Biochimica et Biophysica Acta 1848 (2015) 1436–1449

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Intra-helical salt-bridge and helix destabilizing residues within the same helical turn: Role of functionally important loop E half-helix in channel regulation of major intrinsic proteins

Ravi Kumar Verma [a], Neel Duti Prabh [a], Ramasubbu Sankararamakrishnan [a,b,*]

[a] Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur 208016, India
[b] Centre of Excellence for Chemical Biology, Indian Institute of Technology Kanpur, Kanpur 208016, India

ARTICLE INFO

Article history:
Received 27 August 2014
Received in revised form 8 March 2015
Accepted 13 March 2015
Available online 19 March 2015

Keywords:
Helix stability
Loop flexibility
Aquaporin
Glycine and proline in transmembrane helix
Regulation of water channel

ABSTRACT

The superfamily of major intrinsic proteins (MIPs) includes aquaporin (AQP) and aquaglyceroporin (AQGP) and it is involved in the transport of water and neutral solutes across the membrane. Diverse MIP sequences adopt a unique hour-glass fold with six transmembrane helices (TM1 to TM6) and two half-helices (LB and LE). Loop E contains one of the two conserved NPA motifs and contributes two residues to the aromatic/arginine selectivity filter. Function and regulation of majority of MIP channels are not yet characterized. We have analyzed the loop E region of 1468 MIP sequences and their structural models from six different organism groups. They can be phylogenetically clustered into AQGPs, AQPs, plant MIPs and other MIPs. The LE half-helix in all AQGPs contains an intra-helical salt-bridge and helix-breaking residues Gly/Pro within the same helical turn. All non-AQGPs lack this salt-bridge but have the helix destabilizing Gly and/or Pro in the same positions. However, the segment connecting LE half-helix and TM6 is longer by 10–15 residues in AQGPs compared to all non-AQGPs. We speculate that this longer loop in AQGPs and the LE half-helix of non-AQGPs will be relatively more flexible and this could be functionally important. Molecular dynamics simulations on glycerol-specific GlpF, water-transporting AQP1, its mutant and a fungal AQP channel confirm these predictions. Thus two distinct regions of loop E, one in AQGPs and the other in non-AQGPs, seem to be capable of modulating the transport. These regions can also act in conjunction with other extracellular residues/segments to regulate MIP channel transport.

1. Introduction

Channels constitute an important group of integral membrane proteins that facilitate the transport of molecules and ions across the membranes. They are highly selective and are especially involved in efficient transport of selected molecules or ions. The opening and closing of channels are governed by several factors including phosphorylation or cation regulation. One of the largest groups of channels that transport neutral solutes is the superfamily of major intrinsic proteins (MIPs) [1]. Aquaporins (AQPs) and aquaglyceroporins (AQGPs) are the prototype members of this superfamily [2–5]. These channels are highly selective and are involved in the transport of water, glycerol and other neutral solutes. The gating mechanism of MIP channels has been investigated for several members of the family using biochemical, biophysical, structural and simulation studies [6–9]. Available results show that post-translational modifications [10], interactions with metal ions [11], inhibition by selected drugs [12], lipid environment [13] and protein–protein interactions [14] are some of the factors which can regulate and influence the function of MIP channels. Structural studies have also demonstrated that the conformation of specific loops can regulate the transport activity of some MIP channels [15,16].

Three-dimensional structures of several MIP structures have been determined. More than 20 MIP structures have been deposited in the Protein Data Bank [17]. They include those from mammalian (AQP0, AQP1, AQP2, AQP4 and AQP5), archaeal (AqpM), *Escherichia coli* (AqpZ and GlpF), spinach (SoPIP2;1), yeast (Aqy1) and *Plasmodium falciparum* (PfAQP). Although the sequences are diverse, MIPs from different organisms with different transport properties adopt a unique hour-glass helical fold [18]. The helical bundle is formed by six transmembrane helices (TM1 to TM6) connected by the loops LA to LE. The loops LB (connecting TM2 and TM3) and LE (linking TM5 and TM6) form half-helices and dip into the membrane from opposite directions to form a seventh pseudo-helix (Fig. 1a). LB and LE also possess the highly conserved NPA motifs at the meeting point of the two half-helices. MIP structures have a narrow selectivity filter region formed by four residues near the extracellular side. For the formation of this aromatic/arginine

Abbreviations: MIP, major intrinsic proteins; AQP, aquaporin; AQGP, aquaglyceroporin; PIP, plasma membrane intrinsic proteins; TIP, tonoplast intrinsic proteins; NIP, nodulin 26-like intrinsic proteins; SIP, small and basic intrinsic proteins; XIP, X-intrinsic proteins; MD, molecular dynamics

* Corresponding author at: Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur 208016, India. Tel.: +91 512 2594014; fax: +91 512 2594010.
E-mail address: rsankar@iitk.ac.in (R. Sankararamakrishnan).

http://dx.doi.org/10.1016/j.bbamem.2015.03.013
0005-2736/

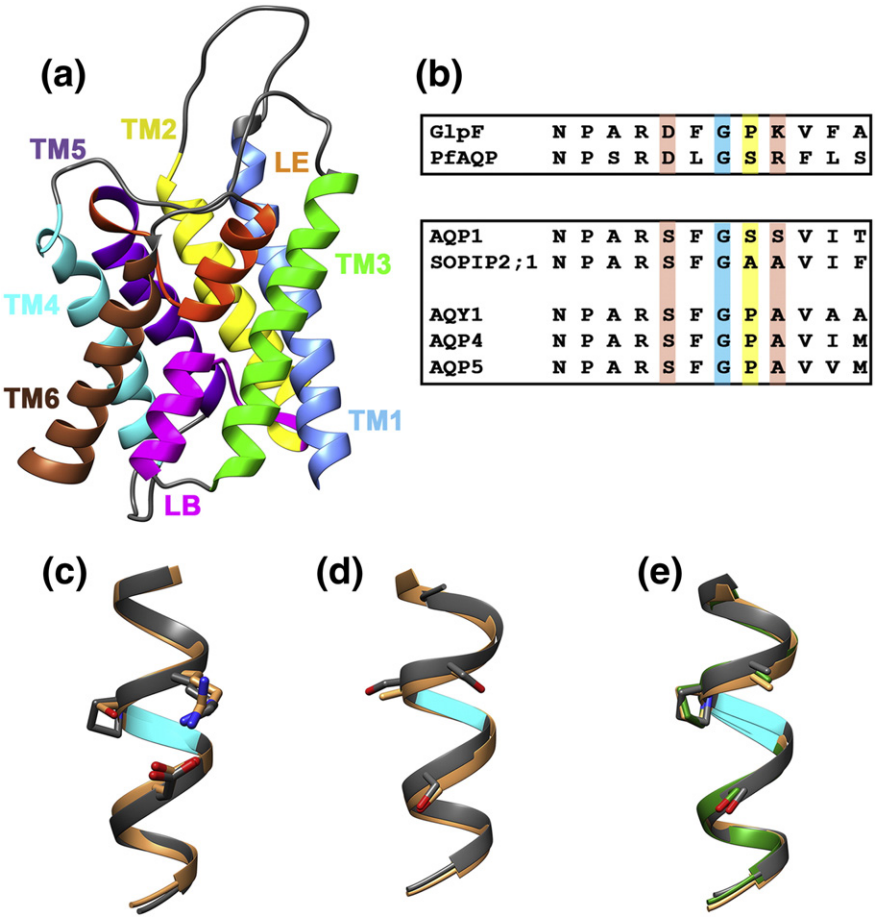

Fig. 1. (a) Structure of a water AQP channel (PDB ID: 1J4N) with a typical hour-glass MIP helical fold. All six transmembrane helices (TM1 to TM6) and the two half-helices (LB and LE) are shown in different colors. Loops connecting the helical segments are displayed in gray. (b) Alignment of sequences in the LE half-helical region in some of the AQGP and AQP channels for which the structures have been determined: GlpF — glycerol facilitator from *E. coli*; PfAQP — *P. falciparum* aquaporin; Aqy1 — yeast aquaporin; SoPIP2;1 — spinach aquaporin; AQP1, AQP4 and AQP5 — mammalian aquaporins. The positions of acidic and basic residues which form intra-helical salt-bridge interaction in AQGP channels and the equivalent positions in AQP channels are shown in brown background. Positions of helix-breaking residues Gly and Pro are displayed respectively in cyan and yellow background in both AQGP and AQP channels. (c) Structural superposition of LE half-helical region of glycerol-conducting GlpF (PDB ID: 1FX8) and PfAQP (PDB ID: 3C02). (d) Superposition of LE half-helices from AQP1 (PDB ID: 1J4N) and SoPIP2;1 (PDB ID: 1Z98). Side-chains of residues at equivalent positions of acidic, basic and Pro residues are substituted by small residues and are displayed in stick representation. (e) LE half-helices from Aqy1 (PDB ID: 2W2E), AQP4 (PDB ID: 3GD8) and AQP5 (DPB ID: 3D9S) are superposed. Side-chains of acidic and basic residues that participate in the intra-helical salt-bridge in AQGP channels are shown and the residues at the equivalent positions are displayed in AQP channels. Side-chain of Pro is also displayed in (c) and (e). Position of Gly within the same helical turn is represented in cyan color.

(Ar/R) selectivity filter, TM2 and TM5 contribute one residue each and the remaining two come from the loop LE. Both the conserved NPA motifs and the Ar/R selectivity filter have been shown to be important for the selectivity of the solutes to be transported [19–22].

In addition to loops LB and LE, structural studies of some MIPs have implicated loop LD in MIP channel gating [16]. LD which connects TM4 and TM5 has been shown to undergo conformational rearrangements and act as a plug to close the channel in spinach aquaporin SoPIP2;1. Loop LD conformation is stabilized by a metal ion and a network of ionic and hydrogen bond interactions. Phosphorylation and pH are also the factors that drive the conformational changes resulting in a transition from closed to open state and vice versa [16,23]. Conformational changes have been reported to be responsible for the gating of several channels. For example, binding of ligands induce conformational changes in GABA receptors that help in activating the channels [24]. Molecular dynamics (MD) simulations exhibit intrinsic flexibility of channel lining helices and this property is attributed to the conformational transitions of the mammalian inward rectifier K+ channels [25]. Reorganization of transmembrane helices and rigid body rotation of an extracellular domain are the important changes observed in the open and closed state structures of pentameric ligand-gated ion channels [26]. Using a series of proline analogs, experiments have suggested that cis–trans isomerization of a single proline residue provides a molecular switch for inter-converting open and closed states in the channels formed by 5HT3 receptors which are members of Cys-loop receptor superfamily [27].

In addition to the conformational changes, specific interactions have been implicated in channel opening and closing. The gating of voltage-gated proton channels has been shown to be regulated by salt-bridge networks [28]. Cross-linking studies demonstrated that the open state conformation of cystic fibrosis transmembrane conductance regulator channels is stabilized by two salt-bridges [29]. On the other hand, in ATP-gated P2X receptor cation channels, salt-bridge interaction stabilizes the closed state and ATP-binding disrupts this salt-bridge so that significant conformational changes can take place to drive the channel to the open state [30]. Similarly, salt-bridge interaction between acidic and basic residues at the interface of ligand-binding domain and transmembrane domain has been shown to be crucial in the gating of glycine receptor channels [31]. Thus for many channels with different structural folds and diverse transport properties, transition from open to closed states (and vice versa) involves conformational changes accompanied by formation or breaking of specific interactions like salt-bridges.

In this paper, we have analyzed more than 1460 sequences from the MIP superfamily from six different organism groups and compared

them. We have specifically focused on the functionally important half-helix formed by loop E. This loop possesses one of the conserved NPA motifs and also contributes two out of four residues for the Ar/R selectivity filter. We demonstrate that the stability of this functionally important half-helix is modulated by a stabilizing intra-helical salt-bridge interaction and/or two helix destabilizing residues glycine and proline. Presence or absence of residues forming the intra-helical salt-bridge and conservation of glycine and proline are analyzed for the MIP members from different organism groups. The hallmark of AQGPs seems to be the presence of intra-helical salt-bridge, whereas both AQPs and AQGPs show high conservation of glycine and/or proline residues. The pattern observed in the half-helix from loop E in plant MIP subfamilies is compared with that found in AQPs and AQGPs from other organism groups. Similarly, the patterns observed in MIP subfamilies which form separate clades that are distinct from AQPs and AQGPs are also analyzed. We also examined the segment connecting the loop E half-helix and TM6. The interplay between the simultaneous presence of stabilizing interactions and helix-destabilizing residues in the half-helix from loop E has been investigated by performing molecular dynamics simulations on three representative MIP channels, one with intra-helical salt-bridge with Gly and Pro in the same helical turn (GlpF). The second channel is AQP1 which does not have an intra-helical salt-bridge in LE half-helix but has helix destabilizing Gly (AQP1). The third channel is from a fungal pathogen with two helix-breaking residues, Gly and Pro in the LE half-helical region and without intra-helical salt-bridge. A mutant simulation of AQP1 has also been carried out in which the Gly in the LE half-helical region has been substituted *in silico* by helix stabilizing Ala. The results of the simulations demonstrate that the helix stability of LE half-helix and the dynamics of the polypeptide segment connecting this half-helix and TM6 can be two important factors that either independently or along with other regions of the channels can regulate the transport properties of MIP channels.

