MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Mar. 1998, p. 1–34 Vol. 62, No. 1
1092-2172/98/$04.00+0

Major Facilitator Superfamily

STEPHANIE S. PAO, IAN T. PAULSEN,† AND MILTON H. SAIER, JR.*

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

"If you do not expect to, you will not discover the unexpected."

Heraclitus

INTRODUCTION

Transport systems allow the uptake of essential nutrients and ions, excretion of end products of metabolism and deleterious substances, and communication between cells and the environment (53). They also provide essential constituents of energy-generating and energy-consuming systems (54). Primary active transporters drive solute accumulation or extrusion by using ATP hydrolysis, photon absorption, electron flow, substrate decarboxylation, or methyl transfer (17). If charged molecules are unidirectionally pumped as a consequence of the consumption of a primary cellular energy source, electrochemical potentials result (54). The consequential chemiosmotic energy generated can then be used to drive the active transport of additional solutes via secondary carriers which merely facilitate the transport of one or more molecular species across the membrane (48, 49).

Recent genome-sequencing data and a wealth of biochemical and molecular genetic investigations have revealed the occurrence of dozens of families of primary and secondary transporters (63). Two such families have been found to occur ubiquitously in all classifications of living organisms. These are the ATP-binding cassette (ABC) superfamily (15, 21, 37, 44) and the major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family (7, 28, 30, 35, 51). While ABC family permeases are in general multicomponent primary active transporters, capable of transporting both small molecules and macromolecules in response to ATP hydrolysis (59), the MFS transporters are single-polypeptide secondary carriers capable only of transporting small solutes in response to chemiosmotic ion gradients. Although well over 100 families of transporters have now been recognized and classified (73), the ABC superfamily and MFS account for nearly half of the solute transporters encoded within the genomes of microorganisms (63). They are also prevalent in higher organisms. The importance of these two families of transport systems to living organisms can therefore not be overestimated.

* Corresponding author. Mailing address: Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116. Phone: (619) 534-4084. Fax: (619) 534-7108. E-mail: msaier@ucsd.edu.

† Present address: School of Biological Sciences, University of Sydney, New South Wales, 2006 Australia.

The MFS was originally believed to function primarily in the uptake of sugars (36, 46). Subsequent studies revealed that drug efflux systems and Krebs cycle metabolites belong to this family (30, 62). The family was then expanded to include organophosphate:phosphate exchangers and oligosaccharide:H^+ symport permeases (51). Reizer et al. (67) noted that a mammalian phosphate:Na^+ symporter is a distant member of this family; Paulsen et al. (60) subdivided the MFS drug efflux pumps into two phylogenetically distinct families with differing topologies; and Goffeau et al. (26) identified a novel MFS family that consists exclusively of functionally uncharacterized proteins from *Saccharomyces cerevisiae* revealed by genome sequencing. Recently, Williams and Shaw (86) noted that a family of bacterial aromatic acid permeases belongs to the MFS. These observations led to the probability that the MFS is far more widespread in nature and far more diverse in function than had been thought previously.

Although isolated reports have allowed recognition of an increasing degree of diversity within the MFS, there has been no recent systematic attempt to identify the sequenced proteins that make up the MFS and to classify these proteins into phylogenetic families. We have therefore undertaken this task in the hopes of allowing (i) recognition of the significance of this family to cell physiology; (ii) extrapolation of biochemical, molecular genetic, and biophysical information obtained from the study of a few such systems to all members of the family; (iii) unification of mechanistic models, to the greatest extent possible, so as to be applicable to a maximal number of transporters; (iv) introduction of a rational system of MFS protein classification; and (v) comprehension of the pathways taken in the development of structural and functional diversity resulting from the evolutionary process used.

In this report, we present analyses that allow us to generalize some previous observations regarding the MFS and to note additional characteristics of this immense superfamily. Thus, based exclusively on degrees of sequence similarity, we have constructed phylogenetic trees which allow us to divide all the recognized members of the MFS into 17 families. The members of each family all proved to be more closely related in sequence to each other than they were to any of the other MFS proteins. This fact presumably reflects the evolutionary histories of these proteins (71, 72), and, remarkably, we find that phylogenetic family correlates with function. Thus, each of the families recognizes and transports a distinct class of structurally related compounds. These observations have allowed us to derive a rational classification system for the MFS based on both phylogeny and function. This classification system has proven applicable to virtually all permeases found in nature (73).

In 1990, Rubin et al. (70) presented evidence that strongly argued in favor of an earlier suggestion (see reference 36), that MFS permeases arose by a tandem intragenic duplication event. In this report, we provide additional statistical evidence in favor of this possibility. This event generated the 12-transmembrane-spanner (TMS) protein topology from a primordial 6-TMS unit. Surprisingly, all currently recognized MFS permeases retain the two six-TMS units within a single polypeptide chain, although in 3 of the 17 MFS families, an additional two TMSs are found (60). Moreover, the well-conserved MFS-specific motif between TMS2 and TMS3 and the related but less well conserved motif between TMS8 and TMS9 (36) prove to be a characteristic of virtually all of the more than 300 MFS proteins identified. The functional significance of this repeated motif has been examined by Jessen-Marshall et al. (39) and by Yamaguchi et al. (87–89).

Many additional observations allowed the identification of highly specific characteristics of individual MFS families as well as general characteristics of the MFS as a whole. We hope that the computational analyses reported will provide a guide for molecular biologists, biochemists, and biophysicists interested in structural, functional, and evolutionary aspects of MFS permeases.

COMPUTER METHODS

The FASTA (64) and BLAST (2) programs were used to screen the peptide and translated nucleotide databases. The statistical significance of sequence similarities between putative members of the various families of the MFS was established by using the RDF2 (64) and GAP (16) programs with at least 200 random shuffles. Binary comparison scores are expressed in standard deviations (SD) (14). A value of 9 SD for a protein segment larger than 60 residues is deemed sufficient to establish homology (18, 71). This criterion was used to establish homology between MFS families (see Table 2).

Multiple-sequence alignments were constructed with the PREALIGN and TREE programs of Feng and Doolittle (22) and the PILEUP program (16). Phylogenetic analyses were routinely performed with the TREE program (22) but were checked with other programs. The different programs generally gave very similar, and often identical, branching orders, and the branch lengths were also strikingly similar. Branch length is approximately proportional to the degree of sequence divergence, which, to a first approximation, is assumed to be proportional to the phylogenetic distance (but see the section Conclusions and Perspectives, below). It is important to emphasize that branch lengths and even branch positions represent approximations to the evolutionary process, allowing facile visualization of the relationships between sequences within families. They reflect relative degrees of sequence divergence and can be considered to represent the evolutionary process only to a first approximation (71, 72).

Average hydropathy, average amphipathicity and average similarity analyses were conducted for all protein families analyzed. They were based on the complete multiple-sequence alignments generated with the TREE program. Only representative, well-conserved portions of these multiple-sequence alignments are presented. The hydropathy analyses were conducted with the assumptions and algorithm described by Kyte and Doolittle (45) with a sliding window of 20 residues. Similarly, a sliding window of 20 residues was used to generate the average similarity and average amphipathicity plots (45a). These latter analyses are not presented but are described in the text (see Table 2).

Charge bias analysis of membrane protein topology was performed with the program TOP PRED (83). Signature sequences were defined by the method of Bairoch et al. (6). The programs MEME and MAST (5) were used to help identify conserved motifs within the protein families of the MFS. Most of the methods used in this study have been applied to numerous transport proteins and have been evaluated (see references 71 and 72 for recent reviews).

SEVENTEEN MFS FAMILIES

Table 1 lists and summarizes the properties of transport protein families found within the current MFS. We have classified current members of the MFS into 17 (possibly 18) distinct families. This number of MFS families represents more than a threefold expansion over that published previously (51). The table provides the family number; the name of the family; the abbreviation of the family to be used in this study; the

TABLE 1. Families within the MFS

Family no.	Family name	Abbreviation	No. of proteins	Source(s)[a]	Size range (aa)	Putative TMSs (no.)	Mechanism(s)	Polarities	Substrates	Representative example(s)
1	Sugar porter	SP	133	Bac, Ar, Pr, Y, An, Pl	404–818	12	Sugar uniport Sugar:proton symport Sugar:sugar antiport	None In Both	Monosaccharides (hexoses, pentoses), disaccharides, quinate, organo-cations, inositols	XylE of *E. coli*
2	Drug:H^+ antiporter (14 TMS)	DHA14	30	Bac, Y		14	Drug:H^+ antiport	Out	Multiple or single drugs	QacA of *S. aureus*
3	Drug:H^+ antiporter (12 TMS)	DHA12	46	Bac, Ar, Y, An		12	Drug:H^+ antiport	Out	Multiple or single drugs	Bmr of *B. subtilis*
4	Organophosphate:P_i antiporter	OPA	12	Bac, An	439–495	12	Organo phosphate:P_i antiport	Both	Sugar-phosphates, glycerol phosphate, phosphoglycerates, phosphoenolpyruvate	UhpT of *E. coli*
5	Oligosaccharide:H^+ symporter	OHS	6	Bac	415–425	12	Sugar:H^+ symport Sugar:sugar antiport	In Both	Di- and trisaccharides	LacY of *E. coli*
6	Metabolite:H^+ symporter	MHS	16	Bac	425–500	12	Solute:H^+ symport	In	Citrate, α-ketoglutarate, proline, betaine, methylphthalate, dicarboxylates	KgtP of *E. coli*
7	Fucose-galactose-glucose:H^+ symporter	FGHS	4	Bac	404–438	12	Hexose uniport Hexose:H^+ symport	None In	L-Fucose, glucose, galactose	FucP of *E. coli*
8	Nitrate/nitrite porter	NNP	13	Bac, Y, Pl	395–547	12	Nitrite uniport? Nitrate:H^+ symport?	Out In	Nitrite, nitrate	NarK of *E. coli*
9	Phosphate:H^+ symporter	PHS	11	Y, Pl	518–587	12	P_i:H^+ symport	In	Inorganic phosphate	Pho-5 of *N. crassa*
10	Nucleoside:H^+ symporter	NHS	2	G− Bac	418	12	Nucleoside:H^+ symport	In	Nucleosides	NupG of *E. coli*
11	Oxalate:formate antiporter	OFA	5	Bac, Ar, An	373–470	12	Anion:anion antiport	Both	Oxalate, formate	OxlT of *O. formigenes*
12	Sialate:H^+ symporter	SHS	3	G− Bac	407–496	14	Substrate:H^+ symport	In	Sialate	NanT of *E. coli*
13	Monocarboxylate porter	MCP	13	Y, An	450–808	12	Substrate:H^+ symport	In	Pyruvate, lactate, mevalonate	Mct of *H. sapiens*
14	Anion:cation symporter	ACS	40	Bac, Y, An	411–596	12	Substrate:H^+ or Na^+ symport	In	Glucarate, hexuronate, tartrate, 4-hydroxyphenyl acetate, inorganic phosphate, allantoate	ExtU of *E. coli*
15	Aromatic acid:H^+ symporter	AAHS	7	Bac	418–460	12	Substrate:H^+ symport	In	Muconate, benzoate; 4-hydroxybenzoate; 2,4-dichlorophenoxy acetate, protocatachurate, 3-hydroxypropionate	PcaK of *P. putida*
16	Unknown major facilitator	UMF	6	Y	600–637	14	Unknown	Unknown	Unknown	Yh1040c of *S. cerevisiae*
17	Cyanate permease	CP	3	Bac	393–402	12	Substrate:H^+ symport?	In	NCO^-	CynX of *E. coli*
18	Proton-dependent oligopeptide transporter	POT	24	Bac, Y, An, Pl	463–783	12	Substrate:H^+ symport	In	Peptides, amino acids, nitrate, chlorate, nitrite	DtpT of *L. lactis*

