J. Membrane Biol. 153, 171–180 (1996)

The Journal of
Membrane
Biology

Topical Review

Phylogenetic Characterization of the MIP Family of Transmembrane Channel Proteins

J.H. Park, M.H. Saier, Jr.
Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116

Received: 19 April 1996/Revised: 3 June 1996

Abstract. The ubiquitous major intrinsic protein (MIP) family includes several transmembrane channel proteins known to exhibit specificity for water and/or neutral solutes. We have identified 84 fully or partially sequenced members of this family, have multiply aligned over 50 representative, divergent, fully sequenced members, have used the resultant multiple alignment to derive current MIP family-specific signature sequences, and have constructed a phylogenetic tree. The tree reveals novel features relevant to the evolutionary history of this protein family. These features plus an evaluation of functional studies lead to the postulates: (i) that all current MIP family proteins derived from two divergent bacterial paralogues, one a glycerol facilitator, the other an aquaporin, and (ii) that most or all current members of the family have retained these or closely related physiological functions.

Key words: Transmembrane channels — Aquaporins — Glycerol facilitators — Nodulins — Evolution — Phylogenetic analyses — MIP family

Introduction

The ubiquitous major intrinsic protein (MIP) family of transmembrane channel proteins is an old, but not ancient family, dating back perhaps 2.5 to 3 billion years in evolutionary time (Pao et al., 1991; Saier, 1994*a*). It arose by an intragenic duplication event that preceded diversification of the family (Pao et al., 1991; Reizer, Reizer & Saier, 1993; Saier, 1994*a*,*b*). These proteins are known to selectively transport water, small neutral solutes, and possibly ions (Maurel et al., 1993, 1994; Ishibashi et al., 1994; Knepper, 1994; Weaver et al., 1994; Kashiwagi, Kanamary & Mizuno, 1995; Luyten et al., 1995). Topological and three-dimensional structural analyses have been reported for one member of this family, the aquaporin CHIP protein (Jung et al., 1994; Preston, 1994; Walz et al., 1994; Jap & Li, 1995).

The recent flurry of research activity designed to elucidate functional and structural attributes of MIP family proteins relates to the physiological importance of animal and plant aquaporins, of which many have been sequenced and characterized. In animals, these proteins are abundant in erythrocytes, renal proximal tubules and several other tissues where their activities are osmotically sensitive and hormonally regulated, allowing cell volume regulation and organismal fluid retention (Knepper, 1994. In plants, homologous channel proteins in vacuolar and plasma membranes in part allow controlled water movement through the tissues, from the roots, up conductive tissues of the stem, to the leaves where water vapor is released (Chrispeels & Agre, 1994). In microorganisms, the physiological functions of recently discovered aquaporins are completely unknown (Calamita et al., 1995).

In 1993, we reported a phylogenetic study of the MIP family proteins that had been sequenced at that time (Reizer et al., 1993). Sequences of eighteen genes and gene fragments were then available for analysis. In the intervening time, that number has increased over 4-fold, largely due to extensive sequencing of bacterial, yeast, plant and animal genomes. This increase in sequence data together with new functional information warranted reinvestigation of the phylogenetic origins of MIP family proteins. In this topical review, we present the results of

Correspondence to: M.H. Saier

Table 1. Proteins of the MIP family analyzed in this study

Abb[a]	Reference #	Accesion #	Protein designation	Source	#AA[b]
ANIMAL 1					
Aqp1(Hsa)	AQP1_HUMAN	P29972	Aquaporin 1	*Homo sapiens* (human)	269
Aqp2(Hsa)	AQP2_HUMAN	P41181	Aquaporin 2	*Homo sapiens*	271
Aqp5(Rno)	A55630	A55630	Aquaporin 5	*Rattus norvegicus* (rat)	265
Aqp(Bma)	BMU22658	U22658	Aquaporin	*Bufo marinus* (giant toad)	273
Aqp4(Rno)	RNU14007	U14007	Aquaporin-4 water channel (AQP4)	*Rattus norvegicus*	324
Aqp(Res)	RANCAQ	L24754	Aquaporin (AQPA)	*Rana esculenta* (frog)	273
Mip(Xle)	JN0557	JN0557	Major intrinsic protein	*Xenopus laevis* (african clawed frog)	262
Mip(Hsa)	MIP-HUMAN	P30301	Major intrinsic protein	*Homo sapiens*	263
Mer(Hsa)	HSU34845	U34845	Mercurial-insensitive water channel	*Homo sapiens*	302
Mer(Mmu)	MMU33012	U33012	Mercurial-insensitive water channel	*Mus musculus* (mouse)	301
Bib(Dme)	BIB_DROME	P23645	Neurogenic protein Big Brain	*Drosophila melanogaster* (fly)	700
ANIMAL 2					
Aqp3(Rno)	RATAQP	L35108	Aquaporin 3	*Rattus norvegicus*	293
Orf1(Cel)	U20864	U20864	F32A5.5 gene product	*Caenorhabditis elegans* (worm)	295
Orf2(Cel)	CELK02G10	U40415	K02G10.7 gene product	*Caenorhabditis elegans*	290
Orf3(Cel)	CELM02F4	U41548	MO2F4.8 gene product	*Caenorhabditis elegans*	302
Orf4(Cel)	Z35595	Z35595	Ce1G6.1 gene product	*Caenorhabditis elegans*	290
PLANT 1					
αTip(Ath)	TIPA_ARATH	P26587	Tonoplast intrinsic protein	*Arabidopsis thaliana*	268
Tip(Pvu)	JQ1106	JQ1106	Tonoplast intrinsic protein	*Phaseolus vulgaris* (kidney beans)	289
rtTip(Ath)	TIPR_ARATH	P21652	Tonoplast intrinsic protein	*Arabidopsis thaliana*	253
Tip1(Nta)	TIP1_TOBAC	P21653	Tonoplast intrinsic protein	*Nicotiana tabacum* (tabacco)	250
Tip(Tre)	TRGTIPLP	Z29946	Tonoplast intrinsic protein	*Trifolium repens* (clover)	248
Cha(Ama)	DIP_ANTMA	P33560	Membrane protein	*Antirrhinum majus* (snapdragon)	250
Cha(Csp)	CUCMP28B	D45078	Membrane protein 28	*Cucurbita sp.*	269
Nod2(Gma)	JQ2288	JQ2288	Nodulin-26	*Glycine max* (soybean)	255
PLANT 2					
MipA(Mcr)	CIPMIPA	L36095	Major intrinsic protein A	*Mesembryanthemum crystallinum*	282
MipB(Mcr)	CIPMIPB	L36097	Major intrinsic protein B	*Mesembryanthemum crystallinum*	286
MipD(Mcr)	MCU26537	U26537	Major intrinsic protein D	*Mesembryanthemum crystallinum*	286
MipE(Mcr)	MCU26538	U26538	Major intrinsic protein E	*Mesembryanthemum crystallinum*	254
Mip(Ath)	544083	s44083	Major intrinsic protein	*Arabidopsis thaliana* (mouse-ear)	286
wTip(Psa)	TIPW_PEA	P25794	Tonoplast intrinsic protein	*Pisum sativum* (garden pea)	289
Tip(Hvu)	S41194	S41194	Transmembrane protein	*Hordeum vulgare* (barley)	288
Cha(Zma)	S49590	S49590	Transmembrane protein	*Zea mays* (maize)	287
Aqp(Aca)	AQUA_ATRCA	U18403	Aquaporin	*Atriplex canescens* (4-winged saltbush)	283
Aqp(Gma)	GMU27347	U27347	Glycine max water channel	*Glycine max* (soybean)	286
Ram(Les)	S34650	S34650	Ripening associated membrane protein	*Lycopersicon esculentum* (tomato)	214
PLANT 3					
Nod(Gma)	NO26_SOYBN	P08995	Nodulin-26	*Glycine max* (soybean)	271
Mip(Nal)	NAU20490	U20490	Major intrinsic protein	*Nicotiana alata* (tabacco)	271
Mip(Osa)	RICYK347	D17443	Major intrinsic protein	*Oryza sativa* (rice)	285