2. Materials and methods

2.1. MIP sequences from MIPModDB database

MIP sequences available in MIPModDB database (http://bioinfo.iitk.ac.in/MIPModDB) [32] have been downloaded for the analysis of this study. MIPModDB database has the homology models of more than 1000 MIPs from bacteria, archaea, fungi, plants and mammals. We have downloaded 79 archaeal, 252 bacterial, 320 plant, 89 non-mammalian metazoan and 96 mammalian MIP sequences from MIPModDB database. Additionally 395 fungal MIP sequences were also added from our recent study [33]. We have also identified new plant MIP sequences by searching the sequence databases and the approach used for this purpose is described below.

2.2. Additional plant MIPs from database search

We performed tBLASTn [34] search available in NCBI (http://www.ncbi.nlm.nih.gov) on non-redundant nr/nt database using five different sets of query sequences and each set corresponds to one of the five plant MIP subfamilies, namely, PIPs (plasma membrane intrinsic proteins), TIPs (tonoplast intrinsic proteins), NIPs (nodulin-26 intrinsic proteins), SIPs (small basic intrinsic proteins) and XIPs (X-intrinsic proteins). The MIPModDB accession codes of the query sequences are as follows. PIP subfamily: POTRIC0506, ORSATI0236, ARTHAL0010 and ZEMAYS0266; TIP subfamily: POTRIC0527, ORSATI0251, ARTHAL0026 and ZEMAYS0282; NIP subfamily: POTRIC0495, ORSATI0223, ARTHAL0001 and ZEMAYS0262; SIP subfamily: POTRIC0522, ORSATI0249, ARTHAL0023 and ZEMAYS0279; XIP subfamily: POTRIC0544, POTRIC0545, POTRIC0546 and POTRIC0547. For PIP, TIP, NIP and SIPs families, one query sequence was chosen each from *Populus tricocarpa*, *Oryza sativa*, *Arabidopsis thaliana* and *Zea mays*. Only in the case of XIPs, all the four query sequences were from *Populus*. The hit sequences thus obtained from the search were first checked for their length. Since all MIPs have the conserved hour-glass fold with six transmembrane segments and two half-helices, the sequences were retained only if the polypeptide length was at least 180 residues long. Redundancy was removed using CD-HIT [35] at 100% level.

We also used the following criteria to validate that the sequences obtained through tBLASTn search belong to the MIP superfamily: (a) the sequences must possess two NPA or NPA-like motifs; (b) they must have six transmembrane helical segments and two half-helices; and (c) there should be conservation of small and weakly polar residues in most of the 17 positions identified in our earlier studies [36,37]. Those sequences that did not satisfy the above criteria were discarded. After screening each sequence with the above stringent criteria, we identified 237 new plant MIPs from the sequence search. In total, we have considered 527 plant MIPs from 69 different plant species for further phylogenetic analysis and homology modeling.

Multiple sequence alignment of plant MIP sequences was performed using Clustal W [38] or Clustal Ω [39] as available in the Clustal X [38] and Seaview software package [40] respectively. The alignment thus produced was verified manually and was given as input for phylogenetic analysis. MEGA 6 [41] was used to produce the phylogenetic trees with maximum parsimony method and neighbor-joining method. The trees produced using two different methods were compared and found to be consistent with each other. Bootstrapping was used to examine the robustness of the trees by generating 1000 replicates and applying 50% majority rule.

We used homology modeling technique to build three-dimensional models of each of the newly identified plant MIP sequences. For the sequences extracted from MIPModDB, the structural models were directly imported from the database. For the new plant MIPs obtained from the database search, we followed the same protocol developed in our laboratory that was used to build homology models [36,37]. We used the software suite Modeller 9.11 [42] and the template structures were that of AQP1 from *Bos taurus* (PDB ID: 1J4N) [43], GlpF from *E. coli* (PDB ID: 1FX8) [44] and AQPM from the archaea *Methanothermobacter marburgensis* (PDB ID: 2F2B) [45]. The target-template alignment of sequences is a crucial step in the modeling procedure and we manually inspected the alignment to ensure that there are no gaps in the transmembrane helical segments and the highly conserved residues in specific transmembrane segments are aligned in the same column. Among the 10 models generated for each sequence, the model with the optimal objective function was selected for further side-chain refinement using SCWRL3 package [46]. Only side-chains of those residues which are not conserved across the three template and the target sequences were modeled. The modeled structure was further refined using energy minimization method with GROMACS 4.5.5 software package [47]. The quality of the final model was examined using PROCHECK [48].

In total, we considered 1468 MIPs from archaea, bacteria, fungi, plants, non-mammalian metazoans and mammals. Sequences and the structural models from the region of the functionally important loop E segment were analyzed within and between the different organism groups.

2.3. Molecular dynamics simulations of representative MIP channels

To investigate the influence of presence/absence of intra-helical salt-bridge along with Gly and Pro in the half-helix of loop LE, we carried out molecular dynamics (MD) simulations of three representative MIP channels in explicit lipid bilayers. The glycerol transporter (GlpF) from *E. coli* has the sequence NPARDF***GP***KVFAWLA in the LE half-helix region. The sequence of the same half-helical region in the water-transporting AQP1 from *Bos tarus* is NPARSF***G***SSVITHNF (acidic and basic residues in GlpF and the equivalent positions in AQP1 are underlined and helix-breaking residues are shown in bold and italic for both sequences). The structures of GlpF and AQP1 have been determined and their respective PDB IDs are 1FX8 [44] and 1J4N [43]. These structures were

downloaded from PDB. The LE half-helix in GlpF structure has both stabilizing intra-helical salt-bridge interaction and helix destabilizing Gly and Pro residues within the same helical turn. In AQP1, no intra-helical salt-bridge was present in the LE half-helical region. However, it has Gly in the same position found in GlpF. In both the structures, only protein atoms were retained and all other molecules including the ligands were deleted. The side-chain of R257 in GlpF was not fully resolved in the experimental structure and it was modeled using the Homology module of InsightII molecular modeling suite of software (Accelerys Inc., San Diego). MIP channels exist as tetramer under physiological conditions. Hence, the biological assemblies of protein tetramers were generated as per the procedure described by de Groot et al. (http://www.mpibpc.mpg.de/267060/practical_05).

The third structure to be simulated is a MIP channel from *Coccidioides posadasii* (CpAQP), a known human fungal pathogen [49]. Its modeled structure was downloaded from the MIPModDB database (MIPModDB accession ID: COPOSA1011). The amino acid sequence of the modeled LE half-helix region is SPARAF***GP***DLVLGDF (the equivalent positions corresponding to acidic and basic residues in GlpF are underlined and the helix-breaking residues are shown in bold and italic). This channel does not have the intra-helical salt-bridge but possesses both Gly and Pro within the same helical turn. Finally, a mutant structure of AQP1 was generated in which the Gly in the LE half-helical region was substituted *in silico* by helix promoting Ala. This mutant model of AQP1 was used for MD simulation for comparison with the wild-type AQP1.

The pre-equilibrated and pre-solvated POPE bilayer containing 340 lipids was used to construct the starting structure of channel-hydrated bilayer complex (http://wcm.ucalgary.ca/tieleman/downloads) [50]. MIP tetramers were inserted in the lipid patch as per the protocol suggested by Peter Tieleman and his colleagues [51]. Molecular dynamics (MD) simulations were carried out using GROMACS 4.5.5 [47] with Berger's united atom force field for lipids [52] and all atom OPLS force field for proteins [53]. The water model TIP3P [54] was used and counter ions were added to neutralize the system. The systems consisted of 81,006, 75,057, 82024 and 75,819 atoms for GlpF, AQP1, AQP1-mutant and CpAQP respectively. All the four systems were minimized using steepest descent and conjugate gradient methods before equilibration.

Each system was equilibrated as per the following procedure. During equilibration, positional restraints were initially imposed on the lipid atoms and were gradually removed in steps of 100 ps using NVT ensemble (constant number of atoms, volume and temperature). During this time, protein atoms were restrained using a harmonic force constant of 10,000 kJ/mol/nm^2. In the second stage of equilibration, the systems were equilibrated for further 1 ns using NPT ensemble (constant number of atoms, pressure and temperature) with restraints applied only on the protein heavy atoms. This is followed by another 10 ns equilibration without any restraints either on protein or lipid atoms and NPT ensemble was used. Semi-isotropic pressure coupling method was applied in which the plane defining the membrane (X–Y) and its normal (Z) were coupled separately.

Long-range interactions were calculated using particle-mesh Ewald (PME) method [55] and VDW interactions were described using a cutoff of 12 Å for all simulations. In NPT ensemble, temperature was coupled using Noose–Hoover coupling algorithm [56] for maintaining constant temperature (T = 310 K) and Parrinello–Rahman algorithm [57] was used for maintaining constant pressure (P = 1 bar). In all the simulations, Periodic Boundary Conditions (PBC) were employed in all three directions. After the equilibration run, the systems were subjected to a production run of 100 ns each. Our analysis of loop E half-helix stability and conformational fluctuations of linker regions are presented for all three MIP wild-type channels and compared.

2.4. Water transport in AQP1 and AQP1 mutant

The water transport properties of wild-type AQP1 and its mutant are compared and are related to the stable nature of LE half-helix. We calculated the number of water permeation events and also potential of mean force (PMF) profiles for both the wild-type and mutant AQP1 channels. PMF was calculated along the pore axis (Z axis) from the average number of water molecules at each position and was evaluated over the entire 100 ns MD trajectories. For each AQP1 monomer, the PMF profile was calculated using the formula

$$G_i(z) = -k_B T \ln \langle n_i(z) \rangle \quad (1)$$

where k_B and T are Boltmann constant and temperature respectively and $\langle n_i(z) \rangle$ represents the average number of water molecules found as a function of pore coordinate along the pore axis. PMF profile was calculated from −30 Å to +30 Å from the cytoplasmic to the extracellular side with the NPA motif positioned at $z = 0$ Å. For each z position, a cylinder of 5 Å radius and thickness of 0.5 Å was considered. To account for the underestimation of bulk free energy of water at the entrance and exit regions of each AQP1 monomer, a trapezoidal correction was applied [58] and the computed correction values for AQP1 and AQP1-mutant are 6.46 and 6.19 kJ/mol respectively.

For each monomer in both AQP1 and AQP1-mutant systems, we also found out the number of water permeation events during the course of 100 ns simulation. For this purpose, we aligned a cylinder of 18 Å length and 5 Å radius along the pore axis. The centroid of the NPA motif corresponds to z = 0 Å and the cylindrical axis varied from z = +13 to z = −5 Å from the extracellular to the cytoplasmic side. A water molecule is considered to be permeated if it completely traverses the cylinder entering from any direction.

3. Results

MIP sequences and structural models from six different organism groups (archaea, bacteria, fungi, plants, non-mammalian metazoans and mammalian) were downloaded from the MIPModDB database (http://bioinfo.iitk.ac.in/MIPModDB) [32]. Additional plant MIPs were found by searching the sequence databases. Their phylogenetic analysis and homology modeling are described in detail in the Materials and methods section. In total, 1468 MIP sequences and their corresponding structural models were considered for analyzing the loop E region with the specific emphasis on the functionally important LE half-helical segment.

3.1. Phylogenetic analysis

We first performed phylogenetic analysis of MIPs and this will help us to find out the number of different MIP clusters in each organism group. Although a minimum of two clusters (AQP and AQGP) is expected, this analysis should also reveal whether any clusters different from AQP and AQGP are found in any of the organism groups. Phylogenetic trees for MIPs from all six organism groups are presented in Fig. 2.