[a] The abbreviations for source organisms used here and in all subsequent tables are as follows: Bac, bacteria; G−, gram negative; Ar, archaea; Pr, eukaryotic protists (usually protozoans); Y, yeasts; F, fungi; Pl, plants; An, animals.

number of currently recognized sequenced members in each family; the range of organisms in which members of the family are found; the size range of the proteins (in numbers of amino acyl residues) for fully sequenced members; the number of putative TMSs in each protein (believed to be uniform for members of a given family); the energy-coupling mechanisms, if any, used by members of the family; the polarities of transport catalyzed by family members; the substrates known to be transported by various members of the family; and a representative and well-characterized member of the family.

The largest family (family 1) is the sugar porter (SP) family, with 133 identified members. These proteins are derived from all of the major groups of living organisms: bacteria, archaea, eukaryotic protists, fungi, mostly yeasts, animals, and plants. These proteins have 12 established or putative TMSs. They can function by uniport, solute:solute antiport, and/or solute:cation

symport, depending on the system and/or conditions. Uniporters exhibit no polarity but can usually catalyze both uniport and antiport depending on whether a substrate is present on the *trans* side of the membrane. The polarity of solute:solute antiporters is indicated in Table 1 by "both." Symporters function with inwardly-direct polarity in the presence of a membrane potential (negative inside), but many of these proteins have also been shown to catalyze antiport when a substrate is present on the *trans* side of the membrane. Substrates transported by SP family members include hexoses, pentoses, disaccharides, quinate, inositols, and organic cations. Most but not all members of the SP family thus catalyze sugar transport.

Family 1 permeases exhibit a size range of 404 to 818 residues. The smaller permeases possess very short hydrophilic N and C termini and short loops connecting the 12 TMSs. As is true of many MFS families, the bacterial sugar porters are usually smaller than the eukaryotic proteins. The larger sizes of the eukaryotic proteins are due to large hydrophilic N and/or C termini or, less frequently, to increased sizes of specific inter-TMS loops. The hydrophilic regions of the eukaryotic proteins may play roles in regulation or in cytoskeletal attachment, and they are frequently subject to phosphorylation by ATP-dependent protein kinases. A representative well-characterized example of the SP family is the arabinose:H^+ symport permease (AraE) of *Escherichia coli* (47).

Families 2 and 3 consist of drug efflux systems which possess 14 and 12 TMSs, respectively (74). Since these permeases uniformly catalyze drug:H^+ antiport, they are referred to as the DHA14 and DHA12 families, respectively. A total of 30 and 46 sequenced members are currently recognized in these two families. Because these permeases have recently been the subject of an extensive review which presented multiple alignments and phylogenetic trees (60), they will not be described or analyzed here. Members of both families are found in bacteria and eukaryotes, and DHA12 family members have also been identified in archaea.

Families 4, 5, and 6, the organophosphate:inorganic phosphate antiporters (OPA), the oligosaccharide:H^+ symporters (OHS), and the metabolite:H^+ symporters (MHS), respectively, were recognized to be families within the MFS in 1993 (30, 51). Since these permeases are restricted to bacteria, it is not surprising that they are all relatively small (400 to 500 residues). All three of these families have become substantially larger and more diverse in function since 1993, due to the sequencing and functional identification of new members.

All the remaining families listed in Table 1 (families 7 to 18) were not recognized in 1993 and are therefore new MFS families. Family 7 (the fucose-galactose-glucose:H^+ symporters [FGHS]) is a small family with four distantly related members. As with most members of the SP family, these proteins are specific for sugars. They all probably function by proton symport. They are relatively small (404 to 438 residues), as expected since they are derived exclusively from bacteria.

The nitrate-nitrite porter (NNP) family (family 8) has members in bacteria, yeasts, and plants. Not surprisingly, these proteins exhibit a larger size range (395 to 547 residues) than was observed for FGHS family members. These proteins catalyze either nitrate uptake or nitrite efflux. The energy-coupling mechanisms are not well defined.

Family 9, the phosphate:H^+ symporter (PHS) family, has sequenced representatives only in yeast and plants. The 11 proteins of the PHS family are fairly uniform in size, but they are substantially larger than most bacterial MFS proteins (518 to 587 residues). The characterized members are uniform in function.

Family 10, the nucleoside:H^+ symporter (NHS) family, has only two bacterial members, and they are of the same size (418 residues each). They are both from *E. coli* and differ in specificity.

Family 11, the oxalate/formate antiporter (OFA) family, is a small but diverse family. Only five members have been sequenced, but these proteins are found in the bacterial, archaeal, and eukaryotic kingdoms. Surprisingly, they are of fairly uniform size (373 to 470 residues). The very small size of one of these proteins (see below) raises the possibility that its sequence is incomplete.

Family 12, the sialate:H^+ symporter (SHS) family, like the NHS family, is very small (with only three members), and, again like the NHS family, the members are all derived from gram-negative bacteria. Their sizes are consistent with those generally observed for bacterial MFS proteins (407 to 496 residues). These proteins differ from most MFS proteins in possessing 14 putative TMSs.

Family 13, with 13 members derived exclusively from yeasts and animals, is the monocarboxylate porter (MCP) family. These permeases transport pyruvate, lactate, and/or mevalonate with inwardly-directed polarity. They all presumably function by proton symport. Their reported sizes range from 450 to 808 residues.

Family 14, the anion:cation symporter (ACS) family, is a relatively large family with 40 sequenced members. The proteins are derived from bacteria, yeasts, and animals, and they exhibit an intermediate range of sizes (411 to 596 residues). They accumulate their substrates in symport with either Na^+ or H^+, depending on the system. They may transport either inorganic anions (e.g., phosphate) or organic anions (e.g., glucarate, hexuronate, tartrate, allantoate, or 4-hydroxylphenyl acetate). Of the functionally characterized porters, the inorganic anion porters of the ACS family cotransport Na^+ while the organic anion porters cotransport H^+.

Family 15, the aromatic acid:H^+ symporter (AAHS) family, consists of seven sequenced proteins, all from bacteria. As expected, these porters show fairly uniform sizes (418 to 460 residues), all on the low end of the scale. They transport a variety of aromatic acids as well as *cis*,*cis*-muconate, as indicated in Table 1. Interestingly, one member of the family has been implicated in chemotaxis, allowing the bacteria to swim up concentration gradients of its substrates (34). This is the only documented case where an MFS protein apparently serves as a chemoreceptor. One of the AAHS proteins (BenK Aca) transports benzoate (11). Two additional (putative) benzoate:H^+ symporters (BenE) have been sequenced. They are both derived from gram-negative bacteria. One is the functionally characterized BenE protein of *Acinetobacter calcoaceticus*, and the other is a closely related protein from *E. coli* (55). These two proteins both contain a single region that exhibits limited sequence similarity to family 15 porters, as might be expected on the basis of the specificity of the *A. calcoaceticus* protein. However, they are very divergent in sequence from the latter proteins and cannot be shown to be homologous to any member of the MFS. They are therefore included in a separate family designated the benzoate:H^+ symporter (BenE; TC #2.46) family (72a).

Six members of a novel family, family 16, the unknown major facilitator (UMF) family, have recently been identified (26). Although it has been proposed that these carriers are drug efflux pumps, no member of this family has been functionally characterized, and consequently the designation UMF has tentatively been assigned to this family. All six currently recognized members of the family are from *Saccharomyces cerevisiae*, and no close homologs are found in other organisms. These proteins exhibit the less common putative 14-TMS topology observed for only two other MFS families. The proteins

TABLE 2. Interfamilial comparison scores for representative members of families 7 to 17[a]

Protein 1	Family	Protein 2	Family	Comparison scores (SD)	% Identity	No. of residues compared	No. of gaps
FucP	7, FGHS	Gtr5	1, SP	8	21	125	0
NasA	8, NNP	YidT	14, ACS	13	21	150	0
Ph84	9, PHS	Itr1	1, SP	10	24	586	19
NupG	10, NHS	CscB	5, OHS	12	19	415	8
YhjX	11, OFA	Ykw1	13, MCP	11	18	402	5
NanT	12, SHS	CitA	6, MHS	10	25	407	11
Mot1	13, MCP	NanT	12, SHS	10	17	494	5
GudT	14, ACS	Bmr1	3, DHA12	12	22	389	8
YheO	15, UMF	Sge1	2, DHA14	12	18	543	5
PcaK	16, AAHS	GudT	14, ACS	13	18	438	5
YycB	17, CP	NarK	8, NNP	10	21	408	6

[a] Comparison scores expressed in standard deviations (SD) were determined with the GAP program with 500 random shuffles. Family and protein abbreviations are as indicated in Table 1 and Tables 3 to 17, respectively, with the exception of proteins from families 2 (DHA14) and 3 (DHA12), which are tabulated in reference 60.

of the UMF family exhibit almost no size variation (range, 606 to 637 residues).

