**Conserved hydrogen-bond motifs of membrane transporters and receptors**

Michalis Lazaratos1, Malte Siemers1, Leonid S. Brown2,

and Ana-Nicoleta Bondar1,3,4\*

1Freie Universität Berlin, Department of Physics, Theoretical Molecular Biophysics, Arnimallee 14, D14195 Berlin, Germany

2University of Guelph, Department of Physics, 50 Stone Road E., Guelph, Ontario, Canada N1G 2W1

3University of Bucharest, Faculty of Physics, Atomiștilor 405,

Măgurele 077125, Romania

4Forschungszentrum Jülich, Institute for Neuroscience and Medicine and Institute for Advanced Simulations (IAS-5/INM-9), Computational Biomedicine, Wilhelm-Johnen Straße, 52428 Jülich, Germany

\*Correspondence: nbondar@fizica.unibuc.ro

**Abstract**

Membrane transporters and receptors often rely on conserved hydrogen bonds to assemble transient paths for ion transfer or long-distance conformational couplings. For transporters and receptors that use proton binding and proton transfer for function, inter-helical hydrogen bonds of titratable protein sidechains that could change protonation are of central interest to formulate hypotheses about reaction mechanisms. Knowledge of hydrogen bonds common at sites of potential interest for proton binding could thus inform and guide studies on functional mechanisms of protonation-coupled membrane proteins. Here we apply graph-theory approaches to identify hydrogen-bond motifs of carboxylate and histidine sidechains in a large data set of static membrane protein structures. We find that carboxylate-hydroxyl hydrogen bonds are present in numerous structures of the dataset, and can be part of more extended H-bond clusters that could be relevant to conformational coupling. Carboxylate-carboxyamide and imidazole-imidazole hydrogen bonds are represented in comparably fewer protein structures of the dataset. Atomistic simulations on two membrane transporters in lipid membranes suggest that many of the hydrogen bond motifs present in static protein structures tend to be robust, and can be part of larger hydrogen-bond clusters that recruit additional hydrogen bonds.

**Introduction**

Proton transfer reactions are fundamental for the functioning of biological cells. Proton transporters bind and transfer protons across cell membranes, and pH-sensing membrane receptors may change protonation state during activation. Description of structure and sequence signatures, or motifs, that proteins use to control conformational dynamics and couplings between protonation and conformational change, could guide the interpretation of experiments on proteins that use proton transfers for function. Here, we present a methodology to extract common hydrogen(H)-bond motifs that involve protein titratable groups from static protein structures. We apply this methodology to common H-bond motifs of carboxylate and histidine sidechains in static structures of membrane transporters and receptors. To evaluate dynamics of internal H-bond networks that contain H-bond motifs of interest, we study the motions of *Pichia pastoris* aquaporin-1 (Aqy1) and *Chlamydomonas reinhardtii* channelrhodopsin-2 (ChR2).

Membrane proteins that bind protons to exert biological function typically rely on carboxylate and histidine groups, as these groups titrate in the physiological range [1, 2]: aspartic and glutamic acids in solution have nominal pKa values of about 3.5-4.5 [3], and histidine, about 6.5 [2]. Cysteine, with a nominal pKa of 8.5 [4], might also be of interest as a potential proton-binding site. The functional role of a titratable protein sidechain, including its potential involvement in a proton-transfer reaction, depends on interactions with the local protein environment, as the environment may impact pKa values significantly [3, 5]. For example, a proton-binding Glu sidechain of the *Escherichia coli* lactose permease LacY has a pKa of ~10.5 [6]; in the *Halobacterium salinarum* proton pump bacteriorhodopsin, the pKa of the Asp sidechain that functions as primary proton acceptor, D85, couples to that of the proton release group [7] such that the pKa of D85 shifts from 2.6 to 7.5 upon deprotonation of the proton release group [8], and to >10.5 during a later long-lived intermediate in which the environment of D85 is thought to be nonpolar [9].

Static structures of a number of membrane transporters and receptors appear to suggest different proteins might have similar interactions at sites where proton binding occurs, or sites that are otherwise important for function. For example, an inter-helical H-bond between a carboxylate group and a Ser/Thr hydroxyl group was noted at functionally important sites of microbial rhodopsins [10, 11], the SERCA calcium pump [10], the multidrug transporter AcrB [10], the brain glutamate transporters EAAT1 and VGLUT2 [12], the ammonia channel Amt/MEP/Rh [13], and proteins from the family of G Protein-Coupled Receptors, GPCRs [14, 15]. The AHA-2 plasma membrane proton pump [16, 17] and the chloride-pumping rhodopsin from *Mastigocladopsis repens* [18] have at proton-transfer sites an inter-helical H-bond between Asp and Asn sidechains; the proton-binding neurotransmitter transporter VGLUT2, the sugar-proton symporter LacY, and the voltage-gated proton channel Hv1, have at functionally important sites a His sidechain interacting with a carboxylate group [12, 19]; H-bonding His sidechains might be involved in the functioning of a pH-sensing GPCR [20], the M2 proton channel [21], and the ammonia channel [13]. More generally, His-Ser pairs are of interest for proton transfer in serine proteases [22], and His-Asp/Glu pairs, in dehydrogenases [23, 24], serine proteases [25], as part of the counterion for the primary proton-binding site in microbial rhodopsins such as proteorhodopsin [26], xanthorhodopsin [27], and *Gloeobacter* rhodopsin [28], and at the cytoplasmic proton donor site of xenorhodopsin [29]. Asp and Glu are preferred H-bond acceptors for Tyr [30].

The examples summarized above are largely based on interactions present in a relatively small number of structures, such that it is often unclear whether particular types of H-bonds are indeed common at functionally important sites of membrane transporters and receptors. Knowledge of common H-bond motifs is valuable, as it could guide the choice of amino acid residues for site-directed mutagenesis studies to decipher protein reaction mechanisms.

To this aim, here we present a methodology to evaluate H-bonds present in static structures of proteins that belong to different families. We rely on a large data set of 683 structures, which we organize in two sub-sets according to the resolution at which structures were solved. We then extract a subset of structures solved at high resolution and which accounts for redundancies in structures of the same protein. Additionally, as proteins from the superfamily of Rhodopsin-like receptors contribute a significant part of the dataset, we separate them into two smaller data sets: one, with structures of microbial and algal rhodopsins, and the second, with G Protein-Coupled Receptors (GPCRs). Each structure in each dataset is subjected to analyses using graph-based algorithms that represent proteins as graphs with protein groups as nodes, and H-bonds, as edges of the graph. We examine the graphs to identify 10 types of H-bonds. Then, for H-bond types of interest, we extract from the H-bond graphs complete H-bond networks for which we characterize their spatial extent.

Inter-helical H-bonds may have complex dynamics difficult to anticipate from static protein structures [31]; to illustrate H-bond dynamics, we augment the analyses of static protein structures with atomistic Molecular Dynamics (MD) simulations to probe H-bond dynamics of *i)* *P. pastoris* Aqy1, which we chose because its structure was solved at subangstrom resolution [32], such that it provides accurate reference for H-bonds sampled during MD simulations; *ii)* a microbial rhodopsin ChR2, as a member of one of the largest protein families represented in our dataset. Both Aqy1 and ChR2 are homo-oligomeric proteins.

The systematic analyses give an overview of H-bond motifs in static membrane protein structures. H-bonds between Asp/Glu and Ser/Thr sidechains are relatively frequent and, when present, tend to locate within the transmembrane region of the protein; H-bonds between His sidechains are rather infrequent – which could be due, at least in part, to His sidechains being relatively infrequent. The protocol we present for analyses of H-bond motifs could be extended to other H-bonds of interest, and to other data sets of protein structures. Most of the H-bonds investigated for the high-resolution structure of Aqy1 remain present in a fluid lipid membrane environment, and an internal H-bond cluster can recruit additional transient H-bonds.

**Methods**

*Datasets Set-A and Set-B of static protein structures.* We first prepared a dataset of 1439 structures of alpha-helical polytopic TM proteins. Coordinate files of proteins pre-oriented in membranes were extracted from the database Orientations of Proteins in Membranes, OPM [33], last accessed on February 16, 2021.

OPM organizes protein structures according to the protein superfamily, which is assigned a unique identifier. For each protein entry, OPM reports the resolution at which the structure was solved, the Protein Data Bank (PDB) [34] identifier, the experimental approach used to solve the structure, the number of TM subunits of the protein, and the number of TM helices. The majority (1426) of the structures from our starting dataset were solved with X-ray crystallography or cryo-Electron Microscopy (cryo-EM); 13 structures were solved with Nuclear Magnetic Resonance (NMR) spectroscopy.

Our initial search for membrane protein structures using a relatively coarse resolution threshold of 5Å led to 1327 entries, of which 1195 structures were X-ray structures solved at a resolution of 4Å or better; the highest resolution, 0.88Å, was reported for Aqy1 [32]. This first set of 1327 protein structures includes proteins from 27 superfamilies. We selected the highest resolution structures and discarded ten structures that lacked TM regions. For several structures we used CHARMM-GUI [35] to ensure coordinate files have standard PDB format.

The resulting dataset contained 683 structures, which we organized into two sub-sets as follows. *Set-A* contains 200 protein structures that belong to 17 superfamilies, and were solved at a resolution of 2.5Å or higher. *Set-B* contains 483 protein structures from 26 different superfamilies, solved at resolutions between 2.5Å and 3.5Å (Tables 1, 2). The PDB entries and corresponding resolutions of structures included in *Set-A* and *Set-B* are listed in Tables S1 and S2.

*Subsets Set-A1, Set-A1m, and Set-A1g of static protein structures*. A number of proteins included in *Set-A* are represented by more than one PDB coordinate set. To find out whether having more than one coordinate set for some of the proteins significantly influences frequencies of the H-bond motifs, we prepared *Set-A1* that accounts for sequence identities between proteins. We used the PDB ID of each coordinate set from *Set-A* in the Macromolecules option of the PDB [34] to find molecules with 100% sequence identity, and checked for their presence in *Set-A*. When the same protein was represented in Set-A by more than one coordinate set in (Table S3), we kept the best-resolution structure (Table S4). Protein homologues from different organisms, such as bacteriorhodopsins from *H. salinarium*, *Haloquadratum walsbyi,* and *Haloarcula marismortui,* were treated as distinct proteins. The resulting subset *Set-A1* contains 147 entries (Tables 1, S4).