Archaeal MIPs clearly exhibit that there are four distinct clusters (Fig. 2a). While two of them correspond to AQP and AQGP clusters, the remaining two are distinct from either AQP or AQGP clusters. We have designated them as MIP-α and MIP-β and MIP-α is the largest among all four clusters with 46 sequences. The sequence pattern (see below) and the Ar/R selectivity filter residues (data not shown) confirm that MIP-α and MIP-β are distinct from the AQP and AQGP clusters. Notably, Zardoya and coworkers in a recently published phylogenetic analysis of MIPs have reported only AQP and AQGP clusters for archaeal MIPs [59]. All the 252 bacterial MIPs are clearly separated into two clades falling into one of the two clusters, namely, AQP or AQGP (Fig. 2b). In our recent study of 395 fungal MIPs [33], we found four major clusters including those belonging to orthodox AQP and AQGP groups (Fig. 2c). Apart from these two groups and XIPs, we found an additional subfamily sharing some characteristic features with the plant SIP subfamily. As reported in previous studies [37,60,61], plant MIPs are divided into five subfamilies, namely, PIPs, TIPs, NIPs, SIPs

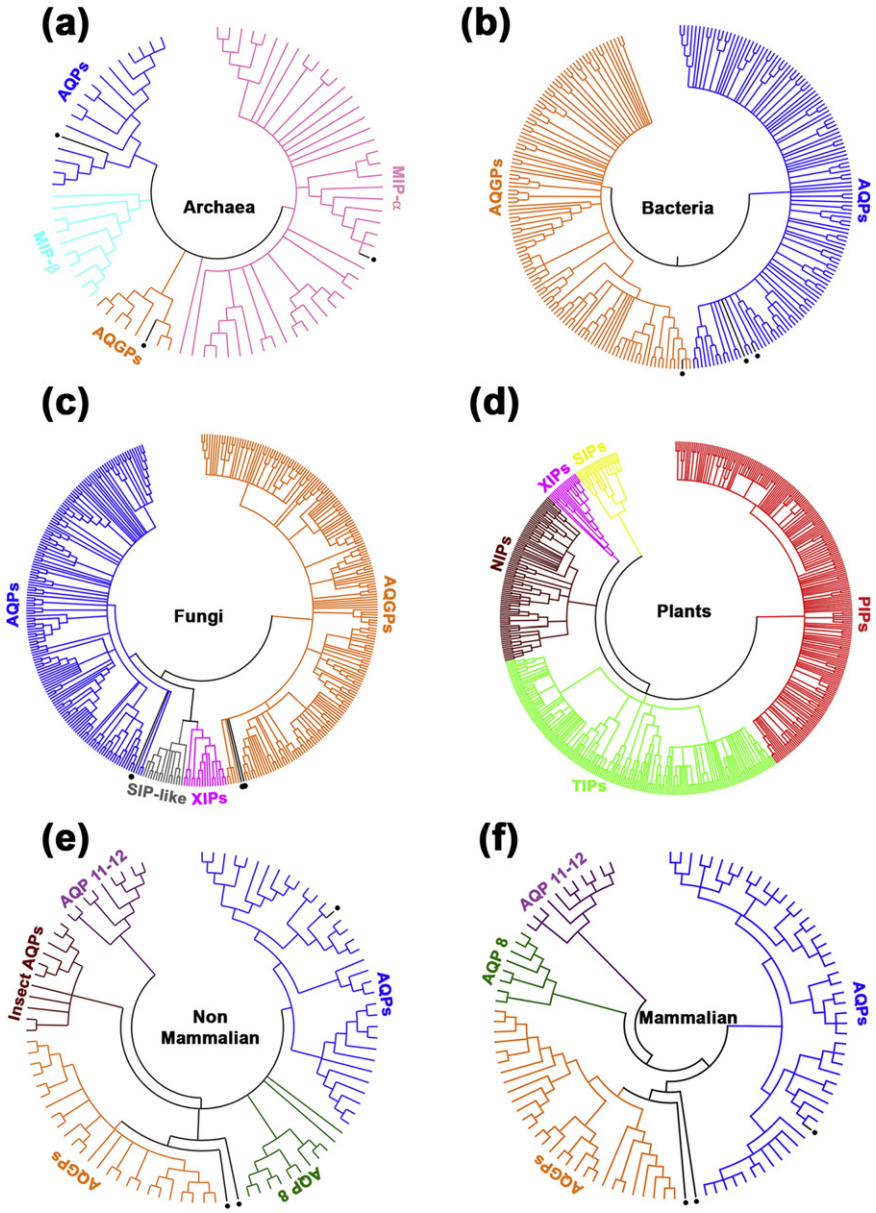

Fig. 2. Phylogenetic analysis of MIP channels for each of the six organism groups. AQP and AQGP clusters are shown in blue and orange colors respectively. Archaeal, fungal, non-mammalian and mammalian MIPs have subfamilies that are distinct from AQPs and AQGPs. All plant subfamilies separately cluster and hence they do not have subfamilies that correspond to AQP or AQGP clusters. Branches shown in black and indicated with a black circle correspond to the reference sequences.

and XIPs (Fig. 2d). In the present study, we found XIPs only in dicot plants confirming earlier results from our laboratory [37]. Phylogenetic analysis of non-mammalian metazoan MIPs showed all four clusters found in the mammalian counterparts and also revealed an additional cluster which is designated as "Insect AQPs" (Fig. 2e). Mammalian MIPs are divided into four clusters and in addition to AQP and AQGP, the other distinct clusters belong to AQP8 and AQP11–12 categories (Fig. 2f). In summary, with the exception of plant MIPs, all organism groups have clearly defined AQP and AQGP clusters. Additional clusters which are distinct from AQP and AQGP clusters are found in archaeal, fungal, mammalian and non-mammalian metazoan MIPs. Only in the case of plant MIPs, the five plant MIP families do not fall into the

Table 1
Summary of MIP sequences from different organism groups.

Organism group	AQP cluster[a]	AQGP cluster[a]	Other MIP groups[a]	Total
Archaea	16	7	56 [MIP-α (46); MIP-β (10)]	79
Bacteria	131	121	– N/A –	252
Non-mammalian metazoan	33	25	31 [AQP8 (11); Insect AQPs (11); AQP11-12 (9)]	89
Mammalian	49	29	18 [AQP8 (8); AQP11-12 (10)]	96
Fungi	163	199	33 [XIPs (17); SIP-like (16)]	395
Plants	– N/A –	– N/A –	557 [PIPs (239); TIPs (181); NIPs (93); SIPs (25); XIPs (19)]	557
Total	392	381	695	1468

[a] MIPs from each organism groups were clustered according to the phylogenetic analysis described in the text and in Fig. 2.

category of AQPs or AQGPs. The summary of all MIP groups are presented in Table 1 and further analysis of the distinct MIP groups are presented below. While there are 393 AQPs, 381 belong to the AQGP cluster from different organisms. Nearly 700 MIPs do not belong to either AQP or AQGP clusters and a bulk of them (557 out of 695) constitute different plant MIP subgroups.

3.2. Analysis of LE half-helix in known MIP structures

Highly conserved Asn residue from the NPA motif initiates the half-helix and it consists of nearly four helical turns. Loop E also contributes two out of four residues (LE1 and LE2) towards formation of the Ar/R selectivity filter. The LE1 position falls just outside the N-terminal end of the half-helix before the NPA motif and LE2 lies well within the half-helical region and just after the NPA motif. The highly conserved arginine in the Ar/R selectivity filter occupies the LE2 position. Hence, sequence conservation and sequence pattern in loop E which contains both the important constriction regions (NPA motif and LE1 and LE2 positions) will be of great importance.

To date, 24 structures from 11 unique MIP sequences have been determined and can be downloaded from the PDB. Among them, GlpF (PDB ID: 1FX8) from *E. coli* is glycerol specific and PfAQP (PDB ID: 3C02) from *P. falciparum* is efficient in transporting both glycerol and water. AqpM (PDB ID: 2F2B) shows moderate efficiency in transporting water. The remaining structures from AQP0 (PDB ID: 1YMG), AQP1 (PDB ID: 1J4N), AQP2 (PDB ID: 4NEF), AQP4 (PDB ID: 3GD8), AQP5 (PDB ID: 3D9S), AQPZ (PDB ID: 1RC2), SoPIP2;1 (PDB ID: 1Z98) and Aqy1 (PDB ID: 3ZOJ) correspond to those of water channels. We have shown the sequences of LE half-helical regions (Fig. 1b) for glycerol-specific channels and selected water-transporting channels. We have also plotted the superposed structures of half-helix from loop E of the corresponding channels (Fig. 1c to e). Two striking features emerged from this analysis. The half-helix in glycerol-specific channels are stabilized by an intra-helical salt-bridge while the equivalent positions are occupied by small residues in water-transporting channels. The second feature is the presence of helix-destabilizing residues, glycine and proline, within the same helical turn. We were intrigued to note that both the stabilizing and destabilizing factors are present at the same time in the same helical turn in AQGP structures. This observation is from a limited number of experimentally determined MIP structures and how widespread the observed pattern among the diverse MIP families is not known. For this purpose, we have analyzed the entire set of more than 1460 MIP sequences. In all MIPs, we carefully looked at the equivalent positions for the possibility of forming an intra-helical salt-bridge and presence of Gly and Pro.

3.3. Intra-helical salt-bridge and helix destabilizing residues in LE half-helix

We examined the potential to form an intra-helical salt-bridge in the loop E half-helical region in AQPs, AQGPs, plant MIPs and MIPs that belong to other groups from different organisms. We have plotted the sequence logos for the loop E region from different MIP groups (Fig. 3). In this plot, the highly conserved Asn of NPA motif is assigned as zero and all other residues are assigned relative to this position. An overwhelming majority of 381 AQGPs contain an acidic residue and a basic residue at +4 and +8 positions respectively within the half-helix of loop LE (Table 2). In an α helix, this arrangement will bring both the residues one above the other that will enable them to form a salt-bridge. An intra-helical salt-bridge will increase the stability of an α-helix. While loop LE half-helix's stability is increased by the presence of an intra-helical salt-bridge interaction in AQGPs, it is conspicuously absent in AQPs, all plant MIPs and other MIP subfamilies from different organisms (Tables 2 and 3; Figs. 3 and 4). The acidic and basic residues are replaced by small neutral residues (Ser/Thr/Ala/Cys) at +4 and +8 positions.

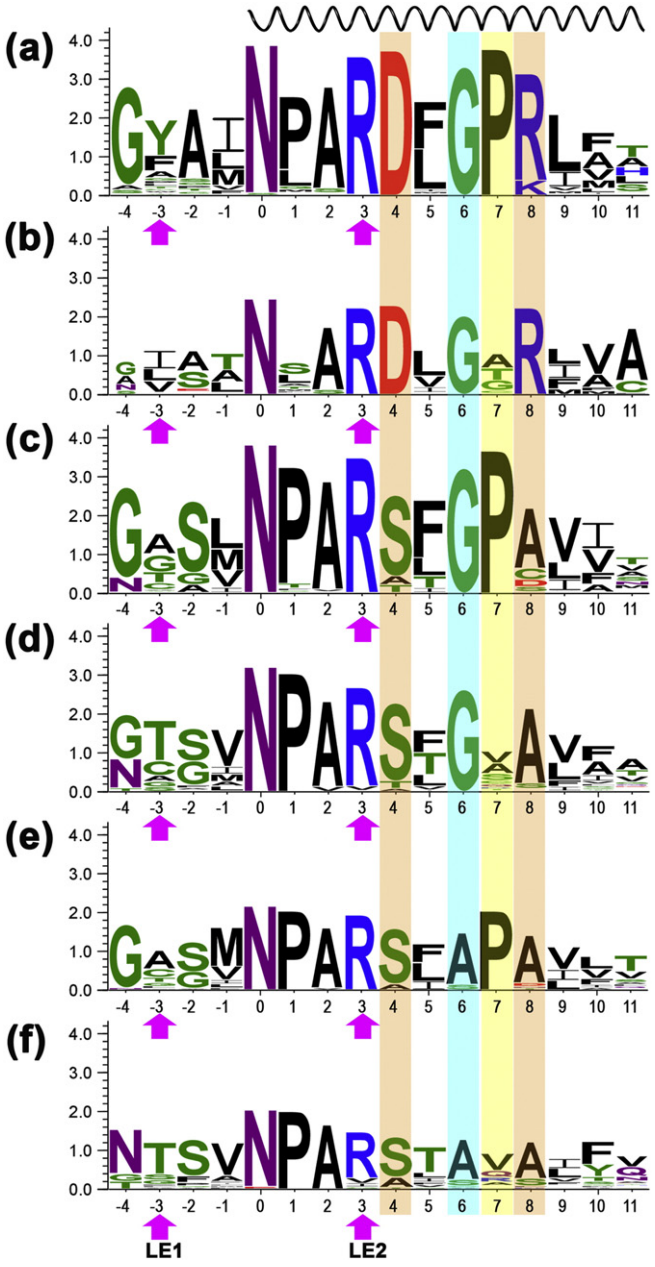

Fig. 3. Sequence logo of AQGP and AQP channels in the loop E region. The position of the highly conserved Asn residues which is part of the NPA motif is designated as zero and all other positions are relative to this residue. The region forming the half-helix within the loop E is indicated. LE1 and LE2 positions (−3 and +3) which form part of the aromatic/arginine selectivity filter are indicated by pink arrows. The positions of acidic and basic residues (+4 and +8) in AQGPs and the equivalent positions in AQP channels are displayed in brown background. Positions of two helix-breaking residues Gly and Pro and the equivalent positions are shown in blue and yellow background respectively. (a) AQGPs with intra-helical salt-bridge and both the helix-destabilizing residues Gly and Pro; (b) AQGPs with intra-helical salt-bridge and Gly; (c) AQPs with both Gly and Pro; (d) AQPs with only Gly; (e) AQPs with only Pro; (f) AQPs in which both Gly and Pro are absent. For the number of MIP sequences used to create sequence logo of each category of AQGP and AQP channels, see Table 2. The web server http://weblogo.berkeley.edu/logo.cgi was used to generate the sequence logos.