Family 17, the cyanate permease (CP) family, includes only three proteins, all from bacteria. They are small proteins (393 to 402 residues with 12 TMSs). The substrate of one of these proteins (CynX of *E. coli*) is believed to be cyanate (NCO^-). The other two members, from *E. coli* and *Bacillus subtilis*, are strikingly divergent in sequence but not in size, as noted above.

The proton-dependent oligopeptide transporter (POT) family has been described previously (62, 78). We have observed sequence similarities of these proteins to members of the SP and DHA14 families (see below). Although this similarity is insufficient to establish homology, the similarities in sequence, mechanism, and topology between proteins of the POT family and those of several MFS families strongly suggest that the POT family is a distant constituent of the MFS.

ESTABLISHMENT OF HOMOLOGY FOR MFS PROTEINS

Proteins within any one family of the MFS exhibit fairly extensive sequence similarities, as revealed by the portions of the multiple alignments shown in Fig. 3 to 17. Intrafamily comparison scores are always in excess of 15 SD, thus easily establishing that the members of any one family are homologous. However, sequence similarity for any two proteins derived from different MFS families is much less extensive. We therefore conducted interfamily binary comparisons to establish homology for all MFS families (18, 71). Homology for families 1 to 6 has been established previously (51). The results of the present comparisons are presented in Table 2 and Figure 1.

As noted above, families 1 to 6 and family 16 have already been shown to be constituent families of the MFS (26, 51, 60). The data presented in Table 2 establish that families 7 to 17 (described above) are all constituents of the MFS. In the case of FucP of family 7 (FGHS), 21% identity (8 SD) was observed with Gtr5 of family 1 in a region of 125 residues that exhibits no gaps in the binary alignment. Gtr5 is an established member of family 1 of the MFS (see Table 3). Of the comparison scores recorded in Table 2, this is the only score below 10 SD, and most of the other sequences compared include all or most of the two proteins compared. Family 7 is therefore the only family included in Table 2 that is not fully established as an MFS constituent. Other considerations provide additional support for the conclusion that the FGHS family is in fact a member of the MFS (see below).

NasA of family 8 exhibits a comparison score of 13 SD with 21% identity to YidT of family 14 for a 150-residue segment exhibiting no gaps, thus linking these two families, and GudT of family 14 exhibits 22% identity and 12 SD to Bmrl of family 3 for the full lengths of the two proteins (eight gaps in the complete binary alignment). Bmrl is an established member of the MFS (60). Thus, families 8 and 14 are members of the MFS as determined by these comparisons. Similarly, the OFA (family 11) member YhjX exhibits 18% identity and 11 SD with five gaps for the full binary alignment with respect to Ykwl of family 13. Family 13 member Motl exhibits 10 SD (17% identity; 5 gaps for the full-length binary alignment) with respect to NanT of family 12, and NanT exhibits 10 SD (25% identity with 11 gaps in the complete binary alignment) with respect to CitA of family 6, a protein shown previously to be a member of the MFS (51). Thus, on the basis of these comparisons and the superfamily principle (18, 71), families 11, 13, and 12 are within the MFS. Using similar logic, the results summarized in Table 2 establish that all 17 families under consideration (with the improbable exception of family 7) are homologous.

Short regions of the binary alignments upon which the comparison scores recorded in Table 2 were based are shown in Fig. 1. These alignments exhibit between 17 and 25% identity with greater than 50% similarity in each case. Most of the regions shown are derived from the N-terminal halves of these proteins. These regions are generally the best-conserved portions of the MFS proteins, as pointed out previously for families 1 to 7 (51, 71).

PHYLOGENETIC TREE FOR THE MFS

A phylogenetic tree for the MFS, which includes representative proteins from most of the families, is shown in Fig. 2. Several features are worthy of note. First, most of the families branch from points near the center of the tree. Second, the DHA14 and DHA12 families (families 2 and 3, respectively) branch off from each other after the initial divergence from the center of the tree, suggesting that they are more closely related to each other than to other MFS families. This is in agreement with their similar specificities. Third, the UMF family (family 16) does not branch from a point near the branch for the two DHA families (families 2 and 3), and thus there is no phylogenetic evidence for the suggestion that they transport drugs, even though the proteins of the UMF and DHA14 families both have 14 putative TMSs (26). Fourth, the MCP and OFA families (families 13 and 11, respectively) branch from each other at a point that is somewhat distant from the center of the tree. A late branching point, suggestive of late divergence, is consistent with the fact that both families transport carboxylates. Fifth, the MHS and SHS families (families 6 and 12,

FIG. 1. Interfamily binary alignments of representative regions within the larger alignments upon which the comparison scores recorded in Table 2 were based. The two proteins compared (families in parentheses) are presented above the alignment. Protein abbreviations are as indicated in Tables 3 to 17 for families 3 to 17, respectively. The number following the protein abbreviation and preceding its sequence is the residue number of the first residue shown. A vertical line specifies an identity, while double and single dots signify close and distant similarities, respectively.

respectively) branch from each other relatively far from the center of the tree, suggesting that they are close familial relatives, having diverged from each other late in the evolutionary process. Most, and perhaps all, of the members of these two families transport anionic compounds. The PHS and SP families (families 9 and 1, respectively) also stem from the primary branch from which the MHS and SHS families stem. However, these families branch off close to the center of the tree. Consequently, close phylogenetic relationships for these families are not suggested. Finally, the OPA, NHS, and OHS families (families 4, 10, and 5, respectively) are found branching from the same trunk. All of the proteins of the NHS and OHS families, and some of the members of the OPA family, transport glycosides.

FAMILY 1: SUGAR PORTER (SP) FAMILY

The SP family was described many years ago, and its description has been repeatedly updated (7, 7a, 28, 30, 35, 36, 46, 58). The present SP family consists of 133 sequenced members derived from bacteria, archaea, and eukarya. The family includes members that are very diverse in sequence and function. As revealed by the information in Table 3, these proteins function under normal physiological conditions either by uniport or by H^+ symport. However, many of these and other MFS permeases can catalyze solute:solute antiport when substrates are present on both sides of the membrane (42). The symporters all function in energized cells with inwardly directed polarity. Substrates of SP family members include ga-

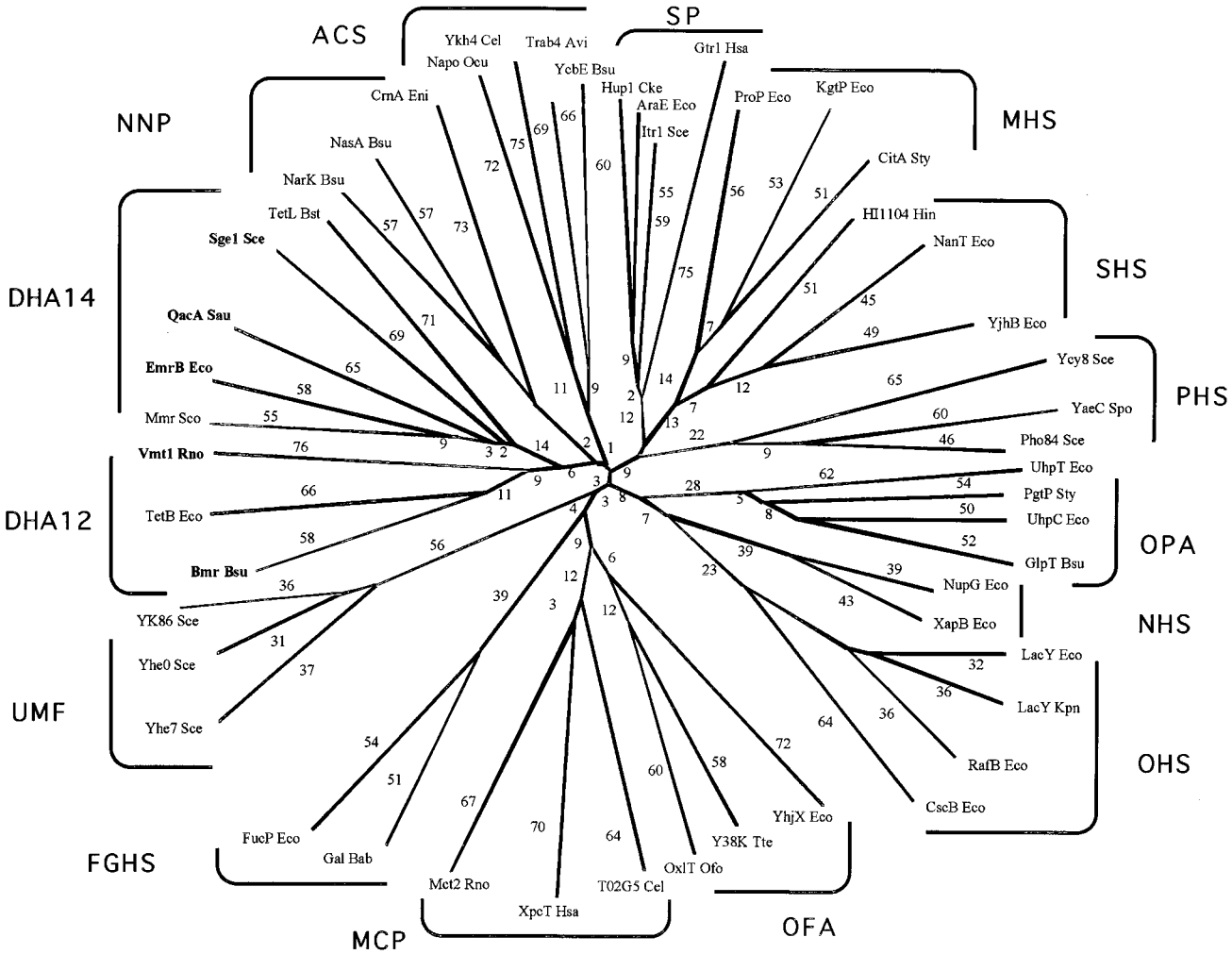

FIG. 2. Phylogenetic tree for the MFS including representative members of most of the currently recognized constituent families. All of the families listed in Table 1 are included except the AAHS family (family 15) and the putative POT family (putative family 18). The TREE program of Feng and Doolittle (22) was used to derive the tree.

