Critical Reviews in Biochemistry and Molecular Biology, 28(3):235–257 (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

*Jonathan Reizer, Aiala Reizer, and Milton H. Saier, Jr.**

Department of Biology, University of California at San Diego, La Jolla, CA 92093–0116

Referee: Tom Wilson, Cell and Molecular Physiology, Harvard Medical School, Boston, MA

ABSTRACT: The major intrinsic protein (MIP) of the bovine lens fiber cell membrane was the first member of the MIP family of proteins to be sequenced and characterized. It is probably a homotetramer with transmembrane channel activity that plays a role in lens biogenesis or maintenance. The polypeptide chain of each subunit may span the membrane six times, and both the N- and C-termini face the cell cytoplasm. Eighteen sequenced or partially sequenced proteins from bacteria, yeast, plants, and animals have now been shown to be members of the MIP family. These proteins appear to function in (1) metazoan development and neurogenesis (MIP and BIB), (2) water transport across the human erythrocyte membrane (ChIP), (3) communication between host plant cells and symbiotic nitrogen-fixing bacteria (NOD), (4) transport across the tonoplast membrane during plant seed development (α-TIP), (5) water stress-induced resistance to desiccation in plants (Wsi-TIP), (6) suppression of a genetic growth defect on fermentable sugars in yeast (FPS1), and (7) transport of glycerol across bacterial cell membranes (GlpF). One other sequenced member of the MIP family (ORF1 of *Lactococcus lactis*) has no known physiological function. The biochemical functions of the eukaryotic proteins are not well established.

Computer analyses have revealed that the first and second halves of all MIP family proteins probably arose by a tandem, intragenic, duplication event. Thus, the primary structure of putative transmembrane helices 1 to 3 is similar to that of putative transmembrane helices 4 to 6 even though they are of opposite orientation in the membrane. Among the most conserved residues in these two repeated halves are a membrane-embedded glutamate (E) in helices 1 and 4, an asparagine-proline-alanine (NPA) sequence in the loops between helices 2 and 3 (cytoplasmically localized) and helices 5 and 6 (extracellularly localized), and a glycine within helices 3 and 6. Statistical analyses suggest that the two halves of these proteins have evolved to serve distinct functions: the first half is more important for the generalized or common functions of these proteins, while the second half of these proteins is more differentiated to provide specific or dissimilar functions of the proteins. The apparent origin of MIP family proteins by duplication of a three-spanner precursor protein suggests an evolutionary origin distinct from other transport proteins with six transmembrane spanners. Based on the phylogenetic tree for the 18 sequenced members of the MIP family, we propose that a single, primordial gene arose in prokaryotes shortly before the emergence of eukaryotes, that this gene was vertically transmitted to the principal eukaryotic kingdoms, and that subsequent gene duplication and divergence events gave rise to kingdom-related subfamilies or clusters of the MIP family.

KEY WORDS: major intrinsic protein, MIP, lens fiber cells, sequence comparisons, phylogenetic tree, evolution, membrane channel proteins, transmembrane transport, intercellular communication

I. INTRODUCTION

In earlier reports, we and others have described a family of integral membrane proteins from bacteria, animals, and plants that function in (1) metazoan development and neurogenesis, (2) communication between host plant cells and symbiotic bacteria, (3) organellar transport during plant seed germination, and (4) nutrient transport in single celled bacteria (Shiels et al., 1988; Baker

* Corresponding author:

1040-9238/93/$.50

and Saier, 1990; Pao et al., 1991). We named this protein family the MIP family, after the first sequenced and best characterized member of this family, the major intrinsic protein or main intrinsic polypeptide (MIP) of the bovine lens fiber cell membrane. Six members had been sequenced and identified at the time of publication of our last effort (Pao et al., 1991). Since that time, new sequence data has allowed expansion of the MIP family from 6 to 18 members. New members of the family include some which are presumed to have functions similar to those described previously, but others are believed to play entirely different physiological roles.

Members of this family have now been identified in both Gram-negative and Gram-positive bacteria as well as in yeast, plants, and animals (see Table 1). Four of these proteins (MIP, ChIP, NOD, and GLP) have been reported to have transmembrane channel activities for various substrates (ions, water, dicarboxylates, and straight chain carbon compounds, respectively) (Heller et al., 1980; Nikaido and Rosenberg, 1985; Ehring et al., 1990, 1993; Ouyang et al., 1991; Preston et al., 1992; van Hoek and Verkman, 1992; Zeidel et al., 1992; see Ehring et al., 1993 for a concise summary of functional studies with MIP family proteins). Recently, the structural relationships of the sequenced plant proteins were reported, and different TIP proteins were found to be localized to tonoplast membranes of either plants or seeds (Höfte et al., 1992).

All MIP family members are integral membrane proteins presumed to have essentially the same topology (Doolittle, 1986). Each subunit of the MIP, ChIP, and NOD polypeptide chains appears to span the membrane as α-helices six times (Takemoto and Takehana, 1986; Horwitz and Bok, 1987; Takemoto et al., 1987, 1988; Zampighi et al., 1989; Verma, 1992), and evidence suggests that the native proteins exist in the membrane as tetramers (Aerts et al., 1990; Smith and Agre, 1991; Verma, 1992). By contrast, the glycerol facilitator of *Escherichia coli* may exist as a homodimer (Voegele et al., 1992). Because the biochemical functions of the eukaryotic proteins are not well established, we here extend our earlier studies to provide (1) a multiple alignment of the 18 sequenced MIP family proteins with identification and discussion of fully or largely conserved residues, (2) statistical analyses of corresponding binary comparisons that establish homology for all members of this protein family, (3) a phylogenetic tree defining the relatedness of the currently known members of the family, (4) a signature sequence allowing rapid identification of new members of this family, (5) a pathway of MIP family protein evolution, and (6) a summary of analyses leading to an expanded understanding of structure-function relationships. In this last

TABLE 1
Major Types of MIP Family Proteins

Abbreviations	Name	Function	Organism
MIP	Major intrinsic protein	Lens biogenesis ? (Ion transport?)	Animal
BIB	Big brain	Brain development	Animal
ChIP	Channel intrinsic protein	Osmoregulation (Water transport)	Animal
NOD	Nodulation protein	Communication (Dicarboxylate transport?)	Plant
TIP	Tonoplast intrinsic protein	α: seed gemination w: Water stress resistance (Water transport)	Plant Plant
FPS	Fermentable sugar Defect suppressor (fdp-1)	?	Yeast
GLP	Glycerol facilitator	Carbohydrate supply (Glycerol transport)	Bacteria

respect, we show that the first and second halves of these proteins, segments 1 and 2, respectively, which probably arose by tandem duplication of a common genetic element in prokaryotes prior to the advent of eukaryotes on Earth, have diverged from each other in ways that suggest that they have acquired distinct functions. We show that in closely related proteins, segments 2 are more similar to each other than are segments 1, but that in distantly related proteins, segments 1 are more similar to each other than are segments 2. These observations suggest that segments 1 are more important for the generalized or common structure/function of these proteins (i.e., biogenesis or channel formation), whereas segments 2 are more important for the specialized or dissimilar functions of the proteins (i.e., channel specificity or regulation).

Based on published reports and the analyses described in this manuscript, we propose that all members of the MIP family serve a common function in transmembrane transport of small molecules, although each class of these proteins may exhibit distinctive specificities and characteristics. Some of these proteins are regulated by protein kinase-mediated phosphorylation, whereas others are posttranslationally modified by palmitoylation. Some of these proteins have distinctive N- and/or C-terminal extensions that are not homologous to other proteins in the database. BIB and FPS, for example, have extensive hydrophilic N- and C-terminal regions of this type (Rao et al., 1990; van Aelst et al., 1991). The degrees of conservation of the proposed sites of phosphorylation between members of the family are reported.

II. COMPUTER METHODS

All sequence analyses and database searches were performed with the GCG package from the University of Wisconsin (Devereux et al., 1984) and the DNASYSTEM package (Smith, 1988). The FASTA program (Pearson and Lipman, 1988) was used to search the current protein identification resource protein databank (PIR, version 34), SWISSPROT (version 23), and GenBank (version 74) databases, and to compare the percent identity of one protein to another. The RDF2 program (Pearson and Lipman, 1988) was used to calculate comparison scores. Homologous sequences were aligned with the TREE program (Feng and Doolittle, 1990) and adjusted by eye when appropriate. Construction of phylogenetic trees and estimation of relative evolutionary distances were as described by Reizer and Reizer (1993) using the progressive alignment (TREE) method of Feng and Doolittle (1990) and the parsimony after progressive alignment (PAPA) method of Doolittle and Feng (1990). Other methods are as described in the table and figure legends.

III. RESULTS

A. Phylogenetic Tree of the MIP Family Protein Members

Figures 1A and B show phylogenetic trees obtained using two different programs of all currently recognized members of the MIP family for which complete sequences are available. Sixteen members are depicted in this figure. The partially sequenced members GLP(Sco) and MIP(Gga) are not included in the tree because artifactually long branch lengths were observed for these members due to incomplete sequences. The major classes of these proteins are described in Table 1, while individual proteins are characterized in Table 2. The members include:

1. MIP (major intrinsic protein of the lens fiber cell of the cow, human, and rat that appears to function in lens biogenesis or maintenance; partial sequence of the chicken protein is also available)
2. BIB (big brain, a neurogenic protein in *Drosophila melanogaster;* partial sequence of the human protein is also available [Adams et al., 1992])
3. ChIP (channel intrinsic protein of the human erythrocyte that can serve as a water channel; also found in kidney tubules (Preston et al., 1992; van Hoek and Verkman, 1992)
4. NOD (nodulin-26, a legume-encoded protein that allows communication between plant cells and bacteria in the nitrogen fixing bacteroid)