The superfamily of Rhodopsin-like receptors and pumps, which contributes the largest number of entries to *Set-A1* (Table 1), includes microbial rhodopsins and GPCRs. To dissect H-bond motifs of these proteins, we generated subsets *Set-A1m*, which contains 28 entries of microbial and algal rhodopsins, and *Set-A1g*, which contains 35 structures of GPCRs (Table S5).

*Groups of static protein structures*. Given the large number of superfamilies represented in the dataset, we used the Transporter Classification Database (TCDB) [36] to group protein superfamilies according to biological function. We relied on the first database index/ classification number that denotes the transporter class. Since the superfamilies of rhodopsin-like receptors and pumps and of the voltage-gated ion channels have diverse subfamilies, we treated separately these two superfamilies (Table 2). The Na+-translocating NADH-quinone reductase is included with the channels/pores group of the ammonia/ urea transporters/ Na+ exporters (Table 2).

According to the TCDB, heliorhodopsin and the human adiponectin receptor 2 of the Rhodopsin-like receptors and pumps superfamily belong to distinct families. We tested the H-bond motifs for these two proteins by analyzing them separately as *Set-A1hm* and *Set-A1hr*.

*Transmembrane (TM) domain and TM protein core*. For simplicity, for static protein structures we restrict H-bond analyses to the protein TM domain defined according to the OPM average membrane boundaries [33]. The average distance between the membrane planes estimated by OPM ranges from 28Å to 36Å, i.e., the maximum distance from the center of the bilayer to the membrane plane is ~18Å. To account for atomic fluctuations that could take place at room temperature in a fluid lipid membrane environment, we add 2.5Å to each side of the membrane plane.

We define as the TM core of a protein the region of the protein within ±6Å from the center of the membrane.

**Table 1**. Summary of the dataset of structures we analyzed. We report the superfamily and the number of structures in *Set-A, Set-A1,* and *Set-B*.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | #*Set-A* | #*Set-A1* | #*Set-B* |
| Sodium/calcium exchanger family | 4 | 3 | 3 |
| Proton or Sodium translocating F-type, V-type and A-type ATPase superfamily | 9 | 9 | 27 |
| Rhodopsin-like receptors and pumps superfamily | 94 | 65 | 141 |
| Proton-translocating pyrophosphatase family | 1 | 1 | 3 |
| Ion channel (VIC) superfamily | 24 | 17 | 118 |
| Major Intrinsic Protein (MIP)/FNT superfamily | 20 | 17 | 5 |
| Chloride transporter (ClC) family | 1 | 1 | 5 |
| Small conductance mechanosensitive ion channel family | 0 | 0 | 4 |
| Large conductance mechanosensitive ion channel family | 0 | 0 | 2 |
| Ammonia/urea transporters/Na+ exporter superfamily | 11 | 10 | 2 |
| Ion transporter superfamily | 0 | 0 | 2 |
| Resistance-nodulation-cell division superfamily | 2 | 1 | 26 |
| ABC transporter superfamily | 7 | 7 | 59 |
| Vacuolar iron transporter family | 0 | 0 | 2 |
| P-type ATPase (P-ATPase) superfamily | 12 | 3 | 46 |
| Proton-translocating transhydrogenase family | 1 | 1 | 1 |
| Copper transporter family | 0 | 0 | 1 |
| Magnesium ion transporter-E (MgtE) family | 1 | 0 | 2 |
| Cation diffusion facilitator family | 0 | 1 | 1 |
| CorA metal ion transporter (MIT) family | 0 | 0 | 3 |
| Drug/Metabolite Transporter (DMT) superfamily | 3 | 3 | 4 |
| Potassium channel TMEM175 family | 1 | 1 | 1 |
| Calcium release-activated calcium (CRAC) channel family | 0 | 0 | 1 |
| Monovalent cation-proton antiporter superfamily | 5 | 0 | 7 |
| Calcium-activated chloride channel family | 0 | 5 | 10 |
| Bestrophin anion channel family | 4 | 0 | 7 |
| *Total number of structures* | 200 | 147 | 483 |

Table 2. Number of protein structures of each group obtained according to the TCDB [36].

|  |  |  |  |
| --- | --- | --- | --- |
| Protein groups | *Set-A* | *Set-A1* | *Set-B* |
| Channels/Pores | 37 | 31 | 38 |
| Electrochemical Potential-driven Transporters | 15 | 13 | 50 |
| Primary Active Transporters | 30 | 21 | 136 |
| Rhodopsin-like receptors and pumps superfamily | 94 | 65 | 141 |
| Voltage-gated Ion Channel (VIC) superfamily | 24 | 17 | 118 |

*H-bond criteria, H-bond graphs and H-bond clusters, shortest path length.* All-H-bond computations were performed with Bridge, a graph-based algorithm that uses geometric criteria to detect H-bond networks in bio-molecules [37]. As *criteria for H-bonding* we used a distance cutoff of 3.5 Å between donor and acceptor heavy atoms, and a H-bond angle cutoff of 60°. In the case of MD simulations, we report *H-bond occupancies* as the percentage of time during which the H-bond criteria are met, normalized by the length of the trajectory segment used for analyses.

An H-bond graph consists of the collection of nodes *ni*, which are groups that H-bond, and edges (H-bonds) connecting nodes *ni*, *nj*. We define an *H-bond cluster* as a H-bond network of amino acid residues (graph nodes) in which each node has at least one edge that connects it to another node of the cluster. Each cluster is characterized by its *size*, defined as the number of nodes, or amino acid residues, that are inter-connected in the cluster. Thus, a singe H-bond between two nodes can be also referred to as an H-bond cluster of size 2.

*The shortest distance path* between nodes *ni* and *nj* connects the two nodes by passing through the least number of intermediate nodes *nk*. *The length of a shortest path* between nodes *ni* and *nj* is given by the number of edges, i.e., H-bonds, that connect the two nodes along the shortest-distance path.

*Betweennees Centrality (BC) and Unique Shortest Paths (USP)*. Centrality measures allow us to assess the relative importance of nodes in a graph. For protein interaction networks, amino acid residues with high centrality values tend to be conserved and/or functionally important [38], including for allosteric conformational coupling [39-41]. The BC of node *ni*  (Scheme 1) gives the number of shortest-distance paths between any two other nodes *nj* and *nk* that pass via node *ni,* divided by the total number of shortest paths that connect *nj*and *nk* irrespective of whether they pass via node *ni.* [42-44]. Thus, the higher the BC value of node *ni*, the larger the number of shortest-distance paths passing through *ni*, suggesting that node *ni* could be important for couplings within the network.

**Diagram

Description automatically generated**

**Scheme 1.** Schematic representation of BC vs. USP values for a hypothetical H-bond graph. Because the USP reports only the unique shortest paths, USP values are smaller and thus might be easier to interpret. In the graph illustrated here, the highest BC value for a node is BC = 66; for the same node, USP = 26.

In the case of crowded networks with large number of nodes, nodes located in the center of a network can have very large BC values which could make it difficult to interpret connections within the network [45]. The USP gives a somewhat more intuitive measure, because it restricts the computation to the unique shortest paths passing through node *ni* (Scheme 1) [45], such that it avoids over-counting intermediate nodes along paths connecting distinct nodes of the network.

*Longest H-bond paths of selected H-bond motifs.* To evaluate the participation of H-bonds of interest in localH-bond clusters, we use Bridge [37] for Connected Component Analyses to extract local H-bond clusters of nodes (protein groups) of interest. To estimate the spatial extent of such local H-bond clusters, we compute the length of the shortest-distance H-bond paths as illustrated in Scheme 2.

Given two H-bonding groups of an H-bond motif, we compute all shortest distance paths that include one of these groups. We take the longest path of the shortest-distance paths thus identified, and report as path length the number of H-bonds of the path. The example given in Scheme 2 shows an Asp-Thr H-bond that is part of a larger H-bond cluster with 12 shortest-distance paths that include the Asp sidechain.

**Diagram

Description automatically generated**

**Scheme 2.** Longest path analysis. (A) The H-bond motif of interest is part of a local H-bond network. Here, the motif consists of an inter-helical Asp-Thr carboxylate-hydroxyl motif. (B) The carboxylate group is chosen as root node to compute H-bond paths. There are 12 shortest-distance paths that include the Asp node, and the length of the longest path is 7.

*Computation of H-bond motifs.* We used Bridge to identify ten types of H-bonds, which we denoted H-bond motifs, chosen based on our observations of typical sidechain H-bond interactions in membrane proteins, as follows: *i)* H-bonds between a Ser/Thr hydroxyl and the *i-3*, *i-4*, or *i-5* backbone carbonyl; *ii)* H-bonds between Asp/Glu and SerThr sidechains; *iii)* combined H-bonds in which the same Ser/Thr hydroxyl group is part of H-bonds at points *i)* and *ii)*; *iv)* H-bonds between Asp and Asn sidechains; *v)* H-bonds between Asp/Glu and Arg sidechains; *vi)* H-bonds between Asp/Glu sidechains and backbone amide groups; *vii)* H-bonds between two His imidazole groups; *viii)* H-bonds between Asn and a Ser/Thr sidechains; *ix)* H-bonds between Asn sidechains and backbone carbonyl groups; *x)* H-bonds between Arg sidechains and backbone carbonyl groups.