TABLE 3. Members of the SP family (family 1)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[b]
AraE Eco	Arabinose-proton symporter	B	*Escherichia coli*	472	P09830	SP
AraE Kox	Arabinose-proton symporter	B	*Klebsiella oxytoca*	472	P45598	SP
GalP Eco	Galactose-proton symporter	B	*Escherichia coli*	464	P37021	SP
Glf Zmo	Glucose facilitated diffusion protein	B	*Zymomonas mobilis*	473	P21906	SP
Gtr Ssp	Glucose transport protein	B	*Synechocystis* sp.	468	P15729	SP
Orf_f469 Eco	Putative sugar transporter	B	*Escherichia coli*	469	U29579	GB
Orf_ZZ Shy	Putative monosaccharide transporter (fragment)	B	*Streptomyces hygroscopicus*	290	X86780	GB
XylE Eco	Xylose-proton symporter	B	*Escherichia coli*	491	P09098	SP
YaaU Eco	Putative sugar transporter (frameshift error?)	B	*Escherichia coli*	220	P31578	SP
YfiG Bsu	Putative sugar transporter	B	*Bacillus subtilis*	482	P54723	SP
YncC Bsu	Putative sugar transporter	B	*Bacillus subtilis*	419	U66480	GB
YxbC Bsu	Putative sugar transporter	B	*Bacillus subtilis*	388	P46333	SP
YxdF Bsu	Putative sugar transporter	B	*Bacillus subtilis*	439	P42417	SP
YyaJ Bsu	SV2 homolog	B	*Bacillus subtilis*	451	P37514	SP
Str Ssp	Sugar transporter	Ar	*Sulfolobus solfataricus*	423	Y08256	GB
Gtr1 Ldo	Putative glucose transporter	Pr	*Leishmania donovani*	547	Q01440	SP
Glut2 Ldo	Putative sugar transporter	Pr	*Leishmania donovani*	558	Q01441	SP
HT1 Tvi	Glucose transporter	Pr	*Trypanosoma vivax*	543	L47540	GB
Pro-1 Ldo	Putative sugar transporter	Pr	*Leishmania donovani*	567	P13865	SP
TcrHT1 Tcr	Hexose transporter	Pr	*Trypanosoma cruzi*	544	U05588	GB
Tht1E Tbr	Glucose transporter 1E	Pr	*Trypanosoma brucei brucei*	528	Q09037	SP
Tht2A Tbr	Glucose transporter 2A	Pr	*Trypanosoma brucei brucei*	529	Q06222	SP
Tht2B Tbr	Glucose transporter 1B/1C/1D/1F/2B	Pr	*Trypanosoma brucei brucei*	527	Q06221	SP
Tht2C Tbr	Glucose transporter 2C (fragment)	Pr	*Trypanosoma brucei brucei*	337	Q09039	SP
Ag1 Sce	α-Glucoside transporter	Y	*Saccharomyces cerevisiae*	617	S59368	GB
Gal2 Sce	Galactose transporter	Y	*Saccharomyces cerevisiae*	574	P13181	SP
Glu Ssp	Putative glucose transporter	Y	*Saccharomyces* sp.	518	L21753	GB
Hgt1 Kla	High-affinity glucose transporter	Y	*Kluyveromyces lactis*	551	P49374	SP
Hxt1 Sce	Low-affinity glucose transporter	Y	*Saccharomyces cerevisiae*	570	P32465	SP
Hxt2 Sce	High-affinity glucose transporter	Y	*Saccharomyces cerevisiae*	541	P23585	SP
Hxt3 Sce	Low-affinity glucose transporter	Y	*Saccharomyces cerevisiae*	567	P32466	SP
Hxt4 Sce	Glucose transporter	Y	*Saccharomyces cerevisiae*	576	P32467	SP
Hxt5 Sce	Glucose transporter	Y	*Saccharomyces cerevisiae*	592	P38695	SP
Hxt6 Sce	Hexose transporter	Y	*Saccharomyces cerevisiae*	570	P39003	SP
Hxt7 Sce	Hexose transporter	Y	*Saccharomyces cerevisiae*	570	P39004	SP
Hxt8 Sce	Putative hexose transporter	Y	*Saccharomyces cerevisiae*	569	P40886	SP
Hxt10 Sce	Putative hexose transporter	Y	*Saccharomyces cerevisiae*	546	P43581	SP
Hxt11 Sce	Putative hexose transporter; drug resistance?	Y	*Saccharomyces cerevisiae*	567	P40885	SP
Hxt13 Sce	Putative hexose transporter	Y	*Saccharomyces cerevisiae*	564	P39924	SP
Hxt14 Sce	Putative hexose transporter	Y	*Saccharomyces cerevisiae*	540	P42833	SP
Hxt16 Sce	Putative sugar transport protein	Y	*Saccharomyces cerevisiae*	567	P47185	SP
Hxt17 Sce	Putative sugar transporter	Y	*Saccharomyces cerevisiae*	343	Z71687	GB
Itr1 Sce	*myo*-Inositol transporter 1	Y	*Saccharomyces cerevisiae*	584	P30605	SP
Itr1 Spo	*myo*-Inositol transporter	Y	*Saccharomyces pombe*	575	X98622	GB
Itr2 Sce	*myo*-Inositol transporter 2	Y	*Saccharomyces cerevisiae*	612	P30606	SP
Kht2 Kla	Putative hexose transporter	Y	*Kluyveromyces marxianus*	566	S51081	PIR
Lac12 Kla	Lactose/galactose transporter	Y	*Kluyveromyces lactis*	587	P07921	SP
Mal3T Sce	Maltose transporter	Y	*Saccharomyces cerevisiae*	614	P38156	SP
Mal6T Sce	Maltose transporter	Y	*Saccharomyces cerevisiae*	614	P15685	SP
Rag1 Kla	Low-affinity glucose transporter	Y	*Kluyveromyces lactis*	567	P18631	SP
Rgt2 Sce	High glucose sensor	Y	*Saccharomyces cerevisiae*	763	X96876	GB
YacI Spo	Putative sugar transporter	Y	*Schizo saccharomyces pombe*	522	Q10710	SP
Ybr241c Sce	Putative sugar transporter	Y	*Saccharomyces cerevisiae*	488	P38142	SP
Yd1199c Sce	Putative sugar transporter	Y	*Saccharomyces cerevisiae*	687	S58778	PIR
Ydr387c Sce	Putative sugar transporter	Y	*Saccharomyces cerevisiae*	555	U32274	GB
Yfl040w Sce	Putative sugar transporter	Y	*Saccharomyces cerevisiae*	540	P43562	SP
Ygk4 Sce	Putative sugar transporter	Y	*Saccharomyces cerevisiae*	486	Z72626	GB
Yil170w Sce	Putative sugar transporter	Y	*Saccharomyces cerevisiae*	457	P40441	SP
Yjr160c Sce	Putative maltose transporter	Y	*Saccharomyces cerevisiae*	602	P47186	SP
Snf3 Sce	Low glucose sensor	Y	*Saccharomyces cerevisiae*	818	P10870	SP
Stl1 Sce	Putative sugar transporter Stl1	Y	*Saccharomyces cerevisiae*	569	P39932	SP
Rco3 Ncr	Sugar transporter	F	*Neurospora crassa*	594	U54768	GB
Qa-Y Ncr	Quinate transporter	F	*Neurospora crassa*	537	P11636	SP
QutD Eni	Quinate transporter	F	*Emericella nidulans*	533	P15325	SP
Glu2 Ssp	Putative glucose transporter (fragment)	Pl	*Saccharum* sp.	287	L21752	GB
Hex6 Rco	Hexose transporter	Pl	*Ricinus communis*	510	L08188	GB
Hup1 Cke	Hexose-proton symporter	Pl	*Chorella kessleri*	533	P15686	SP
Hup2 Cke	Galactose-proton symporter	Pl	*Chorella kessleri*	540	X66855	GB