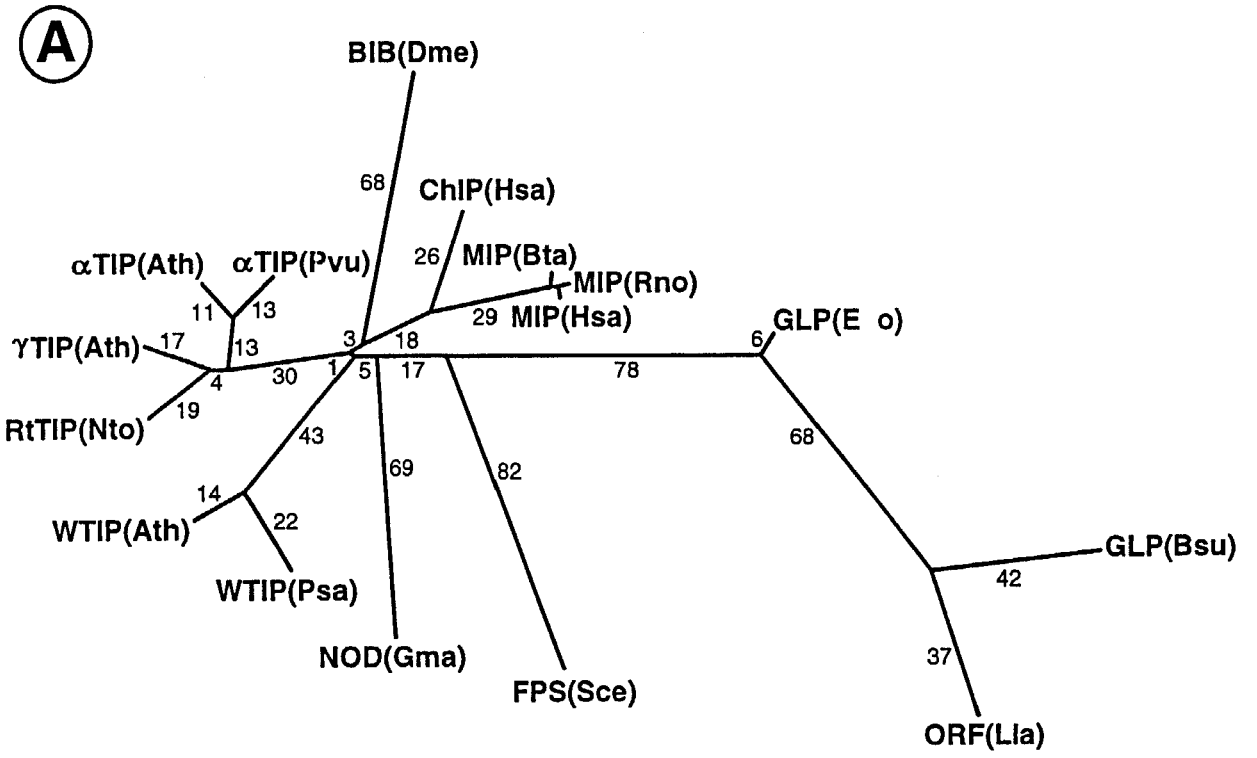

FIGURE 1. Phylogenetic trees of 16 of the 18 current protein members of the MIP family of presumed channel-forming proteins. The 16 proteins depicted are those for which complete sequences are available. Relative evolutionary distances are given adjacent to the branches unless less than one arbitrary unit. The programs of Feng and Doolittle (1990) and Doolittle and Feng (1990) were used to calculate the relative branch lengths. Panel A shows the phylogenetic tree obtained with the TREE program. Panel B (next page) shows the phylogenetic tree obtained with the PAPA program. The abbreviations used are included in Table 2 together with their descriptions, biological sources, number of residues, and accession codes. References of published protein sequences are as follows: WTIP(Psa), the water stress-induced tonoplast intrinsic protein (TIP) of *Pisum sativum* (Guerrero et al., 1990); WTIP(Ath), the water stress-induced TIP of *Arabidopsis thaliana* (Yamaguchi-Shinozaki et al., 1992); RtTIP(Nto), the root-specific TIP of *Nicotiana tabacum* (Yamamoto et al., 1990, 1991); γTIP(Ath), the vegetative tissue-specific TIP of *Arabidopsis thaliana* (Höfte et al., 1992); αTIP(Pvu), the seed-specific TIP of *Phaseolus valgaris* (Johnson et al., 1990); αTIP(Ath), the seed-specific TIP of *Arabidopsis thaliana* (Höfte et al., 1992); NOD(Gma), the plant-encoded nodulin-26 from the peribacteroid membrane of *Glycine max* (soybean)/*Rhizobium* root nodules (Sandal and Marcker, 1988, 1990); FPS(Sce), the *fdp1* suppressor of *Saccharomyces cerevisiae* (Van Aelst et al., 1991); ORF(Lla), hypothetical protein in the 5′ region of the X-prolyl dipeptidyl aminopeptidase-encoding gene of *Lactococcus lactis* ssp. *cremoris* or *Lactococcus lactis* ssp. *lactis* (Mayo et al., 1991); GLP(Eco), the glycerol facilitator of *Escherichia coli* (Muramatsu and Mizuno, 1989); GLP(Bsu), the glycerol facilitator of *Bacillus subtilis* (Holmberg et al., 1990); GLP(Sco), glycerol facilitator of *Streptomyces coelicolor* (Smith and Chater, 1988); MIP(Bta), the major intrinsic protein (MIP) from the bovine *(Bos taurus)* lens fiber cell membrane (Gorin et al., 1984); MIP(Rno), the rat *(Rattus norvegicus)* MIP (Kent and Shiels, 1990); MIP(Hsa), the human *(Homo sapiens)* MIP (Pisano and Chepelinsky, 1988); MIP(Gga), the MIP of chicken *(Gallus gallus)* (Kodama et al., 1990); BIB(Dme), the *Drosophila melanogaster* neurogenic protein termed "Big Brain" (Rao et al., 1990); and ChIP(Hsa), the water channel-forming integral protein of human erythrocytes and kidney (Preston and Agre, 1991). The partially sequenced MIP family proteins, GLP(Sco), and MIP(Gga) are not included in this tree as artifactually long branch lengths reflect their incomplete sequences rather than their evolutionary distances. They nevertheless cluster with other proteins of the same function from other organisms.

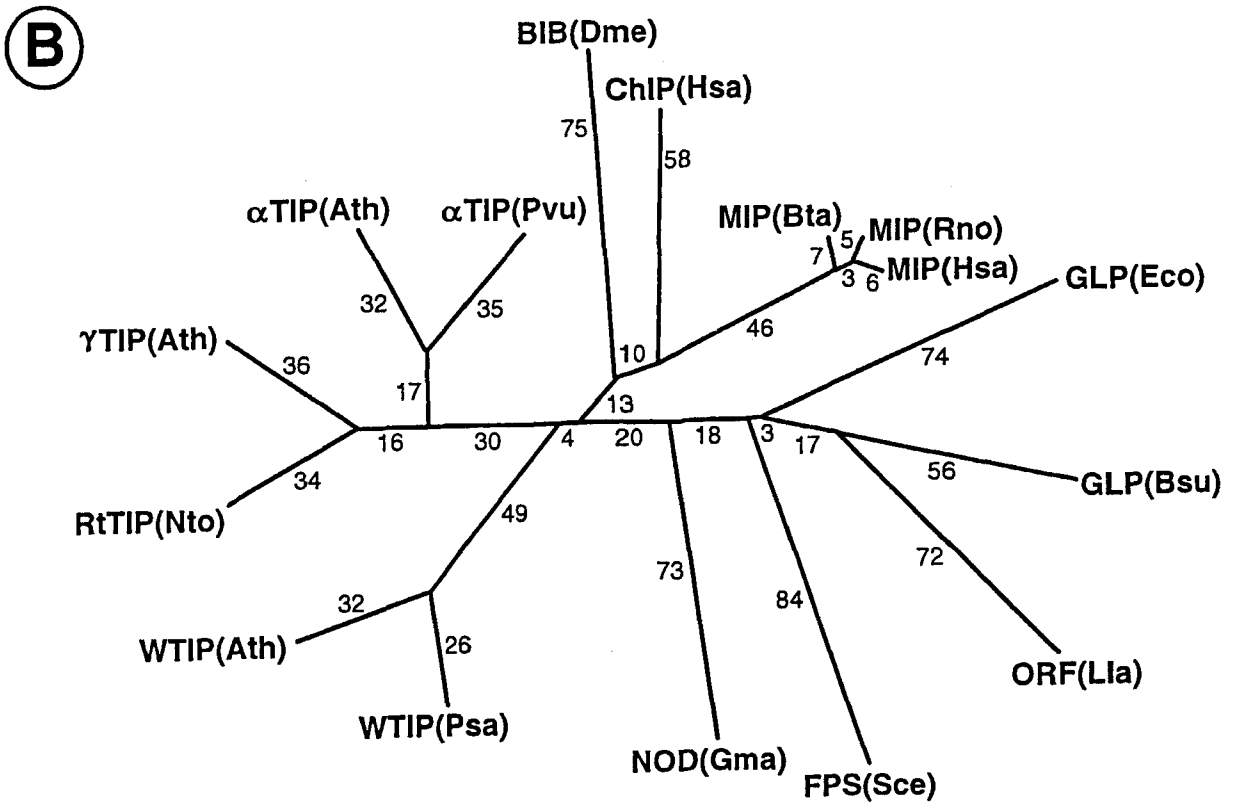

FIGURE 1B

5. TIP (tonoplast intrinsic protein of various plants and plant tissues; the different isoforms are designated Wsi or W [water stress induced], Rt [root], α [seed], and γ [all plant vegetative tissues but not seeds] (Höfte et al., 1992; Ludevid et al., 1992). αTIP may be important in seed germination, possibly by allowing diffusion of amino acids and/or peptides from the tonoplast interior to the cytoplasm
6. FPS (the fdp1 suppressor of *Saccharomyces cerevisiae* (FPS1) that plays a role in sugar fermentation and possibly in carbon catabolite repression
7. GLP (the glycerol facilitator [GlpF] from three evolutionarily divergent bacteria)
8. ORF1 (an open reading frame in *L. lactis* with no known physiological function; possibly a glycerol facilitator)

While the physiological consequences of the genetic loss of some of these proteins have been defined as a result of mutant analyses, evidence for the biochemical functions of only a few of the eukaryotic proteins has been forthcoming. Some evidence suggests that NOD may either function in dicarboxylate transport or regulate this process (Ouyang et al., 1991). Further, as noted here, ChIP is capable of functioning as a water channel (Preston et al., 1992; van Hoek and Verkman, 1992; Zeidel et al., 1992). α- and γ-TIPs may also be capable of functioning as water channels (Maurel et al., 1993), although their potential roles as ion channels has not been eliminated. MIP may transport salt (possibly Na^+) (Nikaido and Rosenberg, 1985).

Binary statistical analyses showed that all protein members of the MIP family are homologous. The complete matrix of the binary statistical analyses is not presented, but statistical analyses of representatives of all major types of MIP family proteins are presented here. Phylogenetic tree construction (Figure 1) revealed the degrees

TABLE 2
Eighteen Sequenced or Partially Sequenced Proteins of the MIP Family Included in this Study

Abbrev.	Description	Biological source	No. of residues	Code/database or references
MIP(Rno)	Major intrinsic protein of the fiber gap junction in eye lens	*Rattus norvegicus* (Rat)	261 (fragment)	MIP_RAT SwissProt
MIP (Hsa)	Major intrinsic protein of the fiber gap junction in eye lens	*Homo sapiens* (Human)	263	S80218 GenBank
MIP(Gga)	Major intrinsic protein of the fiber gap junction in eye lens	*Gallus gallus* (Chicken)	112 (fragment)	A37203 PIR
MIP(Bta)	Major intrinsic protein of the fiber gap junction in eye lens	*Bos taurus* (Bovine)	263	MIP_BOVINE SwissProt
GLP(Bsu)	Glycerol facilitator	*Bacillus subtilis*	274	M99611 GenBank
GLP(Eco)	Glycerol facilitator	*Escherichia coli*	281	GLPF_ECOLI SwissProt
GLP(Sco)	Glycerol facilitator	*Streptomyces coelicolor*	80 (fragment)	GYLA_STRCO SwissProt
NOD(Gma)	Peribacteroid membrane protein, Nodulin-26	*Glycine max* (Soy bean)	271	NO26_SOYN SwissProt
γTIP(Ath)	Tonoplast intrinsic protein (vegetative organ tissues)	*Arabidopsis thaliana*	251	TIPG_ARATH SwissProt
RtTIP(Nto)	Tonoplast intrinsic protein (root specific)	*Nicotiana tobacum* (common tobacco)	250	MEMA_TOBAC SwissProt
αTIP(Ath)	Tonoplast intrinsic protein (seed specific)	*Arabidopsis thaliana*	268	TIPA_ARATH SwissProt
αTIP(Pvu)	Tonoplast intrinsic protein (seed specific)	*Phaseolus vulgaris* (Kidney bean)	256	TIPA_PHAVU SwissProt
WTIP(Psa)	Water stress-induced membrane protein	*Pisum sativum* (Garden pea)	289	TR7A_PEA SwissProt
WTIP(Ath)	Water stress-induced membrane protein	*Arabidopsis thaliana*	285	Yamaguchi-Shinozaki et al (1992)
BIB(Dme)	Big brain neurogenic protein	*Drosophilla melanogaster* (Fruit fly)	700	BIB_DROME SwissProt
Chip(Hsa)	Channel-like integral protein of human erythrocyte and kidney	*Homo sapiens* (Human)	269	HUMCHIP28A_1 GenPept
FPS(Sce)	FDP1 suppressor of sugar fermentation defect	*Saccharomyces cerevisiae* (Baker's yeast)	669	FPS1_YEAST SwissProt
ORF(Lla)	Hypothetical protein in the PEPX 5′ region	*Lactococcus lactis*	289	YDP1_LACLC SwissProt

Note: Eighteen sequenced or partially sequenced proteins of the MIP family included in this study. The table provides the abbreviations (abbrev.) used in this study, a description of these proteins, their biological sources, the number of residues in each protein, and the accession codes or references to the published sequences. References for all published sequences are provided in the legend to Figure 1. The RtTIP(Nto) protein sequence (Yamamoto et al., 1991), shown in this table, is similar (99.2% identity) to another root-specific, putative, membrane channel protein (identification code MEMC_TOBAC in SwissProt) of *Nicotina tobacum.* The latter protein was not included in the multiple alignments shown in Figures 5 and 6. The following errors, which were inadvertently introduced during translation of the genes encoding these two *N. tobacum* proteins (Yamamoto et al., 1991) were corrected before alignment: (1) GAC (residue 11) was originally translated as alanine and was corrected here to aspartate; (2) TTT (residue 13) was originally translated as proline and was corrected here to phenylalanine; (3) GAT (residue 48) was originally translated as alanine and was corrected here to aspartate; (4) AAT (residues 76 and 83) were translated as alanine and were corrected here to asparagine; (5) AAC (residues 95 and 197) were originally translated as alanine and were corrected here to asparagine; and (6) TTC (residue 103) was originally translated as proline and was corrected here to phenylalanine. In addition, the sequence provided from the database for αTIPPvu is incorrect and has been corrected in Johnson et al. (1990).

of relatedness of the various MIP family proteins as well as their clustering patterns. Clustering patterns proved to be independent of the program used for tree construction, and with the exception of the plant proteins, the relative positions observed for the various protein clusters were essentially the same when the TREE and the PAPA methods were employed (compare Figure 1A with Figure 1B).