**Table 3**. Summary of H-bond motifs. For each H-bond motif, we indicate one protein in which the H-bond motif is present, and a corresponding Protein Data Bank (PDB) structure. The resolution at which the structure was solved is indicated in Å. For backbone carbonyl H-bonds of Ser/Thr hydroxyls we include the *i-3*, *i-4*, and *i-5* amino acid residues.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Motif | H-bond pair | | Example | PDB ID, Reference, Resolution (Å) |
| *H-bonds between Ser/Thr hydroxyl groups and backbone carbonyls´groups* | | | | |
| S/T-CO | Ser/Thr hydroxyl | Backbone carbonyl | Bacteriorhodopsin  T46-F42 | 5ZIM [46], 1.25 |
| *H-bond motifs with Asp/Glu sidechains* | | | | |
| S/T-D/E | Ser/Thr hydroxyl | Asp/Glu carboxylate | Aqy1  E51-S107 | 3ZOJ [32], 0.88 |
| S/T―D/E  |  CO | Ser/Thr hydroxyl - Asp/Glu carboxyl | Ser/Thr hydroxyl -  backbone carbonyl | ChR2  E101-T246  V242-T246 | 6EID[47], 2.39 |
| D-N | Asp carboxyl | Asn carboxamide | Proton-translocating pyrophosphatase  N738-D294 | 4A01 [48], 2.35Å |
| D/E-R | Asp/Glu carboxyl | Arg guanidinium | Archaerhodopsin-3  D15-R17 | 6S6C [49], 1.07 |
| D/E-NH | Asp/Glu carboxyl | Backbone amide | Heliorhodopsin  D158-L160 | 6SU3 [50], 1.50 |
| *H-bond motifs with His sidechains* | | | | |
| H-H | His imidazole | His imidazole | Ammonia channel AmtB  H168-H318 | 1U7G [13], 1.40 |
| *Other H-bond motifs with Asn and Arg sidechains* | | | | |
| N-S/T | Asn carboxamide | Ser/Thr hydroxyl | Mep2 Ammonium Transceptor  N88-T95 | 5AEZ [51],1.47 |
| N-CO | Asn carboxamide | Backbone carbonyl | Sodium pumping rhodopsin KR2  N112-S70 | 6YC3 [52], 2.00 |
| R-CO | Arg guanidinium | Sensory rhodopsin II  R66-P182 | 1H2S [53],1.93 |

*Starting coordinates for MD simulations of Aqy1.* As starting coordinates for the Aqy1 tetramer we used the OPM entry for PDB ID:3ZOJ [32]. All 221 water oxygen atoms were included. We used CHARMM [54, 55] to construct coordinates for the water H-atoms, VMD [56] to orient and place the Aqy1 tetramer with its center of mass in the center of coordinates, and CHARMM-GUI [35] to place the protein in hydrated lipid membranes (Table 4).

Being solved at sub-Ångstrom resolution, the crystal structure of Aqy1 contains coordinates for H atoms. All Asp/Glu sidechains of Aqy1 are indicated by the crystal structure as negatively charged. There are in total four His sidechains (Figures S1, S2), of which H44 and H194 are indicated as N2-protonated, and H212 and H242, as -N1 protonated. H212 and H242 have intra-helical H-bonds with the backbone carbonyl of M208 and, respectively, with the W236 sidechain (Table S6, Figures S1, S2). H212 is among amino acid residues that line the water channel, and its strong electron supported the assignment of the tautomeric state [32]. H44 and H194 are both close to the bulk (Figure S1), where dynamic fluctuations in a fluid environment could associate with altered H-bonding of these sidechain. To test whether the tautomeric state of H44 and H194 alters local H-bond interactions, we studied the dynamics of Aqy1 with *i)* the tautomeric state of all four His sidechains as indicated by the crystal structure, i.e., H44 and H194 N2-protonated, and H212 and H242 N1-protonated; *ii)* all four His sidechains N1-protonated. All other titratable sidechains were assigned standard protonation states, that is, all Asp/Glu sidechains are negatively charged, and Arg and Lys sidechains are positively charged.

To probe whether the dynamics of Aqy1 depends on lipids, we performed independent simulations of Aqy1 embedded in *i)* a hydrated lipid bilayer of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), which is often used as model membrane for bacterial membranes; *ii)* a membrane composed of 45%:35%:20% POPE, 1-palmitoyl-2-oleoylphosphatidylcholine (POPC), and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS), which are the major phospholipids of the *P. pastoris* membrane [57] (Table 4).

**Table 4**. Summary of simulations performed. ‘Protonation’ refers to the protonation chosen for selected His, Asp, and Glu sidechains. All other titratable sidechains have standard protonation. All simulations were performed at 300K. The total sampling time is 1.4s.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Simulation** | **Membrane** | **#Atoms** | **Protonation** | **Length (ns)** |
| *ChR2* |  | | | |
| all-trans | POPC | 177,367 | E90, D156  protonated | 200 |
| 13-cis |
| *Aqy1* |  | | | |
| POPE-N1 | POPE | 261,360 | H44-N1, H194-N1 | 261 |
| POPE-N2 | H44-N2, H194-Nε2 | 250 |
| Mixed-HSD | 4.5:3.5:2  POPE:POPC:POPS | 263,094 | H44-N1, H194-N1 | 251 |
| Mixed-HSE | H44-N2  H194-Nε2 | 250 |

*Starting coordinates for MD simulations of ChR2.* Starting coordinates of the wild-type ChR2 homodimer with 52 crystallographic waters and all-*trans* retinal were taken from the OPM entry for PDB ID:6EID [47]. Missing atom coordinates were generated using MODELLER [58-60]; E90 and D156 were considered protonated [61-65], and C34 and C36 were disulfide bridged [47]. ChR2 was embedded in a hydrated lipid membrane with 401 POPC lipid molecules and ions added for charge neutrality (Table 4).

To generate a model of CHR2 with 13-*cis,*15-*anti*-retinal we used adiabatic mapping in CHARMM to twist the C13=C14 bond by 180º.

*Protocol for MD simulations*. All simulations were performed with NAMD [66]. The Smooth Particle Mesh Ewald summation [67, 68] was used to compute electrostatic interactions. For van der Waals interactions in the real space we used a switch function between 10Å and 12Å. Equilibration runs were performed using the standard CHARMM-GUI protocol, during which harmonic restraints placed on atoms of the system are gradually released [35]. We used the *NPT* ensemble (*P*=1bar, *T*=300K) for production runs. Pressure and temperature were controlled using the Nosé-Hoover method and a Langevin piston [69, 70] with a Langevin damping coefficient of 5ps-1. We used CHARMM36 [54, 71-74] parameters used for protein, lipids and ions, TIP3P [75] for water, and the retinal parameters from ref. [10].

Data analyses were performed for the last 100ns of the trajectories. H-bond paths are included in the analyses when present during at least 5% of the last 100ns of the simulations.

**Results**

We applied graph-based searches to identify ten H-bond motifs in static protein structures using a large data set of 200 structures solved at a resolution of 2.5Å or better, and augmented the search with analyses of a larger set of 483 structures solved at resolutions between 2.5 and 3.5Å. To dissect dynamical fluctuations of H-bond networks in a fluid, hydrated lipid membrane environment, we performed 6 sets of independent MD simulations for two membrane proteins.

*H-bonds of Asp/Glu sidechains.* The apparent free energy for membrane insertion of one Asp/Glu residue is ~4-5kcal/mol [76], such that Asp/Glu tend to disfavor being located in the center of membrane proteins [77] (Figure S6). We find that, in protein structures of *Set-A*, Asp/Glu carboxylates often pair with Arg guanidinium groups, or with Ser/Thr hydroxyl groups (Figures 1, 2, 3). Although there are typically more Ser/Thr than Arg groups in the center of a membrane protein (Figures S7, S8), the presence of Ser/Thr-Asp/Glu H-bonds per proteins of a superfamily is relatively similar to Arg-Asp/Glu (Figures 1, 2), suggesting Ser/Thr sidechains have more diverse H-bond partners than Arg.

Consistent with the known propensity of Ser/Thr hydroxyl groups to make intra-helical H-bonds with the backbone carbonyl on the preceding helix [78-80], structures from almost all protein families we analyzed have H-bonds between Ser/Thr hydroxyls and backbone carbonyl groups (Figures 1, 2A, 2B, S3). In a relatively large number of structures that belong to the Superfamily of Rhodopsin-like Receptors and Pumps, the same Ser/Thr hydroxyl group H-bonds to an Asp/Glu sidechain and to a backbone carbonyl group (~48%, 45 entries, see Figures 2C, S3C). Other protein structures in which combined carboxylate-hydroxyl-carbonyl H-bonds are present tend to belong to various transporters, e.g., 9 *Set-A* protein structures from the Major Intrinsic Protein/FNT superfamily (Figures 1A, S3A).

Overall, and consistent with the distribution of Ser/Thr groups (Figure S7), intra-helical H-bonds between Ser/Thr hydroxyl groups and backbone carbonyls may be present along the entire length of the membrane protein (Figures 3, S4). When present, H-bonds between Asp/Glu and Ser/Thr tend to be relatively close to the center of the membrane core (Figures 3, S4).

Examples of proteins with an Asp/Glu-Ser/Thr H-bond within the TM core include Rhodopsin-like Receptors and Pumps, proteins of the superfamily proton or sodium translocating F-type, V-type and A-type ATPases (Figures 3, S4, S5). For some of these proteins, we manually inspected protein structures to verify H-bonds. Among Rhodopsin-like Receptors and Pumps, archaeorhodopsin-2 has H-bond distance between T95 and D120 in structures from refs. [81, 82]; bacteriorhodopsin has an inter-helical H-bond between T46 and D96 [10]; structures of cruxrhodopsin [83], *Coccomyxa subellipsoidea* rhodopsin [84], the thermophilic rhodopsin from *Rubrobacter xylanophilus* [85], sodium pump rhodopsin KR2, channelrhodopsin, *Anabaena* rhodopsin [11], and jumping spider rhodopsin-1 [86], also have H-bonds between Asp/Glu and Ser/Thr sidechains.

ATP synthase protein structures that have inter-helical carboxyl-hydroxyl H-bonds include the c-ring of F1F0-ATP synthase [87], with an inter-subunit H-bond between E65 and S66; the hybrid F/V-rotor ring of the Na+-coupled ATP synthase from *Acetobacterium woodii* [88] and the V-type Na+-coupled ATP synthase from *Enterococcus hirae* [89] have carboxyl-hydroxyl-carbonyl motifs (E79-T63-V60 and E139-T64-L61, respectively).

Proteins of the Voltage-Gated Ion Channel superfamily have numerous Ser/Thr groups in the region from -2 to 6Å, but relatively few carboxylate groups (Figures 3E, S6E), and Asp/Glu-Ser/Thr H-bonds are present in only 5 Set-A structures, which is ~20% for this superfamily (Figure 1A). An example of a protein structure with such an H-bond belongs to the voltage-gated sodium channel NavAB (D81-S112) [90].

Asn sidechains can be present in the center of *Set-A* proteins, though details of their distribution can depend somewhat on the protein superfamily (Figure S9). 27 of the 94 *Set-A* Rhodopsin-like Receptors and Pumps (Figures 1, 2C) have at least one Asp-Asn H-bond. In the proton-translocating pyrophosphatase [48], the Asp-Asn H-bond between D294 and N738 (Figures 2C, S3C) could be relevant for the conformational dynamics of the protein, because N738 stabilizes water molecules of the proton transfer path [48]. The Mep2 fungal transceptor [51] has two Asp-Asn H-bonds, between D365 and N246, and between D273 and N321.