Continued on following page

TABLE 3—*Continued*

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[b]
Hup3 Cke	Glucose transporter	Pl	*Chorella kessleri*	534	S38435	PIR
Int Hvu	Putative sugar transporter	Pl	*Beta vulgaris*	490	U43629	GB
Mst1 Nta	Monosaccharide transport protein	Pl	*Nicotiana tabacum*	523	S25015	PIR
MtSt1 Mtr	Putative sugar transporter	Pl	*Medicago truncatula*	518	U38651	GB
Stc Rco	Sugar transporter	Pl	*Ricinus communis*	523	L08196	GB
Stp1 Ath	Glucose transporter	Pl	*Arabidopsis thaliana*	522	P23586	SP
Stp4 Ath	Monosaccharide transport protein	Pl	*Arabidopsis thaliana*	514	S25009	PIR
B0252 Cel	B0252.3 (putative organic cation transporter)	An	*Caenorhabditis elegans*	435	U23453	GB
C44C10.3 Cel	C44C10.3 (putative organic cation transporter)	An	*Caenorhabditis elegans*	404	Z69787	GB
C46C2.2 Cel	C46C2.2 (putative organic cation transporter)	An	*Caenorhabditis elegans*	517	Z68296	GB
C53B4.1 Cel	C53B4.1 (putative organic cation transporter)	An	*Caenorhabditis elegans*	517	Z68215	GB
F11D5 Cel	F11D5.4/F11D5.5 (putative sugar transporter-frameshift)	An	*Caenorhabditis elegans*	361/192	U41532	GB
F13B12.2 Cel	F13B12.2 (putative glucose transporter)	An	*Caenorhabditis elegans*	431	Z70683	GB
F14B8.3 Cel	Putative organic cation transporter	An	*Caenorhabditis elegans*	524	U28737	GB
F14E5.1 Cel	F14E5.1 (putative sugar transporter)	An	*Caenorhabditis elegans*	466	Z66522	GB
F14E5.1 Cel	F14E5.1 (putative glucose transporter)	An	*Caenorhabditis elegans*	493	Z66522	GB
F23F12.5 Cel	Putative organic cation transporter	An	*Caenorhabditis elegans*	751	P46501	SP
F48E3 Cel	F48E3.2 (putative glucose transporter)	An	*Caenorhabditis elegans*	488	U28735	GB
F53H8 Cel	Similar to glucose transporter	An	*Caenorhabditis elegans*	569	U41023	GB
Glu Dme	Putative glucose transporter	An	*Drosophila melanogaster*	507	U31961	GB
Glut1 Bta	Glucose transporter type 1, erythrocyte	An	*Bos taurus*	492	P27674	SP
Glut1 Gga	Glucose transporter type 1, erythrocyte	An	*Gallus gallus*	490	P46896	SP
Glut1 Hsa	Glucose transporter type 1, erythrocyte	An	*Homo sapiens*	492	P11166	SP
Glut1 Mmu	Glucose transporter type 1, erythrocyte	An	*Mus musculus*	492	P17809	SP
Glut1 Ocu	Glucose transporter type 1, erythrocyte	An	*Oryctolagus cuniculus*	492	P13355	SP
Glut1 Rno	Glucose transporter type 1, erythrocyte	An	*Rattus norvegicus*	492	P11167	SP
Glut1 Ssc	Glucose transporter type 1, erythrocyte	An	*Sus scrofa*	451	P20303	SP
Glut2 Gga	Glucose transporter	An	*Gallus gallus*	533	S37476	PIR
Glut2 Hsa	Glucose transporter type 2, liver	An	*Homo sapiens*	524	P11168	SP
Glut2 Mmu	Glucose transporter type 2, liver	An	*Mus musculus*	523	P14246	SP
Glut2 Rno	Glucose transporter type 2, liver	An	*Rattus norvegicus*	522	P12336	SP
Glut3 Cfa	Glucose transporter type 3, brain	An	*Canis familiaris*	495	P47842	SP
Glut3 Gga	Glucose transporter type 3, brain	An	*Gallus gallus*	496	P28568	SP
Glut3 Hsa	Glucose transporter type 3, brain	An	*Homo sapiens*	496	P11169	SP
Glut3 Mmu	Glucose transporter type 3, brain	An	*Mus musculus*	493	P32037	SP
Glut3 Oar	Glucose transporter type 3, brain	An	*Ovis aries*	494	P47843	SP
Glut3 Rno	Glucose transporter type 3, brain	An	*Rattus norvegicus*	493	Q07647	SP
Glut4 Hsa	Glucose transporter type 4, insulin	An	*Homo sapiens*	509	P14672	SP
Glut4 Mmu	Glucose transporter type 4, insulin	An	*Mus musculus*	510	P14142	SP
Glut4 Rno	Glucose transporter type 4, insulin	An	*Rattus norvegicus*	509	P19357	SP
Glut5 Hsa	Fructose transporter	An	*Homo sapiens*	501	P22732	SP
Glut5 Ocu	Glucose transporter type 5, small intestine	An	*Oryctolagus cuniculus*	486	P46408	SP
Glut5 Rno	Glucose transporter type 5, small intestine	An	*Rattus norvegicus*	502	P43427	SP
Glut7 Rno	Glucose transporter type 7, hepatic	An	*Rattus norvegicus*	528	Q00712	SP
GT2 Mmu	Glucose transporter	An	*Mus musculus*	508	B30310	PIR
GTP1 Tso	Glucose transporter	An	*Taenia solium*	510	U39197	GB
K05F1 Cel	K05F1.6 (putative organic cation transporter)	An	*Caenorhabditis elegans*	745	U29377	GB
K09C4 Cel	Putative glucose transporter	An	*Caenorhabditis elegans*	520	U43375	GB
Ksp Mmu	Putative organic cation transporter	An	*Mus musculus*	545	U52842	GB
M01F1 Cel	M01F1.5 (putative hexose transporter)	An	*Caenorhabditis elegans*	628	Z46381	GB
Nit Rno	Putative organic cation transporter	An	*Rattus norvegicus*	514	L27651	GB
Oct-1 Rno	Organic cation transporter	An	*Rattus norvegicus*	556	X78855	GB
Oct-2 Rno	Organic cation transporter	An	*Rattus norvegicus*	593	D83045	GB
SGTP1 Sma	Glucose transporter	An	*Schistosoma mansoni*	521	A53153	GB
SGTP2 Sma	Glucose transporter	An	*Schistosoma mansoni*	489	B53153	PIR
SGTP4 Sma	Glucose transporter	An	*Schistosoma mansoni*	505	C53153	PIR
SV2 Rno	SV2 synaptic vesicle protein	An	*Rattus norvegicus*	742	S27263	PIR
SV2 Rno	Synaptic vesicle protein 2	An	*Rattus norvegicus*	742	Q02563	SP
T05A1.5 Cel	T05A1.5 (putative organic cation transporter)	An	*Caenorhabditis elegans*	498	Z68219	GB
T05A1.7 Cel	T05A1.7 (putative organic cation transporter)	An	*Caenorhabditis elegans*	251	Z68219	GB
ZK455.8 Cel	ZK455.8 (putative organic cation transporter)	An	*Caenorhabditis elegans*	447	Z66567	GB
ZK563 Cel	ZK563.1 (fragment)	An	*Caenorhabditis elegans*	>357	U40061	GB
ZK637.1 Cel	Putative organic cation transporter	An	*Caenorhabditis elegans*	536	P30638	SP
ZK829.9 Cel	ZK829.9 (putative glucose transporter)	An	*Caenorhabditis elegans*	351	Z73899	GB
ZK892.3 Cel	ZK892.3 (putative organic cation transporter)	An	*Caenorhabditis elegans*	541	Z48638	GB

[a] B, bacterium; Ar, archaea; Pr, protozoan; Y, yeast; F, fungus; Pl, plant; An, animal.
[b] SP, SwissProt; GB, GenBank; PIR, Protein Information Resource.

lactose, arabinose, xylose, and glucose in bacteria; galactose, quinate, myoinositol, lactose, maltose, and α-glucosides in yeasts and fungi; hexoses in trypanosomes and plants; and sugars as well as organic cations and neurotransmitters in animals. Most but by no means all members of the SP family are therefore specific for sugars.

Figure 3A presents a portion of the multiple-sequence alignment of 20 representative members of the SP family. The

A SP

```
Rag1 Kla      (171) G R I I S G L G V G G I T V L S P M L I S E T A P K H L R G T L V S C Y Q L M I T F G I F L G
Gal2 Sce      (179) G R I I S G L G V G G I A V L C P M L I S E I A P K H L R G T L V S C Y Q L M I T A G I F L G
Hxt10 Sce     (157) G R I V S G M G V G G V A V L S P T L I S E I S P K H L R G T C V S F Y Q L M I T L G I F L G
Qa-Y Ncr      (130) G R V L A G I G V G G A S N M V P I Y I S E L A P P A V R G R L V G I Y E L G W Q I G G L V G
Stl1 Sce      (140) G R V V T G V G T G L N T S T I P V W Q S E M S K A E N R G L L V N L E G S T I A F G T M I A
Itr1 Sce      (185) G R L I M G F G V G I G S L I A P L F I S E I A P K M I R G R L T V I N S L W L T G G Q L V A
GalP Eco      (111) S R V L L G L A V G V A S Y T A P L Y L S E I A P E K I R G S M I S M Y Q L M I T I G I L G A
AraE Eco      (118) A R V V L G I A V G I A S Y T A P L Y L S E M A S E N V R G K M I S M Y Q L M V T L G I V L A
Glf Zmo       (122) F R F L A G L G I G V V S T L T P T Y I A E I R P P D K R G Q M V S G Q Q M A I V T G A L T G
XylE Eco      (132) Y R I I G G I G V G L A S M L S P M Y I A E L A P A H I R G K L V S F N Q F A I I F G Q L L V
Stp1 Ath      (139) G R I L L G F G I G F A N Q A V P L Y L S E M A P Y K Y R G A L N I G F Q L S I T I G I L V A
Hup1 Cke      (142) G R V L L G F G V G L G S Q V V P Q Y L S E V A P F S H R G M L N I G Y Q L F V T I G I L I A
Glut3 Rno     (123) G R L I I G I F C G L C T G F V P M Y I G E V S P T A L R G A F G T L N Q L G I V V G I L V A
Glut5 Rno     (130) S R L L V G I C A G I S S N V V P M Y L G E L A P K N L R G A L G V A P Q L F I T V G I L V A
Lac12 Kla     (172) G R W F V A F F A T I A N A A A P T Y C A E V A P A H L R G K V A G L Y N T L W S V G S I V A
Tht2A Tbr     (182) G R V L M G I G L G V V C V I C P M Y V N E N A H P K L S K V D G V L F Q V F I T F G I M L A
Mal6T Sce     (207) G Q A L C G M P W G C F Q C L T V S Y A S E I C P L A L R Y Y L T T Y S N L C W T F G Q L F A
ZK637.1 Cel   (178) F R G L T G F G I G G V   P Q S V T L Y A E F L P T A Q R A K C V V L I E S F W A I G A V F E
Oct-1 Rno     (206) F R L L Q G M V S K G S W V S G Y T L I T E F V G S G Y R R T T A I L Y Q M A F T V G L V G L
Sv2 Rno       (261) C R L L S G V G I G G S I P I V F S Y F S E F L A Q E K R G E H L S W L C M F W M I G G V Y A

Consensus           G R - L - G - G - G - - - - - - P - Y - S E - A P - - - R G - L - - - - Q L - I T - G - - - A
```