Continued on next page

Table 1. Continued

Abb[a]	Reference #	Accesion #	Protein designation	Source	#AA[b]
YEAST 1					
Orf1(Sce)	YSCL9677_7	U25841	P9677.5 gene product	*Saccharomyce cerevisiae* (baker's yeast)	306
YEAST 2					
Fps(Sce)	FPS1_YEAST	P23900	Glycerol uptake/efflux facillitator	*Saccharomyces cerevisiae*	669
YEAST 3					
Orf2(Sce)	TFF4_YEAST	P43549	Hypothetical 70.5 KD protein	*Saccharomyces cerevisiae*	646
BACTERIAL 1					
Orf(Lla)	YDP1_LACLC	P22094	Hypothetical protein	*Lactococcus lactis*	289
Glp2(Hin)	HEAHI1017	L45655	Glycerol uptake facilitator	*Haemophilus influenzae*	226
Glp(Mga)	MGU35010	U35010	Glycerol uptake facilitator	*Mycoplasma gallisepticum*	206
Glp(Mge)	MGU39682	U39682	Glycerol uptake facilitator	*Mycoplasma genitalium*	221
Glp(BsuP	GLPF_BACSU	P18156	Glycerol uptake facilitator	*Bacillus subtilis*	274
Glp(Spn)	SPU12567	U12567	Glycerol uptake facilitator	*Streptococcus pneumoniae*	234
BACTERIAL 2					
Glp(Eco)	P11244	P11244	Glycerol uptake facilitator	*Escherichia coli*	281
Glp1(Hin)	HUI32752	U32752	Glycerol uptake facilitator	*Haemophilus influenzae*	265
Pdf(Sty)	PDUF_SALTY	P37451	Propanediol diffusion facilitator	*Salmonella typhimurium*	264
BACTERIAL 3					
AqpZ(Eco)	ECU38664	U38664	Aquaporin Z protein	*Escherichia coli*	232
BACTERIAL 4					
SmpX(Ssp)	SYOSMPX	D43774	Channel protein	*Synechococcus sp.*	270

[a] Abbreviation used throughout this study
[b] Number of amino acids in the protein

our phylogenetic studies which lead to important predictions regarding the evolutionary origins and functional relationships of MIP family proteins.

Current MIP Family Members

Tables 1 and 2 list all currently sequenced proteins of the MIP family. The fifty-two proteins listed in Table 1, including representatives of all major subclusters of the MIP family, were selected for the phylogenetic analyses reported. Table 2 lists all remaining sequenced MIP family proteins which either exhibit a high degree of sequence similarity with a protein included in Table 1 (*see* last two columns of Table 2) or were incompletely sequenced. As Table 2 lists 32 proteins, the current MIP family includes 84 sequenced or partially sequenced protein members retrieved from the databases.

The information tabulated in Tables 1 and 2 gives the recognized names and organismal origins of the MIP family proteins as well as the protein abbreviations to be used throughout this study. The clustering patterns (*see* Fig. 2 below) and the sizes of the proteins are also indicated. As can be seen, most of these proteins are less than 300 residues in length, although a few of the eukaryotic proteins are substantially longer. For example, two of the three yeast proteins are of greater than 600 residues in length, and Big Brain (Bib) of *Drosophila melanogaster*, the largest protein in the family, is reported to be 700 residues long. The larger sizes of these proteins are due primarily to the presence of extended N-terminal and C-terminal hydrophilic domains.

