al., 2021). Conserved H-bond graph computations are however, of general interest to decipher H-bond networks in protein families other than GPCRs. To directly compare the H-bond graphs of proteins that belong to the same family, but have distinct amino acid residue sequences, a unique numbering scheme, as used here for GPCRs, would need to be implemented, the residues in the protein structures renumbered according to this unique numbering scheme and then the conserved graphs may be computed. The relevance of such computations for the H-bond graphs of a protein family (other than GPCRs) would likely depend on the

overall conservation of both the sequence and structural features among the proteins of the dataset and on the availability of high-resolution static structures for sufficiently many proteins of the family of interest.

A further development we envision is to evaluate the relative contributions of H-bonds to the dynamics of the protein H-bond network. Whereas our current graph computations include all H-bonds that meet the preset H-bond criterion, datasets of high-resolution structures proteins captured along different steps of its reaction cycle bring about the need to evaluate changes in the relative strengths of H-bonds. The relative strength of the H-bonds in structures of two distinct protein intermediates could be evaluated according to the H-bond length changes, which could be computed and mapped onto the conserved H-bond graphs. Alternatively, one could describe intra-protein interactions according to a force-field equation and map the non-bonded interactions between H-bond partners onto the H-bond graph.

7.1 | Nomenclature of targets and ligands

Key protein targets and ligands in this article are hyperlinked to corresponding entries in the IUPHAR/BPS Guide to PHARMACOLOGY http://www.guidetopharmacology.org and are permanently archived in the Concise Guide to PHARMACOLOGY 2023/24 (Alexander et al., 2023).

AUTHOR CONTRIBUTIONS

É. Bertalan: Conceptualization (equal); formal analysis (lead); investigation (equal); methodology (equal); software (lead); validation (equal); visualization (equal); writing—original draft (equal); writing—review and editing (equal). **M. J. Rodrigues:** Investigation (equal); validation (equal); writing—review and editing (equal). **G. F. X. Schertler:** Conceptualization (equal); funding acquisition (equal); supervision (equal); writing—review and editing (equal). **A.-N. Bondar:** Conceptualization (equal); investigation (equal); supervision (lead); validation (equal); writing—original draft (lead); writing—review and editing (lead).

ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 951644 to GFXS and Marie Skłodowska-Curie grant agreement No. 701647 to MJR). Open-access publication funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—491111487. We thank the Central Library of the Forschungszentrum Jülich for making the open access publication possible. The Authors are grateful to Dr. Valerie Panneels for valuable suggestions on this work. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

Gebhard F.X. Schertler is co-founder of the biotechnology companies leadXpro and InterAx.

DATA AVAILABILITY STATEMENT

N/A-Review.

REFERENCES

Alexander, S. P. H., Christopoulos, A., Davenport, A. P., Kelly, E., Mathie, A. A., Peters, J. A., Veale, E. L., Armstrong, J. F., Faccenda, E., Harding, S. D., Davies, J. A., Abbracchio, M. P., Abraham, G., Agoulnik, A., Alexander, W., Aal-hosaini, K., Bäck, M., Baker, J. G., Barnes, N. M., … Ye, R. D. (2023). The Concise Guide to PHARMACOLOGY 2023/24: G protein-coupled receptors. *British Journal of Pharmacology*, *180*, S23–S144. https://doi.org/10.1111/bph.16177

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. https://doi.org/10.1016/S1043-9471(05)80049-7

Bertalan, E., & Bondar, A.-N. (2023). Graphs of protein-water hydrogen-bond networks to dissect structural movies of ion-transfer microbial rhodopsins. *Frontiers in Chemistry*, *10*, 1075648. https://doi.org/10.3389/fchem.2022.1075648

Bertalan, E., Lesca, E., Schertler, G. F. X., & Bondar, A.-N. (2021). C-graphs tool with graphical user interface to dissect conserved hydrogen-bond networks: Applications to visual rhodopsins. *Journal of Chemical Information and Modeling*, *61*, 5692–5707. https://doi.org/10.1021/acs.jcim.1c00827

Bertalan, É., Lešnik, S., Bren, U., & Bondar, A.-N. (2020). Protein-water hydrogen-bond networks of G protein-coupled receptors: Graph-based analyses of static structures and molecular dynamics. *Journal of Structural Biology*, *212*, 107634. https://doi.org/10.1016/j.jsb.2020.107634

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. https://doi.org/10.1016/j.str.2015.09.015

Choe, H.-W., Kim, Y. J., Park, J. H., Morizumi, T., Pai, E. F., Krauß, N., Hofmann, K. P., Scheerer, P., & Ernst, O. P. (2011). Crystal structure of metarhodopsin II. *Nature*, *471*, 651–655. https://doi.org/10.1038/nature09789

Choi, W.-T., Duggineni, S., Xu, Y., Huang, Z., & An, J. (2012). Drug discovery research targeting the CXC chemokine receptor 4 (CXCR4). *Journal of Medicinal Chemistry*, *55*, 974–995.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Sten, C. (2009). *Introduction to algorithms* (third ed.). Massachusetts Institute of Technology.