B

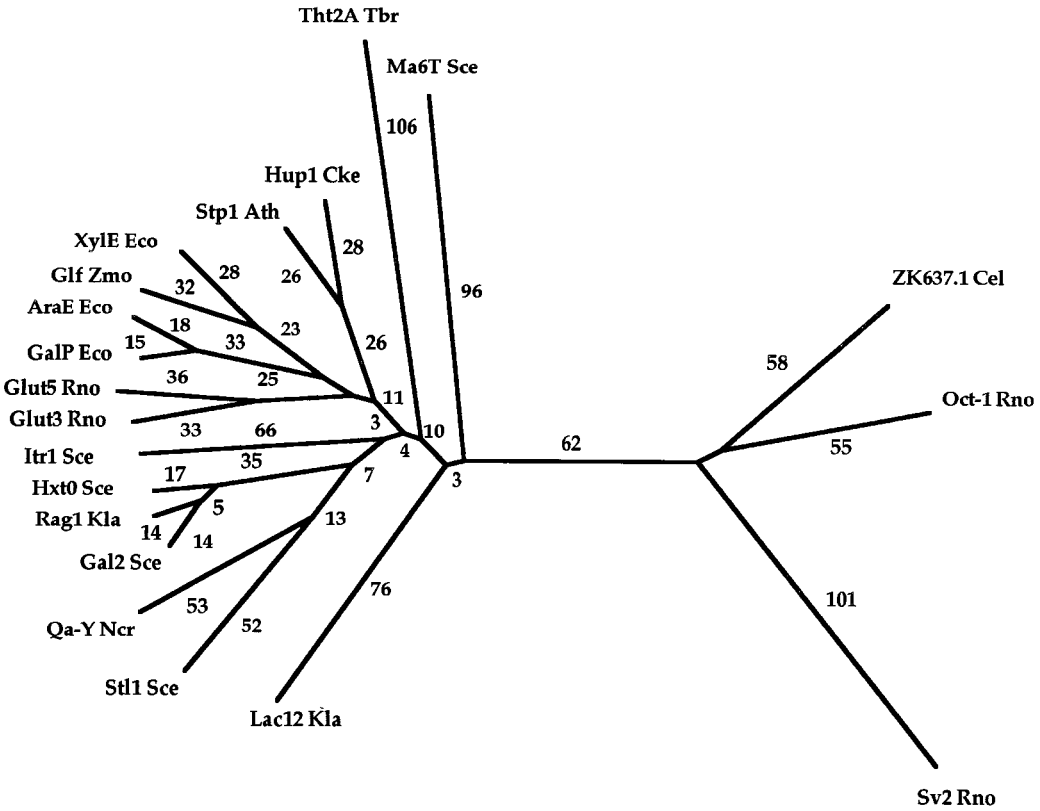

FIG. 3. Partial multiple alignment (A) and phylogenetic tree (B) for representative members of the SP family (family 1) of the MFS. The format of presentation for this figure is essentially the same as for subsequent family-specific figures (Fig. 4 to 17), as follows. (A) The abbreviation of each protein, provided in Table 3, is followed by the residue number in the protein, presented in parentheses, corresponding to the first residue in the alignment shown. Fully conserved residues are highlighted with a line to the right of that residue. A residue appears in the consensus sequence (bottom) if that residue occurs at the specified position in a majority of the aligned proteins. In the phylogenetic tree (B), the branch length (numerical values in arbitrary units) is a measure of sequence divergence and is assumed to be approximately proportional to the phylogenetic distance. The tree was based on the complete multiple-sequence alignment for the proteins included in the study. Both the multiple-sequence alignment and the phylogenetic tree were derived with the TREE program of Feng and Doolittle (22).

proteins included are from bacteria, yeasts, fungi, trypanosomes, plants, and animals. The 47-residue segment shown exhibits no gaps in the multiple alignment, and there are two fully conserved residues. Almost half of the residue positions within this segment exhibit a predominant residue that appears in the consensus sequence. This fact reflects a high degree of conservation. At least 50% of the proteins included in the alignment exhibit the same residue at each of these positions, by definition. While the fully conserved glycine (G) is likely to be of structural importance, the fully conserved glutamate (E) may play a catalytic role. Several of the well-conserved residues (e.g., R's at alignment positions 2 and 29 and G's at alignment positions 6 and 10) are conserved in all but one or two of the proteins depicted. The high degree of conservation of these residues clearly suggests that they play important structural or functional roles.

The phylogenetic tree for the SP family members whose sequences are represented in Fig. 3A is shown in Fig. 3B. On

TABLE 4. Members of the OPA family (family 4)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
GlpT Bsu	Glycerol-3-phosphate transporter	B	*Bacillus subtilis*	444	P37948	SP
GlpT Eco	Glycerol-3-phosphate transporter	B	*Escherichia coli*	452	P08194	SP
IpgH Sfl	IpgH ORFB gene product (truncated by IS insertion)	B	*Shigella flexneri*	333	U28354	GB
PgtP Sty	Phosphoglycerate transporter	B	*Salmonella typhimurium*	406	P12681	SP
UhpC Eco	Hexose phosphate receptor	B	*Escherichia coli*	439	P09836	SP
UhpC Sty	Hexose phosphate receptor	B	*Salmonella typhimurium*	442	P27669	SP
UhpT Eco	Hexose phosphate transporter	B	*Escherichia coli*	463	P13408	SP
UhpT Sty	Hexose phosphate transporter	B	*Salmonella typhimurium*	463	P27670	SP
F47B8.10 Cel	F47B8.10 (unknown function)	An	*Caenorhabditis elegans*	456	Z77662	GB
T11G6.2 Cel	T11G6.2 (unknown function)	An	*Caenorhabditis elegans*	495	Z69384	GB
T11G6.3 Cel	T11G6.3 (unknown function)	An	*Caenorhabditis elegans*	475	Z69384	GB
T11G6.4 Cel	T11G6.4 (unknown function)	An	*Caenorhabditis elegans*	304	Z69384	GB

[a] For abbreviations, see Table 3 footnotes.

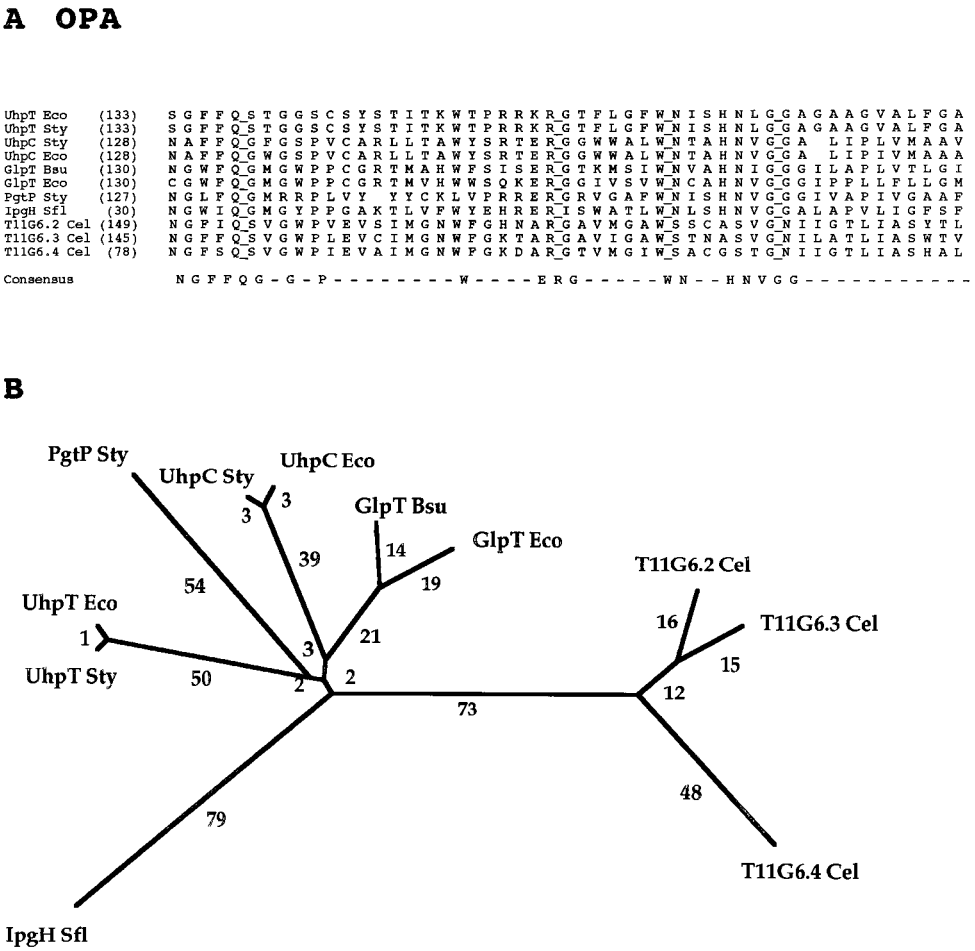

FIG. 4. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the OPA family (family 4). The format of presentation for this figure and Fig. 5 to 17 is essentially as described in the legend to Fig. 3.

the left-hand side of the tree are all of the sugar porters as well as the quinate:H^+ symporter (Qa-Y Ncr). Uniporters and proton symporters are often closely related (e.g., XylE Eco and Glf Zmo). More divergent members of the sugar porter cluster are Tht2A Tbr, Ma6T Sce, and Lac12 Kla of protists and yeasts. It is noteworthy that the bacterial, plant, and animal proteins cluster loosely together but most of the yeast and fungal proteins cluster separately.

Two yeast proteins and one protozoan protein branch from points near the base of the tree. However, the most distant members (right-hand side of the tree) include the synaptic vesicle transporter Sv2 Rno and the organic cation transporter

TABLE 5. Members of the OHS family (family 5)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
CscB Eco	Sucrose transporter	B	*Escherichia coli*	415	P30000	SP
LacY Eco	Lactose-proton symporter	B	*Escherichia coli*	417	P02920	SP
LacY Cfr	Lactose-proton symporter	B	*Citrobacter freundii*	416	P47234	SP
LacY Kpn	Lactose-proton symporter	B	*Klebsiella pneumoniae*	416	P18817	SP
MelY Ecl	Melibiose transporter	B	*Enterobacter cloacae*	425	AB000622	GB
RafB Eco	Raffinose transporter	B	*Escherichia coli*	425	P16552	SP

[a] For abbreviations, see Table 3 footnotes.