Multiple Alignment of MIP Family Proteins

Multiple alignments of the most highly conserved portions of 52 MIP family proteins are shown in Fig. 1*A* and *B*. The alignment shown in Fig. 1*A* includes the well-conserved NPA motif found in the first repeat units of these proteins while that shown in Fig. 1*B* includes the NPA motif found in the second repeat unit (Pao et al., 1991; Reizer et al., 1993). Within the NPA motif of the first repeat unit, the Ns and Ps are fully conserved, and the A is replaced by an S in just one protein, Fps (Sce)

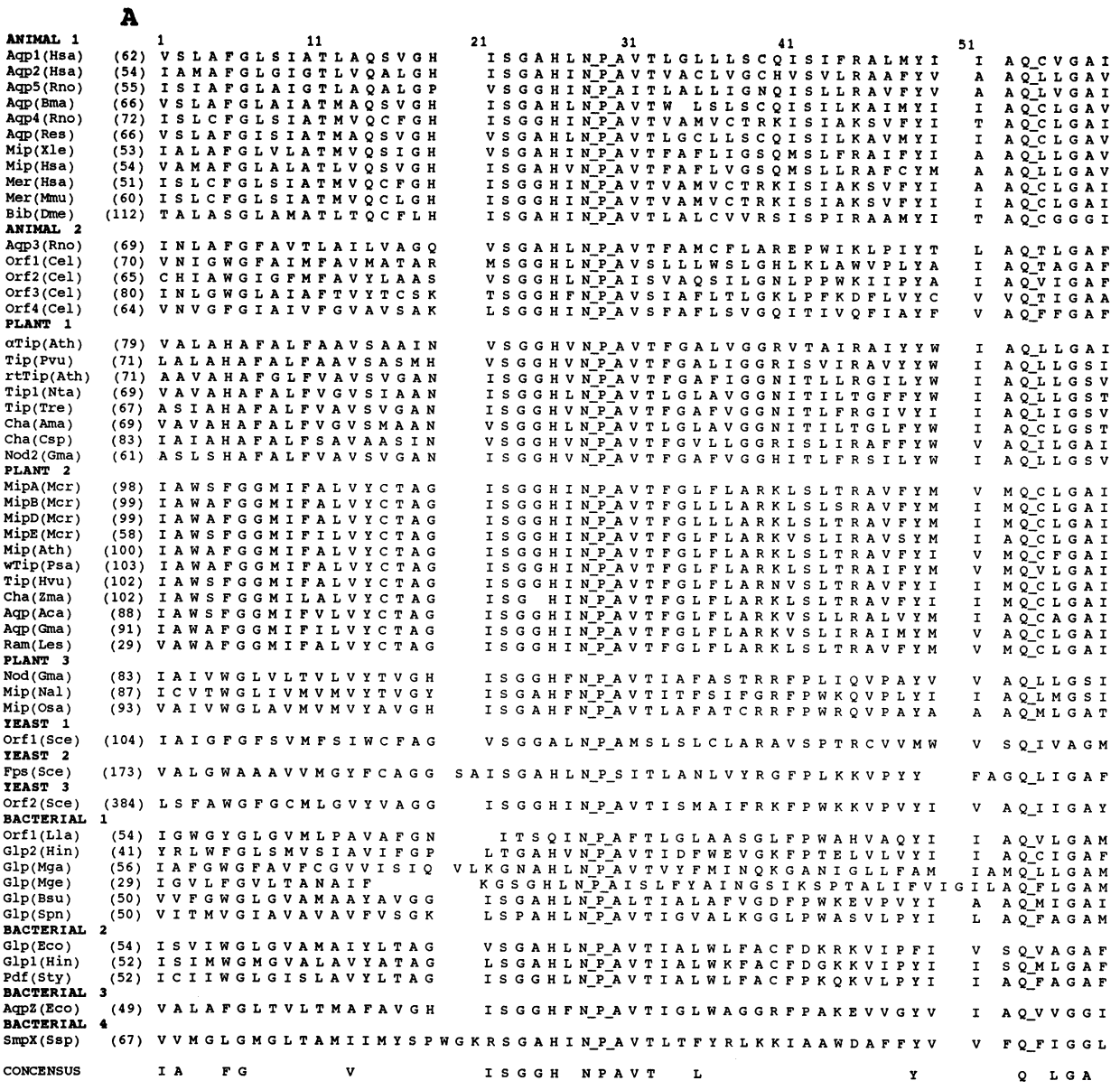

Fig. 1. Multiple alignment of two well-conserved regions of 52 divergent members of the MIP family. The alignment was performed using the PREALIGN program of Feng and Doolittle (1990). Proteins included are indicated by the abbreviations provided in Table 1. Residue number in each protein is indicated in parenthesis at the beginning and end of each aligned sequence. Alignment position and the consensus sequence (CONSENSUS) (27 or more residues at any one position conserved) are provided above and below the multiple alignment, respectively. Part B of Figure 1 is found on the facing page.

(Fig. 1*A*). Within the NPA motif in the second repeat unit, the Ns and As are fully conserved, and the P is replaced by an L in one protein, again Fps(Sce). It will be shown below that Fps(Sce) is one of the most divergent members of the MIP family.

The multiple alignments shown in Fig. 1*A* and *B* revealed two homologous regions of striking sequence similarity which proved to be suitable for signature sequence derivation.

Signature sequence #1 (alignment positions 26-33 in Fig. 1*A*) is:

(HQA)-(LIVF)-N-P-(AS)-(LIVMF)-(TS)-(LIVFW)

Signature sequence #2 (alignment positions 26-36 in Fig. 1*B*) is:

(LIVM)-N-(PL)-A-(LIVR)-(SATD)-(LIVFT)-(GAM)-(LIVPSTA)-(ASKR)-(LIVMF)

Both of these signature sequences were screened against the SwissProt database. They retrieved only protein members of the MIP family. These sequences are an update of those reported previously (Reizer et al., 1993). They should prove useful in identifying new members of the MIP family as these become sequenced.