H-bonds between backbone groups and sidechains of Asp/Glu, Arg, or Asn, are found in proteins of all superfamilies we analyzed, except for the magnesium ion transporter-E that lacks backbone H-bonding of Asn sidechains (Figure 1A). H-bonds between Ser/Thr and Asn sidechains have relatively moderate presence in all structures included in our database (Figure 2C).

Overall, observations presented above on H-bond motifs identified for *Set-A* (Figures 1, 2) remain qualitatively similar when analyses are repeated for *Set-A1* (Figures S10-S14), suggesting that any potential bias from multiple entries of a number of proteins in *Set-A* (Tables S3, S4) is relatively small. For subsets *Set-A1m* and *Set-A1g* (Table S5) we note that Asp-Asn H-bonds are more frequently present in *Set-A1g* (GPCRs) than in *Set-A1m* (microbial and algal rhodopsins) (Figure 1F). Taken together, there are more Asp-Asn and Ser/Thr-Asn than Asp/Glu-Arg H-bond motifs in *Set-A1g* than *Set-A1m* (Figure S11D); likewise, the relative percentage of Asn-backbone carbonyl H-bonds vs. Asp/Glu-backbone amide and Arg-backbone carbonyl H-bonds is larger in *Set-A1g* than in *Set-A1m* (Figure S11F). The smaller percentage of H-bond motifs with Asp/Glu and Arg in *Set-A1g* could be interpreted to suggest fewer strong interactions in structures of *Set-A1g* than *Set-A1m*, which could facilitate conformational dynamics in GPCRs of *Set-A1g*. Such an interpretation would be consistent with the notion that GPCRs have significant conformational dynamics [91]. As found for other microbial and algal rhodopsins, Heliorhodopsin has Asp/Glu-Ser/Thr H-bonding (Figures S10, S11).

*H-bonds of His sidechains*. His sidechains are relatively infrequent in -helical membrane proteins [77] and disfavor locations close to the center of the membrane region [92] (Figure S6). It is then not surprising that H-bonding between His sidechains is rather infrequent in the static protein structures we analyzed: we could identify H-bonded His pairs only in the ammonia transporter AmtB, and in some of the rhodopsin-like proteins (Figures 1, 2C, S11 C&D). When present, H-bonded His pairs tend to locate at sites important for function: in AmtB, the H-bonded His groups H168 and H318 are involved in substrate binding [13] and potentially important for protonation-coupled ammonium transport [93]. H-bonded His sidechains of the adiponectin receptor 2 (H202-H348-H352) and alkaline ceramidase 3 (H81, H222) (Figures 2D, S10C, S11D) might be involved in ion binding [94].

*A screenshot of a computer

Description automatically generated with medium confidence*

**Figure 1.** Summary of all H-bond motifs identified for static protein structures of *Set-A*. The length of the gray bar indicates presence of an H-bond motif in all protein structures included for a superfamily/group. (A) H-bond motifs in each superfamily. (B) H-bond motifs identified in proteins grouped according to their biological function according to TCDB. Analyses for *Set-A1*, *Set-A1m, Set-A1g, Set-A1hm*, and *Set-A1hr*, are presented in Figure S10.

*Other H-bond motifs with Asn and Arg sidechains, and H-bonds of Set-B structures*. H-bonds between backbone groups and sidechains of Arg or Asn are found in proteins of all superfamilies we analyzed, except for the magnesium ion transporter-E that lacks backbone H-bonding of Asn sidechains (Figure 1A). H-bonds between Ser/Thr and Asn sidechains have relatively moderate presence (Figure 2C).

A picture containing writing implement, stationary, pencil, lined

Description automatically generated

**Figure 2**. Summary of H-bond motifs in *Set-A* and *Set-B*. For each superfamily or group of proteins we give the percentage of structures that contain the H-bond motif. The number of structures that contain each motif is indicated on the corresponding vertical bar, and the total number of structures is given in Table 2. (A, B) Ser/Thr-backbone carbonyl, Asp/Glu-Ser/Thr, and Asp/Glu-Ser/Thr-backbone carbonyl H-bonds in *Set-A* (panel A) vs. *Set-B* (panel B). (C, D) Asp/Glu-Arg, Asn-Asp, His-His, Ser/Thr-Asn motifs in *Set-A* (panel C) vs. *Set-B* (panel D). (E, F) Asp/Glu - backbone amide, Arg - backbone carbonyl, and Asn - backbone carbonyl motifs for superfamilies included in *Set-A* (panel E) vs. *Set-B* (panel F). H-bond analyses for *Set-A1*, *Set-A1m, Set-A1g, Set-A1hm*, and *Set-A1hr,* are presented in Figure S11.

Chart, bar chart, surface chart

Description automatically generated

**Figure 3.** Distribution of Asp/Glu-Ser/Thr carboxyl-hydroxyl and carboxyl-hydroxyl-carbonyl H-bond motifs along the membrane normal for *Set-A* protein structures. Results for *Set-B* are presented in Figure S5, and for *Set-A1*, *Set-A1m*, *Set-A1g*, *Set-1Ahm*, and *Set-A1hr*, in Figures S12, S13. Proteins are grouped according to the TCDB (Table 2). (A-E) Distribution along the membrane normal for Channels/Pores (panel A), Electrochemical Potential-driven Transporters (panel B), Primary Active Transporters (panel C), Rhodopsin-like receptors and pumps (Panel D), and Voltage-gated ion channels (panel E).

The presence of H-bond motifs in static structures of *Set-B* is qualitatively very similar to that discussed above for *Set-A* (Figures 2, S3, S5). Overall, the number of *Set-B* protein structures with H-bond motifs can be smaller than for the *Set-A* counterpart: when we sum up the number of motifs according to the protein group (Table 2), 77-93% of the *Set-A* proteins have carboxylate-hydroxyl H-bonds, as compared to ~53-90% in *Set-B*. As *Set-B* structures have lower resolution than *Set-A*, the information about details of H-bonding could be somewhat less accurate.

*H-bond networks with Asp/Glu or His sidechains in static protein structures*. The large number of structures for which we identified various H-bond motifs makes it challenging to verify functional roles that might have been proposed for H-bonding sidechains in each structure. Instead, we relied on graph-based analyses to evaluate the likelihood that H-bond motifs participate in larger H-bond networks that could be of potential interest for structural dynamics of the protein. To this aim, we computed the length *L* of H-bond paths that involve H-bond motifs.

Overall, for all H-bond motifs, most path length values are larger than *L* = 3, i.e., there often is at least one more H-bonding group within H-bond distance of either member of the motif. H-bonds between Asn and Ser/Thr, or between Arg and Asp/Glu, are often part of larger H-bond paths with *L* > 5 (Figures 4C, D). The longest H-bond paths are found for Arg-Asp/Glu and Asn-Ser/Thr H-bonds, which can engage in long paths with *L* > 15; a path with *L* = 27 is present in structure PDB ID:6YC3 of the KR2 pentamer [52], where 44 amino acid residues are part of an H-bond network that includes all protein monomers.

The His pair H171-H326 of the ammonium sensor/transducer Amt [95] participates in an internal H-bond network with 31 other protein groups; in this network, the His-His H-bond is part of 12 linear H-bond paths with *L* = 15, i.e., with 14 other protein groups (Figure S15). But the relatively small number of His-His H-bonds in our dataset (Figures 2C,D) makes it difficult to conclude on their propensity to be part of large H-bond networks.

****

**Figure 4.** Distribution of maximum path lengths for selected H-bonds identified for *Set-A*. Paths were computed using the USP definition illustrated in Scheme 1, and the root node definition from Scheme 2; for simplicity, only the longest paths were kept for analyses. (A-D) The number of paths of maximum path length obtained for carboxylate-hydroxyl-carbonyl H-bonds (panel A), His-His H-bonds (panel B), Asn-Ser/Thr H-bonds (panel C), and Arg-Asp/Glu H-bonds (panel D). Additional analyses are presented in Figure S14.

*Most H-bonds of Aqy1 remain present in a fluid lipid membrane*. During MD simulations the Aqy1 structure remained stable, with C root-mean-squared distances (RMSD) within 1Å for all simulations performed (Figures S16A-D), and with overall similar numbers of internal waters in each monomer (Figures S16E-H).

To evaluate whether the choice of the tautomeric state of H44 and H194 impacts dynamics, we monitored H-bonds of the four His sidechains of Aqy1 and compared to interactions in the starting crystal structure (Table S6). The H-bond between H212-N1 and the L208 backbone carbonyl is sampled in all simulations, although in the mixed membrane simulation with H44/H194-N2 only two Aqy1 monomers preserve the H-bond (Table S7). H242-N2 H-bonds frequently to W246-N1 in all simulations, whereas H-bonds of H44 and H194 appear more dynamic: H44 can H-bond to R129 as indicated by the crystal structure, but only when N-2 is protonated, and only in two protomers (Table S7, Figures S1, S17). The H-bond between H194 and the backbone carbonyl of M188 (Figure S2B) is sampled during both the POPE and mixed membrane simulations with H44/H194-N2 protonated (Table S7, Figure S17); H-bonding between the backbone carbonyl of M188 and the H194 imidazole is absent from both H44/H194-N1 protonated simulations (Table S7, Figure S18).

Taken together, the observations summarized above support the tautomeric assignments for the four His sidechains of Aqy1 proposed with the high-resolution structure [32]. As a consequence, unless otherwise specified, in what follows we focus our analyses on the MD simulations of Aqy1 with H44/H194-N2 protonated tautomers in POPE vs. mixed membrane.

The search for H-bond motifs indicated for the crystal structure of Aqy1 33 H-bonds. There are multiple intra-helical H-bonds between Ser/Thr hydroxyl groups and backbone carbonyls, two Glu-Ser/Thr H-bonds, three Asn-backbone carbonyl H-bonds, multiple H-bonds between Arg sidechains and backbone carbonyl groups, and between Asp/Glu sidechains and backbone amide groups (Table S10). Of these, regardless of the H44/H194 tautomeric state or of the lipid membrane, the same 7 H-bonds are absent or poorly sampled during MD (Tables S8-S10). Most of these 7 H-bonds involve groups directly exposed to the bulk – such as E18, G35 and S38, R195, V233-R236; for S61, the crystal structure indicates two orientations of the sidechain, suggesting structural dynamics; the sidechain of N110 faces a cavity and interacts with water.

