

MIPModDB: a central resource for the superfamily of major intrinsic proteins

Anjali Bansal Gupta¹, Ravi Kumar Verma¹, Vatsal Agarwal², Manu Vajpai¹, Vivek Bansal¹ and Ramasubbu Sankararamakrishnan^{1,*}

¹Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur 208016 and ²Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee 247667, India

Received August 12, 2011; Revised October 5, 2011; Accepted October 8, 2011

ABSTRACT

The channel proteins belonging to the major intrinsic proteins (MIP) superfamily are diverse and are found in all forms of life. Water-transporting aquaporin and glycerol-specific aquaglyceroporin are the prototype members of the MIP superfamily. MIPs have also been shown to transport other neutral molecules and gases across the membrane. They have internal homology and possess conserved sequence motifs. By analyzing a large number of publicly available genome sequences, we have identified more than 1000 MIPs from diverse organisms. We have developed a database MIPModDB which will be a unified resource for all MIPs. For each MIP entry, this database contains information about the source, gene structure, sequence features, substitutions in the conserved NPA motifs, structural model, the residues forming the selectivity filter and channel radius profile. For selected set of MIPs, it is possible to derive structure-based sequence alignment and evolutionary relationship. Sequences and structures of selected MIPs can be downloaded from MIPModDB database which is freely available at <http://bioinfo.iitk.ac.in/MIPModDB>.

INTRODUCTION

Major intrinsic proteins (MIPs) form one of the largest superfamily of channel proteins (1,2). They transport water, neutral solutes such as glycerol and urea, metalloids including antimonite and arsenite and gases like CO₂, nitric oxide and ammonia across the membranes (3–5). Water-transporting aquaporin and glycerol-specific aquaglyceroporin are the prominent members of this family. Members of MIP superfamily are involved in vital physiological processes such as skin moisture, gastrointestinal

fluid transport, fat metabolism, epidermal proliferation, maintaining corneal and lens transparency in eyes and water homeostasis in kidney and central nervous system (6–11). MIPs are found from bacteria to humans and a large number of diverse MIP members have been identified especially in plants (12–17). The abundant MIPs identified in higher plants can be classified into at least five major subfamilies (15). In humans, they are implicated in several diseases such as nephrogenic diabetes insipidus, acute and chronic renal failure, brain edema, cataract and arsenic toxicity (18–24). Sequence analysis of MIP members clearly revealed the presence of highly conserved Asn-Pro-Ala (NPA) signature sequence motif (25,26). MIP sequences also possess internal homology in which the N- and C-terminal halves have significant sequence similarity (25). At the structural level, the N-terminal half is related to the C-terminal half by a pseudo-2-fold symmetry (27,28). Three-dimensional structures of more than 10 MIPs from different organisms such as mammalian (29–31), plant (32), *Escherichia coli* (33), yeast (34) and archaea (35) have been determined. They all adopt a unique hourglass fold even when the sequence identity among them is very low. Two regions of constriction have been identified within the channel. The two NPA motifs form the central constriction and the outer constriction is toward the extracellular side known as aromatic/arginine (ar/R) selectivity filter formed by four residues. Both regions are known to play important role in the solute transport and selectivity (36–39). With their involvement in human physiology and pathophysiology and with the available structural knowledge, members of MIP family are being considered as attractive drug targets (40,41). For example, aquaglyceroporin in intracellular parasite *Plasmodium falciparum* and its structural knowledge (42) has opened new options for novel malaria therapies (43). However, for the majority of MIPs from large number of organisms, the tissue localization, their functional properties and the biological significance are either not known or they are not clearly understood.

*To whom correspondence should be addressed. Tel: +91 512 259 4014; Fax: +91 512 259 4009; Email: rsankar@iitk.ac.in

With the advent of new generation of sequencing technologies (44,45), genome sequences of large number of organisms are available. Using conserved sequence motifs and the internal sequence similarity as constraints, we have previously searched genome sequences of plants to identify and characterize the MIP proteins (15,16). We have extended this approach to genomes of other organism groups and identified more than 1000 MIP sequences in diverse organisms. The wealth of information on MIP sequences is now stored in a database called MIPModDB and various details about the substitutions in the conserved NPA motif, gene structure and the features obtained from structural models are available for each MIP sequence. With diverse MIP sequences, one can also obtain structure-based sequence alignment and evolutionary relationship for a given set of MIP sequences.