As shown in Figures 1A and B, plant members of the MIP family cluster together. Thus, six different TIP proteins have been sequenced, and they fall into two major subclusters. The clustering pattern generally correlates with their cellular locations and induction patterns. The two water stress-induced proteins (W-TIP) cluster tightly together, while the remaining TIP proteins (α- and γ-TIP) form a second loose cluster. NOD26 comprises a branch close to the W-TIP branch (Figure 1). Animal proteins, MIP, ChIP, and BIB, cluster together with the same clustering pattern in both trees. The chicken MIP (not shown) exhibited an artifactually long branch length within the MIP protein cluster, reflecting the fact that the gene encoding this protein has been only partially sequenced (Table 2). ChIP is the closest relative of MIP, while BIB is more distant but still on the same major branch (Figures 1A and B).

All remaining proteins, from yeast and bacteria, are relatively distant from each other, regardless of the program used for tree construction. The yeast FPS1 suppressor protein from *S. cerevisiae* branches from the trunk of the tree after NOD, and it is followed by the three fully sequenced bacterial proteins. The *Streptomyces coelicolor* protein (not shown) exhibited an artifactually long branch length within the GLP cluster, reflecting the fact that only 80 residues of its sequence are available (Table 2 and Smith and Chater, 1988). The function of the ORF from *L. lactis* is not known.

B. Hydropathy Analyses of Eight Representative Members of the MIP Family

As reported previously (Pao et al., 1991), several regions in the aligned MIP family proteins are highly conserved. The dual appearance of the highly conserved NPA sequence in the N- and C-terminal halves of the proteins reflect the presumed tandem intragenic duplication noted previously (Pao et al., 1991) and analyzed in detail here. The topological consequences of this duplication were considered briefly in an earlier report (Saier and Reizer, 1991) and are illustrated in Figure 2.

Figure 3 shows the hydropathy plots of eight dissimilar but representative members of the MIP family, one from each of the major clusters of this protein family (see Figure 1). The hydropathy plots of these representative proteins show striking similarities, but also noteworthy differences. First, BIB shows an extensive hydrophilic C-terminal domain, with tandemly repeated glutamine residues, which is not depicted in Figure 3. This domain is not homologous to any protein in the current databases. Short N-terminal hydrophilic regions are also observed for the plant and yeast proteins, but not for MIP, ChIP, or the bacterial proteins. Second, as indicated by the dashed vertical lines that delineate the two hydrophobic repeat sequences in each protein, there is substantial variation in the hydropathy profiles. In the animal proteins, the first of these repeats (here designated segments or repeats 1) are more hydrophobic than the second of these repeats (here designated segments or repeats 2). The first two putative transmembrane helical spanners are close together, while the third is more distant. In some cases, the intervening region between spanners 2 and 3 is quite hydrophobic (MIP, BIB, ChIP, and GLP) in segments 1, but not in segments 2. Third, the region between the repeated segments 1 and 2 is somewhat variable in length, being very short for αTIPAth, longer for the yeast and bacterial proteins, and intermediate and of fairly constant length for the remaining proteins (Figure 3).

C. Correlation of the Average Hydropathy Plot with the Average Sequence Similarity Plot

As shown in Figure 4A (upper panel), the eight hydropathy plots shown in Figure 3 were averaged following multiple alignment of the sequences. The three putative transmembrane α-helical spanners are clearly revealed for both seg-

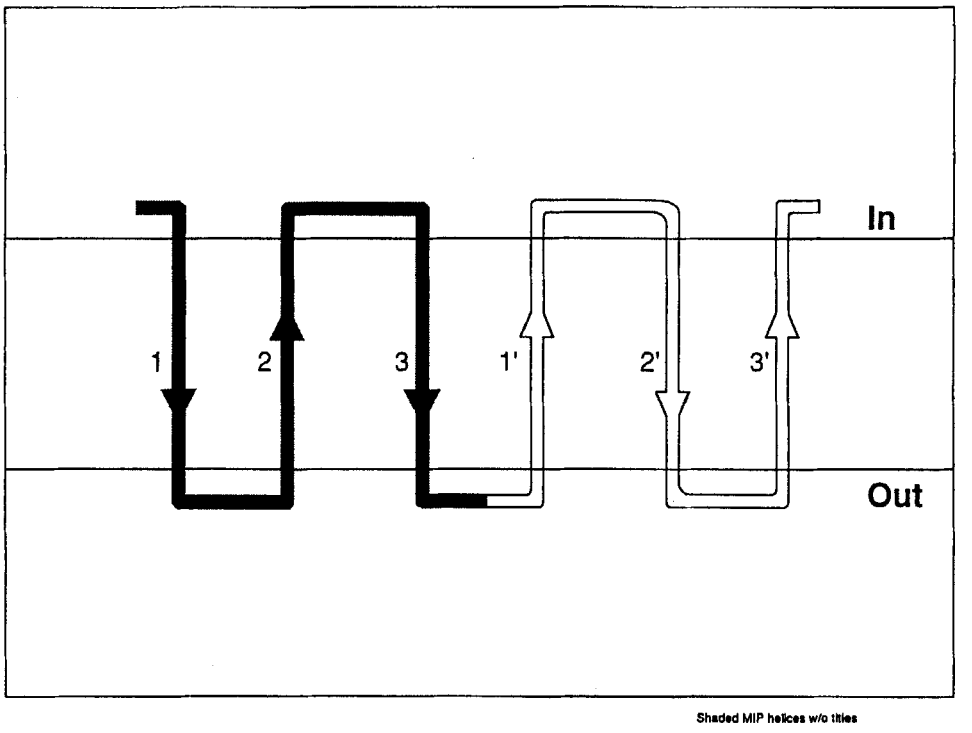

FIGURE 2. Topology (transmembrane orientation) of MIP and other members of the MIP family in the membrane showing the consequence of the tandem intragenic duplication event which is believed to have given rise to the primordial gene of the MIP family. Segments 1 (transmembrane spanners 1 to 3) are homologous to segments 2 (transmembrane spanners 4 to 6); yet homologous transmembrane helices are believed to be of opposite orientation in the membrane (Pao et al., 1991; Saier and Reizer, 1991).

ments 1 (N-terminal halves) and segments 2 (C-terminal halves). Furthermore, the average hydropathy plot, like the individual plots shown in Figure 3, clearly demonstrates that the regions between spanners 2 and 3 in segments 1 are substantially more hydrophobic than the regions between spanners 5 and 6 in segments 2. Otherwise the two halves show remarkable similarity, not only in degree of hydrophobicity, but also in the average spacing between peaks.

In Figure 4B (lower panel) the average similarity scores for the same eight proteins are plotted as a function of position. Comparison of the data shown in Figures 4A and 4B clearly demonstrates that the six hydrophobic spanners consistently show peaks of sequence similarity and that the loop regions between them usually show lower degrees of similarity. However, the hydrophilic regions between spanners 2 and 3 in segments 1 and spanners 5 and 6 in segments 2 show maximal degrees of sequence similarity. The first of these two regions is more conserved than the second, but both exhibit similar sequences that are characteristic of the currently known MIP family members. The sequences shown above these two peaks of sequence similarity in Figure 4B reveal the degree to which they resemble each other.

D. Multiple Alignment of All Sequenced MIP Family Members

The multiple alignment of all 18 sequenced MIP family proteins is shown in Figure 5. Alignment positions are indicated above the multiple alignment, while the residue positions in each protein are indicated at the beginning of each line. Positions with fully conserved residues or with only one nonconserved residue are shaded black, while the consensus sequence (at least nine residues conserved) is provided below the multiple

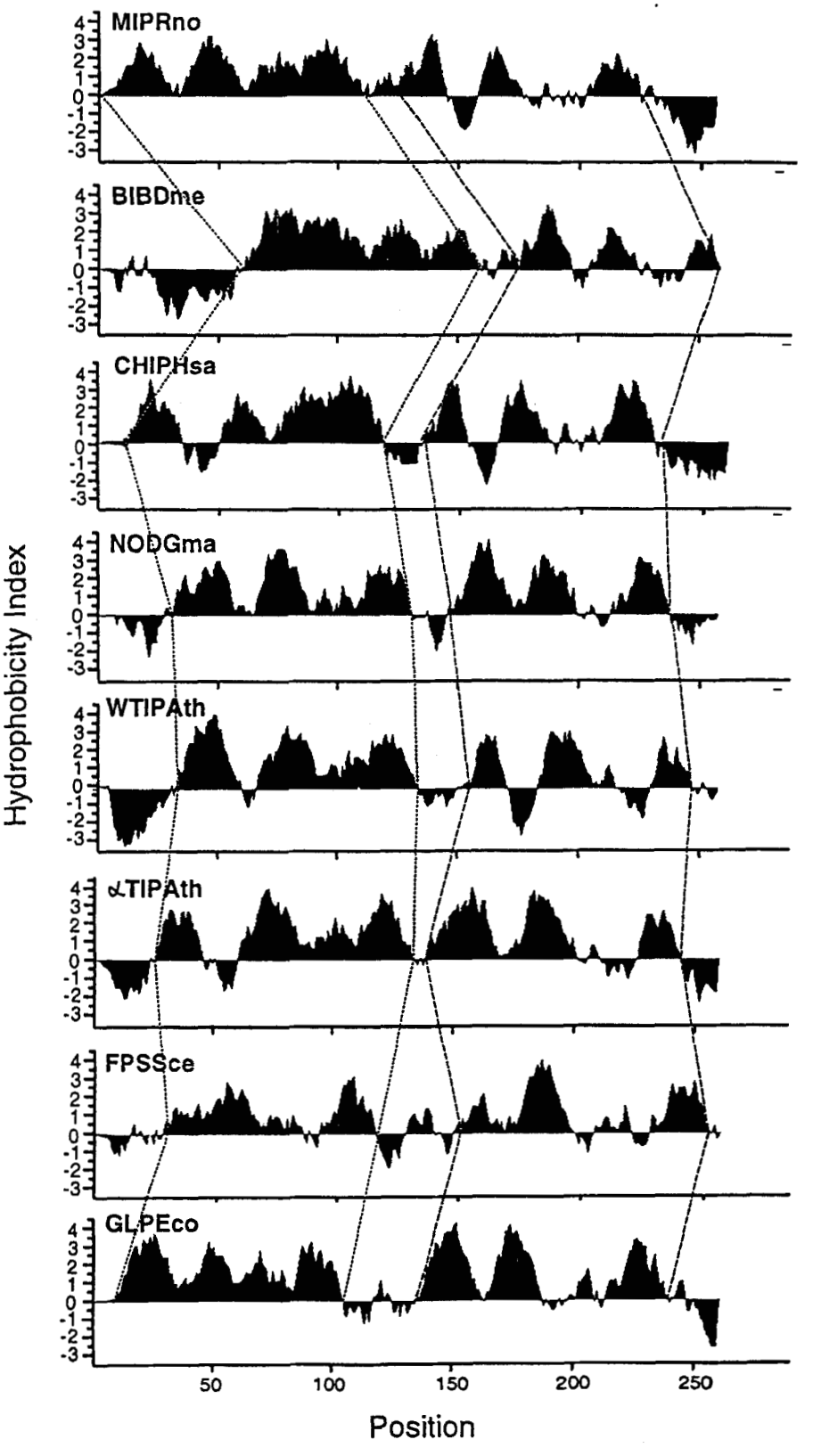

FIGURE 3. Hydropathy analyses of eight representative members of the MIP family. One member of each subgroup (Figure 1) was selected for detailed analysis. The abbreviations of the selected proteins are provided in Table 2. The program of Kyte and Doolittle (1982) was used to generate the hydropathy plots. Dashed lines delineate the boundaries of the three-spanner, repeated segments 1 and 2 within each of these proteins.