The largest H-bond cluster in the crystal structure of Aqy1, extending from N160 to S107 (Figure 5), includes the high-centrality group N112 of the conserved NPA motif in the central channel constriction, Q137, and the backbone carbonyl groups of L223 and E175 (Figures 5, S19C, S20). E175 has an inter-helical H-bond with T219, which further H-bonds to T179; Q137 H-bonds to E51, which has an inter-helical H-bond with S107 (Figures 5, S19C, S20C). H-bonds of this extended cluster tend to be sampled throughout MD, and the cluster recruits additional transient H-bonds (Figures 5, S21, S22).

H-bonding of Aqy1 is qualitatively similar in the two lipid environments we tested (Figures 6, S23, S24, Table S12). Each of the aquaporin monomers uses 6-7 sidechains to anchor frequently to lipids, typically at about the same sites.

Radar chart

Description automatically generated

**Figure 5.** Internal H-bond cluster of Aqy1 in the starting crystal structure and during simulations with H44/H194-Nε2. (A) Comparison of the internal H-bond cluster in the crystal structure vs. MD simulations in POPE. The H-bond graph shows H-bonding groups colored green, except for selected groups of the Aqy1 selectivity filter, which are colored blue. Edges colored gray indicate H-bonds present in both the crystal structure and MD, and magenta edges, H-bonds sampled only during MD. Solid lines indicate H-bonds found in all four Aqy1 monomers, and dotted lines, H-bonds found in 1-3 monomers. (B) The H-bond cluster in a coordinate snapshot from MD in POPE. Qualitatively similar results were obtained for Aqy1 in a mixed lipid membrane and with H44 and H194-N1 (Figures S21, S22).

At the cytoplasmic side, Aqy1 anchors to the membrane mainly at S38/D39/D40 and R123/R129, though Q261/W265 can also H-bond to lipids; at the extracellular side, lipid anchoring may occur at R168/T169/R170 and R236/S237/F238. R168 anchors frequently to lipids in all four monomers during both POPE and mixed membrane simulations with H44/H194 –N2 protonated (Table S12). Extracellular lipid anchoring via D39 and R129 may be part of protein-lipid H-bond clusters with 4-6 other protein groups (Figures 6B, 6C, S24B-D), but most protein groups that H-bond frequently to lipids have only one additional protein H-bond (Figures 6B, S24E).

*Chart, scatter chart, bubble chart

Description automatically generated*

**Figure 6.** Lipid interactions of Aqy1 in a hydrated POPE bilayer. (A) Graph of protein-lipid H-bonds. Green, blue, red, and orange nodes represent groups of the four Aqy1 monomers, and cyan nodes, lipid headgroups. A star indicates an amino acid residue that interacts both with lipids and with a local protein H-bond cluster. Protein-lipid interactions sampled at least 50% are shown. (B-D) Selected protein-protein H-bond clusters that include at least one protein sidechain that H-bonds to lipids. Protein H-bonds shown are sampled during at least 10% of the time. Additional analyses are illustrated in Figures S23, S24.

The analyses above suggest that internal H-bonds of Aqy1 are robust when considering dynamics at room temperature in a fluid lipid membrane, and are rather similar in the two lipid bilayers we considered. A factor that could explain the robustness of the Aqy1 H-bonds is the accuracy of the high-resolution structure; the extended H-bond network of Aqy1, with many high-occupancy H-bonds, could also contribute to structural stability. Whether the robustness of the H-bonds of Aqy1 is generally relevant for any membrane protein whose structure is solved at high resolution is, however, unclear. Moreover, the structure of Aqy1 was solved at a cryogenic temperature [32], as was the vast majority of protein structures [96]. Low temperatures used in cryo-trapping could influence preferred protein conformations [97, 98], interactions at ligand binding sites [98], the strength of H-bonds [99], and the likelihood that water molecules occupy internal protein cavities [100, 101].

For the Aqy1 H-bonds we inspected here, there was only weak sensitivity to the lipid environment in POPE vs. the mixed membrane we tested. This finding is consistent with earlier simulations of the aquaporin AQP0 tetramer, in which the acyl chains of dimyristoyl-phosphatidylcholine (DMPC) were largely responsible for lipid-protein interactions, with headgroups H-bonds being assigned a secondary role [102]. Lipid-protein H-bonds can, however, be key determinants of conformational dynamics of membrane proteins known to couple to lipids. Examples include bovine rhodopsin, whose reaction cycle kinetics depends on the lipid bilayer composition [103], the multidrug transporter LmrP, whose functioning requires an aspartic acid-lipid headgroup interaction [104], and the rhomboid protease of *Escherichia coli*, GlpG, whose enzymatic activity depends on lipids [105], and for which MD simulations revealed lipid-dependent H-bond dynamics [106, 107].

*Inter-helical H-bonds of ChR2 respond to changes in retinal isomeric state*. ChR2 is a member of the superfamily of Rhodopsin-like Receptors and Pumps, which is the best-represented protein superfamily in our dataset (Table 1). A common feature of rhodopsins is that the covalently bound retinal chromophore photo-isomerizes, which triggers a reaction cycle in which the protein passes through a series of intermediate conformations distinguished by the retinal isomeric state, protonation state of selected titratable sidechains, number of internal waters, and protein structure. To probe the room-temperature dynamics of a protein from the superfamily of Rhodopsin-like receptors and pumps, we use here the cation channel ChR2.

The crystal structure of ChR2 was solved at a resolution of 2.39Å [47], which is within the threshold for *Set-A* (Table 1). The inter-helical region of ChR2 hosts numerous H-bonds (Figure S25, Table S13), including an Asp/Glu-Ser/Thr-backbone carbonyl H-bond motif between E101, T246, and V242 close to the extracellular bulk; E123-T127 and S63-N258 are within H-bond distance at the center of the protein, (Figures S25 B&C) and E82, E83, and R268, at the cytoplasmic side (Figure S25A).

During MD the structure of ChR2 remains stable, with C RMSD values within 1.5Å (Figures S26 A,B). Many of the H-bonds between Ser/Thr hydroxyl groups and backbone carbonyls are sampled during both simulations (Table S13). H-bonding between T246 and E101 is absent during MD, as T246 prefers intra-helical interactions with V242 and T250 (Table S13), and E101 prefers to interact with R43. Absence of H-bonding between E101 and T246 during MD is compatible with results on Aqy1, in which H-bonds present in the high-resolution structure and absent from MD tended to be at bulk-exposed sites.

In the starting crystal structure [47] the primary proton acceptor D253 [108] H-bonds to S256; nearby, E123 H-bonds to T127 (Figures S25 B,C), N258 frequently H-bonds to S63, and S63 further connects to A59. An H-bond between E123 and the RSB is stable in one monomer, and sampled with ~30% occupancy in the other monomer (Figures 7A, 8B, S27 A&B, S28 C&D). In monomer-A E123 connects frequently to the RSB and to K93/T127 (Figures 8B, S27 A,C S28 E,G), and the D253-S256 H-bond is sampled frequently (Figure 7A). In monomer-B interactions between the RSB and E123, and between E123 and T127 are less frequently sampled (Figure S28F), and D253 lacks H-bonding with S256 (Figures 7A, 7B).



**Figure 7.** Inter-helical H-bonds in MD simulations of ChR2. (A, B) The RSB connects to E123 in both monomers of all-trans ChR2.(C,D) The 13-cis RSB connects to E82/H134 via H-bonded waters in Monomer-A (C), and to the extracellular side in Monomer-B (panel D).

The TM region of 13-*cis* ChR2 hosts about 40 waters per monomer, as compared to ~25-30 waters in all-*trans* ChR2 (Figure S26 C,D). Internal waters that visit the cytoplasmic half of 13-*cis* ChR2 can bridge the distance of ~12-14Å between the RSB and E82/E83/H134 (Figures 7, S29 E&G, S30). Such a transient water-mediated path is of potential interest for ChR2 function, because H134, which corresponds to the proton donor D96 in bacteriorhodopsin. A water chain between D96 and the RSB is present in a recent structure of a late (N) photocycle intermediate of bacteriorhodopsin [109], and visited transiently in simulations of the D85N mutant [10]. Alternatively, the RSB can H-bond to S256 (Figures 7C&D, 8D, S29E). H-bonding between N258 and S63 occurs ~40-56% of the time when retinal is 13-*cis*, as compared to ~87-96% when retinal is all-*trans*. E123 interacts with T127 and K93, and D253, with K93 (Figures 7D, 8A).

****

**Figure 8**. The dynamics of H-bonds close to the RSB can depend on the isomeric state of the retinal chromophore. (A) The intra-helical H-bond between E123 and T127 is highly stable in both monomers of 13-*cis* ChR2, and in one of the *trans* ChR2 monomers. (B) When all-trans, the RSB prefers to interact with E123 in Monomer-A, and with D253 in Monomer-B. (C) Time series for the frequency with which water molecules bridge the RSB-E83 and to H134, and H134 to E83. (D) Time series of the number of water molecules needed to establish an H-bond path between 13-cis retinal and H134 in Monomer-A.

The finding that H-bond networks at the cytoplasmic side the retinal respond rapidly to changes in the retinal isomeric state, such that H-bond changes are observed on the timescale of relatively short ~200ns simulations, suggests tight coupling between the retinal SB and its immediate protein-water H-bond environment. The rapid assembly of a transient protein-water H-bond path from the RSB to E82/H134 suggests plasticity as anticipated for H-bond dynamic networks.

Taken together, observations above from the two MD simulations of ChR2 indicate that, as noted in previous MD simulations of microbial rhodopsins [10, 110], details of H-bonds near the retinal binding pocket could differ among monomers of homo-oligomeric microbial rhodopsins; likewise, details of some H-bonds near the retinal binding pocket could be distinct in MD vs. crystal structure. Such heterogeneity in H-bond dynamics among monomers, and differences between details of specific H-bonds in the static structure vs. MD, could be due to deficiencies in how atomistic force-fields describe retinal proteins [10, 111], and/or to the limited timescale of the simulations. Moreover, static crystal structures of retinal proteins may have limitations arising from how the refinement software used for X-ray structure modeling describe non-standard interactions such as the -electron system of the retinal, and interactions between the retinal Schiff base and charged protein groups [112].

**Conclusions**

H-bonds and H-bond networks are essential determinants of membrane protein structure and dynamics [31, 113, 114]. In the case of membrane transporters and receptors, H-bonds of internal charged and polar groups are often of central interest in formulating hypotheses about reaction mechanisms. Examples include internal titratable sidechains that could function as proton donor/acceptor groups in proteins that use proton transfer for function, and long-distance H-bond networks in proteins that use allosteric conformational coupling. Knowledge of H-bond motifs that are common to membrane receptors and transporters could help guide considerations of protein sidechains that could be important for protein function, and the choice of amino acid residues for further investigation with site-directed mutagenesis.