Deupi, X. (2014). Relevance of rhodopsin studies for GPCR activation. *BBA - Bioenergetics*, *1837*, 674–682. https://doi.org/10.1016/j.bbabio.2013.09.002

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. *Proceedings of the National Academy of Sciences*, *109*, 119–124. https://doi.org/10.1073/pnas.1114089108

Eilers, M., Reeves, P. J., Ying, W., Khorana, H. G., & Smith, S. O. (1999). Magic angle spinning NMR of the protonated retinylidene Schiff base nitrogen in rhodopsin: Expression of ^{15}N-lysine- and ^{13}C-glycine-labeled opsin in a stable cell line. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 487–492. https://doi.org/10.1073/pnas.96.2.487

Eriksson, U. K., Fischer, G., Friedmann, R., Enkavi, G., Tajkhorshid, E., & Neutze, R. (2013). Subangstrom resolution X-ray structure details aquaporin-water interactions. *Science*, *340*, 1346–1349. https://doi.org/10.1126/science.1234306

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. https://doi.org/10.1038/nature12944

Fredriksson, R., Lagerström, M. C., Lundin, L.-G., & Schöth, H. B. (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogin groups, and fingerprints. *Molecular Pharmaceutics*, *63*, 1256–1272.

Fried, S. D. E., Hewage, K. S. K., Eitel, A. R., Struts, A. V., Weerasinghe, N., Perera, S. M. D. C., & Brown, M. F. (2022). Hydration-mediated G-protein-coupled receptor activation. *PNAS*, *119*, e2117349119. https://doi.org/10.1073/pnas.2117349119

Fritze, O., Filipek, S., Kuksa, V., Palczewski, K., Hoffmann, K. P., & Ernst, O. P. (2003). Role of the conserved NPxxY(x)$_{5,6}$F motif in the rhodopsin ground state and during activation. *PNAS*, *100*, 2290–2295. https://doi.org/10.1073/pnas.0435715100

Goncalves, J. A., South, K., Ahuja, S., Zaitseva, E., Opefi, C. A., Eillers, M., Vogel, R., Reeves, P. J., & Smith, S. O. (2010). Highly conserved tyrosine stabilizes the active state of rhodopsin. *PNAS*, *107*, 19861–19866. https://doi.org/10.1073/pnas.1009405107

Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J., Melo, M. N., Seyler, S. L., Domanski, J., Dotson, D. L., Buchoux, S., Kenney, I. M., & Beckstein, O. (2016). MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. In S. Benthall & S. Rostrup (Eds.), *Proceedings of the 15th Python in Science Conference* (pp. 102–109). SciPy.

Gruhl, T., Weinert, T., Rodrigues, M. J., Milne, C. J., Ortolani, G., Nass, K., Nango, E., Sen, S., Johnson, P. J. M., Cirelli, C., Furrer, A., Mous, S., Skopintsev, P., James, D., Dworkowski, F., Båth, P., Kekilli, D., Ozerov, D., Tanaka, R., … Panneels, V. (2023). Ultrafast structural changes direct the first molecular events of vision. *Nature*, *615*, 939–944. https://doi.org/10.1038/s41586-023-05863-6

Hauser, A., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B., & Gloriam, D. E. (2017). Trends in GPCR drug discovery: New agents, targets and indications. *Nature Reviews Drug Discovery*, *16*, 829–842. https://doi.org/10.1038/nrd.2017.178

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., Traynor, 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. https://doi.org/10.1038/nature14886

Kapur, B., Baldessari, F., Lazaratos, M., Nar, H., Schnapp, G., Giorgetti, A., & Bondar, A.-N. (2023). Protons taken hostage: Dynamic H-bond networks of the pH-sensing GPR68. *Computational and Structural Biotechnology Journal*, *21*, 4370–4384. https://doi.org/10.1016/j.csbj.2023.08.034

Katritch, V., Fenalti, G., Abola, E. E., Roth, B. L., Cherezov, V., & Stevens, R. C. (2014). Allosteric sodium in class A GPCR signaling. *Trends in Biochemical Sciences*, *39*, 233–244. https://doi.org/10.1016/j.tibs.2014.03.002

Kooistra, A. J., Mordalski, S., Pándy-Szekeres, P., Esguerra, M., Mamyrbekov, A., Munk, C., Kererü, G. M., & Gloriam, D. E. (2021). GPCRdb in 2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Research*, *49*, D335–D343. https://doi.org/10.1093/nar/gkaa1080