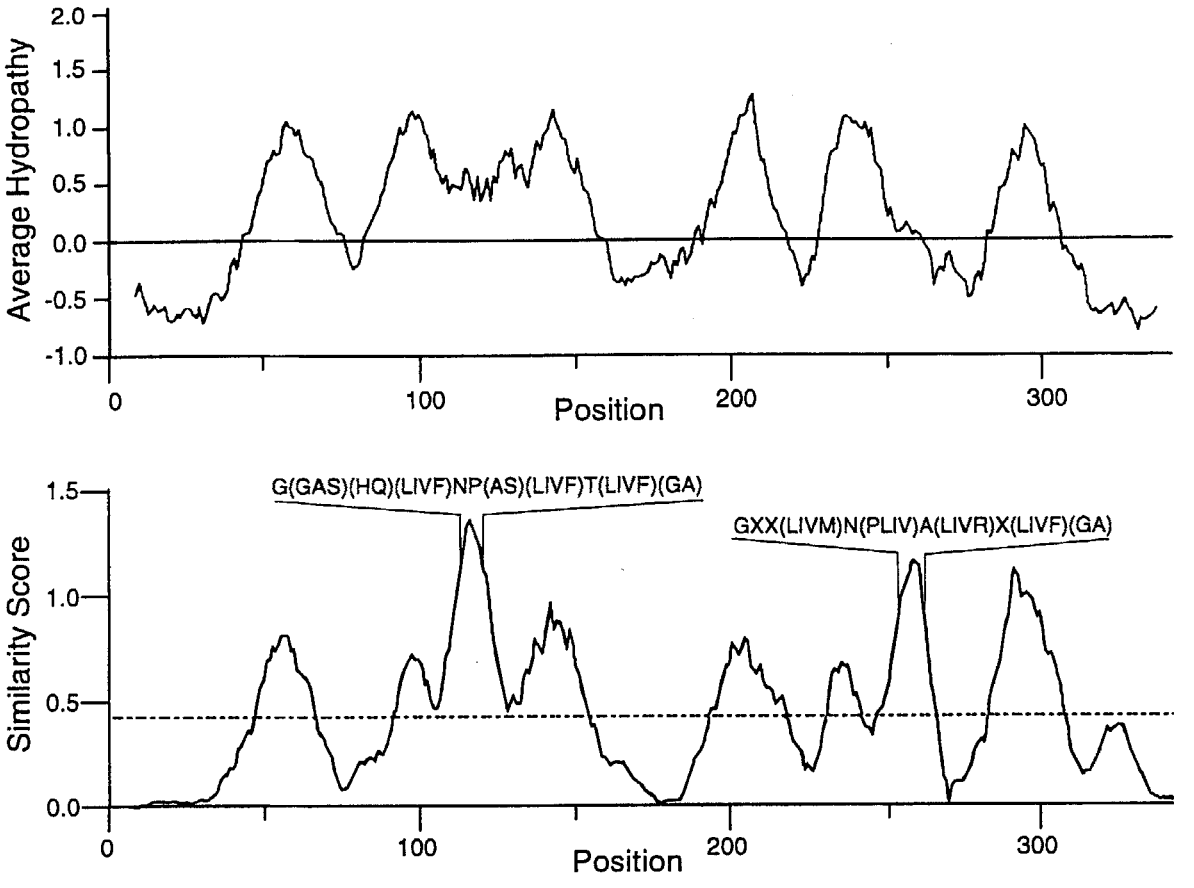

FIGURE 4. Average hydropathy plot (A upper panel) and average similarity plot (B lower panel) for the eight representative proteins of the MIP family depicted in Figure 3. The average hydropathy and average similarity plots were calculated using a sliding window of 20 residues. The average similarity across the entire alignment in (B) is shown as a dotted line. The plot reveals the coincidence of hydropathy and similarity plots except for the peaks of similarity between spanners 2 and 3 and spanners 5 and 6. The degenerate sequences of these two hydrophilic regions were derived from all protein members of the MIP family rather than from the seven representative proteins shown in Table 1 and are provided above the similarity plot. In these sequences, ambiguous residues at a particular position are given in parentheses, whereas an X indicates that any residue can occur at the specified position.

alignment. Following a variable N-terminal region is a conserved, hydrophobic stretch of about 18 residues containing a single fully conserved glutamyl residue (E) at position 3 in the multiple alignment. These 18 residues correspond to putative transmembrane spanner #1, and the conserved glutamyl residue is in the N-terminal portion of this spanner, near the cytoplasmic side of the membrane. A largely conserved threonyl residue (T) occurs four residues after the fully conserved glutamyl residue (position 7). The loop region between spanners 1 and 2 shows many gaps, but the entire region from the beginning of putative transmembrane spanner 2 to the end of spanner 3 is well conserved. The consensus sequence reveals two five-residue sequences that are largely conserved: I S G A H and N P A V T, separated by a single variable hydrophobic residue (positions 95 to 105 in the multiple alignment). The N P A V T sequence is particularly well conserved. Within the putative transmembrane spanner 3, at positions 126 and 129, are fully conserved glutaminyl and glycyl residues, respectively.

Following the central loop region where numerous gaps occur, putative transmembrane spanner #4 with its fully conserved glutamyl (E) resi-

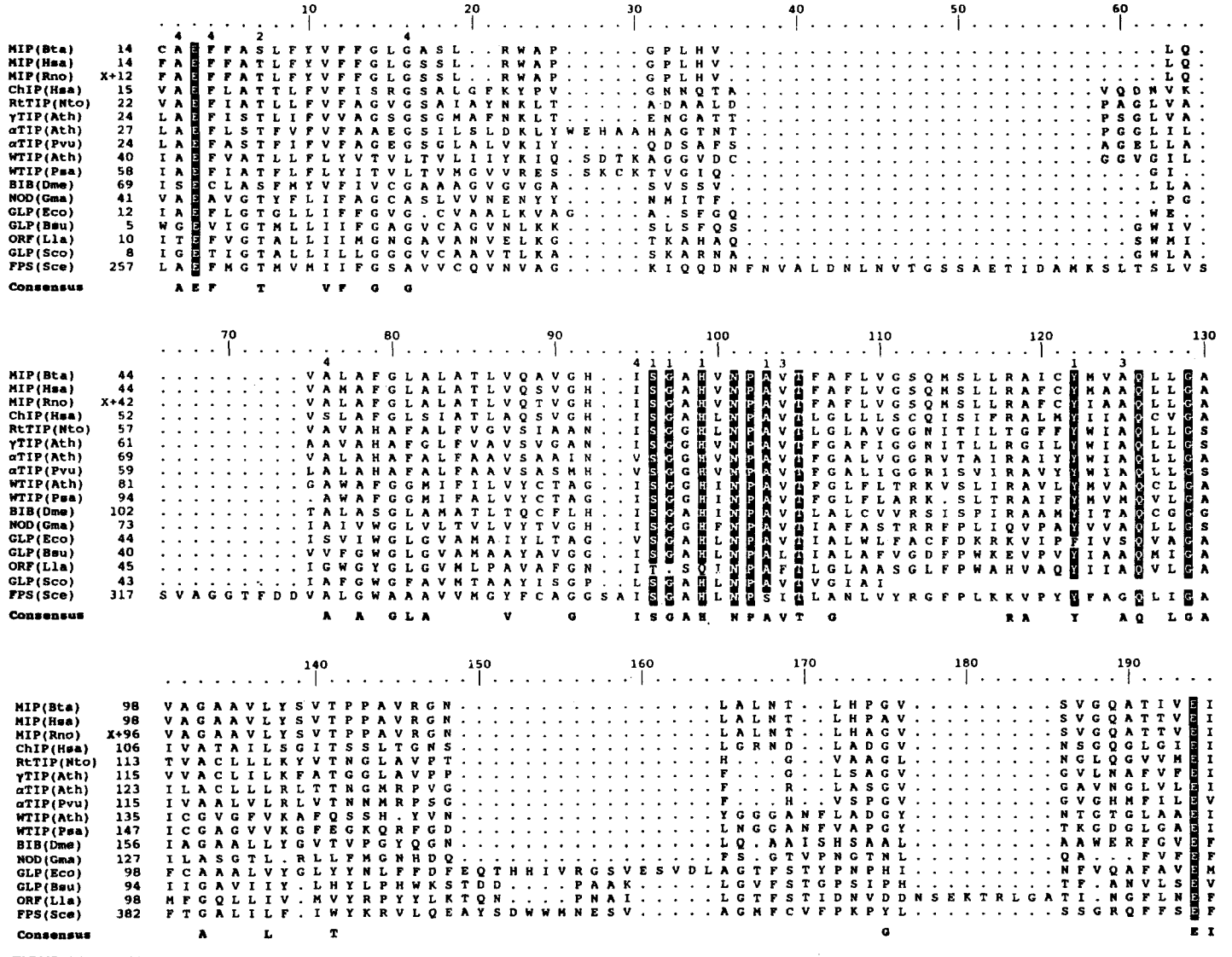

FIGURE 5. Multiple alignment of 18 fully or partially sequenced proteins of the MIP family. Numbers at the top of the aligned sequences denote the residue position in the multiple alignment. The residue number in each protein is provided at the beginning of each line. In the cases of MIP(Rno) and MIP(Gga), an X indicates uncertainty of the residue number due to incompleteness of the N-terminal sequence. Numbers (1, 2, 3, or 4) immediately above a position in the multiple alignment indicate that the residue at that position is conserved with 1, 2, 3, or 4 exceptions, respectively. Positions that are fully conserved or that are conserved with only one exception are highlighted in black. The consensus sequence (at least nine residues conserved) is provided below the multiple alignment. Abbreviations for the various proteins are as indicated in Table 2.