DATABASE CONSTRUCTION

MIP genes were identified using BLAST (46) from completed and partial genomes available in NCBI database and the protocol used for this purpose is the same as described in our previous studies (15,16). Thirteen human aquaporins and five plant MIPs belonging to the five major plant subfamilies were used as query sequences. Sequences of short length and those sequences with missing transmembrane segments and important loop regions were discarded. We also found that some MIPs have been wrongly annotated. For example, an MIP from *Salinispora tropica* is annotated as ‘low molecular weight phosphotyrosine protein phosphatase’ (NCBI accession: YP_001157491). MIPs from plants identified from previous work (13–16,47) were also included. We have also searched the motif-oriented database, MIPDB (48), and considered those MIP sequences not identified in the above BLAST search. The final data set contained 1008 MIP sequences from 341 different organisms. For each MIP sequence, presence of conserved NPA sequence motif and internal sequence similarity were examined and the sequences were also confirmed by pattern databases like Pfam (49) and PROSITE (50). The three-dimensional structure of each MIP sequence was modeled by homology modeling procedure using the same protocol that was applied earlier for plant MIPs (15,16). Structures of bovine AQP1, *E. coli* GlpF and archeal AQPM were used as template structures and their PDB (51) IDs are 1J4N, 1FX8 and 2F2B, respectively. The channel radius profile was calculated using the HOLE program (52) as described previously (16). Thus the contents of this database can be largely categorized into sequence and structural data and are explained in more details in the following sections. The important statistics related to sequence and structure data of MIPModDB is given in Table 1. Both the sequence and structure data for a representative MIP sequence is shown in Figure 1.

MIP SEQUENCE DATA

In general, only a single MIP has been identified in most of the microbial genomes. Plants have large number of

Table 1. Important statistics of MIPModDB

Number of MIP sequences	1008
Number of organisms	341
Substitutions in the NPA motifs	219
Only in the first NPA motif (loop B)	74
Only in the second NPA motif (loop E)	82
Both NPA motifs	63
MIPs with selectivity filter similar to water-channels ^a (FHTR + FHAR + FHCR + FHSR)	349
MIPs with selectivity filter similar to glycerol channels ^a (WGYR + WGFR + WGWR)	170
Experimentally determined MIP structures	38

^aThe selectivity filter is formed by four residues and the corresponding amino acids are given in one letter codes. The first and second residues come from the second and the fifth transmembrane segments, respectively. The other two residues are contributed by the loop E. See text for details.

MIPs in comparison to animals. For each MIP, its source and the NCBI accession ID are given. Each MIP sequence is also given a unique identifier derived from its scientific name. The first two characters are taken from its genus and the next four characters are from its species name followed by a four digit unique number. Wherever it is available, UNIPROT (53) accession ID is also provided. Apart from the primary structure information, sequence data includes exon–intron organization of the gene, substitution (if any) in the conserved NPA motif and percentage sequence similarity with the template sequences that are used in the homology modeling procedure to build three-dimensional models. Sequence similarity between a given MIP sequence and the template sequences was calculated using the program NEEDLE as available in the EMBOSS suite of programs (54). Only the modeled part of the target MIP sequence was considered for this purpose.

Gene structure

For each MIP, gene structure is represented in the form of a graphical diagram. It gives the length and the positions of the introns with respect to the secondary structures of the corresponding MIP. The red and blue vertical bars indicate the starting positions of helices and loops B/E, respectively. The information regarding the positions of each individual introns was extracted from the NCBI database annotations. Transmembrane segments are marked based on the modeled structures (see below). Knowledge of intron–exon organization helps to understand the evolution of MIPs across different organisms and MIP subfamilies within a same species. For example, in plants it has been shown that the number and positions of introns are conserved within a given MIP subfamily (14–16).

Substitutions in NPA motif

In addition to its role in solute transport and selectivity (36,55–57), substitutions in the highly conserved NPA motifs seem to be important in other functional roles such as protein targeting (58) and full expression of the protein (59). In our data set, substitutions in only the first

MIPModDB

Home Search Statistics Publications About us Feedback Help

HOSAPI0405

ACCESSION NUMBER : [NP_932766.1](#)
 ORGANISM : *Homo sapiens*
 ANNOTATION : aquaporin 1
 UNIPROT ACCESSION : [AQPI_HUMAN](#)
 PDB ID : [1HG1](#)

GENE STRUCTURE

Exons Introns Helix start position Loop B and Loop E start position

SEQUENCE BASED INFORMATION

Length of sequence	Length modeled		Similarity with templates (as calculated by NEEDLE program from EMBOSS Package)		
	from	to	1J4N	1FX8	2F2B
269	1	247	93.6	39.5	46.3

[View FASTA sequence](#)

STRUCTURE BASED INFORMATION

Residues in selectivity filter	NPA motif		Root mean squared deviation with templates (as calculated by DALI program)		
	I	II	1J4N	1FX8	2F2B
FHCR	NPA	NPA	1.3	2.1	1.8

[Structure based alignment](#)

View:
[Structure superimposed with 1FX8](#)
[Structure superimposed with 1J4N](#)

Downloads:
[Alignment in PIR Format](#)
[Alignment in PAP Format](#)
[Modelled Structure in PDB Format](#)
[Diameter profile along the channel](#)

Figure 1. Screenshot of a representative MIP protein, human aquaporin 1. Information about gene structure, substitutions in NPA motif, residues forming the ar/R selectivity filter, sequence similarity with the templates, RMSD calculated for the modeled structure with the three template structures are some of the features reported for a given MIP in the protein page.