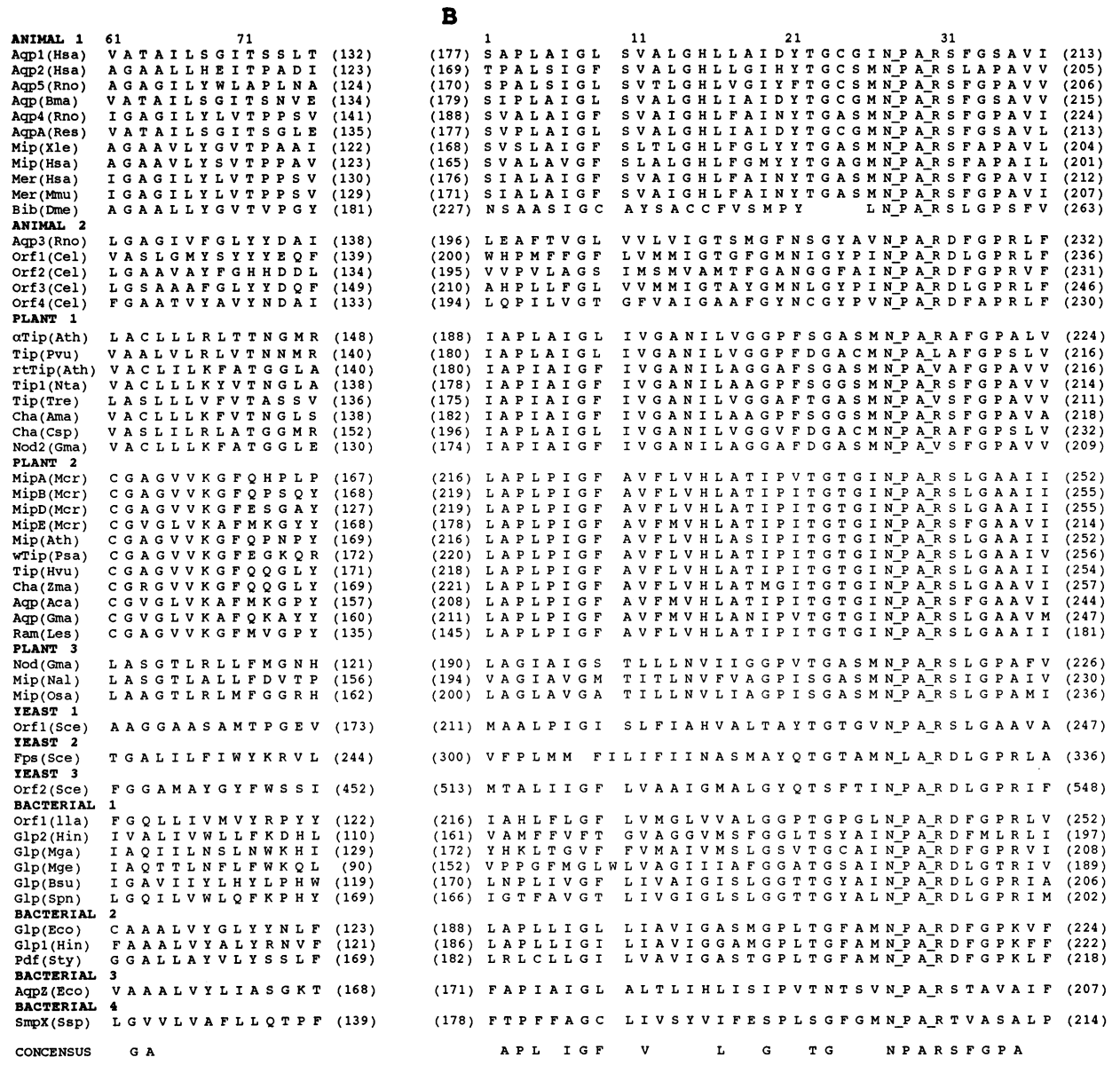

Fig. 1. *Continued*

Phylogenetic Characterization of the MIP Family Proteins

The phylogenetic tree for the 52 MIP family proteins listed in Table 1 is shown in Fig. 2. All currently sequenced MIP family proteins (including those listed in Table 2) fall into twelve subfamilies, four of bacterial origin, three of yeast origin, three of plant origin, and two of animal origin. The subfamilies from each of these four phylogenetically divergent groups of organisms will be considered separately.

Bacterial Proteins

There are four major types of bacterial isoforms of MIP family proteins. All of the six proteins included in bacterial subfamily #1 (plus three fragments included in Table 2) are probably glycerol facilitators. Five of these proteins are from Gram-positive bacteria, and only Glp2(Hin) is from a Gram-negative bacterium (*Haemophilus influenzae*). Of these proteins, only Glp2(Hin) is encoded by a gene that is not included within an operon that also encodes proteins concerned with glycerol catabolism. Perhaps the gene encoding Glp2(Hin) was acquired from a Gram-positive bacterium as a result of horizontal gene transmission. The gene encoding Glp2(Hin) was found to exhibit 37% G+C while Glp1(Hin) proved to possess 43% G+C. The overall G+C content of the *H. influenzae* genome is 38% (Fleischmann et al., 1995), similar to that of many low G+C Gram-positive bacteria.