The curated data set we used for H-bond analyses contains 683 protein structures, of which 200 were solved at a resolution of 2.5Å or better. An advantage of the large number of structures included in our dataset is that it allows an overview of H-bonds present in static structures of many different proteins. A caveat of such a dataset is that proteins are chemically diverse entities [99], and H-bond dynamics depends on the local hydrophobic packing [10, 115] –such that, e.g., two inter-helical Asp-Thr H-bonds of bacteriorhodopsin, D96-T46 and D115-T90, have distinct dynamics [10]. Moreover, membrane proteins of the same superfamily may have distinct functions and different conformational dynamics: The large superfamily of Rhodopsin-like receptors and pumps includes GPCRs, known to have significant conformational dynamics [91], and microbial rhodopsins – which can function as ion channels, ion pums, or as sensors, and tend to be much more rigid.

The protocol we presented here allowed us to evaluate efficiently the presence of H-bonds in membrane protein structures. We evaluated, for transmembrane regions of each static protein structure, the presence of ten types of H-bonds, which we selected based on previous observations of H-bonds at sites important for the functioning of model proteins. To find out whether H-bonds we identified could be part of long-distance H-bond networks of potential interest for conformational coupling, we searched for the local H-bond clusters that include H-bonds of interest, and computed linear path lengths that characterize the spatial extent of the clusters.

With the protocol we present and the data sets we assembled for static protein structures, we found that structures from almost all protein families have an H-bond motif whereby the same Ser/Thr hydroxyl group H-bonds to an Asp/Glu sidechain and to a backbone carbonyl group (Figures 1, 2A, 2B). When present, such Asp/Glu-Ser/Thr-backbone carbonyl H-bonds tend to locate in the transmembrane core of the protein (Figure 3), and may participate in more extended H-bond clusters (Figure 4A). Such H-bonds could be particularly important for the protein conformational dynamics: intra-helical H-bonds between Ser/Thr hydroxyl groups and backbone carbonyl groups could contribute to an enhanced local dynamics of the helix [80], and inter-helical Asp/Glu-Ser/Thr H-bonds could help couple protonation change with protein conformational dynamics [11].

Whereas only a subset of protein structures have H-bonds between Asn and Asp sidechains, many proteins have Asp-Ser/Thr and Asn-Ser/Thr H-bonds (Figure 2). To the extent that an Asn sidechain resembles Asp H-bonding, the findings on Asn and Asp H-bonding could be interpreted to suggest that Asp-Ser H-bonding may occur for both charged and neutral Asp sidechains. Pairs of H-bonded His sidechains are rarely present, which could be due to the relatively small frequency of His sidechains being present within the transmembrane core of membrane proteins, and/or to limitations of the protein datasets we used.

At room temperature in a fluid lipid membrane environment, protein H-bonds which are part of local H-bond clusters can break and reform rapidly, such that H-bonds other than those indicated by the static structure can be sampled, whereas other H-bonds present in the crystal structure might be sampled infrequently; such complex dynamics can be difficult to anticipate based on inspection of static protein structures [31]. Most of the H-bonds we investigated for the crystal structure of Aqy1 were robust during dynamics at room temperature, being sampled in all four Aqy1 monomers for two different protonation states of two His sidechains, and in two different lipid membrane environments. By contrast, selected H-bonds near the retinal chromophore of ChR2 sampled somewhat different dynamics in the two protein copies of the homo-dimer; moreover, the dynamics of active-site H-bonds were altered when the geometry of the chromophore was changed. These results from MD simulations of ChR2 could be interpreted to suggest a potentially reduced robustness of H-bonds observed in static structures of proteins that bind cofactor molecules. For such proteins, accurate description of H-bonding may require improved representation of cofactors in software used for crystallographic refinement, and in force field simulations.

The data sets of static structures provide an overview of H-bond motifs in a relatively large number of static protein structures. The methodology we implemented here could be extended to other H-bond types and protein data sets of interest to derive a more complete picture of relationships between H-bonds and H-bond networks in static protein structures and room temperature simulations. The protocol we used to eliminate multiple structure entries of the same protein accounts for redundancies in protein structure entries; a future improvement of the protocol could include an automated sequence alignments such that the length of the protein construct used to solve the structure is accounted for, and only the structure of the longest, i.e., most complete construct, is considered.

Our analyses focused on ten types of H-bonds typical for membrane proteins. Other types of interactions of potential interest include weak CH…O interactions, important for protein-protein interactions [116], CH… interactions, thought to be important for protein stability [117], water-mediated H-bond bridges in membrane receptors such as GPCRs [15, 118, 119], lipid-protein interactions [120, 121], and interactions between membrane proteins and ligands.

**Acknowledgements**

A-NB acknowledges support in part by the Excellence Initiative of the German Federal and State Governments via the Freie Universität Berlin and from the German Research Foundation DFG SFB 1078, Project C4. Computing time was provided by computing allocations from the HLRN, the North-German Supercomputing Alliance, and by the CURTA High-Performance Computing System of ZEDAT of the Freie Universität. LSB acknowledges support by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC, discovery grant RGPIN-2018-04397) and the University of Guelph. ML and A-NB thank Jens Dreger (FU Berlin), Boris Proppe (ZEDAT HPC, FU Berlin), and Christian Tuma (HLRN) for excellent technical support. ML acknowledges Krzysztof Buzar for valuable technical discussions during early stages of this work.

**References**

[1] M. Damaghi, J.W. Wojktowiak, R.J. Gillies, pH sensing and regulation in cancer, Frontiers in Physiology, 4 (2013) 1-10.

[2] A. Schönichen, B.A. Webb, M.P. Jacobson, D.L. Barber, Considering protonation as a posttranslational modification regulating protein structure and function, Annu. Rev. Biophys., 42 (2013) 289-314.

[3] R.A. Thurlkill, G.R. Grimsley, J.M. Scholtz, C.N. Pace, pK values of the ionizable groups of proteins, Prot. Sci., 15 (2006) 1214-1218.

[4] J. Srivastava, D.L. Barber, M.P. Jacbson, Intracellular pH sensors: design principles and functional significance, Physiol., 22 (2007) 30-39.

[5] R.A. Thurlkill, G.R. Grimsley, J.M. Schlotz, C.N. Pace, Hydrogen bonding markedly reduces the pK of burried carboxyl groups in proteins, J. Mol. Biol., 362 (2006) 594-604.

[6] N. Grytsyk, J. Sugihara, H.R. Kaback, P. Hellwig, pKa of Glu325 in LacY, PNAS, 114 (2017) 1530-1535.

[7] S.P. Balashov, R. Govindjee, E.S. Imasheva, S. Misra, T.G. Ebrey, Y. Feng, R.K. Crouch, D.R. Menick, The two pKa's of aspartate-85 and control of thermal isomerization and proton release in the arginine-82 to lysine mutant of bacteriorhodopsin, Biochemistry, 34 (1995) 8820-8834.

[8] S.P. Balashov, E.S. Imasheva, R. Govindjee, T.G. Ebrey, Titration of aspartate-85 in bacteriorhodopsin: what is says about chromophore isomerization and proton release, Biophys. J., 70 (1996) 473-481.

[9] M.S. Braiman, A.K. Dioumaev, J. Lewis, A large photolysis-induced pKa increase of the chromophore counterion in bacteriorhodopsin: Implications for ion transport mechanisms of retinal proteins, Biophys. J., 70 (1996) 939-947.

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

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

[12] A.-N. Bondar, Proton-binding motifs of membrame-bound proteins: from bacteriorhodopsin to spike protein S, Frontiers in Chemistry, 9 (2021) 685761.

[13] S. Khademi, J. O'Connel, J. Remis, Y. Robles-Colmenares, L.J. Miercke, R.M. Stroud, Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35Å, Science, 305 (2004) 1587-1594.

[14] A.-N. Bondar, H.J. Lemieux, Reactions at membrane interfaces, Chem. Rev., 119 (2019) 6162-6183.

[15] É. Bertalan, S. Lešnik, U. Bren, A.-N. Bondar, Protein-water hydrogen-bond networks of G protein-coupled receptors: Graph-based analyses of static structures and molecular dynamics, J. Struct. Biol., 212 (2020) 107634.

[16] B.P. Pedersen, M.J. Buch-Pedersen, J.B. Morth, M.G. Palmgren, P. Nissen, Crystal structure of the plasma membrane proton pump, Nature, 450 (2007) 1111-1114.

[17] F. Guerra, A.-N. Bondar, Dynamics of the plasma membrane proton pump, J. Membr. Biol., 248 (2015) 443-453.

[18] J.E. Besaw, W.-L. Ou, T. Morizumi, B.T. Eger, J.D. Sanchez Vasquez, J.H.Y. Chi, A. Harris, L.S. Brown, R.J.D. Miller, O.P. Ernst, The crystal structues of a chloride-pumping microbial rhodopsin and its proton-pumping mutant illuminate proton transfer determinants, J. Biol. Chem., 295 (2020) 14793-14804.

[19] A.-N. Bondar, Mechanisms of long-distance allosteric couplings in proton-binding membrane transporters, Advances in Protein Chemistry and Structural Biology, 128 (2022) 199-239.

[20] M.-G. Ludwig, M. Vanek, D. Guerini, J.A. Gasser, C.E. Jones, U. Junker, H. Hofstetter, R.M. Wolf, K. Seuwen, Proton-sensing G-protein-coupled receptors, Nature, 425 (2003) 93-98.

[21] R. Fu, Y. Miao, H. Qin, T.A. Cross, Observation of the imidazole-imidazolium hydrogen bonds responsible for selective proton conductance in the influenza A M2 channel, J. Am. Chem. Soc., 142 (2020) 2115-2119.

[22] M.W. Hunkapiller, M.D. Forgac, J.H. Richards, Mechanism of action of serine proteases: tetrahedral intermediate and concerted proton transfer, Biochemistry, 15 (1976) 5581-5588.

[23] J.J. Birktoft, L.J. Banaszak, The presence of a histidine-aspartic acid pair in the active site of 2-hydroxyacid dehydrogenases, J. Biol. Chem., 258 (1983) 472-482.