NPA motif occur in 74 examples. In 82 cases, substitutions are found only in the second NPA. Both NPA motifs are substituted in 63 MIPs (Table 1). In total, substitutions in at least one of the NPA motifs are found in about 22% of the total MIPs in our data set. Majority of the substitutions involve mutation of either Pro or Ala of NPA motif. Only a handful of examples (less than 16) are found in which Asn is mutated indicating its important role in both in structure as a helix capping residue and function as a residue responsible for cation exclusion as demonstrated in recent studies (60).

DATA FROM MIP STRUCTURAL MODELS

Structure-based data includes the atomic model obtained using the homology modeling procedure, residues that

form the ar/R selectivity filter, structure-based sequence alignment, conservation of residues at the helix–helix interface and the HOLE radius profile. The structure-based details also include the root mean square deviation (RMSD) calculated for the modeled MIP and each of the template structure using the program DALI (61). The superposed figures are available in two different orientations (Figure 2A). Many MIPs have long N- and C-terminal extensions and hence the start and end positions of the polypeptide segment used in the homology modeling method are given.

Aromatic/arginine selectivity filter

Four residues form the outer constriction nearly 8 Å from the conserved NPA motif toward the extracellular side. These residues are contributed by second and fifth TM

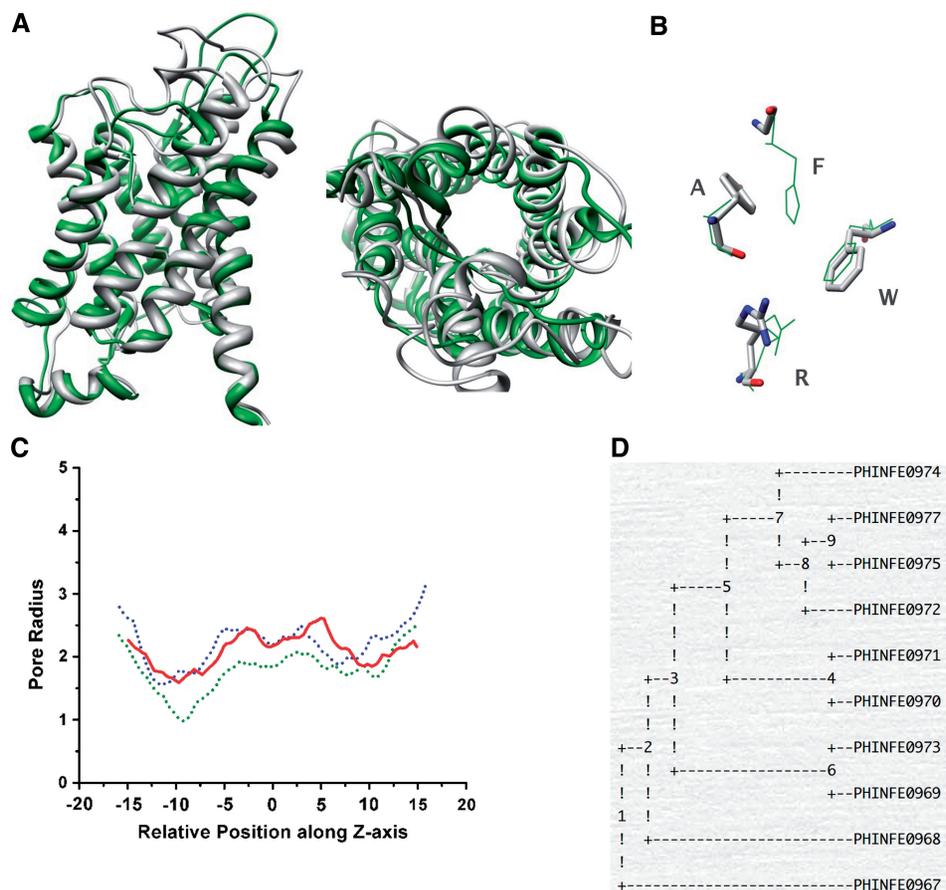


Figure 2. (A) Superposition of the modeled MIP structure with 1J4N, the structure of bovine aquaporin shown in two different orientations, namely, parallel (left) and perpendicular (right) to the channel axis. (B) The four residues of the ar/R selectivity filter superposed on that of 1J4N structure. (C) Comparison of HOLE radius profiles plotted for the water channel (green), glycerol facilitator (blue) and the modeled MIP structure (red). The ar/R selectivity filter is approximately located at -10 Å. (D) Phylogenetic tree calculated for all MIPs of a representative organism *Phytophthora infestans* using parsimony method.