Table 2. Sequenced proteins of the MIP family not included in this study

Reference #	Accession #	Protein designation	Source	#AA[a]	Similar protein[b]	% Identity[b]
ANIMAL 1						
AQP1_MOUSE	Q02013	Aquaporin	*Mus musculus* (mouse)	269	Aqp1(Hsa)	94%
APQ1_RAT	P29975	Aquaporin	*Rattus norvegicus* (rat)	269	Aqp1(Hsa)	92%
JT0749	JT0749	Water channel protein	*Rattus norvegicus*	269	Aqp1(Hsa)	92%
S37639	S37639	Chip-28 protein	*Rattus norvegicus*	269	Aqp1(Hsa)	92%
A44395	A44395	Proximal tubule water transporter	*Rattus norvegicus*	269	Aqp1(Hsa)	93%
JC1320	JC1320	Chip 28	*Rattus norvegicus*	269	Aqp1(Hsa)	90%
JC2348	JC2348	Chip 29	*Bos primigenius taurus* (cattle)	271	Aqp1(Hsa)	90%
AQP2_RAT	P34080	Aquaporin	*Rattus norvegicus*	271	Aqp2(Hsa)	90%
MIP_RANPI	Q06019	Major intrinsic protein	*Rana pipien* (northern leopard frog)	263	Mip(Xle)	100%
MIP_BOVINE	P06624	Major intrinsic protein	*Bos taurus*	263	Mip(Hsa)	92%
MIP_RAT	P09011	Major intrinsic protein	*Rattus norvegicus*	261	Mip(Hsa)	93%
HSU34846	U34846	Mercurial-insensitive water channel	*Homo sapien* (human)	342	Mer(Hsa)	99%
PLANT 1						
JQ1106	JQ1106	Tonoplast 27K intrinsic protein	*Phaseolus vulgaris* (kidney bean)	256	Tip(Pvu)	96%
TIPG_ARATH	P25818	Tonoplast intrinsic protein	*Arabidopsis thaliana*	251	rtTip(Ath)	96%
RICYK333	D25534	Tonoplast intrinsic protein	*Oryza sativa*	251	rtTip(Ath)	78%
TIP2_TOBAC	P24422	Tonoplast intrinsic protein	*Nicotiana tabacum* (tabacco)	250	Tip1(Nta)	99%
S45406	S45406	Tonoplast intrinsic protein	*Nicotiana tabacum* (tabacco)	251	Tip1(Nta)	96%
CUCMP23A	D45077	Membrane protein 23	*Oucurbita sp.*	280	Cha(Csp)	88%
PLANT 2						
ATHPTP	D26609	Putitive transmembrane protein	*Arabidopsis thaliana*	288	MipB(Mcr)	90%
WC1A_ARATH	P43285	Plasma membrane intrinsic prot. 1A	*Arabidopsis thaliana*	286	MipB(Mcr)	86%
WC1B_ARATH	Q06611	Plasma membrane intrinsic prot. 1B	*Arabidopsis thaliana*	286	MipB(Mcr)	86%
WC1C_ARATH	Q08733	Plasma membrane intrinsic prot. 1C	*Arabidopsis thaliana*	286	MipB(Mcr)	86%
WC2A_ARATH	P43286	Plasma membrane intrinsic prot. 2A	*Arabidopsis thaliana*	287	MipD(Mcr)	76%
WC2B_ARATH	P43287	Plasma membrane intrinsic prot. 2B	*Arabidopsis thaliana*	285	MipD(Mcr)	78%
WC2C_ARATH	P30302	Plasma membrane intrinsic prot. 2C	*Arabidopsis thaliana*	285	MipD(Mcr)	78%
S42556	S42556	Hypothetical protein	*Arabidopsis thaliana*	286	Mip(Ath)	93%
FRAGMENTS						
MIP_CHICK	A37203	Major intrinsic protein	*Gallus gallus* (Fragment)	112	Mip(Hsa)	80%
CIPMIPC	L36096	Major intrinsic protein C	*Mesembryanthemum crystallinum* (Fragment)	194	MipA(Mcr)	78%
GLPF_SHIFL	P31140	Glycerol facilitator	*Shigella flexneri* (Fragment)	214	Glp(Eco)	100%
GYLA_STRCO	P19255	Glycerol facilitator	*Streptomyces coelicolor* (Fragment)	80	Glp(Bsu)	54%
MC140	Z33098	Glycerol facilitator	*Mycoplasma capricolum* (Fragment)	90	Glp(Bsu)	42%
CPGLGF	X86492	Glycerol facilitator	*Clostridium perfringens* (Fragment)	149	Glp(Bsu)	60%

[a] Number of amino acids in the protein

[b] The proteins listed in this column are included in Table 1 and are the closest homologues of the proteins indicated in column 1. These two proteins exhibit the percent identity indicated in the last column.

Bacterial subfamily #2 includes three proteins (Table 1) plus one fragment (Table 2), all from closely related Gram-negative bacteria. They are all believed to be glycerol facilitators. This subfamily includes the well-characterized glycerol facilitator of *E. coli* (Glp Eco) (Heller, Lin & Wilson, 1980) which when incorporated into frog oocyte membranes efficiently transports glycerol but not water (Maurel et al., 1993, 1994). Glp1(Hin) is encoded by a gene that is cotranscribed with a glycerol kinase-encoding gene. The sequenced protein from *S. typhimurium* (Pdf(Sty)) presumably transports the close glycerol analogue, propanediol, as it is found within the *pdu* operon, concerned with the catabolism of this compound (Chen, Andersson & Roth, 1994).

The third bacterial subfamily includes a single protein, AqpZ(Eco). This protein has been shown to be a mercury-insensitive aquaporin. It cannot transport glycerol or urea as does the *E. coli* glycerol facilitator (Calamita et al., 1995). The genes encoding the two *E. coli* MIP family proteins (*aqpZ* and *glpF*) proved to exhibit 51-52% G+C content, the same as for the *E. coli* genome.

The fourth bacterial subfamily includes a single sequenced protein from the blue-green bacterium *Synechococcus* (Kashiwagi et al., 1995). The biochemical function of this protein is not known, but on the basis of an *in vivo* genetic analysis, it has been proposed to play a role in Cu^{2+} homeostasis (Kashiwagi et al., 1995). This fact might suggest that it plays a role in ion transport.