[24] M. Gondry, F. Lederer, Functional properties of the histidine-aspartate ion pair of flavocytochrome b2 (L-lactate dehydrogenase): Substitution of Asp282 with asparagine, Biochemistry, 35 (1996) 8587-8594.

[25] A.A. Kossiakoff, S.A. Spencer, Direct determination of the protonation states of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine proteases: neutron structure of trypsin, Biochemistry, 20 (1981) 6462-6474.

[26] S. Hirschi, D. Kalbermatter, Z. Ucurcum, T. Lemmin, D. Fotiadis, Cryo-EM structure and dynamics of the green-light absorbin proteorhodopsin, Nature Comm., 12 (2021) 4107.

[27] H. Luecke, B. Schobert, J. Stagno, E.S. Imasheva, J.M. Wang, S.P. Balashov, J.K. Lanyi, Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore, Proc. Natl. Acad. Sci. USA, 105 (2008) 16561-16565.

[28] T. Morizumi, W.-L. Ou, N. van Eps, K. Inoue, H. Kandori, L.S. Brown, O.P. Ernst, X-ray crystallographic structure and oligomerization of *Gloeobacter* rhodopsin, Scientific Reports, 9 (2019) 11283.

[29] A. Harris, M. Ljumovic, A.-N. Bondar, Y. Shibata, S. Ito, K. Inoue, H. Kandori, L.S. Brown, A new group of eubacterial light-driven retinal-binding proton pumps with an unusual cytoplasmic proton donor, Biochim. Biophys. Acta, 1847 (2015) 1518-1529.

[30] S. Rajagopal, S. Vishveshwara, Short hydrogen bonds in proteins, FEBS J., 272 (2005) 1819-1832.

[31] A.-N. Bondar, S.H. White, Hydrogen bond dynamics in membrane protein function, Biochim. Biophys. Acta, 1818 (2012) 942-950.

[32] U.K. Eriksson, G. Fischer, R. Friedmann, G. Enkavi, E. Tajkhorshid, R. Neutze, Subangstrom resolution X-ray structure details aquaporin-water interactions, Science, 340 (2013) 1346-1349.

[33] M. Lomize, I.D. Pogozheva, H. Joo, H.I. Mosberg, A.L. Lomize, OPM database and PPM server: resources for positioning of proteins in membranes, Nucleic Acid Research, 40 (2011) D370-D376.

[34] H.M. Berman, J. Westbrook, G. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acid Research, 28 (2000) 235-242.

[35] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, Journal of Computational Chemistry, 29 (2008) 1859-1865.

[36] M.H. Saier Jr., V.S. Reddy, D.G. Tamang, Å. Västermark, The transporter classification database, Nucleic Acids Research, 42 (2013) D251-D258.

[37] M. Siemers, M. Lazaratos, K. Karathanou, F. Guerra, L.S. Brown, A.-N. Bondar, Bridge: A graph-based algorithm to analyze dynamic H-bond networks in membrane proteins, Journal of Chemical Theory and Computation, 15 (2019) 6781-6798.

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

[39] A.S. Fokas, D.J. Cole, S.E. Ahnert, A.W. Chin, Residue geometry networks: a rigidity-based approach to the amino acid network and evolutionary rate, Sci. Rep., 6 (2016) 33213.

[40] K. Karathanou, A.-N. Bondar, Using graphs of dynamic hydrogen-bond networks to dissect conformational coupling in a protein motor, Journal of Chemical Information and Modeling, 59 (2019) 1882-1896.

[41] S. Krishnamurthy, N. Elefteriadis, K. Karathanou, J.H. Smit, A.G. Portaliou, K.E. Chatzi, S. Karamanou, A.-N. Bondar, G. Gouridis, A. Economou, A nexus of intrinsic dynamics underlies translocase priming, Structure, 29 (2021) 846-858.e847.

[42] L.C. Freeman, A set of measures of centrality based on betweenness, Sociometry, 40 (1977) 35-41.

[43] L.C. Freeman, Centrality in social networks. Conceptual clarification, Social Networks, 1 (1979) 215-239.

[44] U. Brandes, A faster algorithm for betweenness centrality, Journal of Mathematical Sociology, 25 (2001) 163-177.

[45] K. Karathanou, M. Lazaratos, E. Bertalan, M. Siemers, K. Buzar, G.F.X. Schertler, C. del Val, A.-N. Bondar, A graph-based approach identifies dynamic H-bond communication networks in spike protein S of SARS-CoV-2, J. Struct. Biol., 212 (2020) 107617.

[46] N. Hasegawa, H. Jonotsuka, K. Miki, K. Takeda, X-ray structure analysis of bacteriorhodopsin at 1.3Å resolution, Scientific Reports, 8 (2018) 13123.

[47] O. Volkov, K. Kovalev, V. Polovinkin, V. Borschchevskiy, C. Bamann, R. Astashkin, E. Marin, A. Popov, T. Balandin, D. Willbold, G. Büldt, E. Bamberg, V. Gordeliy, Structural insights into ion conduction by channelrhodopsin 2, Science, 358 (2017) eaan8862.

[48] S.-M. Lin, J.-Y. Tsai, H. C.-D., C.-L. Huang, C.-L. Chiu, M.-H. Liu, J.-Y. Tung, T.-H. Liu, R.-L. Pan, Y.-J. Sun, Crystal structure of a membrane-embedded H+-translocating pyrophosphatase, Nature, 484 (2012) 399-403.

[49] J.F. Bada Juarez, P.J. Judge, S. Adam, D. Axford, J. Vinals, J. Birch, T.O.C. Kwan, K.H. JHoi, H.-Y. Yen, A. Vial, P.-E. Milhiet, C.V. Robinson, I. Schapiro, I. Moraes, A. Watts, Structures of the archaerhodopsin-3 transporter reveal that disordering of internal water networks underpins receptor sensitization, Nature Comm., 12 (2021) 629.

[50] K. Kovalev, D. Volkov, R. Astashkin, A. Alekseev, I. Guschkin, J.M. Haro.Moreno, I. Chizhov, S. Siletsky, M. Mamedov, A. Rogachev, T. Baladin, V. Borschchevskiy, A. Popov, G. Bourenkov, E. Bamberg, F. Rodriguez-Valera, G. Büldt, V. Gordeliy, High-resolution structural insights into the heliorhodopsin family, Proc. Natl. Acad. Sci., 117 (2020) 4131-4141.

[51] B. van den Berg, A. Chembath, D. Jefferies, A. Basle, S. Khalid, J.C. Rutherford, Structural basis for Mep2 ammonium trnsceptor activation by phosphorylation, Nature Comm., 7 (2015) 11337.

[52] K. Kovalev, V. Polovinkin, I. Gushchin, A. Alekseev, V. Shevchenko, V. Borschchevskiy, R. Astashkin, T. Balandin, D. Bratanov, S. Vaganova, A. Popov, V. Chupin, E. Bamberg, V. Gordeliy, Structure and mechanisms of sodium-pumping KR2 rhodopsin, Science Advances, 5 (2019) eaav2671.

[53] V.I. Gordeliy, J. Labahn, R. Moukhametzianov, R. Efremov, J. Granzin, R. Schlesinger, G. Büldt, A.J. Scheidig, J.P. Klare, M. Engelhardt, Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex, Nature, 419 (2002) 484-487.

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

[55] B.R. Brooks, C.L.I. Brooks, A.D. MacKerell Jr., L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.D. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: the biomolecular simulation program, J. Comput. Chem, 30 (2009) 1545-1614.

[56] W. Humphrey, W. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph., 14 (1996) 33-38.

[57] K. Grillitisch, P. Tarazona, L. HKlug, T. Weiessnegger, G. Zellnig, E. Leitner, I. Feussner, G. Daum, Isolation and characterization of the plasma membrane from the yeast *Pichia pastoris*, Biochim. Biophys. Acta, 1838 (2014) 1889-1897.

[58] A. Fiser, R.K. Do, A. Sali, Modeling of loops in protein structures, Protein Science, 9 (2000) 1753-1773.

[59] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol., 234 (1993) 779-815.

[60] M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes, Annu. Rev. Biomol. Struct., 29 (2000) 291-325.

[61] M. Nack, I. Radu, M. Gossing, C. Bamann, E. Bamberg, G.F. von Mollard, J. Heberle, The DC gate in channelrhodopsin-2: crucial hydrogen bonding interaction between C128 and D156, Photochem. Photobiol. Sci., 9 (2010) 194-198.

[62] K. Eisenhauer, J. Kuhne, E. Ritter, A. Berndt, S. Wolf, E. Freier, F. Bartl, P. Hegemann, K. Gerwert, In channelrhodopsin-2 Glu-90 is crucial for ion selectivity and is deprotonated during the photocycle, J. Biol. Chem., 287 (2012) 6904-6911.

[63] E. Ritter, K. Stehfest, A. Berndt, P. Hegemann, F.J. Bartl, Monitoring light-induced structural changes of channlerhodopsin-2 by UV-visible and Fourier transform infrared spectroscopy, J. Biol. Chem., 283 (2008) 35033-35041.

[64] J. Kuhne, K. Eisenhauer, E. Ritter, P. Hegemann, K. Gerwert, F. Bartl, Early formation of the ion-conducting pore in channelrhodopsin-2, Angew. Chem. Int. Ed., 54 (2015) 4953-4957.

[65] V.A. Lórenz-Fonfría, V. Muders, R. Schlesinger, J. Heberle, Changes in the hydrogen-bonding strength of water molecules and cysteine residues in the conductive state of channelrhodopsin-1, J. Chem. Phys., 141 (2014) 22D507.

[66] J.C. Phillips, B. Braun, W. Wang, J. Gumbart, E. Takjkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem, 26 (2005) 1781-1802.

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

[68] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys., 103 (1995) 8577-8593.

[69] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant-pressure molecular-dynamics algorithms, J. Chem. Phys., 101 (1994) 4177-4189.

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

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

[72] A.D. MacKerell Jr., D. Bashford, M. Bellot, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E.I. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B, 102 (1998) 3586-3616.

[73] A.D. MacKerell Jr., M. Feig, C.L.I. Brooks, 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 (2004) 1400-1415.

[74] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Votrobyov, A.D. MacKerell Jr., R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, J. Phys. Chem. B, 114 (2010) 7830-7843.

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

[76] K. Xie, T. Hessa, S. Seppälä, M. Rapp, G. von Heijne, R.E. Dalbey, Features of transmembrane segments that promote the lateral release from the translocase into the lipid phase, Biochemistry, 46 (2007) 15153-15161.