segments and loop E. This aromatic/arginine (ar/R) selectivity filter has been implicated in obstructing the proton conduction (62) and efficient solute transport (37,63,64). The four ar/R selectivity filter residues, represented by their one letter codes, are given for each MIP. For example, 'FIIR' indicates that the first and second residues Phe and Ile are from TM2 and TM5, respectively, and the last two are the loop E residues. Analysis of selectivity filter residues indicates that there are 349 MIPs, which have selectivity filter (FHTR, FHAR, FHCR or FHSR) similar to that found in water-selective aquaporin channels (FHTR or FHCR). The number of MIPs having selectivity filter (WGYR, WGFR or WGWR) similar to the glycerol-specific aquaglyceroporin (WGFR) is 170. Thus, about 50% of the total MIPs in the database have selectivity filter typical of aquaporin or aquaglyceroporin. The remaining half has substitutions that can alter the size and chemical nature of the outer constriction. This will have major influence in the nature of solute that is being transported by the channel. The channel diameters of water channel and the aquaglyceroporin at the ar/R constriction are 2.0 and 3.5 Å, respectively. The selectivity filter residues of the predicted MIP model superposed on that

of the pure water channel from bovine aquaporin and that of aquaglyceroporin from *E. coli* are available (Figure 2B). The HOLE radius profiles of all three MIPs, water channel, glycerol channel and the modeled MIP channel, can be compared (Figure 2C). This will give an idea about the size of the region around the ar/R selectivity filter region with respect to both water channel and aquaglyceroporin and it will help the user to predict the possible size of a solute that can pass through this constriction.

Structure-based sequence alignment

As mentioned earlier, MIP sequences are diverse. For example, the sequence identities between some of the plant MIP subfamilies are as low as 20% (13,15). In such cases, programs such as ClustalW (65) that are used to generate multiple sequence alignment of a given set of sequences are unlikely to produce meaningful alignment for a diverse set of MIP sequences. Instead, if the structurally equivalent positions belonging to the TM helical segments are aligned, they are likely to result in high conservation and indicate the importance of residues in certain positions. Structure-based sequence alignment

for all the TM segments and the functionally important loops B and E is provided for all the MIPs present in the same organism along with the template sequences. We have also previously identified 17 positions that occur in the helix–helix interface. They are small and weakly polar residues and have been shown to be highly conserved if the amino acids Ala, Thr, Ser, Cys and Gly are considered as a group (15,16). These positions along with the residues that form the ar/R selectivity filter are highlighted in the structure-based sequence alignment. An example of structure-based sequence alignment as obtained from the MIPModDB database is shown in Figure 3.

Construction of phylogenetic tree for selected MIPs

The database provides an interface whereby the user will be able to analyze the evolutionary relationship among the selected group of MIPs. For constructing a phylogenetic tree, the user needs to select at least three sequences. The user will have the option to choose one of the three different methods, namely, neighbor-joining method, maximum likelihood method and maximum parsimony method to construct a phylogenetic tree. In addition to the complication in generating a multiple sequence alignment of diverse MIP sequences, many MIPs have long N- and C-termini as well as long loops connecting the TM segments. Hence to avoid errors, while constructing the evolutionary tree, the input used for this purpose is the structure-based sequence alignment of TM helical regions and the loops B and E. The program PHYLIP which is part of the EMBOSS suite of programs (54) is being utilized by the server to create the phylogenetic trees (Figure 2D).

IMPLEMENTATION

The information content on MIPs is maintained as a relational database using MySQL (<http://www.mysql.com>). This allows easy access and storage. The database is hosted on a web server running apache (<http://www.apache.org/>) on Fedora Core Linux platform and can be queried through the web interface which is implemented in PHP (v5.2) scripting language (<http://www.php.net/>).

ACCESS TO MIP DATA

MIPModDB allows the users to browse, retrieve and query the database. The Statistics page lists all the MIPs in three different categories: (i) based on the NPA sequence motif, (ii) selectivity filter residues and (iii) organism-wise grouping. Users can browse MIPs according to the conservation or substitutions that are found in the NPA boxes. One can also look for MIPs with particular residues that form the selectivity filter. MIPs are also organized as per the organism in which they occur. They are arranged in the descending order in which the model tree *Populus trichocarpa* is present on the top with the maximum number of 55 MIPs.

The database interface allows the user to retrieve and identify MIPs using various features of MIPs as query.

It can be searched by unique accession and complete or partial amino acid sequence. The search facility also allows the user to select MIP(s) with particular residues in the selectivity filter. Alternatively, MIPs with specific substitutions in the NPA motif also can be queried. A query based on an organism name will retrieve all MIP sequences from a particular species. More than one MIP features can also be used to narrow down the search. More detailed information can be retrieved by following the associated links. In addition to searching specific MIPs, the database enables users to download all the MIP sequences from a given organism or all the sequences that have the same ar/R selectivity filter residues in FASTA format. MIP sequences with specific substitutions in the NPA motif can also be downloaded. Sequence alignments used to generate the three-dimensional models can be downloaded in PIR or PAP format. For each MIP, the coordinates of the model are available in the PDB format. Similarly, the phylogenetic tree for a selected set of MIPs can be downloaded.