The presence of four bacterial subfamilies, three of which are represented by the proteins found in *E. coli* and *H. influenzae*, argues that a single bacterium may possess two, and possibly three distantly related MIP family paralogues that diverged from each other early in evolutionary time. The proteins that comprise subfamilies 1 and 2 are probably glycerol facilitator orthologues that derived from a functionally similar ancestral protein common to Gram-positive and Gram-negative bacteria. AqpZ(Eco) is an aquaporin of unknown physiological function, while SmpX(Ssp), implicated in the physiological process of Cu^{2+} homeostasis, has not been characterized biochemically.

Yeast Proteins

Most (>90%) of the *Saccharomyces cerevisiae* genome had been sequenced and deposited into the databases as of the completion of the analyses reported here. Three highly divergent yeast MIP family isoforms were identified. Because of their distant relatedness, we have put each into a distinct subfamily (yeast subfamilies #1, #2, and #3; Fig. 2). One of these, Fps(Sce), has been shown to exhibit glycerol facilitator activity and may play a role in osmoregulation (Luyten et al., 1995). Expression of the FPSI gene encoding Fps(Sce) suppresses the growth defect on fermentable sugars of the *fdp1* mutant which exhibits defective control of glycolysis (Van Aelst et al., 1991; Luyten et al., 1995). The other yeast homologues have not been functionally characterized. Because these three proteins are nearly as distant from each other as they are from the four subfamilies of bacterial proteins, they probably did not arise as a result of gene duplication events that occurred in yeast. They may have been acquired vertically from one or more ancestral organism(s) or horizontally from other organisms. The genes encoding these three proteins proved to exhibit 51% (Orf1), 43% (Fps) and 43% (Orf2) G+C content. The overall G+C content of *S. cerevisiae* is 40% (Lloyd & Sharp, 1992). Thus, Orf1 may have been obtained by horizontal transmission. Based on their phylogenetic positions relative to proteins of known function, we suggest that Fps and Orf2 are glycerol facilitator isoforms while Orf1 is an aquaporin.

Plant Proteins

Plants possess three MIP subfamilies, all of which have many members. Members of subfamilies #1 and #2 are found in various plant tissues, and several have been localized to the vacuolar (tonoplast) or plasma membrane (Höfte et al., 1992; Chrispeels & Agre, 1994; Chrispeels & Maurel, 1994; Engel, Walz & Agre, 1994). Most of these proteins are believed to be aquaporins. Since several of the proteins in plant subfamilies #1 and #2 are from *Arabidopsis thaliana* (Tables 1 and 2), several of those in plant subfamily #2 are from *Mesembryanthemum crystallinum* (MipA-E Mcr; Tables 1 and 2), and representatives in all three plant subfamilies are from *Glycine max* (Table 1), it is clear that higher plants possess a number of closely related as well as distantly related MIP family paralogues. The closely related paralogues undoubtedly arose by gene duplication events that occurred after the plant kingdom diverged from the animal and fungal kingdoms.

One well-characterized protein of plant subfamily #3, the nodule-specific symbiosome protein from the soybean, Nod(Gma) (Verma, 1992), has been reported to be capable of transporting ions (Weaver et al., 1994). However, another member of this cluster (Mip Nal) is probably an aquaporin (M.J. Chrispeels, *personal communication*). The distant relationship between plant subfamily #3 and subfamilies #1 and #2 suggests that the former proteins may be functionally dissimilar from the latter proteins. It is interesting to note that close homologues of Nod(Gma) are found in rice and tobacco. Consequently, this subfamily is not restricted to nitrogen fixing legumes and cannot be considered to be nodulation-specific.

Animal Proteins

Sequenced animal MIP family proteins fall into just two subfamilies. Animal subfamily #1 is large and diverse

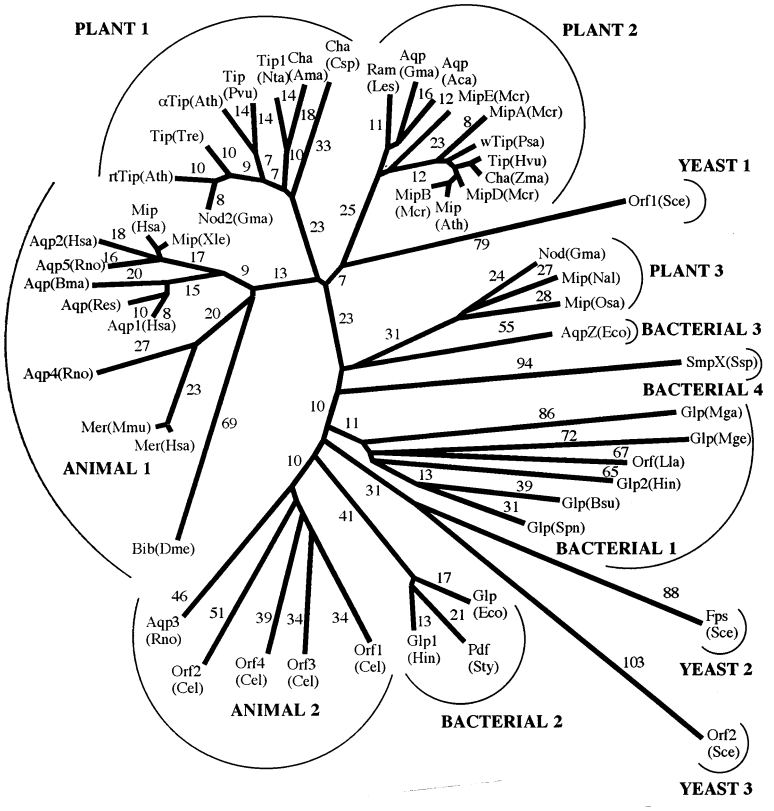

Fig. 2. Phylogenetic tree of the 52 MIP family proteins included in Table 1. The TREE program of Feng and Doolittle (1990) was used for tree construction. Branch length is approximately proportional to phylogenetic distance. Clustering patterns are indicated as discussed in the text.

including the *Drosophila* neurogenic protein Big Brain (Bib(Dme)) (Rao, Jan & Jan, 1990), major intrinsic proteins of the mammalian lens (e.g., Mip Hsa) (Ehring, Zampighi & Hall, 1993) and aquaporins from a variety of vertebrate tissues (*see* Tables 1 and 2; Knepper, 1994). The phylogenetic grouping of Bib and Mip with known aquaporins leads us to suggest that the functionally uncharacterized Bib and Mip proteins are also aquaporins. Although Mip has been reported to exhibit anion selective conductance properties with symmetrical voltage dependence following reconstitution in planar lipid bilayers (Ehring et al., 1993), the physiological significance of this observation is not clear.