Lesca, E., Panneels, V., & Schertler, G. F. X. (2018). The role of water molecules in phototransduction of retinal proteins and G protein-coupled receptors. *Faraday Discussions*, *207*, 27–37. https://doi.org/10.1039/C7FD00207F

Lesnik, S., Bren, U., Domratcheva, T., & Bondar, A.-N. (2023). Fentanyl and the fluorinated fentanyl derivative NFEPP elicit distinct hydrogen-bond dynamics of the opioid receptor. *Journal of Chemical Information and Modeling*, *63*, 4732–4748. https://doi.org/10.1021/acs.jcim.3c00197

Lessen, H. J., Majumdar, A., & Fleming, K. G. (2020). Backbone hydrogen bond energies in membrane proteins are insensitive to large changes in local water concentration. *Journal of the American Chemical Society*, *142*, 6227–6235. https://doi.org/10.1021/jacs.0c00290

Li, J., Edwards, P. C., Burghammer, M., Villa, C., & Schertler, G. F. X. (2004). Structure of bovine rhodopsin in a trigonal crystal form. *Journal of Molecular Biology*, *343*, 1409–1438. https://doi.org/10.1016/j.jmb.2004.08.090

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 Acids Research*, *40*, D370–D376.

Mayol, E., García-Recio, A., Tiemann, J. K. S., Hildebrand, P. W., Guixà-González, R., Olivella, M., & Cordomi, A. (2020). HomolWat: A web server tool to incorporate 'homologous' water molecules into GPCR structures. *Nucleic Acids Research*, *48*, W54–W59. https://doi.org/10.1093/nar/gkaa440

Michaud-Agrawal, N., Denning, E. J., & Woolf, T. B. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, *32*, 2319–2327. https://doi.org/10.1002/jcc.21787

Morizumi, T., Kim, D., Li, H., Govorunova, E. G., Sineshchekov, O. A., Wang, Y., Zheng, L., Bertalan, É., Bondar, A. N., Askari, A., Brown, L. S., Spudich, J. L., & Ernst, O. P. (2023). Structures of channelrhodopsin paralogs in peptidics explain their contrasting K^+ and Na^+ selectivities. *Nature Communications*, *14*, 4365. https://doi.org/10.1038/s41467-023-40041-2

Morozenko, A., & Stuchebrukhov, A. A. (2016). Dowser++, a new method of hydrating protein structures. *Proteins*, *84*, 1347–1357. https://doi.org/10.1002/prot.25081

Murakami, M., & Kouyama, T. (2008). Crystal structure of squid rhodopsin. *Nature*, *453*, 363–367. https://doi.org/10.1038/nature06925

Nagata, T., Terakita, A., Kandori, H., Kojima, D., Schichida, Y., & Maeda, A. (1997). Water and peptide backbone structure in the active center of bovine rhodopsin. *Biochemistry*, *36*, 6164–6170. https://doi.org/10.1021/bi962920t

Nakamichi, H., Buß, V., & Okada, T. (2007). Photoisomerization mechanism of rhodopsin and 9-cis rhodopsin revealed by X-ray crystallography. *Biophysical Journal*, *92*, 106–108. https://doi.org/10.1529/biophysj.107.108225

Nakamichi, H., & Okada, T. (2006). Local peptide movement in the photoreaction intermediate of rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 12729–12734. https://doi.org/10.1073/pnas.0601765103

Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E. M., & Shichida, Y. (2001). Functional role of internal water molecules in rhodopsin revelealed by X-ray crystallography. *PNAS*, *99*, 5982–5987. https://doi.org/10.1073/pnas.082666399

Okada, T., Sugihara, M., Bondar, A.-N., Elstner, M., Entel, P., & Buss, V. (2004). The retinal configuration and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *Journal of Molecular Biology*, *342*, 571–583. https://doi.org/10.1016/j.jmb.2004.07.044

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Trong, I. L., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*, *289*, 739–745. https://doi.org/10.1126/science.289.5480.739

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. *Journal of Computational Chemistry*, *25*, 1605–1612. https://doi.org/10.1002/jcc.20084

Poornima, C. S., & Dean, P. M. (1995). Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. *Journal of Computer-Aided Molecular Design*, *9*, 500–512. https://doi.org/10.1007/BF00124321

Qu, Q., Huang, W., Aydin, D., Paggi, J. M., Seven, A. B., Wang, H., Chakraborty, S., Che, T., DiBerto, J. F., Robertson, M. J., Inoue, A.,