COMPARISON WITH OTHER TRANSPORTER DATABASES

There are databases which are developed specifically for membrane proteins that are involved in transporting the solutes across the membranes. The database TransportDB (66) provides information about the complete list of transporters for a given organism. For example, in the case of humans, the type of transporters listed include those that are ATP-dependent, ion channels, secondary transporters and unclassified. It also lists all outer membrane porins and channels from different organisms. The other database TCDB is a Transporter Classification Database (67) and is a classification system for membrane transporters. The superfamilies of transporters in this database include channel-forming toxins and peptides, transporters, symporters, antiporters, porins, carriers and ion channels. Both the above databases do not include all the MIP superfamily members whose members are known to predominantly transport neutral solutes. For example, while TransportDB lists human MIPs, MIP members in other organisms are not found. TCDB does not seem to include MIP superfamily although some MIP sequences are found. Moreover, while the above two databases largely contain sequence information and known PDB structures, MIPModDB provides structural models and associated information for more than 1000 MIP sequences.

FUTURE DIRECTIONS

Few MIPs have been functionally very well characterized and experimental studies are being carried out to determine the solute transport properties of many more MIPs. In the next version of MIPModDB database, whenever it is available, functional properties of MIPs, post-translational modification and cellular localization will be annotated and the related literature will be linked through PUBMED. As new MIPs are being recognized,

STRUCTURE BASED ALIGNMENT

A conserved positions at helix-helix interface **A** Ar/R selectivity filter

Protein	Helix 1	Helix 2
1J4N	FWRVAEFLAIFLFIITIGSALGF	DNVKVLAFLCLSIATLAQSVGH
1FX8	LKGQCIAEFLGGLLIFFVGCVAAL	GQWEIIVITGLGVAMAIYLTAG
1RC2	MFRKLAECFGFWLVFGCGSAVLA	GFAGVLAFLCLTLTMAFAVGH
1Z98	FWRAAIAEFIALLLFLYIVATVIGH	GLLGIWAFLGMIFVLVYCTAG
2B60	FWRAIFAEFFALFYVFFCLGASLRW	HVLQVLAFLCLLALATLVQAVGH
2F2B	LTKRCIAEFIGFILVFFCAGSAAVT	DWVAI LAFLFAIAASIYALGN
SIMELI0072	MFKKLCAEFLGCLWLVGCGSAVLA	GLLGVLAFLCLTLTMAIYTVGG
SIMELI0172	LTRRCVAEALGGLLVAAVWGSIGMA	ALVANLAFLCLLAILVVLVTLGP

Protein	Helix 3	Helix 4
1J4N	SVLRAIMYIIAQCVAIVATAILSG	SGQGLGIEIIGLQLVLCVLTAT
1FX8	DKRKVIPFIVSQVAFCMAALVYG	FVQAFAVEMVIAAILMGLILALT
1RC2	PAKEVGVYIAQVVGIVMAALLYL	MLSALVVELVLVAGFLLVIHICAT
1Z98	SLLRALVYMIACQLGICVGLVKA	KGTALGAEIIGFVLVYTVFAT
2B60	SLLRAICYVVAQLLQVAIAAVLYS	VGQATIVEIFLQLQFVLCIFATY
2F2B	PGREVPYIIAQLLQVAFCSFIFLQ	YWQAMLAEVVGLLMLITIMICIA
SIMELI0072	PAASLVSYVIAQVAQIIIAAVLYV	LTAALVTEVVMFFFLIILICST
SIMELI0172	TRRDCAAYVIAQVAQIACTALAHL	GPAQWLSEGVAVFGLVATILVGI

Protein	Helix 5	Helix 6
1J4N	SGPLAICFSVALGGLLAID	WIFWVCPFICGALVLIYDFI
1FX8	LAPLLIQLLAVICASMGP	LVPLFCPIVGVIVAFAYRKL
1RC2	FAPIAICLALTLICLISIP	WFFWVCPIVGCIIGLIYRTL
1Z98	LAPLPICFAVFMVCLATIP	WIFWVCPFICGAVVAAHYQYV
2B60	SVALAVCFSLTLGLFGLMY	WYVWVCPVICGGLSLLYDFL
2F2B	FAGIICLTVAGICLTLGN	SIYVIVPIVGAVLALTYQYL
SIMELI0072	FAPIAICLALTLICLVSIP	WLFWVCPVFCGAVIGIVVKS
SIMELI0172	AVPWLVLVYITAAVFT--	PGFVIVELLCVCLALMRWL

Protein	Loop B	Loop E
1J4N	AHLNPAVTLLLS	GGINPARFGSSVIT
1FX8	AHLNPAVTLLWLF	GAMNPARFGPKVFA
1RC2	GHFNPVTLLWAG	NSVNPARFATAI-F
1Z98	GHNPAVTLLFLA	GGINPARFGAAVIF
2B60	AHVNPVTLLFLVG	GAMNPARFAPAILT
2F2B	CHINPAVTLLWSV	GSLNPARFGPYLND
SIMELI0072	GHFNPVSLFLAVA	NSVNPARFGQALFV
SIMELI0172	AHFNPVSLFLALS	TFANPARFLARSFTN