Animal subfamily #2 includes only one sequenced mammalian protein (Aqp3) as well as four functionally uncharacterized ORFs from the invertebrate, *C. elegans.* Aqp3(Rno) has been shown to be capable of transporting glycerol and urea in addition to water (Ishibashi et al., 1994). In this capacity it differs from the characterized aquaporins of animal subfamily #1 or from the γ-Tip protein from *A. thaliana* (Maurel et al., 1993, 1994).

Animal subfamily #1 clearly includes many paralogues. For example, Mer(Hsa), Aqp1(Hsa), Aqp2(Hsa) and Mip(Hsa) are all from humans. Further, a human orthologue of Bib(Dme) has been reported (Adams et al., 1992). Thus, humans possess multiple MIP family paralogues. The actual number of paralogues to be found in any one mammalian species will undoubtedly prove to be large.

Multiple paralogues are also found in the invertebrate, *C. elegans*, which has at least four paralogues within animal subfamily #2. No *C. elegans* gene encoding a subfamily #1 protein has yet been sequenced, even though about one-third of the *C. elegans* genome and half of the genes of this organism had been sequenced at the time of this analysis (Waterston & Sulston, 1995).

Conclusions

The results summarized in this topical review reveal the phylogenetic relationships between currently sequenced MIP family members. These proteins fall into twelve subfamilies, each including proteins that are exclusively derived from a single eukaryotic kingdom or from bacteria. It is interesting to note that all recognized glycerol facilitators (Aqp3(Rno), Fps(Sce) and bacterial glycerol facilitators) cluster loosely together at the bottom part of the tree shown in Fig. 2 while all established aquaporins (including both animal and plant aquaporins as well as AqpZ(Eco)) cluster loosely together in the top part of the tree. This observation leads us to suggest that most if not all MIP family proteins will prove to be of two physiologically relevant functional types: (i) aquaporins and

(ii) channels for small neutral solutes such as glycerol. The observations that some members of the MIP family are capable of catalyzing ion transport following reconstitution in lipid bilayers (Ehring et al., 1993; Weaver et al., 1994) while the *Synechococcus* homologue apparently plays a role in copper homeostasis (Kashiwagi et al., 1995) are provocative, but do not establish physiologically relevant functions in ion transport. Interestingly, the *Synechococcus* homologue, SmpX(Ssp), radiates from a central position in the tree shown in Fig. 2, between the functionally characterized aquaporins and the functionally characterized glycerol facilitators. SmpX(Ssp) function cannot therefore be inferred from its phylogeny.

With the exception of a single *Haemophilus* protein, Glp2(Hin), and the *E. coli* aquaporin, AqpZ(Eco), the bacterial proteins cluster in accordance with the phylogenies of the organisms of origin. True bacterial paralogues have been identified only in *E. coli* (GlpF and AqpZ) and possibly in *H. influenzae* (Glp1 and Glp2). If the Glp2(Hin) gene was obtained horizontally from a Gram-positive bacterium as suggested by its phylogeny (Fig. 2), then bacteria may possess only two types of MIP family paralogues, aquaporins (e.g., AqpZ(Eco)) and glycerol facilitators (e.g., Glp(Eco) and its close homologue Pdf(Sty)). Thus, bacteria possess either two or three functionally dissimilar MIP family paralogues, depending on the as yet unknown biochemical function of the *Synechococcus* protein. Similarly, yeast (with 90% of the *S. cerevisiae* genome sequenced) possess only three recognized MIP family paralogues, presumably serving two or three dissimilar functions. We therefore conclude that most of the gene duplication events that gave rise to the large number of MIP family isoforms in animals and plants occurred late in the evolutionary process, after the divergence of these two major kingdoms from each other. This proliferation of gene duplication events may have occurred in response to a need for the maintenance of strict intracellular homeostasis and intercellular communication in complex, multicellular organisms.

We thank Drs. Maarten J. Chrispeels and Ian T. Paulsen for critically reading the manuscript, Nerses Sanossian for help in the retrieval of protein sequences from the databases, and Mary Beth Hiller for assistance in the preparation of this manuscript. This work was supported by USPHS grants 5RO1AI 21702 and 2RO1AI 14176 from the National Institute of Allergy and Infectious Diseases.

References

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C., Venter, J.C. 1992. Sequence identification of 2,375 human brain genes. *Nature* **355:**632–634

Calamita, G., Bishai, W.R., Preson, G.M., Guggino, W.B., Agre, P. 1995. Molecular cloning and characterization of AqpZ, a water channel from *Escherichia coli*. *J. Biol. Chem.* **270:**29063–29066

Chen, P., Andersson, D.I., Roth, J.R. 1994. The control region of the *pdu/cob* regulon in *Salmonella typhimurium*. *J. Bacteriol.* **176:**5474–5482

Chrispeels, M.J., Agre, P. 1994. Aquaporins: Water channel proteins of plant and animal cells. *Trends Biochem. Sci.* **19:**421–425

Chrispeels, M.J., Maurel, C. 1994. Aquaporins: The molecular basis of facilitated water movement through living plant cells? *Plant Physiol.* **105:**9–13