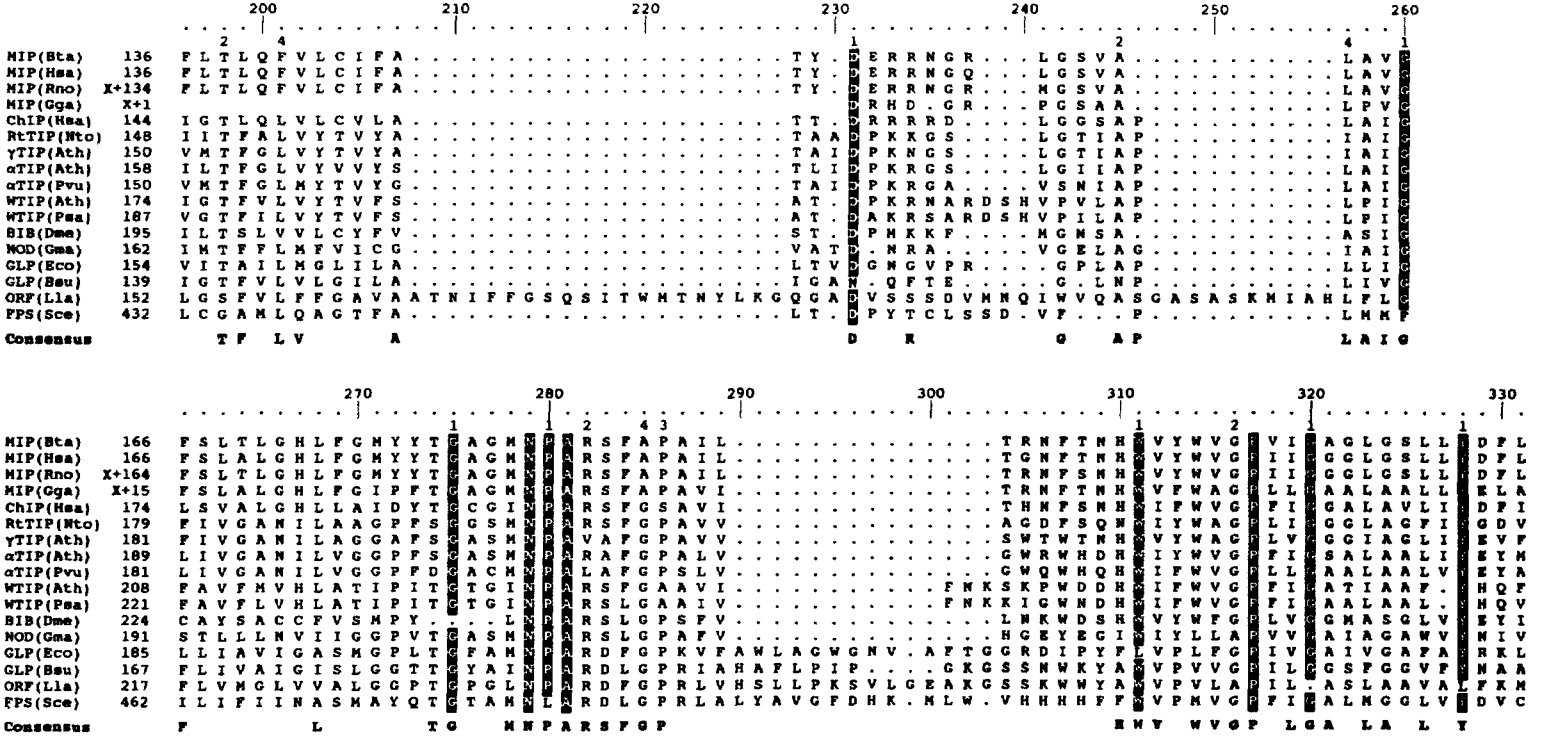

FIGURE 5. (continued)

due is found (position 194). Like the homologous spanner 1 in segment 1, a largely conserved threonyl residue (T) occurs four residues after the conserved glutamyl residue (position 198).

At the beginning of the loop region between spanners 4 and 5 is a nearly fully conserved aspartyl (D) residue (position 231), not found between spanners 1 and 2 in segment 1. However, greater variability is observed in the regions before and after spanners 4 and 5 when compared with the corresponding regions in segments 1. In the highly conserved loop regions between spanners 5 and 6 (on the outer surface of the membrane), only a T G sequence (positions 274 and 275), corresponding in position to the I S G A H sequence in segment 1, is recorded in the consensus sequence (Figure 5); however, the N P A V T sequence of segment 1 is largely conserved as N P A R S. Four additional fully or nearly fully conserved residues are seen within or following spanner 6: a tryptophanyl (W) residue at position 311, a prolyl (P) residue at position 317, a glycyl (G) residue at position 320, and a tyrosyl (Y) residue at position 328. The well-conserved P Hy Hy G sequence (positions 317 to 320) corresponds to the Q Hy Hy G sequence of segment 1 (Hy denotes any hydrophobic residue).

The presence of a largely conserved arginyl (R) residue in the external loop between spanners 5 and 6 (position 282 in the multiple alignment shown in Figure 5) is surprising in view of the "positive charge inside" rule of von Heijne that states that basic residues (lysyl and arginyl residues) occur with higher frequency on cytoplasmic loops than on external loops (von Heijne, 1991). The computer program of von Heijne (1992) for prediction of transmembrane topology was applied to several sequences shown in Figure 5, and six to eight putative transmembrane spanners were predicted for the various proteins. The "inside-out" rule (von Heijne, 1992) was, on the average, followed for these proteins. For example, assuming six transmembrane spanners for the eight divergent, representative proteins depicted in Figure 3, the average number of positive charges found outside of the putative transmembrane segments were as follows: N-termini, 1.5; loops 1 (outside), 0.9; loops 2 (inside), 2.1; loops 3 (outside), 1.25; loops 4 (inside), 1.75; loops 5 (outside), 1.1, and C-termini, 3.6.

E. Sites of Phosphorylation of MIP Family Proteins

A site of phosphorylation of MIPBta is serine residue 243 near the C-terminus of the protein (Johnson et al., 1986; Lampe et al., 1986, 1988; Lampe and Johnson, 1989). This residue is not conserved in other MIP family members (not shown). The phosphorylation of this residue occurs by a cyclic AMP-dependent (A-type) kinase, and its phosphorylation by protein kinase A converts a voltage-independent Na^+ channel into a voltage-dependent channel (Ehring et al., 1991). MIPBta can also be phosphorylated by a Ca^{2+}-calmodulin-type (C-type) kinase, but the position of this latter phosphorylation is not known (Lampe et al., 1986; Lampe and Johnson, 1989). α-TIP from the kidney bean (TIPPvu) (Johnson and Chrispeels, 1992) and NOD from soybeans (NODGma) (Weaver et al., 1991) are also known to be phosphorylated. The position of phosphorylation in NOD by a Ca^{2+}-dependent calmodulin-independent kinase (Weaver et al., 1991; Weaver and Roberts, 1992; Miao et al., 1992) is a C-terminal seryl residue (S262). α-TIP from kidney bean is phosphorylated on an N-terminal serine (S7) (Johnson and Chrispeels, 1992). These phosphorylated residues are always present in the N-terminal and C-terminal portions of the MIP family proteins that are not well conserved in the proteins whose sequences are depicted in Figure 5, and consequently these portions of the alignment are not shown. In the case of NOD, phosphorylation correlated with stimulation of malate uptake across the peribacteroid membrane of the soybean nodule (Ouyang et al., 1991). Site-specific mutagenesis analyses of these phosphorylation sites in only one member of the MIP family, α-TIP of kidney bean, have been reported, and in this case, the S7A mutation abolished phosphorylation of the protein as expected (M. Chrispeels, personal communication). These results establish that the poorly conserved N- and C-termini of MIP family proteins can serve in regulatory capacities rather than catalytic capacities.

F. Signature Sequence for the MIP Family

Employing the multiple alignment of the 18 proteins shown in Figure 5, the following signature sequence was derived: (HQ) (LIVMF) N P (AST) (LIVMF) T (LIVMF) (GA). Ambiguous residues at a specific position are given in parentheses. This signature sequence corresponds to residues 99 through 107 in the multiple alignment shown in Figure 5 (between transmembrane spanners 2 and 3). It recognizes all currently known protein members of the MIP family, but not other protein(s) currently present in the PIR (version 34) and SWISSPROT (version 23) databanks. The signature sequence proposed here should assist in the identification of new protein members of this family and should replace the signature sequence of the MIP family proposed by Bairoch (1992).

G. Multiple Alignment of Segments 1 with Segments 2 and Generation of the Corresponding Phylogenetic Tree

When segments 1 and 2 are multiply aligned (Figure 6), only five residues are fully conserved or conserved in all but one position. These residues, highlighted in black, are the glutamyl (E) residue at the beginning of spanners 1 and 4, the NPA sequence between spanners 2 and 3 (segments 1) as well as spanners 5 and 6 (segments 2), and the glycyl (G) residue found just after spanners 3 and 6. However, the fourth residues following the conserved glutamyl residues at the beginnings of spanners 1 in segments 1 and spanners 4 in segments 2 are threonyl residues in all but four of the proteins. In three of these proteins, a seryl residue replaces the conserved threonyl residue. Only in FPS(Sce)2 is a nonhydroxyl amino acid present at this position. Because segments 1 are oriented in the membrane from inside to out, and segments 4 are of the opposite orientation, these hydrophilic residues could partially comprise the hydrophilic lining of a transmembrane channel with the symmetric transmembrane sequence: E T T E. No other conserved transmembrane hydrophilic residues seem to serve as candidates for potential hydrophilic channel liners.

The phylogenetic tree corresponding to the multiple alignment shown in Figure 6 is shown in Figure 7. Although branch lengths vary substantially, the tree is largely symmetrical with all segment 2 sequences depicted on the left and all segment 1 sequences depicted on the right. Thus, the order of the sequences depicted on the left is the same as those shown on the right except for the plant proteins where the three major clusters are inverted relative to each other in the segment 1 vs. segment 2 sides of the tree. This difference may help to explain the discrepancies in the clustering patterns observed for intact plant proteins depicted in Figure 1A vs. Figure 1B (see earlier). The obvious near "mirror image" appearance observed in Figure 7 strongly argues for the parallel evolution of segments 1 and 2 from a common ancestral sequence at rates which were at least comparable. The configuration of the tree implies that the tandem intragenic duplication that gave rise to the primordial gene encoding the six-spanner precursor of the MIP family occurred before any of the duplication and divergence events that led to the different members of the family.

H. Statistical Analyses of the Binary Comparisons for Segments 1 and Segments 2

Table 3 presents a matrix of binary comparison scores for representative segments 1 against each other (top left), for the corresponding segments 2 against each other (bottom right), and for these segments 1 against the corresponding segments 2 (top right, boxed). Data for the eight divergent, representative, MIP family proteins depicted in Figure 3 are presented. When segment 1 is compared with segment 2 of the same protein, the value obtained is highlighted with a black background. Examination of the values obtained when segments 1 or segments 2 are compared with segments 1 or segments 2, respectively, reveals clear homology (comparison scores [in brackets] frequently above 9 SD). By contrast, when segments 1 are compared with segments 2, values of this magnitude are seldom obtained, even when the two segments compared are within the same protein (highlighted values, black back-

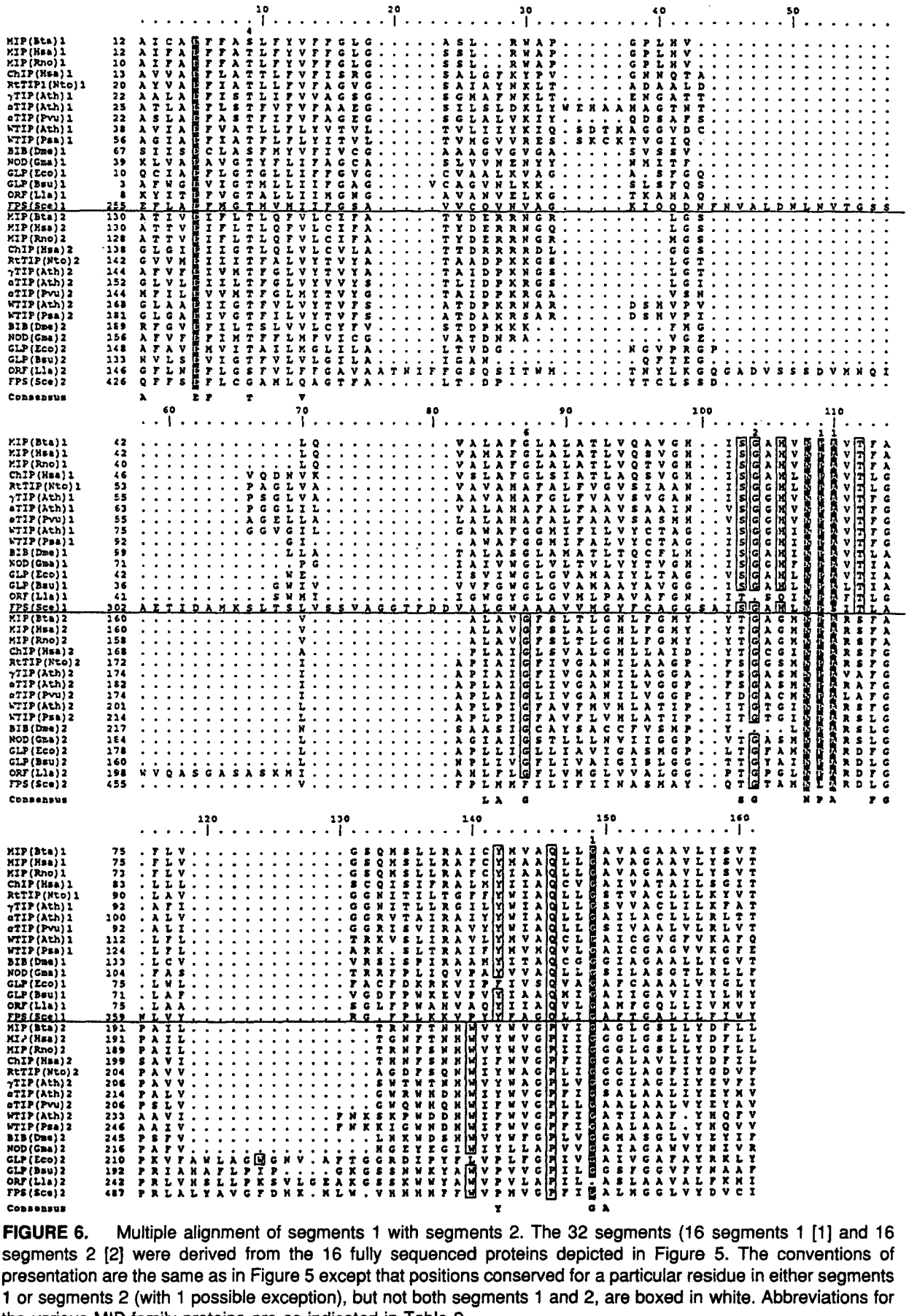

FIGURE 6. Multiple alignment of segments 1 with segments 2. The 32 segments (16 segments 1 [1] and 16 segments 2 [2] were derived from the 16 fully sequenced proteins depicted in Figure 5. The conventions of presentation are the same as in Figure 5 except that positions conserved for a particular residue in either segments 1 or segments 2 (with 1 possible exception), but not both segments 1 and 2, are boxed in white. Abbreviations for the various MIP family proteins are as indicated in Table 2.