It is also intriguing to note the presence of at least one helix destabilizing residues within the same helical turn where an intra-helical salt-bridge interaction is present in AQGPs (Fig. 3). In 99.7% of AQGPs and 98.4 to 99.6% of plant MIPs, either Gly or Pro or both are present within the same helical turn at +6 and +7 positions respectively (Figs. 3 and 4). Only in the case of AQPs and other MIP subfamilies from different organisms, a small but significant fraction of MIPs (8 to

Table 2
Stabilizing intra-helical salt-bridge interaction and helix destabilizing residues (glycine and proline) within the same helical turn of half-helix LE in non-plant MIP subfamilies.

Salt-bridge[a]	Glycine[a]	Proline[a]	Motif[b]	AQPs[c]	AQGPs[c]	Other MIP groups[c]
✓	✓	✓	DXGP(R/K)	0	301 (79%)	0
✓	✓	X	DXGX(R/K)	0	43 (11.3%)	0
✓	X	✓	DXXP(R/K)	0	34 (8.9%)	0
✓	X	X	DXXX(R/K)	0	1 (0.3%)	0
X	✓	✓	(S/T/A/C)XGP(S/T/A/C)	237 (60%)	0	63 (45.6%)
X	✓	X	(S/T/A/C)XGX(S/T/A/C)	92 (23.5%)	1 (0.3%)	15 (10.9%)
X	X	✓	(S/T/A/C)XXP(S/T/A/C)	32 (8.2%)	1 (0.3%)	36 (26.1%)
X	X	X	(S/T/A/C)XXX(S/T/A/C)	31 (7.9%)	0	24 (17.4%)

[a] The presence or absence of intra-helical salt-bridge formed by acidic and basic residues at +4 and +8 positions and Gly and Pro residues at +6 and +7 positions of LE half-helix are indicated respectively by ✓ and X respectively. For details regarding the relative positions of the residues, see Figs. 3 and 4.
[b] The motif corresponds to the region +4 to +8 positions in the loop E half-helix.
[c] Number of AQPs, AQGPs and other MIPs correspond to five different organism groups (archaea, bacteria, fungi, non-mammalian metazoans and mammalian). For details, see Table 1.

17%) have neither Gly nor Pro in the same positions (Table 2). The following sections describe the presence or absence of intra-helical salt-bridge and helix destabilizing residues for each major category of MIP channels.

3.3.1. AQGPs

Almost all MIP members classified as AQGPs in our phylogenetic analysis possess the intra-helical salt-bridge which is observed within the half-helix (Fig. 3a and b). While this interaction will provide extra stability to the helical segment, we also recognized the fact that in nearly 99.1% of all AQGPs, the same helical turn also contains Gly (+6 position) and/or Pro (+7 position) which are known to destabilize an α-helix (Table 2). We found that 301 (79%) out of 381 AQGPs contain both Gly and Pro (Fig. 3a) while 43 (11.2%) and 34 (8.9%) members have either Gly or Pro respectively. The 43 AQGPs belong to the newly identified δ-subgroup of fungal AQGPs. This group has only Gly (Fig. 3b) and Pro at the +7 position is substituted by small residues (Ala, Thr, Gly, Cys and Ser). AQP7 and AQP9 belonging to mammalian and non-mammalian species (total number: 34) have only Pro and the equivalent position in Gly at the +6 position is replaced by Ala, Ser or Pro.

3.3.2. AQPs

In all 392 MIPs classified as AQPs, an intra-helical salt-bridge is absent and the positions corresponding to acidic and basic residues at the +4 and +8 positions are substituted by small neutral residues such as Ser, Thr, Ala and Cys (Fig. 3c to e). We found that about 92% of the members possess Gly and/or Pro within the same helical turn at the +6 and +7 positions respectively (Table 2). However, it should be noted that only 60% contain both Gly and Pro (Fig. 3c) and this is a smaller fraction compared to AQGPs where almost 80% of all AQGPs contain both Gly and Pro within the same helical turn. In 92 examples within the AQP cluster, only Gly is found at the +6 position (Fig. 3d). These MIPs include mammalian AQP1 and plant SoPIP2;1 channels. In these cases, the Pro residue at the +7 position is substituted by a large number of diverse residues (Ala, Ser, Cys, Thr, Gln, Val, Arg). This is in contrast to AQGPs where in the helical turn with only Gly residue, Pro in the succeeding position is replaced by small neutral residues. In another 32 AQPs, while Pro is present, Gly at +6 is substituted mostly by Ala (Fig. 3e). Most of the members from this group are homologs of AQP0 or AQP2. AqpZ homologs in bacteria have neither Gly nor Pro and these 31 members have small neutral residues in the place of Gly and the Pro residue at +7 is replaced by bulky Val, Arg and Gln in majority of the examples (Fig. 3f).

3.3.3. Other MIP clusters from organism groups other than plants

There are 138 MIPs from archaea, non-mammalian metazoans, and mammalian and fungal groups and they clustered separately from AQPs and AQGPs in the respective organism groups. Analysis of half-helix region from loop LE exhibited characteristics similar to AQPs. These MIPs do not have acidic and basic residues after the conserved NPA motif in the LE half-helix segment. As in AQPs, these +4 and +8 positions are occupied by small neutral residues such as Ala, Thr, Ser and Cys. Within the same helical turn, about 83% of them contain Gly and/or Pro at the +6 and +7 positions respectively (Table 2). We identified 24 out of 138 MIPs that do not have either Gly or Pro in the corresponding positions. This is about 17% of all the MIPs from this category.

3.3.4. Plant MIPs

Among the 557 plant MIPs belonging to five different subgroups, none have the capability to form intra-helical salt-bridge in the loop LE half-helical region. This is evident from the sequence logo plotted for the loop LE region for all plant MIP subfamilies (Fig. 4). The positions corresponding to acidic and basic residues are occupied by small neutral residues as in the case of AQPs. However, there are differences between the conventional AQPs and plant MIPs. Each subfamily seems to have specific preference for one of the helix destabilizing residues within the same helical turn (Table 3). An overwhelming majority of PIPs and SIPs have only Gly at +6 and the succeeding residue is occupied by Ala and Trp in almost all the PIPs and SIPs respectively (Fig. 4a and c). The remaining three subfamilies, namely TIPs, NIPs and XIPs, show clear preference (84 to 95%) to have both Gly and Pro within the helical turn of interest (Table 3 and Fig. 4b, d and e).

Table 3
Stabilizing intra-helical salt-bridge interaction and helix destabilizing residues (glycine and proline) within the same helical turn of half-helix LE in different plant MIP subfamilies.

Salt-bridge[a]	Glycine[a]	Proline[a]	Motif[b]	PIPs[c]	TIPs[c]	NIPs[c]	SIPs[c]	XIPs[c]
X	✓	✓	(S/T/A/C)XGP(S/T/A/C)	7 (2.9%)	172 (95.0%)	78 (83.9%)	0	16 (84.2%)
X	✓	X	(S/T/A/C)XGX(S/T/A/C)	231 (96.6%)	4 (2.2%)	5 (5.4%)	22 (88.0%)	3 (15.8%)
X	X	✓	(S/T/A/C)XXP(S/T/A/C)	0	2 (1.1%)	9 (9.7%)	0	0
X	X	X	(S/T/A/C)XXX(S/T/A/C)	1 (0.4%)	3 (1.6%)	1 (1.1%)	3 (12.0%)	0

[a] See footnote a of Table 2.
[b] See footnote b of Table 2.
[c] The phylogenetic analysis of 557 plant MIPs considered in this study clustered them into five different subfamilies as reported in earlier studies [36,58,59]. For other details, see Fig. 2 and Table 2.

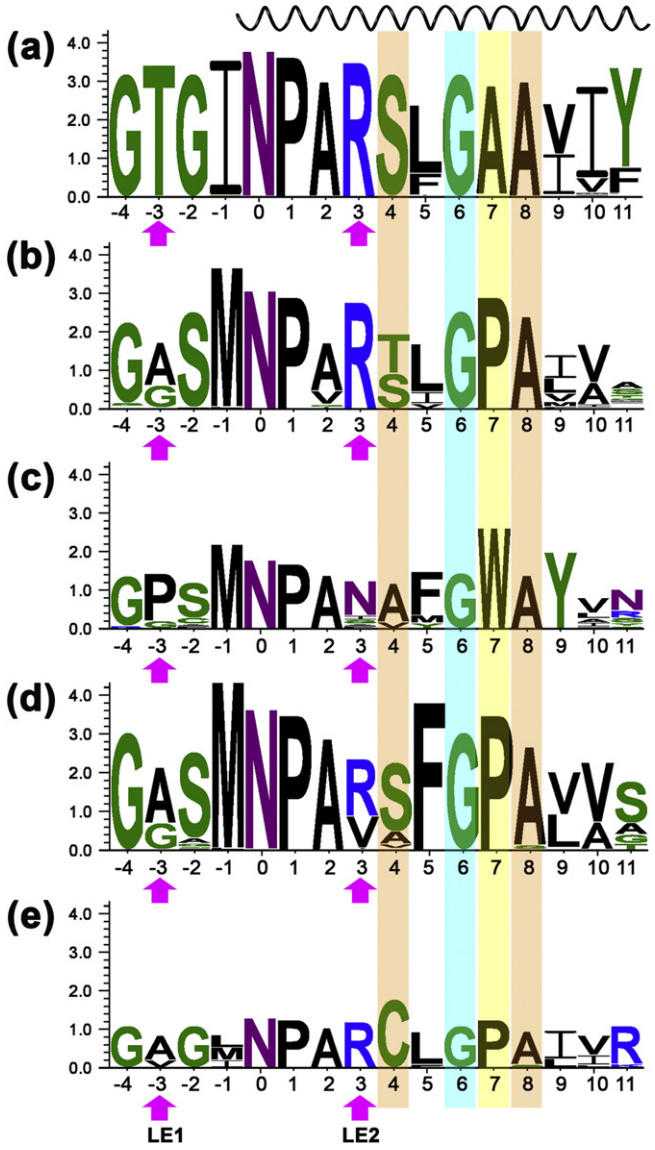

Fig. 4. Sequence logs of plant MIPs in the loop E region. The positions of residues are relative with respect to the highly conserved Asn residue of NPA motif and this is explained in Fig. 3. The positions equivalent to acidic and basic residues of AQGP channels are shown in brown background. Positions of Gly and Pro and their equivalent positions are displayed in cyan and yellow color respectively. Sequence logos of (a) PIPs, (b) NIPs, (c) SIPs, (d) TIPs and (e) XIPs indicate that both Gly and Pro are present in NIPs, TIPs and XIPs. In PIPs and SIPs, only Gly is present and Pro is absent. For the number of sequences used to produce each sequence logo, see Table 3. For all other details, see Fig. 3.

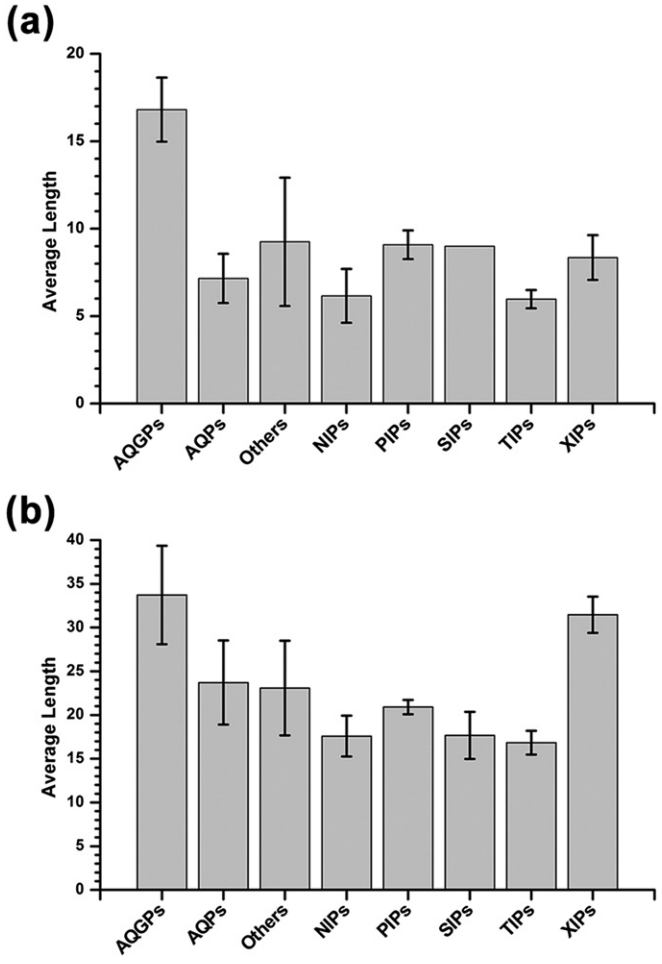

Fig. 5. Histograms of average lengths of the two linker segments for different MIP channel families. Average lengths of (a) polypeptide segment connecting the LE half-helix and TM6 transmembrane helix and (b) loop C connecting the symmetrically related two halves of the helical bundle are displayed.