Figure 3. Structure-based sequence alignment for a selected set of MIPs. This alignment is always produced with the six high-resolution MIP structures and their PDB IDs are also shown. This alignment is produced for the six transmembrane segments and the two functionally important loops B and E. The residues forming the ar/R selectivity filter are shown in the dark brown background. Seventeen positions previously identified to occur in the helix-helix interface (16) are highly group conserved when small and weakly polar residues are considered together as a group. They are displayed in cyan background.

the users will have the option to upload the sequences in the future version. Ultimately, our software will follow a pipeline procedure that will take an MIP sequence to series of steps which will include extracting the sequence features, building a homology model, identifying the selectivity filter residues, generating the HOLE radius profile and a possible prediction of the solutes that are likely to be transported.

ACKNOWLEDGEMENTS

We thank Vivek Modi for his help and all our lab members for the useful input. A.B.G. and R.K.V. would like to thank CSIR for a senior research fellowship.

FUNDING

The Department of Biotechnology, Government of India (BT/HRD/34/17/2008, partial); CSIR (senior research

fellowship to A.B.G. and R.K.V.). Funding for open access charge: Indian Institute of Technology Kanpur.

Conflict of interest statement. None declared.

REFERENCES

- Benga, G. (2009) Water channel proteins (later called aquaporins) and relatives: past, present and future. *IUBMB Life*, **61**, 112–133.
- Danielson, J.A.H. and Johanson, U. (2010) Phylogeny of major intrinsic proteins. In: Jahn, T.P. and Bienert, G.P. (eds), *MIPs and Their Role in the Exchange of Metalloids*, Vol. 679. Springer-Verlag, Berlin, pp. 19–32.
- Hachez, C. and Chaumont, F. (2010) Aquaporins: A family of highly regulated multifunctional channels. In: Jahn, T.P. and Bienert, G.P. (eds), *MIPs and their Role in the Exchange of Metalloids*, Vol. 679. Springer-Verlag, Berlin, pp. 1–18.
- Wu, B. and Beitz, E. (2007) Aquaporins with selectivity for unconventional permeants. *Cell. Mol. Life Sci.*, **64**, 2413–2421.
- Bienert, G.P., Bienert, M.D., Jahn, T.P., Boutry, M. and Chaumont, F. (2011) Solanaceae XIPs are plasma membrane