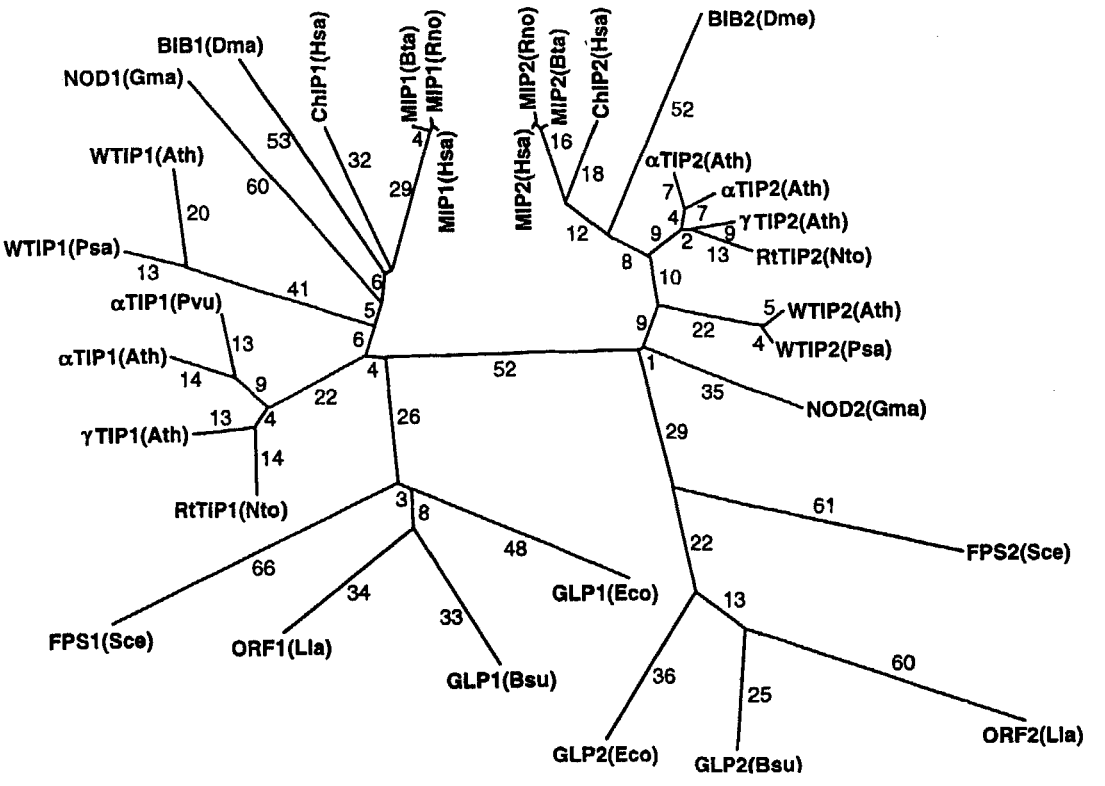

FIGURE 7. Phylogenetic tree of the MIP family segments 1 (1) and segments 2 (2) presented in Figure 6. The tree was constructed as described in the legend to Figure 1 employing the TREE program. Branch lengths, proportional to evolutionary distances, are provided except when less than 1 arbitrary unit. Abbreviations for the various MIP family proteins are as presented in Table 2.

ground). Thus, values of 9 SD (W-TIPAth2 vs. αTIPAth1) and 7 SD (αTIPAth1 vs. αTIPAth2) were among the largest comparison scores obtained. A binary comparison matrix of segments 1 and segments 2 of all currently sequenced members of the MIP protein family confirmed this conclusion, indicating that the data shown in Table 3 faithfully represent the similarity scores of the two repeats in all members of the MIP family (data not presented). These observations further substantiate the conclusion that divergence of segments 1 from segments 2 following the intragenic duplication event which gave rise to the intact protein of six transmembrane spanners occurred substantially before the divergence events that gave rise to the different members of the MIP family.

Distance scores for all segments 1 against each other, for all segments 2 against each other, and for all segments 1 against all segments 2 were calculated and tabulated in decreasing order of the distance scores for the intact proteins being compared. One hundred and twenty distance scores for the segment 1:segment 1 or segment 2:segment 2 comparisons were obtained, while 240 segment 1:segment 2 comparisons were obtained. These distance scores were then subdivided into ten groups of decreasing distance scores for the intact proteins being compared (12 or 24 values in each group) and averaged. As shown in Figure 8, when the intact protein distance scores were large (i.e., for distantly related proteins), the corresponding values for the segment 1 vs. segment 1 comparisons were of lesser magnitude than those for the segment 2 vs. segment 2 scores. By contrast, when the intact protein distance scores were small (i.e., for closely related proteins), the distance scores for segments 1 vs. segments 1 were generally larger than the segment 2 vs. segment 2 scores. The values presented in Figure 8 reveal that seg-

TABLE 3
Binary Comparisons of the First and Second Segments of the Eight Representative MIP Family Proteins

	BIBDme1 (89)	CHIPHsa1 (93)	NODGma1 (88)	WTIPAth1 (97)	αTIPAth1 (98)	FPSSce1 (83)	GLPEco1 (88)	MIPRno2 (89)	BIBDme2 (84)	CHIPHsa2 (89)	NODGma2 (88)	WTIPAth2 (96)	αTIPAth2 (90)	FPSSce2 (101)	GLPEco2 (103)
MIPRno1 (86)	46(84) [13]	48(93) [19]	51(55) [11]	45(62) [9]	47(64) [11]	33(63) [9]	31(86) [9]	(59)31 [4]	28(18) [0.4]	25(59) [2]	54(13) [-1]	20(10) [-1]	32(57) [4]	19(36) [0]	20(90) [2]
BIBDme1 (89)		48(93) [18]	30(89) [10]	38(80) [12]	35(65) [7]	44(32) [5]	30(88) [6]	23(86) [2]	41(17) [-0.4]	31(36) [0.3]	19(37) [3]	26(69) [6]	23(69) [2]	29(38) [1]	18(62) [1]
CHIPHsa1 (93)			36(55) [10]	40(97) [20]	36(97) [14]	31(59) [8]	32(93) [12]	19(43) [2]	54(13) [0.3]	39(31) [1]	25(85) [4]	27(11) [-1]	36(36) [4]	19(63) [1]	20(50) [4]
NODGma1 (88)				47(58) [10]	32(97) [6]	38(61) [11]	36(90) [11]	15(34) [0.1]	50(10) [-1]	17(47) [-1]	27(88) [6]	100(6) [-0.5]	25(64) [3]	17(36) [-0.5]	21(43) [1]
WTIPAth1 (97)					39(97) [17]	30(56) [6]	32(81) [7]	25(24) [-0.1]	67(9) [-0.6]	58(12) [-0.6]	50(8) [-0.3]	31(58) [5]	25(63) [2]	33(18) [-1]	29(34) [0.4]
αTIPAth1 (98)						33(42) [5]	26(55) [5]	28(54) [5]	50(12) [2]	26(53) [4]	33(57) [2]	23(83) [9]	24(79) [7]	29(14) [0.3]	40(10) [-1]
FPSSce1 (83)							36(67) [12]	19(26) [-0.4]	25(20) [-0.1]	57(7) [-0.3]	25(40) [2]	57(7) [-0.2]	33(9) [0.1]	19(37) [0.2]	24(41) [-0.1]
GLPEco1 (88)								12(33) [0.1]	100(2) [-1]	35(26) [1]	23(22) [0.1]	24(62) [1]	26(61) [1]	29(14) [-0.2]	22(23) [0.3]
MIPRno2 (89)									37(87) [15]	59(87) [38]	37(87) [16]	36(93) [18]	41(88) [24]	38(55) [5]	19(58) [2]
BIBDme2 (84)										34(89) [20]	30(88) [14]	42(38) [12]	37(90) [15]	50(18) [4]	43(14) [3]
CHIPHsa2 (89)											32(81) [15]	52(94) [25]	51(90) [28]	50(18) [5]	34(61) [5]
NODGma2 (88)												30(90) [11]	40(90) [18]	37(19) [3]	34(96) [6]
WTIPAth2 (96)													47(95) [23]	35(69) [7]	38(45) [5]
αTIPAth2 (90)														38(34) [6]	34(67) [9]
FPSSce2 (101)															35(103) [12]

Binary comparisons of the first and second segments of the eight representative MIP family proteins. The solid thick lines segregate comparisons of (A) segments 1 with themselves (top left), (B) segments 1 with segments 2 (top right), and (C) segments 2 with themselves (bottom right). Highlighted boxes (black background) indicate comparisons of segments 1 with segments 2 from the same proteins. The number in parentheses below the abbreviation for a particular segment indicates the length in residues of that segment. The FASTA program using the dipeptide identities mode (ktup2; Pearson and Lipman, 1988) was used to assess similarities of the indicated proteins. Values in the table which are not in parentheses or brackets represent percent identities for segments having the number of compared residues indicated in parentheses. Comparison scores in standard deviations, using the RDF2 program (Pearson and Lipman, 1988) and 300 shuffles, are given in brackets with the values for percent identity.

ments 1 are more similar to each other than are the segments 2 when they are derived from distantly related proteins, but that segments 2 are more similar to each other than are segments 1 when they are derived from closely related proteins.

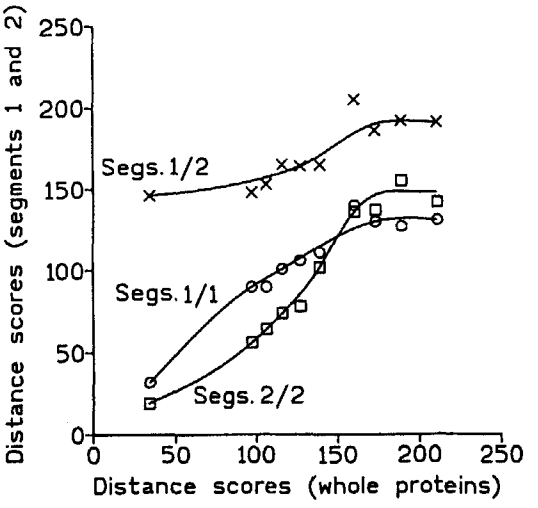

FIGURE 8. Correlation of the averaged distance scores between intact protein members of the MIP family (plotted on the X axis) with the corresponding segment 1:segment 1 comparisons (Segments 1/1; 120 total), the corresponding segment 2:segment 2 comparisons (segments 2/2; 120 total), and the corresponding segment 1:segment 2 comparisons (segments 1/2; 240 total), plotted on the Y axis. See text for details.