3.4. Loop connecting the LE half-helix and TM6 and the loop LC

One notable difference we found between the crystal structures of aquaglyceroporins (GlpF and PfAQP) and aquaporins (AQP0, AQP1, AQP2, AQP4, AQP5, SoPIP2;1 and Aqy1) is the length of the loop connecting the loop LE half-helix and the TM6 helix. While the length of this loop is almost 16 to 19 residues in aquaglyceroporins, the same loop region in aquaporin varies from 6 to 9 residues in most of the cases where structures have been determined. We examined the length of this loop region for all MIP groups from different organisms. A histogram of the average length of this loop region is plotted in Fig. 5a for various MIP groups. While this length varies from 5.9 to 9.2 residues in AQPs and various plant MIP subgroups, the average length of the same region is 16.8 residues in AQGPs (Fig. 5a). Similarly, loop LC which connects the two halves of the hour-glass fold also seems to be longer in AQGPs compared to non-AQGPs with the exception of the plant XIP subfamily. The longer loop LC in the plant XIP subfamily has been noted in our previous studies on *Populus* MIPs [37]. The average length of LC in AQGPs is 34 residues while the same region in AQPs, plant MIPs excluding XIPs and other MIPs varies from 18 to 24 residues (Fig. 5b). Loop LC in XIPs has an average length of 31 residues which is comparable to AQGPs. Thus in addition to the presence of intra-helical salt-bridge, the lengths of two linker regions (the segment connecting the half-helix in loop LE and the transmembrane helix TM6 and loop LC) seem to be another distinguishing feature between AQGPs and non-AQGP groups.

3.5. Stability of LE half-helix in MD simulations of MIP channels

MD simulations of three representative MIP channels (GlpF, AQP1 and the fungal CpAQP channel) were carried out in explicit POPE bilayers each for a period of 100 ns after more than 11 ns of equilibration. We first examined the stability of all transmembrane helices and the two half-helices from loops LB and LE. In all four monomers of GlpF, AQP1 and CpAQP channels, the six transmembrane helices remained stable (data not shown). The half-helix from loop LB also stayed intact as observed in the DSSP plots. However, the DSSP plots of LE half-helix clearly demonstrate the influence of intra-helical salt-bridge in GlpF channel and the helix destabilizing residues in GlpF, AQP1 and CpAQP channels (Supplementary Fig. S1). Only one out of

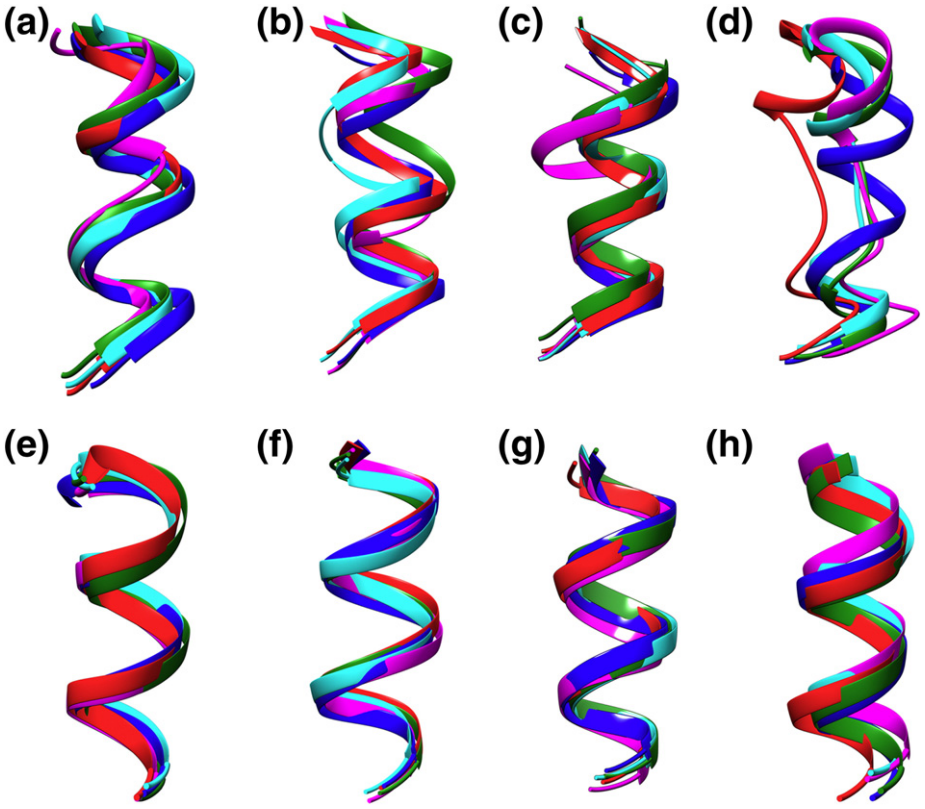

Fig. 6. The starting structure of loop E half-helix (blue) is superposed on the LE half-helices of each of the four monomers at the end of 100 ns production run for (a) GlpF, (b) wild-type AQP1, (c) AQP1 mutant and (d) CpAQP channels. Superposition of the starting structure of loop B half-helical region (blue) is shown with the same region from each monomer at the end of production run for (d) GlpF, (e) wild-type AQP1, (f) AQP1 mutant and (g) CpAQP channels.

four monomers exhibits some disruption of loop E helix in GlpF simulation. On the other hand, the half-helix LE shows significant unwinding in both AQP1 and CpAQP simulations. With the two destabilizing residues, Gly and Pro, the fungal channel displays the most severe disturbance of helical character in all four monomers. The LE and LB half-helices from all four monomers at the end of 100 ns MD simulations were superposed on the respective starting structures (Fig. 6). In GlpF simulation, LE half-helix remains close to the starting structure in all four monomers at the end of a 100 ns production run. In AQP1 and CpAQP channels, it is evident that the half-helix in loop E is destabilized in three out of four monomers with maximum disruption of helical character observed in the fungal channels. In all three simulations, the loop LB maintained its helical character throughout the simulation period. The stability of loop LE half-helix in GlpF can be attributed to the stable intra-helical salt-bridge interaction. Although two helix destabilizing residues, Gly and Pro, are present within the same helical turn, the intra-helical salt-bridge interaction is strictly maintained in all four monomers in the GlpF channel (Fig. 7). This interaction minimizes the loss of helical character and more than compensates the destabilizing effects of Gly and Pro. In the case of AQP1 and CpAQP channels, no intra-helical salt-bridge interaction is present in LE half-helix. AQP1 with one helix destabilizing residue and CpAQP with two helix-destabilizing residues showed significant disruption of helical character in at least three out of four monomers. To further validate the above findings, we carried out an additional simulation of AQP1 mutant in which the Gly residue in LE half-helix was substituted by Ala. Comparison of DSSP plots of LE half-helical region between AQP1 wild-type and mutant clearly demonstrated that AQP1 mutant has the most stable helical region in LE (Supplementary Fig. S1). Hence, it would be interesting to find whether there is any correlation between the stable nature of LE half-helix and the channel's transport properties.

3.6. Water transport in AQP1 wild-type and mutant channels

To establish a possible link between the stable nature of LE half-helix and the channel's ability to transport water efficiently, we calculated the water permeation events across the channel during the course of the 100 ns simulation. The potential of mean force profiles for both wild-type and mutant AQP1 channels were also evaluated. The number of water molecules permeating the channel in all four monomers is 41 and 84 for the wild-type and mutant AQP1 channels respectively. Thus the helical stability due to the substitution of Gly by Ala in the

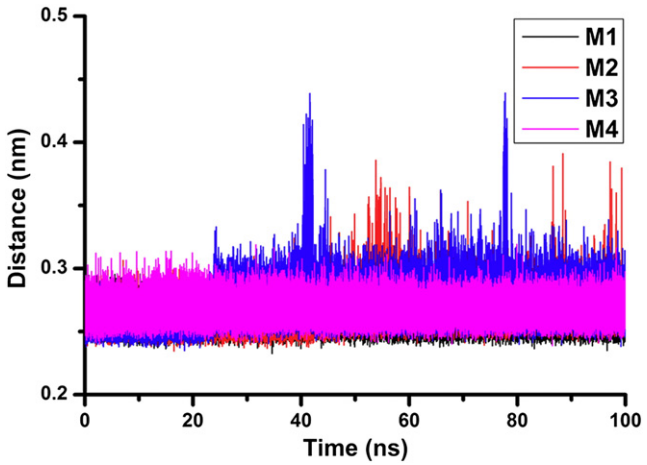

Fig. 7. MD trajectories of the distances between the acidic and basic residues present respectively in the +4 and +8 positions of the LE half-helix for each of the monomer in the GlpF channel.

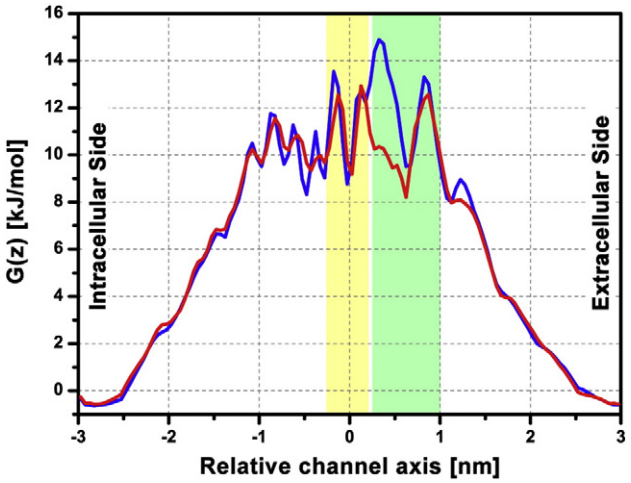

Fig. 8. Potential of mean force (PMF) profiles calculated for AQP1 wild-type (blue) and AQP1 mutant (red). The constrictions due to NPA motifs and the aromatic/arginine selectivity filter region are shown in yellow and green bands respectively.

LE half-helical region helps to transport more than double the number of water molecules in AQP1 mutant channel compared to the wild-type. The PMF profiles calculated as described in the Materials and methods section is presented in Fig. 8. When Gly in LE half-helix was mutated to helix promoting Ala, the free energy barrier in the aromatic/arginine selectivity filter is reduced by almost 6 kJ/mol. This explains the higher rate of water transport in the mutant channel compared to wild-type AQP1.

3.7. Conformational fluctuations of the segment connecting the LE half-helix and TM6 and loop C

Analysis of MIP sequences in the region linking the LE half-helix and the N-terminus of the transmembrane helix segment TM6 indicates that the loop length is longer in AQGP channels compared to AQPs, plant MIPs and MIPs belonging to other groups. The length of loop LC is also longer in AQGPs compared to non-AQGPs with the exception of the XIP subfamily members from plants and fungi. The molecular plots of the region comprising the LE half-helix and TM6 regions along with the loop LC are shown in Fig. 9 and Supplementary Fig. S2. It is obvious that the longer loop regions present in the GlpF channel sample larger number of conformations. The same loop regions in AQP1 and CpAQP channels are shorter and their conformational flexibility is restricted. The region connecting the LE half-helix and TM6 and loop LC are both located on the same side of the channel defining the vestibule of pore mouth at the extracellular side. Since both of them are longer in GlpF compared to AQP1 and CpAQP, we wanted to know if these two loop regions interact with each other. We calculated the average number of hydrogen bonds formed between these two regions during the 100 ns simulations in all four monomers. The average number ± standard deviation of hydrogen bonds between these two loop regions is 1.78 ± 0.90 in GlpF and 0.44 ± 0.49 in AQP1. This very clearly illustrates that the linker segment connecting the LE half-helix and TM6 and loop LC influence each other significantly through hydrogen bond interactions in GlpF. This is possible since the longer loop regions with 38 (loop C) and 19 (region connecting LE half-helix and TM6) residues in GlpF have greater conformational flexibility and can approach each other through conformational transitions. This is not possible with AQP1 and CpAQP channels with both loop LC (17 to 20 residues) and the segment linking LE half-helix and TM6 (6 to 7 residues) are much shorter in length. The longer loop LC in aquaglyceroporins has been shown to interact with the selectivity filter residue Arg located at the LE2 position [62] and mutational studies have also suggested functional importance for this region [63]. Thus the interactions between the loop LC and loop LE regions and LC's interactions in turn with the selectivity filter residues have the potential to impact the transport properties of the GlpF channel.