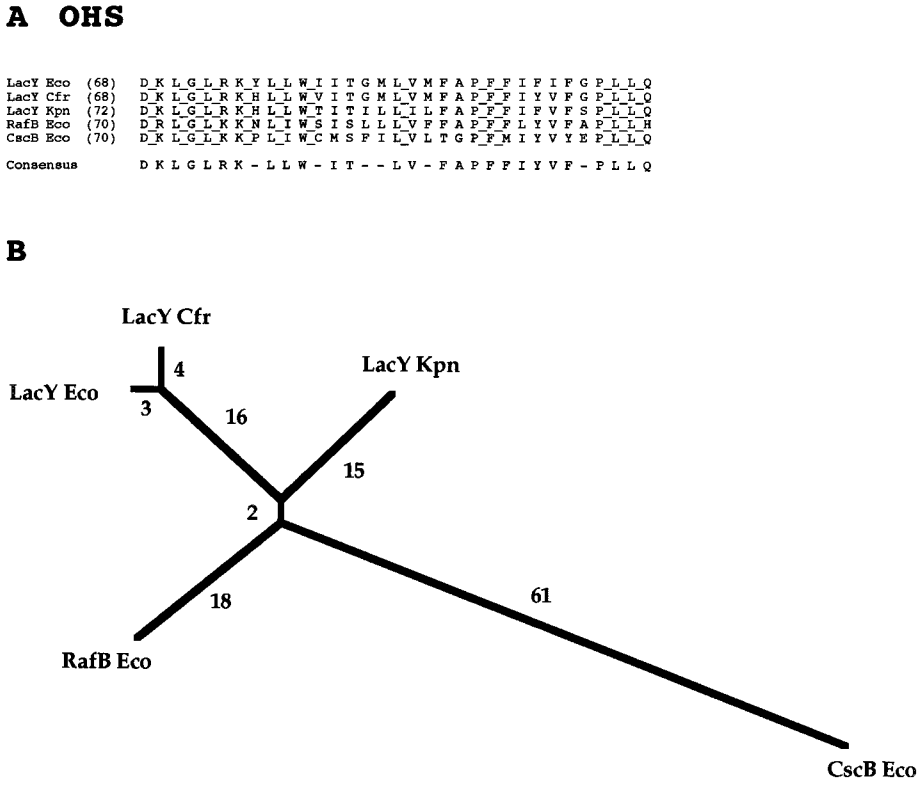

FIG. 5. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the OHS family (family 5).

Oct-1 Rno, both from the rat. The great phylogenetic distances observed between these proteins and the sugar porters correlate roughly with the divergent substrate specificity of these transporters relative to other members of the SP family.

FAMILY 2: DRUG:H$^+$ ANTIPORTER (14-SPANNER) (DHA14) DRUG EFFLUX FAMILY

The DHA14 drug efflux family has been described recently (60). Thirty members of the family were identified in that study. All functionally characterized members of the DHA14 family have been found to catalyze drug efflux. Of these functionally characterized permeases, 7 are multidrug resistance pumps from gram-negative and gram-positive bacteria as well as yeasts, 12 are putative drug-specific pumps from gram-positive bacteria, and 11 are hypothetical or uncharacterized proteins from gram-negative bacteria, yeasts, and fungi. A multiple alignment and a dendogram for the family were presented in that study (60). The multidrug resistance pumps, drug-specific permeases, and uncharacterized proteins did not group together on the DHA14 family dendogram but instead proved to be scattered in an apparently random fashion relative to each other. This fact suggests that drug-specific and multidrug efflux pumps arose repeatedly by narrowing and broadening of their specificities and that the functionally uncharacterized members of the family are also probably involved in drug efflux (74). Because of the extensive treatment of this family by Paulsen et al. (60), these proteins will not be considered further here.

FAMILY 3: DRUG:H$^+$ ANTIPORTER (12-SPANNER) (DHA12) DRUG EFFLUX FAMILY

The DHA12 drug efflux family, also described by Paulsen et al. (60), consists of 46 proteins. Of these, 9 have been shown to be multidrug resistance pumps, 15 are probably drug-specific efflux pumps, and 22 are hypothetical or uncharacterized proteins. Like the DHA14 family, functionally characterized members of the DHA12 family exhibit specificities only for drugs, although the range of drugs transported is remarkable (60). Interestingly, the range of organisms in which DHA12 family members are found is wider than that for the DHA14 family. Thus, the DHA12 MDR pumps are found in animals as well as in yeasts and a variety of gram-negative and gram-positive

TABLE 6. Members of the MHS family (family 6)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
CitA Eco	Citrate transporter (plasmid)	B	*Escherichia coli*	431	P07860	SP
CitA Sty	Citrate transporter	B	*Salmonella typhimurium*	434	P24115	SP
CitH Kpn	Citrate transporter	B	*Klebsiella pneumoniae*	444	P16482	SP
f427 Eco	Putative metabolite transporter	B	*Escherichia coli*	427	D90798	GB
HI0281 Hin	Putative metabolite transporter	B	*Haemophilus influenzae*	438	P44610	SP
HI0418 Hin	Putative metabolite transporter	B	*Haemophilus influenzae*	447	P44699	SP
I364 Mtu	MTCI364.12 (putative metabolite transporter)	B	*Mycobacterium tuberculosis*	425	Z93777	GB
KgtP Eco	α-Ketoglutarate transporter	B	*Escherichia coli*	432	P17448	SP
MopB Bce	4-Methyl-*O*-phthalate transporter	B	*Burkholderia cepacia*	449	U29532	GB
o438 Eco	Putative metabolite transporter	B	*Escherichia coli*	438	D90837	GB
Orf3 Shy	Involved in bialophos biosynthesis	B	*Streptomyces hygroscopicus*	447	E47031	PIR
OusA Ech	Osmoprotectant uptake system	B	*Erwinia chrysanthemi*	498	X82267	GB
PcaT Ppu	Dicarboxylic acid transport	B	*Pseudomonas putida*	429	U48776	GB
ProP Eco	Proline/betaine transporter	B	*Escherichia coli*	500	P30848	SP
YhjE Eco	Putative metabolite transporter	B	*Escherichia coli*	440	P37643	SP

[a] For abbreviations, see Table 3 footnotes.

A MHS

```
CitA Eco   (67)  P_I G_A V V L G_A Y I D_R I G_R R K G L M I T L A I M_G C G T L L I_A L V P_G Y Q T I G_L L A_P V L V
CitA Sty   (70)  P_V G_A I V L G_A Y I D_R I G_R R K G L M V T L A I M_G C G T L L I_A L V P_G Y Q T I G_L A A_P A L V
CitH Kpn   (83)  P_I G_A I V L G_A Y I D_K V G_R R K G L I V T L S I M_A T G T F L I_V L I P_S Y Q T I G_L W A_P L L V
KgtP Eco   (77)  P_I G_G W L F G_R I A D_K H G_R K K S M L L S V C M M_C F G S L V I_A C L P_G Y E T I G_T W A_P A L L
PcaT Ppu   (71)  P_I G_G W I F G_R L A D_R H G_R K N S L M I S V L M M_C F G S L M I_A C L P_T Y G S I G_T W A_P A L L
Orf3 Shy   (75)  P_V G_A T V M G_W Y A D_R Y G_R R S A L I V T I L L M_G L G S L M I_G L T P_S Y A T A G_P V A_P V V L
ProP Eco   (80)  P_L G_G L F F G_M L G D_K Y G_R Q K I L A I T I V I M_S I S T F C I_G L I P_S Y D T I G_I W A_P I L L
OusA Ech   (79)  P_L G_G V F F G_A L G D_K Y G_R Q K I L A I T I I I M_S I S T F C I_G L I P_S Y E R I G_I W A_P I L L
HI0418 Hin (78)  P_L G_A I L F G_H F G D_R F G_R K N T F V M S L L L M_G I S T V V I_G L L P_T Y D S I G_I W A_T I L L
YhjE Eco   (78)  P_I G_S A V F G_H F G D_R V G_R K A T L V A S L L T M_G I S T V V I_G L L P_G Y A T I G_I F A_P L L L
MopB Bce   (88)  P_L G_G L V F G_H F G D_K I G_R K S M M L V T L L M M_G I G T A A I_G L L P_S V A Q I G_M S A_T C L L

Consensus        P - G - - V F G - - - D R - G R - K - L - - T L - - M - - G T - - I G L - P - Y - T I G - W A P - L L
```

B

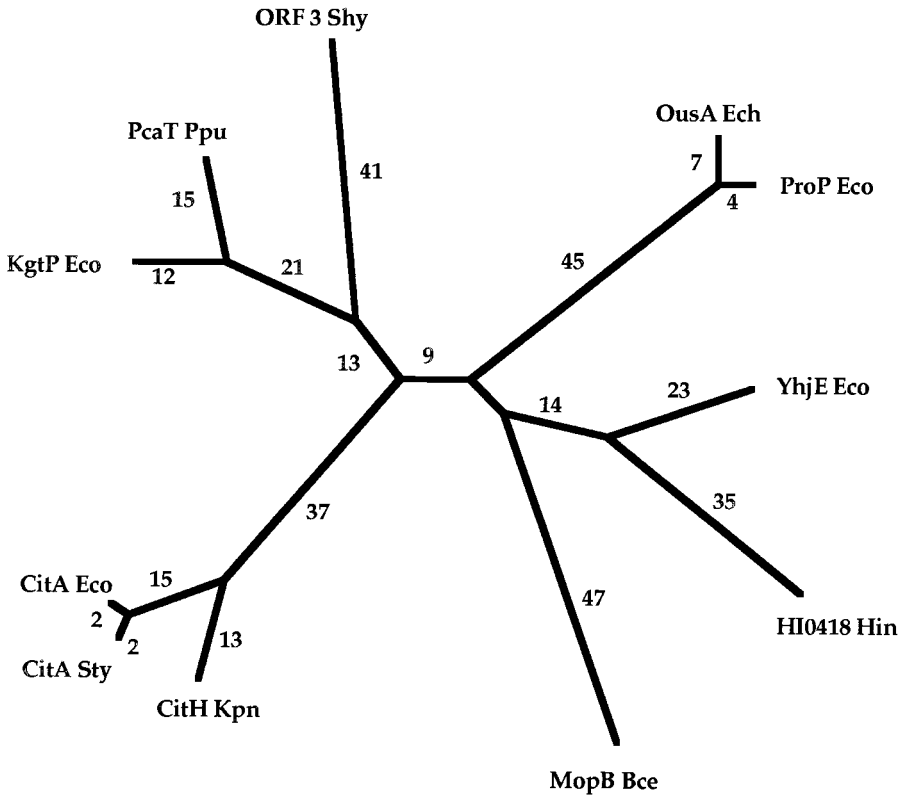

FIG. 6. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the MHS family (family 6).

bacteria. The proven and putative drug-specific efflux pumps are also found in a wide range of gram-negative and gram-positive bacteria, yeasts, and animals. Uncharacterized members of this family include an even wider range of organisms, including humans and archaea (reference 60 and unpublished results).