- aquaporins that facilitate the transport of many uncharged substrates. *Plant J.*, **66**, 306–317.
6. Ishibashi, K., Kondo, S., Hara, S. and Morishita, Y. (2011) The evolutionary aspects of aquaporin family. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, **300**, R566–R576.
 7. Ding, T., Gu, F., Fu, L. and Ma, Y. J. (2010) Aquaporin-4 in glioma invasion and an analysis of molecular mechanisms. *J. Clin. Neurosci.*, **17**, 1359–1361.
 8. Noda, Y., Sohara, E., Ohta, E. and Sasaki, S. (2010) Aquaporins in kidney pathophysiology. *Nat. Rev. Nephrol.*, **6**, 168–178.
 9. Hara-Chikuma, M. and Verkman, A. S. (2008) Roles of aquaporin-3 in the epidermis. *J. Invest. Dermatol.*, **128**, 2145–2151.
 10. Verkman, A. S., Ruiz-Ederra, J. and Levin, M. H. (2008) Functions of aquaporins in eye. *Prog. Retin. Eye Res.*, **27**, 420–433.
 11. Ma, T. H. and Verkman, A. S. (1999) Aquaporin water channels in gastrointestinal physiology. *J. Physiol. London*, **517**, 317–326.
 12. Zardoya, R. (2005) Phylogeny and evolution of the major intrinsic protein family. *Biol. Cell*, **97**, 397–414.
 13. Chaumont, F., Barrieu, F., Wojcik, E., Chrispeels, M. J. and Jung, R. (2001) Aquaporins constitute a large and highly divergent protein family in maize. *Plant Physiol.*, **125**, 1206–1215.
 14. Johanson, U., Karlsson, M., Johansson, I., Gustavsson, S., Sjövall, S., Frayssé, L., Weig, A. R. and Kjellbom, P. (2001) 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**, 1358–1369.
 15. Gupta, A. B. and Sankararamkrishnan, R. (2009) 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**, Art. No. 134.
 16. Bansal, A. and Sankararamkrishnan, R. (2007) 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**, Art. No. 27.
 17. Park, W., Scheffler, B. E., Bauer, P. J. and Campbell, B. T. (2010) Identification of the family of aquaporin genes and their expression in upland cotton (*Gossypium hirsutum* L.). *BMC Plant Biol.*, **10**, Art. No. 142.
 18. Loreto, C. and Reggio, E. (2010) Aquaporin and vascular diseases. *Curr. Neuropharmacol.*, **8**, 105–111.
 19. Verkman, A. S. (2009) Aquaporins: translating bench research to human disease. *J. Exp. Biol.*, **212**, 1707–1715.
 20. Morishita, Y., Sakube, Y., Sasaki, S. and Ishibashi, K. (2004) Molecular mechanisms and drug development in aquaporin water channel diseases. Aquaporin superfamily (superaquaporins): expansion of aquaporins restricted to multicellular organisms. *J. Pharmacol. Sci.*, **96**, 276–279.
 21. King, L. S., Kozono, D. and Agre, P. (2004) From structure to disease: the evolving tale of aquaporin biology. *Nat. Rev. Mol. Cell Biol.*, **5**, 687–698.
 22. Agre, P. and Kozono, D. (2003) Aquaporin water channels: molecular mechanisms for human diseases. *FEBS Lett.*, **555**, 72–78.
 23. Bedford, J. J., Leader, J. P. and Walker, R. J. (2003) Aquaporin expression in normal human kidney and in renal disease. *J. Am. Soc. Nephrol.*, **14**, 2581–2587.
 24. Loonen, A. J. M., Knoers, N. V. A. M., van Os, C. H. and Deen, P. M. T. (2008) Aquaporin 2 mutations in nephrogenic diabetes insipidus. *Semin. Nephrol.*, **28**, 252–265.
 25. Heymann, J. B. and Engel, A. (2000) Structural clues in the sequences of the aquaporins. *J. Mol. Biol.*, **295**, 1039–1053.
 26. Kruse, E., Uehlein, N. and Kaldenhoff, R. (2006) The aquaporins. *Genome Biol.*, **7**, Art. no. 206.
 27. Gonen, T. and Walz, T. (2006) The structure of aquaporins. *Quart. Rev. Biophys.*, **39**, 361–396.
 28. Fujiyoshi, Y., Mitsuoka, K., de Groot, B. L., Philippson, A., Grubmüller, H., Agre, P. and Engel, A. (2002) Structure and function of water channels. *Curr. Opin. Struct. Biol.*, **12**, 509–515.
 29. Ho, J. D., Yeh, R., Sandstrom, A., Chorny, I., Harries, W. E., Robbins, R. A., Miercke, L. J. and Stroud, R. M. (2009) Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance. *Proc. Natl Acad. Sci. USA*, **106**, 7437–7442.
 30. Sui, H., Han, B.-G., Lee, J. K., Walian, P. and Jap, B. K. (2001) Structural basis of water-specific transport through the AQP1 water channel. *Nature*, **414**, 872–878.
 31. Horsefield, R., Norden, K., Fellert, M., Backmark, A., Tornroth-Horsefield, S., van Scheltinga, T. A. C., Kvassman, J., Kjellbom, P., Johanson, U. and Neutze, R. (2008) High-resolution X-ray structure of human aquaporin 5. *Proc. Natl Acad. Sci. USA*, **105**, 13327–13332.
 32. Tornroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R. and Kjellbom, P. (2006) Structural mechanism of plant aquaporin gating. *Nature*, **439**, 688–694.
 33. Savage, D. F., Egea, P. F., Robles-Colmenares, Y., O'Connell, J. D. III and Stroud, R. M. (2003) Architecture and selectivity in aquaporins: 2.5 Å structure of aquaporin Z. *PLoS Biol.*, **1**, 334–340.
 34. Fischer, G., Kosinska-Eriksson, U., Aponte-Santamaria, C., Palmgren, M., Geijer, C., Hedfalk, K., Hohmann, S., de Groot, B. L., Neutze, R. and Lindkvist-Petersson, K. (2009) Crystal structure of a yeast aquaporin at 1.15 Å reveals a novel gating mechanism. *PLoS Biol.*, **7**, Art. No. e1000130.
 35. Lee, J. K., Kozono, D., Remis, J., Kitagawa, Y., Agre, P. and Stroud, R. M. (2005) Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 Å. *Proc. Natl Acad. Sci. USA*, **102**, 18932–18937.
 36. de Groot, B. L., Frigato, T., Helms, V. and Grubmüller, H. (2003) The mechanism of proton exclusion in the aquaporin-1 water channel. *J. Mol. Biol.*, **333**, 279–293.
 37. Beitz, E., Wu, B., Holm, L. M., Schultz, J. E. and Zeuthen, T. (2006) Point mutations in the aromatic/arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia and protons. *Proc. Natl Acad. Sci. USA*, **103**, 269–274.
 38. Wu, B. H., Steinbronn, C., Alsterfjord, M., Zeuthen, T. and Beitz, E. (2009) Concerted action of two cation filters in the aquaporin water channel. *EMBO J.*, **28**, 2188–2194.
 39. Fu, D., Libson, A., Miercke, L. J. W., Weitzman, C., Nollert, P., Krucinski, J. and Stroud, R. M. (2000) Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*, **290**, 481–486.
 40. Frigeri, A., Nicchia, G. P. and Svelto, M. (2007) Aquaporins as targets for drug discovery. *Curr. Pharm. Des.*, **13**, 2421–2427.
 41. Castle, N. A. (2005) Aquaporins as targets for drug discovery. *Drug Discov. Today*, **10**, 485–493.
 42. Newby, Z. E. R., O'Connell, J. D. III, Robles-Colmenares, Y., Khademi, S., Miercke, L. J. W. and Stroud, R. M. (2008) Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum*. *Nat. Struct. Mol. Biol.*, **15**, 619–625.
 43. Kun, J. F. and de Carvalho, E. G. (2009) Novel therapeutic targets in *Plasmodium falciparum*: aquaglyceroporins. *Expert Opin. Ther. Targets*, **13**, 385–394.
 44. Morozova, O. and Marra, M. A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics*, **92**, 255–264.
 45. Lister, R., Gregory, B. D. and Ecker, J. R. (2009) Next is now: new technologies for sequencing genomes, transcriptomes, and beyond. *Curr. Opin. Plant Biol.*, **12**, 107–118.
 46. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
 47. Danielson, J. A. H. and Johanson, U. (2008) Unexpected complexity of the aquaporin gene family in the moss *Physcomitrella patens*. *BMC Plant Biol.*, **8**, Art. No. 45.
 48. El Karkouri, K., Gueune, H. and Delamarque, C. (2005) MIPDB: a relational database dedicated to MIP family proteins. *Biol. Cell*, **97**, 535–543.
 49. Finn, R. D., Mistry, J., Tate, J., Cogill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K. et al. (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–D222.