4. Discussion

4.1. Intra-helical salt-bridge can enhance the stability of LE half-helix in AQGPs

We have analyzed the sequence and structural features of the functionally important loop LE region from more than 1460 MIP sequences belonging to six organism groups. Since loop LE contributes two residues to the aromatic/arginine selectivity filter, understanding the sequence, structural and dynamic features of this region is extremely important to understand the mechanism of the MIP channel's transport and selectivity. Three important features are observed in the LE region. The first observation is the presence of an intra-helical salt-bridge in the LE half-helix in all AQGPs. This interaction is absent in all non-AQGP groups including the orthodox AQPs from all non-plant organisms and the clusters belonging to AQP8, AQP11–12, insect AQPs and archaeal MIP-α and MIP-β groups. Interestingly, none of the plant MIP groups show the presence of an intra-helical salt-bridge in the LE half-helix although several reports suggest that plant MIPs from specific subgroups like NIPs and TIPs are involved in the transport of glycerol (Supplementary Table S1). The conservation of acidic and basic residues at +4 and +8 positions in the loop E region of AQGPs has been recognized in earlier studies. Froger et al. [64] analyzed a small set of MIP sequences and five positions (P1 to P5) were identified to have different physico-chemical properties between AQGP channels that transport small neutral solutes and the water transporting AQPs. The conserved

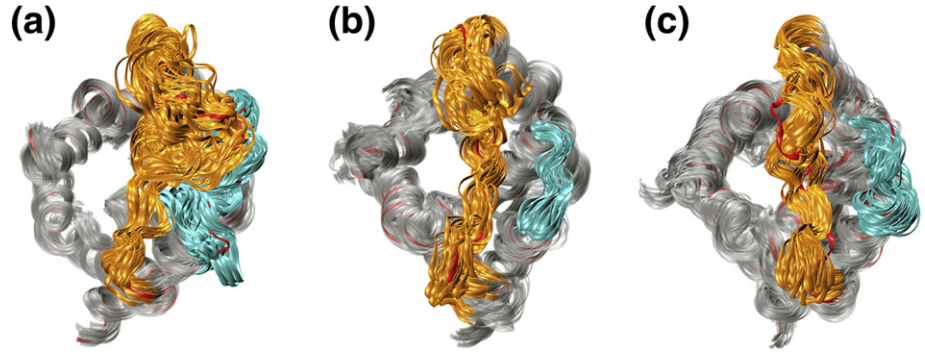

Fig. 9. Superposition of MD snap-shots with the starting structure (red) shown for (a) GlpF, (b) AQP1 and (c) CpAQP channels. The MD simulated structures were saved at the end of each 10 ns. The transmembrane helical regions of each monomer (gray) were superposed on that of the starting structure. The two linker regions are highlighted in different colors. The region connecting the half-helix and the TM6 helix is displayed in cyan and the loop C region that links the two halves of the channel is exhibited in orange.

acidic and basic residue positions in AQGPs correspond to P2 and P3 positions as defined by Froger et al. [64]. However, the importance of these two positions in enhancing the stability of half-helix has not been explicitly stated in earlier studies. The available AQGP crystal structures clearly indicate that these residues participate in intra-helical salt-bridge interaction. Analysis of protein crystal structures suggests that such intra-helical ion-pairs may provide stability to helical structures [65] and this conclusion is supported by the current MD simulations on three MIP channels. With this salt-bridge oriented away from the channel interior, it is anticipated that this interaction of the LE half-helix is not likely to be disrupted. Our MD studies confirm that the intra-helical salt-bridge is maintained in all monomers of GlpF (Fig. 7) which could explain the stable nature of the LE half-helix throughout the simulations. Mutation and functional studies have implicated these two positions (+4 and +8) to be important for transport and oligomerization [66,67].

Although intra-helical salt-bridges are invariably observed in all AQGP channels, the absence of intra-helical salt-bridge in MIP channels does not automatically imply that they are water-specific channels. Several plant MIPs which are reported to transport glycerol (Supplementary Table S1) do not possess the intra-helical salt-bridge. Additionally, the substrate specificities of MIP subfamilies which belong to neither the AQGP cluster nor the AQP group are not known and they do not have this stabilizing interaction.

4.2. Glycine and/or proline can introduce flexibility in LE half-helix of MIPs

While the intra-helical salt-bridge can enhance the stability of LE half-helix in AQGPs, what is most intriguing is the second and the most important observation found in the same half-helix in almost all the MIPs. At least one helix breaking residues, Gly and/or Pro at the +6 and +7 positions, is observed in 99.5% of AQGPs and 92% of all AQPs. Only in the case of non-plant MIPs that cluster distinctly from AQGPs and AQPs, a small fraction is devoid of both Gly and Pro. Even in plant MIPs, 99.1% of all the plant channels contain Gly and/or Pro at these positions. Crystal structure analysis and experimental and computational studies have demonstrated that Gly and Pro are strong helix breakers [68–70]. Glycine and proline residues when they are present in the middle of an α-helix are also known to introduce kink and distort the helical structure [71–74], a feature commonly observed in transmembrane helices. The importance of such kinks and distortion of pore-lining helices produced by Gly and Pro in the function of some channel proteins have been demonstrated [75–77]. In AQGPs, these two helix-breaking residues are observed within the same helical turn (between the +4 and +8 positions) in which intra-helical salt-bridge is also present. Both helix stabilizing interaction and helix destabilizing residues are simultaneously present in the LE half-helix of AQGPs. In all AQPs, other MIPs and plant MIPs, these helix-breaking residues are present and no stabilizing intra-helical salt-bridge interaction is observed indicating that the LE half-helix in non-AQGPs is likely to be less stable compared to that in AQGPs. Proline-introduced kinks in α-helices are observed if proline residues are preceded by a non-Gly residue [71,72]. In majority of the examples of MIP channels, the preceding residue of proline is glycine in the LE half-helix and hence no kink is observed in the experimentally determined MIP structures. However, since proline's backbone nitrogen cannot participate in the intra-helical hydrogen bond, in addition to glycine at the +6 position, the absence of this intra-helical hydrogen bond due to proline residue will confer more flexibility to this region.

This is evident in our molecular dynamics simulations of GlpF, AQP1 and CpAQP. While the LE half-helix is stable in GlpF with an intra-helical salt-bridge, the same region shows unwinding in AQP1 which has Gly and no intra-helical salt-bridge. The destabilization of the same half-helical region in CpAQP is more dramatic than that observed in AQP1. With both Gly and Pro in LE half-helix and no stabilizing intra-helical salt-bridge present, significant loss of helical character is observed in three out of four monomers of the CpAQP channel. When the helix-destabilizing residue is mutated to Ala, the LE half-helix remained mostly stable in the mutant AQP1 indicating the role of Gly in providing flexibility to the half-helical region. The half-helix LE possesses the most conserved arginine residue at the LE2 position (+3 position in the sequence logo; see Figs. 3 and 4) that is part of the Ar/R selectivity filter. The fact that this half-helix is more dynamically flexible in non-AQGPs implies that this feature may have functional significance.

If intra-helical salt bridge is to impart stability in the LE half-helix of AQGPs, then the intriguing question is why the helix destabilizing residues should also be present in the same helical turn in AQGPs. The degree of flexibility or the ability for the helix to be destabilized seems to be modulated by one or two helix destabilizing residues and presence or absence of intra-helical salt-bridge. Only rarely helix breaking residues are absent in MIP channels. The fact that overwhelming majority of MIP channels has at least one helix destabilizing residue implies that LE half-helix flexibility appears to be absolutely essential for the transport of solutes. Even in AQGPs, while intra-helical salt-bridge can preserve the helical character of LE half-helix, the presence of helix destabilizing residues in the same helical turn ensures that the half-helix in LE is not rigid and is flexible to some extent. Such property could play a vital role in the transport of glycerol and other solutes in AQGPs. To our knowledge, we have not come across any mutational studies that substituted either the Gly or the Pro present at the +6 and +7 positions respectively. Hence, our present simulation study of AQP1 mutant in which the Gly was substituted *in silico* to Ala at the +6 position demonstrated that Gly is indeed a vital residue that can modulate the helical character of LE half-helix. Mutation of one or both residues can possibly reveal the role of these residues in the efficient transport of solutes in MIP channels. The present MD simulations of the three MIP channels have demonstrated the potential role of these residues in the stability of LE half-helix which in turn will have influence on the function of MIP channels.

Several MIP channels have been investigated using MD simulation technique including AQP1 [9,78], *E. coli* GlpF [79], AQP0 [80], AQP4 [81], AQP5 [82], *E. coli* AqpZ, yeast Aqy1 [7], plant SoPIP2;1 [16], PfAQP from *P. falciparum* [58] and archaeal AqpM [83]. To the best of our knowledge, the previous MD simulation studies of MIP channels have not explicitly addressed the loop E half-helix stability. By having both stabilizing intra-helical salt-bridge and helix destabilizing residues simultaneously in the same helical turn of LE half-helix in GlpF and a majority of AQGPs, the stability of this helical segment could be easily modulated by these two competing factors. It appears that the presence of intra-helical salt-bridge alone will make the helix highly stable which may not be desired for the function of AQGPs. To introduce some amount of flexibility, helix destabilizing residues are also present in the same helical turn. This is also evident in the DSSP plots (Supplementary Fig. S1). While the LE half-helical region of GlpF remained most stable, DSSP plots also showed unwinding of this helix in at least one monomer. In spite of the intra-helical salt-bridge, LE half-helix in the M4 monomer of GlpF is severely disrupted. Minor disturbance of helical character in the same region is also seen in M1 and M2 monomers implying that the presence of helix destabilizing residues in GlpF is to ensure that some flexibility is introduced in this helical region.

However, in the case of non-AQGPs, no intra-helical salt-bridge is present and most of them have Gly and/or Pro within the LE half-helix. It can be clearly speculated that this helix will be flexible and is most likely to be destabilized in non-AQGP channels. This is also observed in the current AQP1 and CpAQP simulations. The loop E region with two selectivity filter residues has exhibited conformational changes that included partial unwinding of the half-helix. This region may again adopt helical conformation as evidenced in the case of M4 and M1 monomers of AQP1 and CpAQP respectively (see Supplementary Fig. S1). Such property could be one way to regulate the transport of molecules across the membrane. The conformational properties of longer linker regions (loop LC and the segment connecting the LE half-

helix and TM6) could also play a significant role in influencing the transport of solutes in AQGP channels as discussed below.

4.3. Conformational flexibility of longer linker regions could be important in the regulation of AQGP transport

The third interesting observation is that the two linker regions, loop C and the segment connecting the LE half-helix and TM6, are longer in AQGPs than that found in AQPs. Our analysis of more than 1400 MIP sequences demonstrates that the length of the polypeptide segment connecting the LE half-helix and the TM6 region is at least 10 residues longer in AQGPs compared to non-AQGP channels. This longer loop can easily adopt many different conformations and the same is observed in GlpF simulation (Fig. 9). In non-AQGPs, the average length of the same segment is 6 to 9 residues compared to 17 residues in AQGPs. Such a small loop connecting two helical segments in non-AQGPs will have constraints and can sample only limited conformations. This is also obvious from both the AQP1 and CpAQP simulations (Fig. 9). We hypothesize that this loop region will have a greater role in the regulation of AQGPs compared to non-AQGPs. Thus we see a correlation between the presence of intra-helical salt-bridge in LE half-helix and the length of the polypeptide segment connecting LE half-helix and TM6. All AQGPs have invariably intra-helical salt-bridge in the LE-half-helix and the linker region connecting this half-helix and TM6 is more than 15 residues long. In the case of non-AQGPs including plant MIPs, no intra-helical salt-bridge is present in the LE half-helix but the region linking the LE half-helix and TM6 is much shorter with an average length of 6 to 9 residues. Similarly, loop LC which connects the symmetrically related two halves of the helical bundle is longer in AQGPs with an average length of 34 residues. In non-AQGPs except XIPs, this loop is much shorter with an average length of 17 to 23 residues. The present MD simulations show that these two linker regions interact in GlpF channels. Since loop LC is also known to interact with the important Arg residue from the selectivity filter region in aquaglyceroporins [62], we hypothesize that the two longer linker regions in AQGPs can exert influence in the function of AQGP channels. The role of these linker regions can be established by constructing chimeras of AQPs and AQGPs in which the two linker regions (loop LC and LE half-helix–TM6 connecting segment) can be interchanged between AQP1 and GlpF. Functional studies of these chimeras along with substitution of selectivity filter residues are likely to reveal the importance of these loop regions in the transport properties of MIP channels.