Ehring, G.R., Zampighi, G.A., Hall, J.E. 1993. Does MIP play a role in cell-cell communication? *In:* Progress in Cell Research, Vol. 3, J.E. Hall, G.A. Zampighi and R.M. Davis, editors. pp. 153–162. Elsevier Science Publishers B.V., Amsterdam, The Netherlands

Engel, A., Walz, T., Agre, P. 1994. The aquaporin family of membrane water channels. *Curr. Opin. Struct. Biol.* **4:**545–553

Feng, D.-F., Doolittle, R.F. 1990. Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol.* **183:**375–387

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., Venter, J.C. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:**496–512

Heller, K.B., Lin, E.C.C., Wilson, T.H. 1980. Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. *J. Bacteriol.* **144:**274–278

Höfte, H., Hubbard, L., Reizer, J., Ludevid, D., Herman, E.M., Chrispeels, M.J. 1992. Vegetative and seed-specific forms of tonoplast intrinsic protein in the vacuolar membrane of *Arabidopsis thaliana*. *Plant Physiol.* **99:**561–570

Ishibashi, K., Sasaki, S., Fushimi, K., Uchida, S., Kuwahara, M., Saito, H., Furukawa, T., Nakajima, K., Yamaguchi, Y., Gojobori, T., Marumo, F. 1994. Molecular cloning and expression of a member of the aquaporin family with permeability to glycerol and urea in addition to water expressed at the basolateral membrane of kidney collecting duct cells. *Proc. Natl. Acad. Sci. USA* **91:**6269–6273

Jap, B.K., Li, H. 1995. Structure of the osmo-regulated H_2O-channel, AQP-CHIP, in projection at 3.5 Å resolution. *J. Mol. Biol.* **251:**413–420

Jung, J.S., Preston, G.M., Smith, B.L., Guggino, W.B., Agre, P. 1994. Molecular structure of the water channel through aquaporin CHIP. The hourglass model. *J. Biol. Chem.* **269:**14686–14654

Kashiwagi, S., Kanamaru, K., Mizuno, T. 1995. A *Synechococcus* gene encoding a putative pore-forming intrinsic membrane protein. *Biochim. Biophys. Acta* **1237:**189–192

Knepper, M.A. 1994. The aquaporin family of molecular water channels. *Proc. Natl. Acad. Sci. USA* **91:**6255–6258

Lloyd, A.T., Sharp, P.M. 1992. Evolution of codon usage patterns: The extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **20:**5289–5295

Luyten, K., Albertyn, J, Skibbe, S.F., Prior, A.B., Ramos, J., Thevelien, J.M., Hohmann, S. 1995. Fps1, a yeast member of the MIP family of channel proteins, is a facilitator for glycerol uptake and efflux and is inactive under osmotic stress. *EMBO J.* **14:**1360–1371

Maurel, C., Reizer, J., Schroeder, J.I., Chrispeels, M.J. 1993. The vacu-

olar membrane protein γ-TIP creates water specific channels in *Xenopus* oocytes. *EMBO J.* **12:**2241–2247

Maurel, C., Reizer, J., Schroeder, J.I., Chrispeels, M.J., Saier, M.H., Jr. 1994. Functional characterization of the *Escherichia coli* glycerol facilitator, GlpF, in *Xenopus* oocytes. *J. Biol. Chem.* **269:**11869–11872

Pao, G.M., Wu, L.-F., Johnson, K.D., Höfte, H., Chrispeels, M.J., Sweet, G., Sandal, N.N., Saier, M.H., Jr. 1991. Evolution of the MIP family of integral membrane transport proteins. *Mol. Microbiol.* **5:**33–37

Preston, G.M., Jung, J.S., Guggino, W.B., Agre, P. 1994. Membrane topology of aquaporin CHIP. Analysis of functional epitope-scanning mutants by vectorial proteolysis. *J. Biol. Chem.* **269:**1668–1673

Rao, Y., Jan, L.Y., Jan, Y.N. 1990. Similarity of the product of the *Drosophila* neurogenic gene *big brain* to transmembrane channel proteins. *Nature* **345:**163–167

Reizer, J., Reizer, A., Saier, M.H., Jr. 1993. The MIP family of integral membrane channel proteins: Sequence comparisons, evolutionary relationships, reconstructed pathway of evolution, and proposed functional differentiation of the two repeated halves of the proteins. *Crit. Rev. Biochem. Mol. Biol.* **28:**235–257

Saier, M.H., Jr. 1994*a*. Computer-aided analyses of transport protein sequences: Gleaning evidence concerning function, structure, biogenesis, and evolution. *Microbiol. Rev.* **58:**71–93

Saier, M.H., Jr. 1994*b*. Convergence and divergence in the evolution of transport proteins. *Bioessays* **16:**23–29

Van Aelst, L., Hohmann, S., Zimmerman, F.K., Jans, A.W.H., Thevelein, J.M. 1991. A yeast homologue of the bovine lens fibre MIP gene family complements the growth defect of a *Saccharomyces cerevisiae* mutant on fermentable sugars but not its defect in glucose-induced RAS-mediated cAMP signalling. *EMBO J.* **10:**2095–2104

Verma, D.P.S. 1992. Soybean Nodulin-26: A channel protein conserved from bacteria to mammals. *In:* Transport and Receptor Proteins of Plant Membranes. D.T. Cooke and D.T. Clarkson, editors. pp. 113–117, Plenum Press, New York

Walz, T., Smith, B.L., Agre, P., Engel, A. 1994. The three-dimensional structure of human erythrocyte aquaporin CHIP. *EMBO J.* **13:**2985–2993

Waterston, R., Sulston, J. 1995. The genome of *Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA* **92:**10836–10840

Weaver, C.D., Shomer, N.H., Louis, C.F., Roberts, D.M. 1994. Nodulin 26, a nodule-specific symbiosome membrane protein from soybean, is an ion channel. *J. Biol. Chem.* **269:**17858–17862