Suomivuori, C. M., Roth, B. L., Majumdar, S., Dror, R. O., Kobilka, B. K., & Skiniotis, G. (2022). Insights into distinct signaling profiles of the μOR activated by diverse agonists. *Nature Chemical Biology*, *19*, 423–430. https://doi.org/10.1038/s41589-022-01208-y

Rasmussen, S. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T. A., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., & Kobilka, B. K. (2011). Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature*, *477*, 549–555. https://doi.org/10.1038/nature10361

Rosenbaum, D. M., Rasmussen, S. G. F., & Kobilka, B. K. (2009). The structure and function of G-protein-coupled receptors. *Nature*, *459*, 356–363. https://doi.org/10.1038/nature08144

Sanchez-Reyes, O. B., Cooke, A. L. G., Tranter, D. B., Rashid, D., Eilers, M., Reeves, P. J., & Smith, S. O. (2017). G protein-coupled receptors contain two conserved packing clusters. *Biophysical Journal*, *112*, 2315–2326. https://doi.org/10.1016/j.bpj.2017.04.051

Santos, R., Ursu, O., Bento, A. P., Donadi, R. S., Oprea, T. I., & Overington, J. P. (2017). A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery*, *16*, 19–34. https://doi.org/10.1038/nrd.2016.230

Siemers, M., & Bondar, A.-N. (2021). Interactive interface for graph-based analyses of dynamic H-bond networks: Application to spike protein S. *Journal of Chemical Information and Modeling*, *61*, 2998–3014. https://doi.org/10.1021/acs.jcim.1c00306

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. *Journal of Chemical Theory and Computation*, *15*, 6781–6798. https://doi.org/10.1021/acs.jctc.9b00697

Soper, A. K., Bruni, F., & Ricci, M. A. (1997). Site-site pair correlation functions of water from 25 to 400°C: Revised analysis of new and old diffraction data. *The Journal of Chemical Physics*, *106*, 247–254. https://doi.org/10.1063/1.473030

Standfuss, J., Edwards, P. C., D'Antona, A., Fransen, M., Xie, G., Oprian, D. D., & Schertler, G. F. X. (2011). The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature*, *471*, 656–660. https://doi.org/10.1038/nature09795

Teller, D. C., Okada, T., Behnke, C. A., Palczewski, K., & Stenkamp, R. E. (2001). Advances in determination of a high-resolution structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). *Biochemistry*, *40*, 7761–7772. https://doi.org/10.1021/bi0155091

Trzaskowski, B., Latek, D., Yuan, S., Ghoshdastider, U., Debinski, A., & Filipek, S. (2012). Action of molecular switches in GPCRs—Theoretical and experimental studies. *Current Medicinal Chemistry*, *19*, 1090–1109. https://doi.org/10.2174/092986712799320556

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. *Proceedings of the National Academy of Sciences*, *116*, 14547–14556. https://doi.org/10.1073/pnas.1902192116

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. https://doi.org/10.1038/nature11896

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. *Proceedings of the National Academy of Sciences*, *116*, 3288–3293. https://doi.org/10.1073/pnas.1809251116

Verhoeven, M. A., Creemers, A. F. L., Bovee-Geurts, P. H. M., De Grip, W. J., Lugtenburg, J., & de Groot, H. J. M. (2001). Ultra-high-field MAS NMR assay of a multispin labeled ligand bound to its G-protein receptor target the natural membrane environment: Electronic structure of the retinylidene chromophore in rhodopsin. *Biochemistry*, *40*, 3282–3288. https://doi.org/10.1021/bi0023798

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. https://doi.org/10.1016/j.str.2017.11.013

Vogel, R., Mahalingam, M., Lüdeke, S., Huber, T., Siebert, F., & Sakmar, T. P. (2008). Functional role of the "ionic lock" - an interhelical hydrogen-bond network in family A heptahelical receptors. *Journal of Molecular Biology*, *380*, 648–655. https://doi.org/10.1016/j.jmb.2008.05.022

Wacker, D., Stevens, R. C., & Roth, B. L. (2017). How ligands illuminate GPCR molecular pharmacology. *Cell*, *170*, 414–427. https://doi.org/10.1016/j.cell.2017.07.009

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 Communications*, *5*, 4733. https://doi.org/10.1038/ncomms5733

Zhou, Q., Yang, D., Wu, M., Guo, Y., Guo, W., Zhong, L., Cai, X., Dai, A., Jang, W., Shakhnovich, 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. https://doi.org/10.7554/eLife.50279

How to cite this article: Bertalan, É., Rodrigues, M. J., Schertler, G. F. X., & Bondar, A.-N. (2025). Graph-based algorithms to dissect long-distance water-mediated H-bond networks for conformational couplings in GPCRs. *British Journal of Pharmacology*, *182*(14), 3163–3177. https://doi.org/10.1111/bph.16387