4.4. Flexible LE half-helix and linker regions could regulate MIP channel transport in tandem with other extracellular residues/regions

Several reports have implicated individual residues and segments from the extracellular region in gating and regulation of MIP channel's function. Experimental and computational studies have suggested that conformation of the highly conserved Arg side-chain in the Ar/R selectivity filter could be important in the opening and closing of the channels [84,85]. This selectivity filter Arg in *E. coli* AqpZ has been shown to adopt two different conformational states corresponding to open and closed states. For the same AqpZ channel, MD simulation studies and free energy calculations illustrated the influence of different protonation states of histidine (His 174), another selectivity filter residue in the Ar/R selectivity filter [86]. Interaction of a Tyr residue located near the extracellular side of the TM2 helix with the Ar/R selectivity filter Arg has been shown to stabilize the closed state of an *Arabidopsis* MIP channel [87]. Structural studies of AQP0 exhibited different conformations of extracellular loop LA that connect TM1 and TM2 helices [88,89]. These studies demonstrated that loop LA conformation can displace certain residues and constrain the pore near the Ar/R selectivity filter implying a potential role for loop LA in gating the channel. Recent MD studies on human AQP4 also highlighted the role of residues in the Ar/R selectivity filter in gating [90]. Thus, different regions and residues near the extracellular side are known to be involved in the gating of MIP channels. Using bioinformatics approach and MD simulations of representative MIP channels, the present study has identified specific residue positions and linker regions that are likely to influence the MIP channel transport and these are not recognized in any of the previous studies. The helix destabilizing residues in the LE half-helical region of all MIP channels, the intra-helical salt-bridge in the same region in AQGPs and the linker region connecting the LE half-helix and the TM6 helix are some of the elements that can independently influence the transport of MIP channels. They can also regulate the MIP channel function in conjunction with other extracellular regions/residues that have been already recognized in the previous studies [84–90]. These specific residues and/or regions can undergo smaller conformational changes which can "pinch" in upon the Ar/R constriction region and thus can regulate the water permeability [8].

Results of this study still leave one question open regarding plant MIPs. If the intra-helical salt-bridge in the LE half-helix and the longer loop region that connects this half-helix and TM6 can be considered as characteristic features of AQGPs, what about plant MIPs? All plant MIP subfamilies lack intra-helical salt-bridge in the LE half-helix and the lengths of the linkers (loop LC and the region connecting the LE half-helix and TM6) are similar to those observed in AQPs. However, several reports suggest that many plant NIPs and few TIPs are shown to be involved in the transport of glycerol even though they lack the AQGP features reported in this study (Supplementary Table S1). Perhaps, there are other features that may influence the selectivity, transport and regulation of plant MIP function and further investigation is necessary to unravel the distinct mechanisms adopted by plant MIPs that are involved in glycerol transport.

5. Conclusions

In this study, we have analyzed more than 1400 sequences of major intrinsic protein channels from six organism groups. The sequences and structural features of the loop LE region were examined. Loop LE contributes two out of four residues to aromatic/arginine selectivity filter. Since it possesses one of the two highly conserved NPA motifs, it also forms part of the narrow constriction region formed by the conserved NPA motifs. In all AQGP channels, the half-helix region of loop LE has both the stabilizing intra-helical salt-bridge and helix destabilizing residues within the same helical turn. In non-AQGPs, the intra-helical salt-bridge is absent but the helix breaking residues Gly and/or Pro are present at the equivalent positions. The sequence and structural features clearly indicate that the loop LE half-helix is relatively more flexible in non-AQGP channels compared to AQGPs. Another feature which distinguishes AQGPs from non-AQGPs is the longer polypeptide segment connecting the LE half-helix and the sixth transmembrane helical segment. The average length of this region is about 17 residues in AQGPs compared to 6 to 9 residues in non-AQGPs. Hence this linker region can adopt multiple conformations in AQGPs. This is in contrast to all non-AQGPs in which this region is likely to be constrained with limited conformations. A similar trend is observed in loop LC which connects the symmetrically related two halves of the helical bundle. The average length of loop LC is about 34 residues in AQGPs while the same loop is 17 to 24 residues long in non-AQGPs. Only XIPs have lengths (average 31 residues) comparable to AQGPs. Molecular dynamics simulations of glycerol facilitator GlpF, AQP1 water channel and a fungal AQP channel demonstrate that the loop E half-helix is destabilized in at least three out of four monomers of the AQP1 and CpAQP channels during the 100 ns simulation. The same helical region remains relatively more stable in GlpF simulation in spite of the presence of two helix-breaking residues. This is attributed to the stable intra-helical salt-bridge interaction present in the LE half-helix in GlpF. With loop LE contributing two out of four residues for the aromatic/arginine selectivity filter, the flexibility observed in the LE half-helix is

likely to influence the transport of non-AQGP channels. In the case of AQGPs, the conformational heterogeneity of the longer polypeptide linker region loop LC and the segment connecting LE half-helix and TM6 and the interactions between the two linker regions could have a great effect in the transport of these channels. In both AQGPs and non-AQGPs, the loop LE region can be important in regulating the channel function. The regulation can be achieved by the flexible LE half-helix in non-AQGPs. In the case of AQGPs, the multiple conformations of the two long linker segments can either occlude or allow the solutes to pass through the channel. LE half-helix and the identified linker regions could also act in association with other extracellular regions that have been recognized in earlier studies and are implicated in the regulation of MIP channel transport. The conclusions presented in this study are testable experimentally. Mutation of the helix destabilizing residues, introduction of salt-bridge in the LE half-helix of non-AQGPs or chimeric studies involving the loop segment between the LE half-helix and TM6 are some of the experimental or computational approaches that can shed light on the specific regulatory role of loop LE in the transport of MIP channels.

DSSP plots of loop E helical region of GlpF, AQP1 and CpAQP channels (Fig. S1); molecular plots of MD simulated structures displaying the region from the LE half-helix to the TM6 helix (Fig. S2). Plant MIP channels that are shown to be involved in glycerol transport (Table S1). Supplementary data related to this article can be found online at doi: http://dx.doi.org/10.1016/j.bbamem.2015.03.013.

Transparency document

The Transparency document associated with this article can be found in the online version.

Acknowledgements

RS acknowledges the Department of Biotechnology (DBT) for the funding received from the National Bioscience Award for Career Development (BT/HRD/34/17/2008). We thank the Ministry of Human Resources and Development (MHRD), Government of India for establishing the Center of Excellence for Chemical Biology at IIT-Kanpur. We gratefully acknowledge the Computer Centre, IIT-Kanpur for the High-Performance Computing facility supported by funding from DST and MHRD. RS is USV Chair Professor at IIT-Kanpur. RKV acknowledges CSIR for the Senior Research Fellowship. NDP thanks IIT-Kanpur for the fellowship. We thank one anonymous referee for the useful comments and suggestions that vastly improved the manuscript. Dr. Alok Jain is gratefully acknowledged for his help in MD simulations. We thank our lab members, especially Manu Vajpai and Mishtu Mukherjee, for the fruitful discussions.

References

[1] S.D. Tyerman, H.J. Bohnert, C. Maurel, E. Steudle, J.A.C. Smith, Plant aquaporins: their molecular biology, biophysics and significance for plant water relations, J. Exp. Bot. 50 (1999) 1055–1071.
[2] P. Agre, L.S. King, M. Yasui, W.B. Guggino, O.P. Otterson, Y. Fujiyoshi, A. Engel, S. Nielsen, Aquaporin water channels — from atomic structure to clinical medicine, J. Physiol. Lond. 542 (2002) 3–16.
[3] J.M. Carbrey, D.A. Gorelick-Feldman, D. Kozono, J. Praetorius, S. Nielsen, P. Agre, Aquaglyceroporin AQP9: solute permeation and metabolic control of expression in liver, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 2945–2950.
[4] A. Rojek, J. Praetorius, J. Frokiaer, S. Nielsen, R.A. Fenton, A current view of the mammalian aquaglyceroporins, Annu. Rev. Physiol. 70 (2008) 301–327.
[5] D. Gomes, A. Agasse, P. Thiebaud, S. Delrot, H. Geros, F. Chaumont, Aquaporins are multifunctional water and solute transporters highly divergent in living organisms, Biochim. Biophys. Acta 1788 (2009) 1213–1228.
[6] F. Chaumont, S.D. Tyerman, Aquaporins: highly regulated channels controlling plant water relations, Plant Physiol. 164 (2014) 1600–1618.
[7] G. Fischer, U. Kosinska-Eriksson, C. Aponte-Santamaria, M. Palmgren, C. Geijer, K. Hedfalk, S. Hohmann, B.L. de Groot, R. Neutze, K. Lindkvist-Petersson, Crystal structure of a yeast aquaporin at 1.15 Angstrom reveals a novel gating mechanism, PLoS Biol. 7 (2009) (Art. No. e1000130).
[8] K. Hedfalk, S. Tornroth-Horsefield, M. Nyblom, U. Johanson, P. Kjellbom, R. Neutze, Aquaporin gating, Curr. Opin. Struct. Biol. 16 (2006) 447–456.
[9] B.L. de Groot, H. Grubmuller, Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF, Science 294 (2001) 2353–2357.
[10] Y. Yukutake, M. Yasui, Regulation of water permeability through aquaporin-4, Neuroscience 168 (2010) 885–891.
[11] L. Verdoucq, A. Grondin, C. Maurel, Structure–function analysis of plant aquaporin AtPIP2;1 gating by divalent cations and protons, Biochem. J. 415 (2008) 409–416.
[12] J. Kato, M.K. Hayashi, S. Aizu, Y. Yukutake, J. Takeda, M. Yasui, A general anaesthetic propofol inhibits aquaporin-4 in the presence of Zn2 +, Biochem. J. 454 (2013) 275–282.
[13] J.H. Tong, J.T. Canty, M.M. Briggs, T.J. McIntosh, The water permeability of lens aquaporin-0 depends on its lipid bilayer environment, Exp. Eye Res. 113 (2013) 32–40.
[14] J. Sjohamn, K. Hedfalk, Unraveling aquaporin interaction partners, Biochim. Biophys. Acta 1840 (2014) 1614–1623.
[15] C. Jozefkowicz, P. Rosi, L. Sigaut, G. Soto, L. Pietrasanta, G. Amodeo, K. Alleva, Loop A is critical for the functional interaction of two *Beta vulgaris* PIP aquaporins, PLoS One 8 (2013) (Art. No. e57993).
[16] S. Tornroth-Horsefield, Y. Wang, K. Hedfalk, U. Johanson, M. Karlsson, E. Tajkhorshid, R. Neutze, P. Kjellbom, Structural mechanism of plant aquaporin gating, Nature 439 (2006) 688–694.
[17] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.
[18] T. Gonen, T. Walz, The structure of aquaporins, Q. Rev. Biophys. 39 (2006) 361–396.
[19] N. Mitani-Ueno, N. Yamaji, F.J. Zhao, J.F. Ma, The aromatic/arginine selectivity filter of NIP aquaporins plays a critical role in substrate selectivity for silicon, boron, and arsenic, J. Exp. Bot. 62 (2011) 4391–4398.
[20] H. Li, H.N. Chen, C. Steinbronn, B.H. Wu, E. Beitz, T. Zeuthen, G.A. Voth, Enhancement of proton conductance by mutations of the selectivity filter of aquaporin-1, J. Mol. Biol. 407 (2011) 607–620.
[21] E. Beitz, B. Wu, L.M. Holm, J.E. Schultz, T. Zeuthen, Point mutations in the aromatic/arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia, and protons, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 269–274.
[22] M.O. Jensen, E. Tajkhorshid, K. Schulten, Electrostatic tuning of permeation and selectivity in aquaporin water channels, Biophys. J. 85 (2003) 2884–2899.
[23] A. Frick, M. Jarva, S. Tornroth-Horsefield, Structural basis for pH gating of plant aquaporins, FEBS Lett. 587 (2013) 989–993.
[24] T.L. Kash, A. Jenkins, J.C. Kelley, J.R. Trudell, N.L. Harrison, Coupling of agonist binding to channel gating in the GABA(A) receptor, Nature 421 (2003) 272–275.
[25] A. Grottesi, C. Domene, B. Hall, M.S.P. Sansom, Conformational dynamics of M2 helices in KirBac channels: helix flexibility in relation to gating via molecular dynamics simulations, Biochemistry 44 (2005) 14586–14594.
[26] N. Bocquet, H. Nury, M. Baaden, C. Le Poupon, J.P. Changeux, M. Delarue, P.J. Corringer, X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation, Nature 457 (2009) 111–114.
[27] S.C.R. Lummis, D.L. Beene, L.W. Lee, H.A. Lester, R.W. Broadhurst, D.A. Dougherty, Cis–trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel, Nature 438 (2005) 248–252.
[28] A. Chamberlin, F. Qiu, S. Rebolledo, Y.B. Wang, S.Y. Noskov, H.P. Larsson, Hydrophobic plug functions as a gate in voltage-gated proton channels, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) E273–E282.
[29] G. Cui, C.S. Freeman, T. Knotts, C.Z. Prince, C. Kuang, N.A. McCarty, Two salt bridges differentially contribute to the maintenance of cystic fibrosis transmembrane conductance regulator (CFTR) channel function, J. Biol. Chem. 288 (2013) 20758–20767.
[30] R. Hausmann, J. Gunther, A. Kless, D. Kuhlmann, M.U. Kassack, G. Bahrenberg, F. Markwardt, G. Schmalzing, Salt bridge switching from Arg290/Glu167 to Arg290/ATP promotes the closed-to-open transition of the P2X2 receptor, Mol. Pharmacol. 83 (2013) 73–84.
[31] S.A. Pless, A.W.Y. Leung, J.D. Galpin, C.A. Ahern, Contributions of conserved residues at the gating interface of glycine receptors, J. Biol. Chem. 286 (2011) 35129–35136.
[32] A.B. Gupta, R.K. Verma, V. Agarwal, M. Vajpai, V. Bansal, R. Sankararamakrishnan, MIPModDB: a central resource for the superfamily of major intrinsic proteins, Nucleic Acids Res. 40 (2012) D362–D369.
[33] R.K. Verma, N.D. Prabh, R. Sankararamakrishnan, New subfamilies of major intrinsic proteins in fungi suggest novel transport properties in fungal channels: Implications for the host–fungal interactions, BMC Evol. Biol. 14 (2014) (Art. No. 173).
[34] S.F. Altschul, T.L. Madden, A.A. Schaffer, J.H. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[35] W.Z. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics 22 (2006) 1658–1659.
[36] A. Bansal, R. Sankararamakrishnan, Homology modeling of major intrinsic proteins in rice, maize and *Arabidopsis*: comparative analysis of transmembrane helix association and aromatic/arginine selectivity filters, BMC Struct. Biol. 7 (2007) (Art. No. 27).
[37] A.B. Gupta, R. Sankararamakrishnan, Genome-wide analysis of major intrinsic proteins in the tree plant *Populus trichocarpa*: characterization of XIP subfamily of aquaporins from evolutionary perspective, BMC Plant Biol. 9 (2009) (Art. no. 134).
[38] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thomson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics 23 (2007) 2947–2948.
[39] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W.Z. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thomson, D.G. Higgins, Fast, scalable generation of high-

quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) (Art. No. 539).
[40] M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building, Mol. Biol. Evol. 27 (2010) 221–224.
[41] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.
[42] N. Eswar, D. Eramian, B. Webb, M.-Y. Shen, A. Sali, Protein structure modeling with MODELLER, Methods Mol. Biol. 426 (2008) 145–159.
[43] H. Sui, B.G. Han, J.K. Lee, P. Walian, B.K. Jap, Structural basis of water-specific transport through the AQP1 water channel, Nature 414 (2001) 872–878.
[44] D. Fu, A. Libson, L.J. Miercke, C. Weitzman, P. Nollert, J. Krucinski, R.M. Stroud, Structure of a glycerol-conducting channel and the basis for its selectivity, Science 290 (2000) 481–486.
[45] J.K. Lee, D. Kozono, J. Remis, Y. Kitagawa, P. Agre, R.M. Stroud, Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 A, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 18932–18937.
[46] A.A. Canutescu, A.A. Shelenkov, R.L. Dunbrack, A graph-theory algorithm for rapid protein side-chain prediction, Prot. Sci. 12 (2003) 2001–2014.
[47] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447.
[48] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[49] T.J. Sharpton, J.E. Stajich, S.D. Rounsley, M.J. Gardner, J.R. Wortman, V.S. Jordar, R. Maiti, C.D. Kodira, D.E. Neafsey, Q. Zeng, C.Y. Hung, C. McMahan, A. Muszewska, M. Grynberg, M.A. Mandel, E.M. JKellner, B.M. Barker, J.N. Galgiani, M.J. Orbach, T.N. Kirkland, G.T. Cole, M.R. Henn, B.W. Birren, J.W. Taylor, Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives, Genome Res. 19 (2009) 1722–1731.
[50] D.P. Tieleman, H.J.C. Berendsen, A molecular dynamics study of the pores formed by *Escherichia coli* OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer, Biophys. J. 74 (1998) 2786–2801.
[51] C. Kandt, W.L. Ash, D.P. Tieleman, Setting up and running molecular dynamics simulations of membrane proteins, Methods 41 (2007) 475–488.
[52] O. Berger, O. Edholm, F. Jahnig, Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature, Biophys. J. 72 (1997) 2002–2013.
[53] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236.
[54] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating water, J. Chem. Phys. 79 (1983) 926–935.
[55] 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.
[56] A.L. Cheng, K.M. Merz, Application of the Nose–Hoover chain algorithm to the study of protein dynamics, J. Phys. Chem. 100 (1996) 1927–1937.
[57] M. Parrinello, A. Rahman, Polymorphic transitions in single-crystals — a new molecular dynamics method, J. Appl. Phys. 52 (1981) 7182–7190.
[58] C. Aponte-Santamaria, J.S. Hub, B.L. de Groot, Dynamics and energetics of solute permeation through *Plasmodium falciparum* aquaglyceroporin, Phys. Chem. Chem. Phys. 12 (2010) 10246–10254.
[59] F. Abascal, I. Irisarri, R. Zardoya, Diversity and evolution of membrane intrinsic proteins, Biochim. Biophys. Acta 1840 (2014) 1468–1481.
[60] U. Johanson, S. Gustavsson, A new subfamily of major intrinsic proteins in plants, Mol. Biol. Evol. 19 (2002) 456–461.
[61] U. Johanson, M. Karlsson, I. Johansson, S. Gustavsson, S. Sjovall, L. Fraysse, A.R. Weig, P. Kjellbom, The complete set of genes encoding major intrinsic proteins in Arabidopsis provides a framework for a new nomenclature for major intrinsic proteins in plants, Plant Physiol. 126 (2001) 1358–1369.
[62] Z.E.R. Newby, J. O'Connell III, Y. Robles-Colmenares, S. Khademi, L.J. Miercke, R.M. Stroud, Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum*, Nat. Struct. Mol. Biol. 15 (2008) 619–625.
[63] E. Beitz, S. Pavlovic-Djuranovic, M. Yasui, P. Agre, J.E. Schultz, Molecular dissection of water and glycerol permeability of the aquaglyceroporin from *Plasmodium falciparum* by mutational analysis, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 1153–1158.
[64] A. Froger, B. Tallur, D. Thomas, C. Delamarche, Prediction of functional residues in water channels and related proteins, Protein Sci. 7 (1998) 1458–1468.
[65] M. Sundaralingam, Y.C. Sekharudu, N. Yathindra, V. Ravichandran, Ion pairs in alpha-helices, Proteins Struct. Funct. Genet. 2 (1987) 64–71.
[66] V. Lagree, A. Froger, S. Deschamps, I. Pellerin, C. Delamarche, G. Bonnec, J. Gouranton, D. Thomas, J.-F. Hubert, Oligomerization state of water channels and glycerol facilitators, J. Biol. Chem. 273 (1998) 33949–33953.
[67] L. Duchesne, S. Deschamps, I. Pellerin, V. Lagree, A. Froger, D. Thomas, P. Bron, C. Delamarche, J.-F. Hubert, Oligomerization of water and solute channels of the major intrinsic protein (MIP) family, Kidney Int. 60 (2001) 422–426.
[68] C.A. Rohl, A. Chakrabartty, R.L. Baldwin, Helix propagation and N-cap propensities of the amino acids in alanine-based peptides in 40 volume percent trifluoroethanol, Protein Sci. 5 (1996) 2623–2637.
[69] J.A. D'Aquino, J. Gomez, V.J. Hilser, K.H. Lee, L.M. Amzel, E. Freire, The magnitude of the backbone conformational entropy change in protein folding, Proteins Struct. Funct. Genet. 25 (1996) 143–156.
[70] K. Gunasekaran, H.A. Nagarajaram, C. Ramakrishnan, P. Balaram, Stereochemical punctuation marks in protein structures: glycine and proline containing helix stop signals, J. Mol. Biol. 275 (1998) 917–932.
[71] R. Sankararamakrishnan, S. Vishveshwara, Geometry of proline-containing alpha-helices in proteins, Int. J. Pept. Protein Res. 39 (1992) 356–363.
[72] R. Sankararamakrishnan, S. Vishveshwara, Conformational studies on peptides with proline in the right-handed alpha-helical region, Biopolymers 30 (1990) 287–298.
[73] F.S. Cordes, J.N. Bright, M.S.P. Sansom, Proline-induced distortions of transmembrane helices, J. Mol. Biol. 323 (2002) 951–960.
[74] B. Kneissl, S.C. Mueller, C.S. Tautermann, A. Hildebrandt, String kernels and high-quality data set for improved prediction of kinked helices in alpha-helical membrane proteins, J. Chem. Inf. Model. 51 (2011) 3017–3025.
[75] B. Akitake, A. Anishkin, N. Liu, S. Sukharev, Straightening and sequential buckling of the pore-lining helices define the gating cycle of MscS, Nat. Struct. Mol. Biol. 14 (2007) 1141–1149.
[76] S.H. Ding, L. Ingleby, C.A. Ahern, R. Horn, Investigating the putative glycine hinge in shaker potassium channel, J. Gen. Physiol. 126 (2005) 213–226.
[77] S. Thouta, S. Sokolov, Y. Abe, S.J. Clark, Y.M. Cheng, T.W. VClaydon, Pro line scan of the hERG channel S6 helix reveals the location of the intracellular pore gate, Biophys. J. 106 (2014) 1057–1069.
[78] J.S. Hub, B.L. de Groot, Mechanism of selectivity in aquaporins and aquaglyceroporins, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 1198–1203.
[79] E. Tajkhorshid, P. Nollert, M.O. Jensen, L.J.W. Miercke, J.D. O'Connell, R.M. Stroud, K. Schulten, Control of the selectivity of the aquaporin water channel family by global orientational tuning, Science 296 (2002) 525–530.
[80] M.O. Jensen, R.O. Dror, H.F. Xu, D.W. Borhani, I.T. Arkin, M.P. Eastwood, D.E. Shaw, Dynamic control of slow water transport by aquaporin 0: Implications for hydration and junction stability in the eye lens, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 14430–14435.
[81] J.D. Ho, R. Yeh, A. Sandstrom, I. Chorny, W.E.C. Harries, R.A. Robbins, L.J.W. Miercke, R.M. Stroud, Crystal structure of human aquaporin 4 at 1.8 Angstrom and its mechanism of conductance, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 7437–7442.
[82] L. Janosi, M. Ceccarelli, The gating mechanism of the human aquaporin 5 revealed by molecular dynamics simulations, PLoS One 8 (2013) (Art. No. e59897).
[83] R. Araya-Secchi, J.A. Garate, D.S. Holmes, T. Perez-Acle, Molecular dynamics study of the archaeal aquaporin AqpM, BMC Genomics 12 (Suppl. 4) (2011) (Art. No. S8).
[84] J.S. Jiang, B.V. Daniels, D. Fu, Crystal structure of AqpZ tetramer reveals two distinct Arg-189 conformations associated with water permeation through the narrowest constriction of the water-conducting channel, J. Biol. Chem. 281 (2006) 454–460.
[85] L. Xin, H.B. Su, C.H. Nielsen, C.Y. Tang, J. Torres, Y.G. Mu, Water permeation dynamics of AqpZ: a tale of two states, Biochim. Biophys. Acta 1808 (2011) 1581–1586.
[86] G.D. Hu, L.Y. Chen, J. Wang, Insights into the mechanisms of the selectivity filter of *Escherichia coli* aquaporin Z, J. Mol. Model. 18 (2012) 3731–3741.
[87] T. Li, W.G. Choi, I.S. Wallace, J. Baudry, D.M. Roberts, Arabidopsis thaliana NIP7;1: an anther-specific boric acid transporter of the aquaporin superfamily regulated by an unusual tyrosine in helix 2 of the transport pore, Biochemistry 50 (2011) 6633–6641.
[88] T. Gonen, P. Sliz, J. Kistler, Y. Cheng, T. Walz, Aquaporin-0 membrane junctions reveal the structure of a closed water pore, Nature 429 (2004) 193–197.
[89] T. Gonen, Y. Cheng, P. Sliz, Y. Hiroaki, Y. Fujiyoshi, S.C. Harrison, T. Walz, Lipid–protein interactions in double-layered two-dimensional AQP0 crystals, Nature 438 (2005) 633–638.
[90] D. Alberga, O. Nicolotti, G. Lattanzi, G.P. Nicchia, A. Frigeri, F. Pisani, V. Benfenati, G.F. Mangiatordi, A new gating site in human aquaporin-4: insights from molecular dynamics simulations, Biochim. Biophys. Acta 1838 (2014) 3052–3060.