Also presented in Figure 8 are the results of a similar averaging process when segments 1 are compared with segments 2. It can be seen that segments 1 are more similar to segments 2 for similar proteins than for highly divergent proteins. This observation suggests that while segments 1 and 2 play different functional roles, they nevertheless function in some cooperative capacity. This proposed common function must account for the evolutionary pressure which caused similar proteins to exhibit segment 1 sequences that more closely resemble segment 2 sequences than for the evolutionarily divergent proteins of the MIP family.

These observations lead to the conclusion that segments 1 must be more important for the generalized or common structure/function of the entire family of MIP proteins, while segments 2 are more important for the specific or dissimilar functions of these proteins, shared only by closely related proteins. A common function primarily involving segments 1 might be proper membrane insertion or channel formation, while a dissimilar function, determined more by the segments 2, might be solute specificity or channel regulation. Both segments might function cooperatively, for example, to create the channel, but with a greater dependency on segment 1 than on segment 2. The results of site-specific mutagenesis studies are likely to provide answers as to the nature of the functional differentiation that these repeat sequences have undergone during evolutionary history.

IV. DISCUSSION

The two halves of all MIP family proteins were apparently derived from a common ancestral gene, half as big as the current genes, and this ancestral gene tandemly duplicated internally to give rise to the primordial gene of the MIP family proteins. We presume that this event occurred in prokaryotes, well before divergence of the species. MIP family proteins are roughly related to each other as are the organisms in which they are found (Figures 1 and 7). Because only one such gene has been found in any one bacterium, but several have been found in single eukaryotes, we propose that a single MIP family gene was transmitted vertically from the prokaryotic ancestor to each of the eukaryotic kingdoms (animals, plants, and fungi) and that these genes then duplicated and diverged within the eukaryotes to yield current MIP protein subfamilies.

Initial statistical analyses revealed that segments 2 exhibit a higher degree of average similarity (i.e., they have lower average distance scores) than do the segments 1, but that the segments 2 also exhibit more variability (data not presented). Comparison of the segment 1 vs. segment 1 distance scores with the corresponding segment 2 vs. segment 2 distance scores, both plotted as a function of the distance scores for the intact proteins (Figure 8), revealed that when the proteins are closely related, the segment 2 scores are usually smaller than the segment 1 scores, but

when the proteins are distantly related, the segment 1 scores are generally smaller than the segment 2 scores (see Figure 8). This observation led us to propose that the N-terminal halves evolved for a function largely common to all MIP family proteins, such as channel formation or membrane insertion, while the second halves differentiated to play a subfamily specific role such as solute specificity or regulation. Our statistical analyses appear to provide the first evidence for regional differentiation of permease proteins derived from repeated internal structural elements.

The picture emerging from the statistical analysis of the earlier-mentioned distance scores is, however, incomplete. When segment 1 vs. segment 2 distance scores are compared with each other for structurally and functionally divergent MIP family proteins when compared with MIP family proteins of similar structure and function, we found that segments 1 are more similar to segments 2 for the similar or identical proteins. This observation suggests that the two segments have retained at least one common function that depends on some degree of sequence similarity between the two segments.

We have recently conducted comparative functional analyses employing two of the MIP family proteins, i.e., γ-TIPAth and GLPEco, following expression in *Xenopus oocytes* (Maurel et al., 1993). The results of [^{14}C]glycerol uptake measurements have established that GLPEco, but not γ-TIPAth, can transport this straight chain carbon compound. The results show, however, that of these two proteins, only γ-TIPAth readily transports water. Detailed analyses of [^{14}C]glycerol uptake via GLPEco has confirmed virtually all of the functional characteristics reported by Heller et al. (1980). Thus, GLPEco is a simple channel protein capable of facilitating the nonspecific passage of straight chain compounds across the membrane. We anticipate that all protein members of the MIP family will prove to exhibit simple channel characteristics, but each for differing substrates. Current evidence favors this notion although more extensive experimental work will be required to establish the nature of the substrates transported in each individual case.

The analyses reported in an earlier paper (Pao et al., 1991) and presented more extensively in Table 2 and Figure 7 have also led to the proposal that MIP family proteins evolved by duplication of a triplex, that is, a three-spanner precursor (Figure 2), about 2.5 billion years ago, shortly before the advent of eukaryotes on Earth (Figure 9). By contrast, the mitochondrial carrier family (MCF) apparently evolved more recently by triplication of a duplex, that is, a two-spanner precursor, probably about 1.5 billion years ago, *after* the advent of mitochondria within eukaryotes (Kuan and Saier, 1993; see accompanying paper). All members of the latter family occur in eukaryotic organelles (most are within mitochondria), and none have yet been found in prokaryotes. Comparison scores obtained when comparing the internal repeat sequences within protein members of the MCF are much larger than those obtained when corresponding analyses are conducted with the MIP family (Figure 9 and Kuan and Saier, 1993). These findings indicate that the internal repeats within the MCF protein members are less divergent than the internal repeats within various members of the MIP protein family. Finally, it has been suggested that the major facilitator superfamily (MFS) (Marger and Saier, 1993; Reizer et al., 1993) evolved much earlier, possibly by duplication of a six-spanner precursor (Henderson and Maiden, 1990; Henderson, 1991; Griffith et al., 1992). This last-mentioned duplication event probably occurred greater than 3.5 billion years ago, long before the advent of eukaryotes on Earth. It is probably only in this last family, the MFS, that the duplication and divergence events, giving rise to currently recognized protein members of the MFS, occurred long before the prokaryotic-eukaryotic divergence (Marger and Saier, 1993). Thus, although these three transport protein families (the MIP family, the MCF, and the MFS) evolved independently of each other, by distinct routes, and at different times in evolutionary history, they all generated fundamental structural units of six transmembrane spanners. This observation disproves the proposal of several investigators, suggesting that transport proteins exhibiting the characteristic six-spanner unit (Saier and Reizer, 1991; Nikaido and Saier, 1992) might share a common evolutionary origin (Krupinski et al., 1989; Maloney, 1990; Saier, 1990; Jan and Jan, 1992). Why this six transmembrane structural element is so often used for transport function has yet to be ascertained.

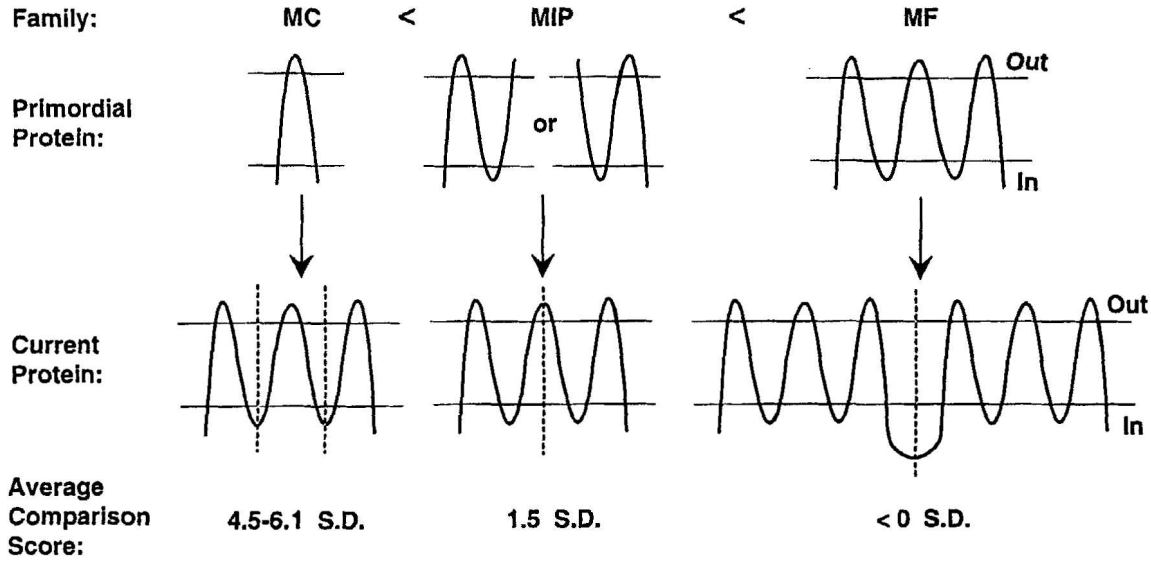

FIGURE 9. Proposed dissimilar evolutionary pathways for three families of transport proteins all of which exhibit the six spanner structural unit (Saier and Reizer, 1991; Nikaido and Saier, 1992). The mitochondrial carrier (MC) family arose by triplication of a two-spanner segment, estimated to have occurred about 1.5 billion years ago (average comparison scores between segments of 4.5 to 6.1 SD; Kuan and Saier, 1993). The MIP family arose by duplication of a three-spanner segment, about 2.5 billion years ago (average comparison scores between segments 1 and 2 of 1.5 SD), while the major facilitator (MF) family may have arisen by duplication of a six-spanner segment, an event which occurred more than an estimated 3.5 billion years ago (comparison scores between six-spanner segments of <0 SD) (Henderson and Maiden, 1990; Henderson, 1990, 1991; Griffith et al., 1992; Marger and Saier, 1993).

REFERENCES

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

Aerts, T., Xia, J.-Z., Slegers, H., de Block, J., and Clauwaert, J. (1990), Hydrodynamic characterization of the major intrinsic protein from the bovine lens fiber membranes, *J. Biol. Chem.*, **265:** 8675–8680.

Bairoch, A. (1992), Prosite — a dictionary of sites and patterns in proteins, *Nucl. Acids Res.*, **20** (Suppl.): S2013–S2018.

Baker, M. E. and Saier, M. H., Jr., (1990), A common ancestor for bovine lens fiber major intrinsic protein, soybean nodulin-26 protein, and *E. coli* glycerol facilitator, *Cell*, **60:** 185–186.

Devereux, J., Haeberli, P., and Smithies, O., (1984), A comprehensive set of sequence analyses for the VAX, *Nucl. Acids Res.*, **12:** 387–395.

Doolittle, R. F. (1986) *Of Urfs and Orfs. A Primer on How to Analyze Derived Amino Acid Sequences*, University Science Books, Mill Valley, CA.

Doolittle, R. F. and Feng, D.-F. (1990), Nearest neighbor procedure for relating progressively aligned amino acid sequences, *Meth. Enzymol.*, **183:** 659–669.

Ehring, G. R., Zampighi, G., Horwitz, J., Bok, D., and Hall, J. E. (1990), Properties of channels reconstituted from the major intrinsic protein of lens fiber membranes, *J. Gen. Physiol.*, **96:** 631–664.

Ehring, G. R., Lagos, N., Zampighi, G. A., and Hall, J. E. (1991), Phosphorylation modulates the voltage dependence of channels reconstituted from the major intrinsic protein of lens fiber membranes, *J. Membr. Biol.*, **126:** 75–88.

Ehring, G. R., Zampighi, G. A., and Hall, J. E. (1993), Does MIP play a role in cell-cell communication? in *Gap Junctions, Progress in Cell Research, III*, Hall, J. E., Zampighi, G. A., and Davis, R. M., Eds., Elsevier Publishers, Amsterdam; pp. 143–152.

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

Gorin, M. B., Yancey, S. B., Cline, J., Revel, J.-P., and Horwitz, J. (1984), The major intrinsic protein (MIP) of the bovine lens fiber membrane: characterization and structure based on cDNA cloning, *Cell*, **39:** 49–59.

Griffith, J. K., Baker, M. E., Rouch, D. A., Page, M. G. P., Skurray, R. A., Paulsen, I. T., Chater, K. F., Baldwin, S. A., and Henderson P. J. F. (1992), Membrane transport proteins: implications of sequence comparisons, *Curr. Opin. Cell Biol.*, **4:** 684–695.

Guerrero, F. D., Jones, J. T., and Mullet, J. E. (1990), Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted. Sequence and expression of three inducible genes, *Plant Mol. Biol.*, **15:** 11–26.

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

Henderson, P. J. F. (1990), The homologous glucose transport proteins of prokaryotes and eukaryotes, *Res. Microbiol.*, **141:** 316–328.