50. Sigrist,C.J.A., Cerutti,L., de Castro,E., Langendijk-Genevaux,P.S., Bulliard,V., Bairoch,A. and Hulo,N. (2010) PROSITE: a protein domain database for functional characterization and annotation. *Nucleic Acids Res.*, **38**, D161–D166.
51. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
52. Smart,O.S., Neduvilil,J.G., Wang,X., Wallace,B.A. and Sansom,M.S.P. (1996) HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.*, **14**, 354–360.
53. The UniProt Consortium. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.*, **39**, D214–D219.
54. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European molecular biology open software suite. *Trends Genet.*, **16**, 276–277.
55. de Groot,B.L. and Grubmuller,H. (2001) Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science*, **294**, 2353–2357.
56. Chakrabarti,N., Tajkhorshid,E., Roux,B. and Pomes,R. (2004) Molecular basis of proton blockage in aquaporins. *Structure*, **12**, 65–74.
57. Jensen,M.O., Tajkhorshid,E. and Schulten,K. (2001) The mechanism of glycerol conduction in aquaglyceroporins. *Structure*, **9**, 1083–1093.
58. Guan,X.G., Su,W.H., Yi,F., Zhang,D., Hao,F., Zhang,H.G., Liu,Y.J., Feng,X.C. and Ma,T.H. (2010) NPA motifs play a key role in plasma membrane targeting of aquaporin-4. *IUBMB Life*, **62**, 222–226.
59. Ikeda,M., Andoo,A., Shimono,M., Takamatsu,N., Taki,A., Muta,K., Matsushita,W., Uechi,T., Matsuzaki,T., Kenmochi,N. *et al.* (2011) The NPC motif of aquaporin-11, unlike the NPA motif of known aquaporins, is essential for full expression of molecular function. *J. Biol. Chem.*, **286**, 3342–3350.
60. Wree,D., Wu,B.H., Zeuthen,T. and Beitz,E. (2011) Requirement of asparagine in the aquaporin NPA sequence signature motifs for cation exclusion. *FEBS J.*, **278**, 740–748.
61. Holm,L. and Rosenstrom,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
62. Li,H., Chen,H.N., Steinbronn,C., Wu,B.H., Beitz,E., Zeuthen,T. and Voth,G.A. (2011) Enhancement of proton conductance by mutations of the selectivity filter of aquaporin-1. *J. Mol. Biol.*, **407**, 607–620.
63. Aponte-Santamaria,C., Hub,J.S. and de Groot,B.L. (2010) Dynamics and energetics of solute permeation through the *Plasmodium falciparum* aquaglyceroporin. *Phys. Chem. Chem. Phys.*, **12**, 10246–10254.
64. Savage,D.F., O'Connell,J.D. III, Miercke,L.J.W., Finer-Moore,J. and Stroud,R.M. (2010) Structural context shapes the aquaporin selectivity filter. *Proc. Natl Acad. Sci. USA*, **107**, 17164–17169.
65. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) Clustal W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
66. Ren,Q., Chen,K. and Paulsen,I.T. (2007) TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.*, **35**, D274–D279.
67. Saier,M.H.J., Yen,M.R., Noto,K., Tamang,D.G. and Elkan,C. (2009) The transporter classification database: recent advances. *Nucleic Acids Res.*, **37**, D274–D278.