[77] J.P. Ulmschneider, M.S.P. Sansom, Amino acid distributions in integral membrane protein structures, Biochim. Biophys. Acta, 1512 (2001) 1-14.

[78] L.G. Presta, G.D. Rose, Helix signals in proteins, Science, 240 (1988) 1632-1641.

[79] T.M. Gray, B.W. Matthews, J. Mol. Biol., 175 (1984) 75-81.

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

[81] N. Enami, K. Yoshimura, M. Murakami, H. Okumura, K. Ihara, T. Kouyama, Crystal structures of archaeorhodopsin-1 and -2: common structural motif in archaeal light-driven proton pumps, J. Mol. Biol., 358 (2006) 675-685.

[82] T. Kouyama, R. Fujii, S. Kanada, T. Nakanishi, S.K. Chan, M. Murakami, Structure of archaeorhodopsin-2 and 1.8Å resolution, Acta Cryst., D70 (2014) 2692-2701.

[83] S.K. Chan, T. Kitajima-Ihara, R. Fujii, T. Gotoh, M. Murakami, K. Ihara, T. Kouyama, Crystal structure of cruxrhodopsin-3 from *Haloarcula vallismortis*, PLoS One, 9 (2014) e108362.

[84] R. Fudim, M. Szczepek, J. Vierock, A. Vogt, A. Schmidt, G. Kleinau, P. Fischer, F. Bartl, P. Scheerer, P. Hegemann, Design of a light-gated proton channel based on the crystal structure of Coccomyxarhodopsin, Sci. Signal., 12 (2019).

[85] T. Hayashi, S. Yasuda, K. Suzuki, T. Akiyama, K. Kanehara, K. Kojima, M. Tanabe, T. Kato, T. Senda, Y. Sudo, T. Murata, M. Kinoshita, How does a microbial rhodopsin RxR realize its exceptionally high thermostability with the proton-pumping function being retained?, J. Phys. Chem. B, 124 (2020) 990-1000.

[86] A. Varma, E. Mutt, J. Mühle, V. Panneels, A. Terakita, X. Deupi, P. Nogly, G.X. Schertler, E. Lesca, Crystal structure of jumping spider rhodopsin-1 as a light sensitive GPCR, Proc. Natl. Acad. Sci., 116 (2019) 14547-14556.

[87] S. Schulz, M. Iglesias-Cans, A. Krah, O. Yildiz, V. Leone, D. Matthies, G.M. Cook, J.D. Faraldo.Gomez, T. Meier, A new type of Na(+)-driven ATP synthase membrane rotor with a two-carboxylate ion-coupling motif, PLoS Biol., 11 (2013) 01596.

[88] D. Matthies, W. Zhou, A.L. Klyszejko, C. Anselmi, O. Yildiz, K. Brandt, V. Muller, J.D. Faraldo-Gomez, T. Meier, High-resolution structure and mechanism of an F/V-hybrid rotor ring in a Na+-coupled ATP synthase, Nature Comm., 5 (2014) 5286.

[89] T. Murata, I. Yamato, Y. Kakinuma, A.G.W. Leslie, J.E. Walker, Structure of the rotor of the vacuolar-type Na-ATPase from *Enterococcus hirae*, Science, 308 (2005) 654-659.

[90] T.M. Gamal El-Din, M.J. Lenaeus, K. Ramanadane, N. Zheng, W.A. Catterall, Molecular dissection of multiphase inactivation of the bacterial sodium channel, J. Gen. Physiol., 151 (2019) 174-185.

[91] X. Deupi, B. Kobilka, Energy landscapes as a tool to integrate GPCR structure, dynamics, and function, Physiol., 25 (2010) 293-303.

[92] M.B. Ulmschneider, M.S.P. Sansom, A. Di Nola, Properties of integral membrane protein structures: Derivation of an implicit membrane potential, Proteins: Structure, Function, and Bioinformatics, 252-265 (2005).

[93] S. Wang, E.A. Orabi, S. Baday, S. Bernèche, G. Lamoreux, Ammonium transporters achieve charge transfer by fragmenting their substrate, J. Am. Chem. Soc., 134 (2012) 14019-10427.

[94] I. Vasiliauskaité-Brooks, R. Sounier, P. Rochaisx, G. Bellot, M. Fortier, F. Hoh, L. De Colibus, C. CBechara, E.M. Saied, C. Arenz, C. Leyrat, S. Grenier, Structural insight into adiponectin receptors suggest ceramidse activity, Nature, 544 (2016) 120-123.

[95] T. Pflüger, C.F. Hernández, P. Lewe, F. Frank, H. Mertens, D.I. Svergun, M.W. Baumstark, V.Y. Lunin, M.S.M. Jetten, S.L.A. Andrade, Signaling ammonium across membranes through an ammonium sensor histidine kinase, Nature Comm., 9 (2018) 164.

[96] M. Fischer, Macromolecular room temperature crystallography, Quarterly Reviews of Biophysics, 54 (2021) e1.

[97] J.S. Fraser, H. van den Bedem, A.J. Samelson, P.T. Lang, J.M. Holton, N. Echols, T. Alber, Accessing protein conformational ensembles using room-temperature X-ray crystallography, Proc. Natl. Acad. Sci. USA, 108 (2011) 16247-16252.

[98] S.Y.C. Bradford, L. El Khoury, Y. Ge, M. Osato, D.L. Mobley, M. Fischer, Temperature artifacts in protein structures bias ligand-binding predictions, Chem. Sci., 12 (2021) 11275.

[99] R. Jaenicke, Protein structure and function at low temperatures, Phil. Trans. R. Soc. Lond. B, 326 (1990) 535-553.

[100] M. Gottschalk, N.A. Dencher, B. Halle, Microsecond exchange of internal water molecules in bacteriorhodopsin, Journal of Molecular Biology, 311 (2001) 605-621.

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

[102] C. Aponte-Santamaría, R. Briones, A.D. Schenk, T. Walz, B. de Groot, Molecular driving forces defining lipid positions around aquaporin-0, Proc. Natl. Acad. Sci., 109 (2012) 9887-9892.

[103] M.F. Brown, Modulation of rhodopsin function by properties of the membrane bilayer, Chemistry and Physics of Lipids, 73 (1994) 159-180.

[104] P. Hakizimana, M. Masureel, B. Gbaguidi, J.-M. Ruysschaert, C. Govaerts, Interactions between phosphatidylethanolamine headgroup and LmrP, a multidrug transporter, J. Biol. Chem., 283 (2008) 9369-9376.

[105] S. Urban, M.S. Wolfe, Reconstitution of intramembrane proteolysis *in vitro* reveals that pure rhomboid is sufficient for catalysis and specificity, Proc. Natl. Acad. Sci. USA, 102 (2005) 1883-1888.

[106] A.-N. Bondar, C. del Val, S.H. White, Rhomboid protease dynamics and lipid interactions, Structure, 17 (2009) 395-405.

[107] A.-N. Bondar, Mechanisms by which lipids influence conformational dynamics of the GlpG intramembrane rhomboid protease, J. Phys. Chem. B, 123 (2019) 4159-4172.

[108] V.A. Lórenz-Fonfría, T. Resler, N. Krause, M. Nack, M. Gossing, G.F. von Mollard, C. Bamann, E. Bamberg, R. Schlesinger, J. Heberle, Transient protonation changes in channelrhodpsin-2 and their relevance to channel gating, Proc. Natl. Acad. Sci. USA, 110 (2013) 1273-1281.

[109] T. Weinert, P. Skopintsev, D. James, F. Dworkowski, E. Panepucci, D. Kekilli, A. Furrer, S. Brünle, S. Mous, D. Ozerov, P. Nogly, M. Wang, J. Standfuss, Proton uptake mechanism in bacteriorhodopsin captured by serial synchrothron crystallography, Science, 365 (2019).

[110] S. Adam, A.-N. Bondar, Mechanism by which water and protein electrostatic interactions control proton transfer at the active site of channelrhodopsin, Plos One, 13 (2018) e0201298.

[111] A.-N. Bondar, M. Knapp-Mohammady, S. Suhai, S. Fischer, J.C. Smith, Ground-state properties of the retinal molecule: from quantum mechanical to classical mechanical computations of retinal proteins, Theor. Chem. Acc., 130 (2011) 1169-1183.

[112] T. Okada, M. Sugihara, A.-N. Bondar, M. Elstner, P. Entel, V. Buss, The retinal configuration and its environment in rhodopsin in light of a new 2.2 Å crystal structure, J. Mol. Biol., 342 (2004) 571-583.

[113] N.H. Joh, A. Min, S. Faham, J.P. Whitelegge, D. Yang, V.L. Woods Jr., J.U. Bowie, Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins Nature, 453 (2008) 1266-1270.

[114] S.H. White, W.C. Wimley, Membrane protein folding and stability: physical principles, Annu. Rev. Biomol. Struct., 28 (1999) 319-365.

[115] G. Nemethy, I.Z. Steinberg, H.A. Scheraga, Influence of water structure and of hydrophobic interactions on the strength of side-chain hydrogen bonds in proteins, Biopolymers, 1 (1963) 43-69.

[116] L. Jiang, L. Lai, CH...O hydrogen bonds at protein-protein interfaces, J. Biol. Chem., 277 (2002) 37732-37740.

[117] M. Brandl, M.S. Weiss, A. Jabs, J. Sühnel, R. Hilgenfeld, J. Mol. Biol., 307 (2001) 357-377.

[118] E. Bertalan, E. Lesca, G.F.X. Schertler, A.-N. Bondar, C-Graphs tool with graphical user interface to dissect conserved hydrogen-bond networks: applications to visual rhodopsins, Journal of Chemical Information and Modeling, 61 (2021) 5692-5707.

[119] A.J. Venkatakrishnan, A.K. Ma, R. Fonseca, N.R. Latorraca, B. Kelly, R.M. Betz, C. Asawa, B.K. Kobilka, R.O. Dror, Diverse GPCRs exhibit conserved water networks for stabilization and activation, Proc. Natl. Acad. Sci., 116 (2019) 3288-3293.

[120] A. Ben-Shem, D. Fass, E. Bibi, Structural basis for intramembrane proteolysis by rhomboid serine proteases, Proc. Natl. Acad. Sci. USA, 104 (2007) 462-466.

[121] H. Palsoditir, C. Hunte, Lipids in membrane protein structures, Biochim. Biophys. Acta, 1666 (2004) 2-18.

**Graphical Abstract**