Henderson, P. J. F. (1991), Sugar transport proteins, *Curr. Opin. Struct. Biol.*, **1:** 590–601.

Henderson, P. J. F. and Maiden, M. C. J. (1990), Homologous sugar transport proteins in *Escherichia coli* and their relatives in both prokaryotes and eukaryotes, *Philos. Trans. R. Soc. London Ser. B*, **326:** 391–410.

Höfte, H., Hubbard, L., Reizer, J., Ludevid, D., Herman, E. M., and 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.

Holmberg, C., Beijer, L., Rutberg, B., and Rutberg, L. (1990), Glycerol catabolism in *Bacillus subtilis:* nucleotide sequence of the genes encoding glycerol kinase (glpK) and glycerol-3-phosphate dehydrogenase (glpD), *J. Gen. Microbiol.*, **136:** 2367–2375.

Horwitz, J. and Bok, D. (1987), Conformational properties of the main intrinsic polypeptide (MIP26) isolated from lens plasma membranes, *Biochemistry*, **26:** 8092–8098.

Jan, L. Y. and Jan, Y. N. (1992), Tracing the roots of ion channels, *Cell*, **69:** 715–718.

Johnson, K. D. and Chrispeels, M. J. (1992), A tonoplast-bound protein kinase phosphorylates tonoplast intrinsic protein, *Plant Physiol.*, **100:** 1787–1795.

Johnson, K. D. Höfte, H., and Chrispeels, M. J. (1990), An intrinsic tonoplast protein of protein storage vacuoles in seeds is structurally related to a bacterial solute transporter (GlpF), *Plant Cell*, **2:** 525–532.

Johnson, K. R., Lampe, P. D., Hur, K. C., Louis, C. F., and Johnson, R. G. (1986), A lens intracellular junction protein, MP26, is a phosphoprotein, *J. Cell. Biol.*, **102:** 1334–1343.

Kent, N. A. and Shiels, A. (1990), Nucleotide and derived amino-acid sequence of the major intrinsic protein of rat eye-lens, *Nucl. Acids Res.* **18:** 4256.

Kodama, R., Agata, K., Mochii, M., and Eguchi, G. (1990), Partial amino acid sequence of the major intrinsic protein (MIP) of the chicken lens deduced from the nucleotide sequence of a cDNA clone, *Exp. Eye Res.*, **50:** 737–741.

Krupinski, J., Coussen, F., Bakalyar, H. A., Tang, W.-J., Feinstein, P. G., Orth, K., Slaughter, C., Reed, R. R., and Gilman, A. G. (1989), Adenylyl cyclase amino acid sequence: possible channel- or transporter-like structure, *Science*, **244:** 1558–1564.

Kuan, J. and Saier, M. H., Jr. (1993), The mitochondrial carrier family of transport proteins: structural, functional and evolutionary relationships, *Crit. Revs. Biochem. Mol. Biol.*, **28:** 209–233.

Kyte, J. and Doolittle, R. F. (1982), A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.*, **157:** 105–132.

Lampe, P. D. and Johnson, R. G. (1989), Phosphorylation of MP26, a lens junction protein, is enhanced by activators of protein kinase C., *J. Membr. Biol.*, **107:** 145–155.

Lampe, P. D., Bazzi, M. D., Nelsestuen, G. L., and Johnson, R. G. (1986), Phosphorylation of lens intrinsic membrane proteins by protein kinase C., *Eur. J. Biochem.*, **156:** 351–357.

Lampe, P. D., Eccleston, E. D., Howard, J. B., and Johnson, R. G. (1988), The amino acid sequence of *in vivo* phosphorylation site of lens MP26 *J. Cell. Biol.* (Abstr.)., **107:** 558.

Ludevid, D., Höfte, H., Himelblau, E., and Chrispeels, M. J. (1992), The expression pattern of the tonoplast intrinsic protein γ-TIP in *Arabidopsis thaliana* is correlated with cell enlargement, *Plant Physiol.*, **100:** 1633–1639.

Maeshima, M. (1992), Characterization of the major integral protein of vacuolar membrane, *Plant Physiol.*, **98:** 1248–1254.

Maloney, P. C. (1990), A consensus structure for membrane transport, *Res. Microbiol.*, **141:** 374–383.

Marger, M. D. and Saier, M. H., Jr. (1993), A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport, *TIBS*, **18:** 13–20.

Maurel, C., Reizer, Schroeder, J. I., and Chrispeels, M. J. (1993), The vacuolar membrane protein γ-TIP creates water specific channels in *Xenopus oocytes, EMBO. J.*, in press.

Mayo, B., Kok, J., Venema, K., Brockelmann, W., Teuber, M., Reinke, H., and Venema, G. (1991), Molecular cloning and sequence analysis of the X-prolyl dipeptidyl aminopeptidase gene from *Lactococcus lactis* subsp. *cremoris., Appl. Environ. Microbiol.*, **57:** 38–44.

Miao, G.-H., Hong, Z., and Verma, D. P. S. (1992), Topology and phosphorylation of soybean nodulin-26, an intrinsic protein of the peribacteroid membrane, *J. Cell Biol.*, **118:** 481–490.

Muramatsu, S. and Mizuno, T. (1989), Nucleotide sequence of the region encompassing the *glpKF* operon and its upstream region containing a bent DNA sequence of *Escherichia coli, Nucl. Acids Res.*, **17:** 4378.

Nikaido, H. and Rosenberg, E. Y. (1985), Functional reconstitution of lens gap junction proteins into proteoliposomes, *J. Membr. Biol.*, **85:** 87–92.

Nikaido, H. and Saier, M. H., Jr. (1992), Transport proteins in bacteria: common themes in their design, *Science*, **258:** 936–942.

Ouyang, L., Whelan, J., Weaver, C. D., Roberts, D. M., and Day, D. A. (1991), Protein phosphorylation stimulates the rate of malate uptake across the peribacteroid membrane of soybean nodules, *FEBS Letts.*, **293:** 188–190.

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

Pearson, W. R. and Lipman, D. J. (1988), Improved tools for biological sequence comparison, *Proc. Natl. Acad. Sci. U.S.A.*, **85:** 2444–2448.

Pisano, M. M. and Chepelinsky, A. B. (1991), Genomic cloning, complete nucleotide sequence, and structure of the human gene encoding the major intrinsic protein (MIP) of the lens, *Genomics*, **11:** 981–990.

Preston, G. M. and Agre, P. (1991), Isolation of the cDNA for erythrocyte integral membrane protein of 28 kilodaltons: member of an ancient channel family, *Proc. Natl. Acad. Sci. U.S.A.*, **88:** 11110–11114.

Preston, G. M., Carroll, T. P., Guggino, W. B., and Agre, P. (1992), Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein, *Science*, **256:** 385–387.

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

Reizer, A. and Reizer, J. (1993), Progressive multiple alignment of protein sequences and the construction of phylogenetic trees, in *Methods in Molecular Biology: Computer Analysis of Sequence Data*, Griffin, A. M. and Griffin, H. G., Eds., Humana Press, Totowa, NJ, in press.

Reizer, J., Finley, K., Kakuda, D., MacLeod, C. L., Reizer, A., and Saier, M. H., Jr. (1993), Mammalian integral membrane receptors are homologous to facilitators and antiporters of yeast, fungi, and eubacteria, *Prot. Sci.*, **2:** 20–30.

Saier, M. H., Jr. (1990), Evolution of permease diversity and energy-coupling mechanisms: an introduction, *Res. Microbiol.*, **141:** 281–286.

Saier, M. H., Jr. and Reizer, J. (1991), Families and superfamilies of transport proteins common to prokaryotes and eukaryotes, *Curr. Opin. Struct. Biol.*, **1:** 362–368.

Sandal, N. N. and Marcker, K. A. (1988), Soybean nodulin 26 is homologous to the major intrinsic protein of the bovine lens fiber membrane, *Nucl. Acids Res.*, **16:** 9347.

Sandal, N. N. and Marcker, K. A. (1990), Some nodulin and Nod proteins show similarity to specific animal proteins, in *Proceedings of the 8th International Congress on Nitrogen Fixation*, Greshoff, P. M., Ed., Chapman and Hall, New York, pp. 687–692.

Shiels, A., Kent, N. A., McHale, M., and Bangham, J. A. (1988), Homology of MIP26 to Nod26, *Nucl. Acids Res.*, **16:** 9348.

Smith, B. L. and Agre, P. (1991), Erythrocyte M28,000 transmembrane protein exists as a multi-subunit oligomer similar to a channel protein, *J. Biol. Chem.*, **266:** 6407–6415.

Smith, C. P. and Chater, K. F. (1988), Structure and regulation of controlling sequences for the *Streptomyces coelicolor* glycerol operon, *J. Mol. Biol.*, **204:** 569–580.

Smith, D. W. (1988), A complete and integrated system for DNA protein sequence analysis using VAX/VMS computers, *Comp. Appl. BioSci.*, **4:** 212.

Takemoto, L. and Takehana, M. (1986), Major intrinsic polypeptide (MIP26K) from human lens membrane: characterization of low-molecular-weight forms in the aging human lens, *Exp. Eye Res.*, **43:** 661–667.

Takemoto, L., Kodama, T., and Takemoto, D. (1987), Antisera to synthetic peptides of MIP26K as probes of changes in opaque vs. transparent regions within the same human cataractous lens, *Exp. Eye Res.*, **45:** 179–183.

Takemoto, L., Kuck, J., and Kuck, K. (1988), Changes in the major intrinsic polypeptide (MIP26K) during opacification of the Emory mouse lens, *Exp. Eye Res.*, **47:** 329–336.

Van Aelst, L., Hohmann, S., Zimmermann, F. K., Jans, A. W. H., and 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.

van Hoek, A. N. and Verkman, A. S. (1992), Functional reconstitution of the isolated erythrocyte water channel CHIP28, *J. Biol. Chem.*, **267:** 18267–18269.

Verma, D. P. S. (1992), Soybean nodulin-26: a channel protein conserved from bacteria to mammals, in *Transport and Receptor Proteins of Plant Membranes*, Cooke, D. T. and Clarkson, D. T., Eds., Plenum Press, New York, pp. 113–117.

Voegele, R., Sweet, G., and Boos, W. (1992), Isolation and characterization of the glycerol transport protein, GlpF, from *Escherichia coli*, *Bio Engineering* (abst P212), February, p. 57.

von Heijne, G. (1992), Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule, *J. Mol. Biol.*, **225:** 487–494.

Weaver, C. D. and Roberts, D. M. (1992), Determination of the site of phosphorylation of nodulin 26 by the calcium-dependent protein kinase from soybean nodules, *Biochemistry*, **31:** 8954–8959.

Weaver, C. D., Crombie, B., Stacey, G., and Roberts, D. M. (1991), Calcium-dependent phosphorylation of symbiosome membrane proteins from nitrogen-fixing soybean nodules, *Plant Physiol.*, **95:** 222–227.

Yamaguchi-Shinozaki, K., Koizumi, M., Urao, S., and Shinozaki, K. (1992), Molecular cloning and characterization of 9 cDNAs for genes that are responsive to desiccation in *Arabidopsis thaliana* — sequence analysis of one cDNA clone that encodes a putative trans-

membrane channel protein, *Plant Cell Physiol.*, **33:** 217–224.

Yamamoto, Y. T., Cheng, C. L., and Conkling, M. A. (1990), Root-specific genes from tobacco and *Arabidopsis* homologous to an evolutionary conserved gene family of membrane channel proteins, *Nucl. Acids Res.*, **18:** 7449.

Yamamoto, Y. T., Taylor, C. G., Acedo, G. N., Cheng, C. L., and Conkling, M. A. (1991), Characterization of *cis*-acting sequences regulating root-specific gene expression in tobacco, *Plant Cell*, **3:** 371–382.

Zampighi, G. A., Hall, J. E., Ehring, G. R., and Simon, S. A. (1989), The structural organization and protein composition of lens fiber junctions, *J. Cell Biol.*, **108:** 2255–2276.

Zeidel, M. L., Ambudkar, S. V., Smith, B. L., and Agre, P. (1992), Reconstitution of functional water channels in liposomes containing purified red cell CHIP28 protein, *Biochemistry*, **31:** 7436–7440.