The dendogram for the DHA12 family (60) resembles that for the DHA14 family in that multidrug resistance and drug-specific pumps are interspersed. Thus, in both families, phylogeny does not appear to provide an indication of drug specificity.

FAMILY 4: ORGANOPHOSPHATE:INORGANIC PHOSPHATE ANTIPORTER (OPA) FAMILY

Table 4 lists the members of the OPA family of the MFS. The seven functionally characterized members are derived from gram-negative and gram-positive bacteria and function in the transport of either sugar phosphates, glycerol phosphate, or phosphoglycerates and phosphoenolpyruvate. The UhpC proteins are believed to function in the regulation of hexose phosphate transporter synthesis. UhpC presumably serves as a receptor for glucose-6-phosphate the inducer, in controlling transcription of the *uhpT* operon (38). It is a rare example of an MFS member which does not serve a primary transport function. However, it is not known whether or not it has the capacity to transport its ligand, glucose-6-phosphate.

The predominant mechanism of transport catalyzed by permeases of the OPA family under normal physiological conditions appears to be antiport of an organophosphate ester for inorganic phosphate (49, 50). These permeases may also be capable of catalyzing substrate:H^+ symport (19). The best-characterized members of the family are UhpT and GlpT, both of *E. coli*, for which detailed topological models have been presented (29, 90, 91). The OPA family includes several proteins from the worm *Caenorhabditis elegans*. Thus, this family includes members derived from eukaryotes as well as prokaryotes.

A well-conserved segment of the complete multiple alignment for the OPA family is presented in Fig. 4A. The 52-residue multiple-sequence alignment reveals only three single-residue gaps and shows four fully conserved residues (Q, R, W, and G). Of the 52 alignment positions, 19 are conserved in a majority of the proteins and hence appear in the consensus sequence. Structural (G, P), hydrophobic (F, V), semipolar (W), and strongly polar (N, Q, E, R and H) residues occur in the consensus sequence, with the last group being overrepresented.

The phylogenetic tree for the OPA family shows all bacterial proteins clustering together, as do the uncharacterized *C. elegans* proteins. Surprisingly, the UhpT transport proteins are as distant from the UhpC receptor proteins as these proteins are from the phosphoglycerate transporter (PgtP) or the glycerol phosphate transporters (GlpT). It is therefore of interest that UhpC is apparently specific for glucose-6-phosphate and 2-deoxyglucose-6-phosphate whereas UhpT recognizes a wide spectrum of sugar phosphates (3, 38). The large separation observed for the bacterial and animal proteins suggests that a primordial gene encoding one of the latter proteins was transferred to eukaryotes by vertical transmission from their prokaryotic progenitors and that gene duplication events in the developing eukaryote gave rise to the three paralogs found in *C. elegans*. Similarly, the configuration of the tree suggests that the gene duplication and divergence events that gave rise to the functionally dissimilar members of the bacterium-specific subfamilies occurred after the divergence of eukaryotes from prokaryotes.

FAMILY 5: OLIGOSACCHARIDE:H^+ SYMPORTER (OHS) FAMILY

The current OHS family consists of six proteins, three of which are β-galactoside permeases from closely related bacteria (Table 5). The lactose permease of *E. coli* is not only the best-characterized member of this family but also probably the most extensively studied permease in the MFS (41, 82). An experimentally verified 12-TMS topological model for LacY has been published (10), and extensive data provide evidence for the nature of the substrate binding sites within the transmembrane region of the permease (8, 12, 23, 27, 43, 57). A detailed mechanistic model that incorporates information obtained using many different experimental approaches has recently been proposed (40).

The other members of the OHS are specific for (i) the trisaccharide, raffinose; (ii) the α,β-nonreducing glucoside-fructoside, sucrose; and (iii) the α-galactoside, melibiose. All of these putative proton symporters are from gram-negative bacteria. The sequence of the melibiose permease of *Enterobacter cloacae* was deposited in the database after the completion of our phylogenetic analysis of the OHS family and is therefore not represented in Fig. 5. However, this protein proved to resemble RafB Eco (75% identity) more closely than it resembles one of the lactose permeases (50% identity) or the sucrose permease (<40% identity).

Figure 5A presents an alignment of a well-conserved 33-residue portion of the OHS family proteins. Over one-third of the residues shown in this gap-free alignment are fully conserved, and a large majority of the residues appear in the consensus sequence.

The phylogenetic tree (Fig. 5B) reveals that the raffinose permease clusters tightly with the lactose permeases, but that the sucrose permease is much more distant. Raffinose is a trisaccharide which incorporates the structural elements of sucrose, and melibiose in a single molecule. The degree to which these different permeases overlap in specificity is not known, but the broad specificity of the lactose permease of *E. coli* is noteworthy (56, 82).

FAMILY 6: METABOLITE:H^+ SYMPORTER (MHS) FAMILY

The MHS family includes 16 currently sequenced members of widely differing specificities (Table 6). Those of known transport function recognize (i) citrate, (ii) α-ketoglutarate, (iii) proline and betaine, (iv) 4-methyl-*O*-phthalate, and (v) dicarboxylates. The α-ketoglutarate:H^+ symport permease of *E. coli* (KgtP) is probably the best-characterized member of this family (75). An experimentally documented 12-TMS topological model has been proposed for this permease (76).

Metabolites transported by members of the MHS family have little in common, except that they all possess at least one carboxyl group. Several protein members of the MHS family are specific for Krebs cycle intermediates. All are from bacteria, and all characterized members of the MHS family function by proton symport.

A 51-residue segment of the multiple-sequence alignment of the MHS family, including all functionally characterized members, is shown in Fig. 6A. There are no gaps in the aligned sequences, and 11 residues are fully conserved. The majority of the fully conserved residues are probably of structural significance. These residues include four G's, two P's and an A. Two fully conserved residues (M and I) are hydrophobic, and two (D and R) are hydrophilic charged residues. Over half of the positions appear in the consensus sequence, illustrating the high degree of conservation observed for the members of this family.

TABLE 7. Members of the FGHS family (family 7)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
FucP Eco	L-Fucose permease	B	*Escherichia coli*	438	P11551	SP
FucP Hin	Putative L-fucose permease	B	*Haemophilus influenzae*	428	P44776	SP
GluP Bab	Galactose/glucose transporter	B	*Brucella abortus*	412	U43785	GB
GlcP Bsu	Glucose/mannose:H^+ symporter	B	*Bacillus subtilis*	404	AF002191	GB

[a] For abbreviations, see Table 3 footnotes.

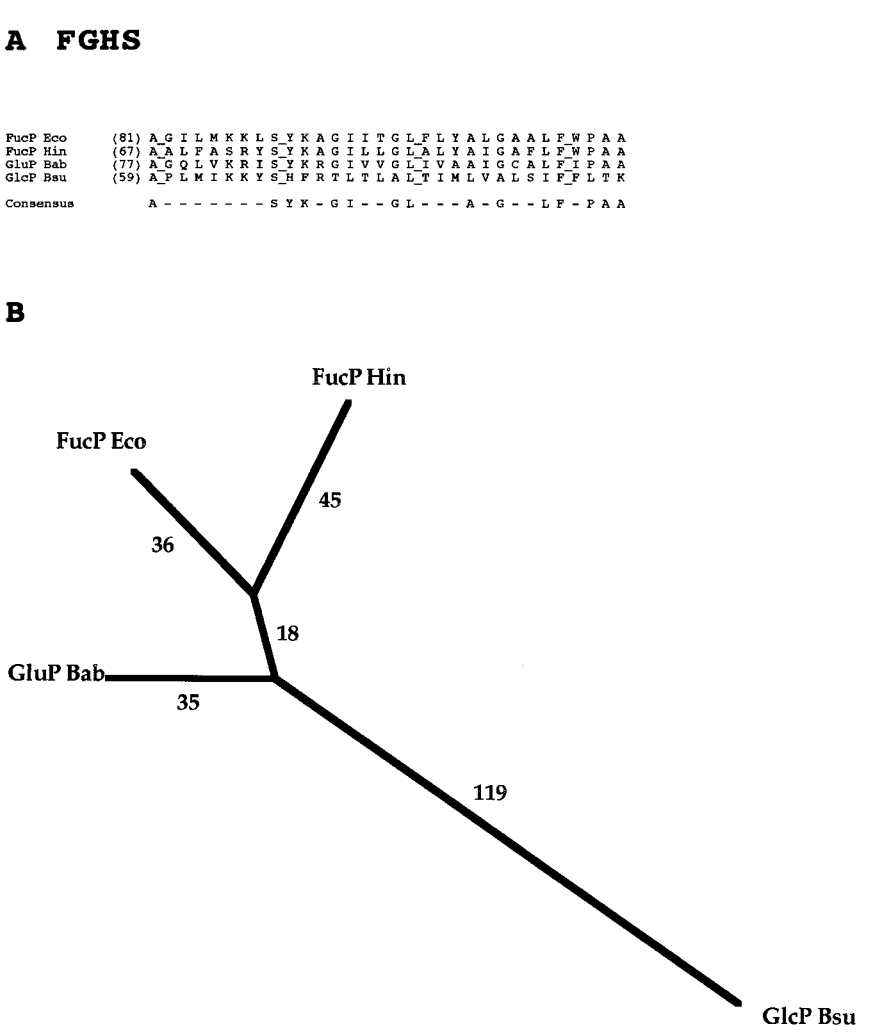

FIG. 7. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the FGHS family (family 7).

The phylogenetic tree for the MHS family reveals clustering according to substrate specificity. Thus, the α-ketoglutarate permease of *E. coli* and the dicarboxylate permease of *Pseudomonas putida* cluster together, the osmoprotectant (proline/betaine) permeases of *E. coli* and *Erwinia chrysanthemi* cluster tightly together and are undoubtedly orthologs, and all three citrate permeases cluster tightly together. These three proteins are also undoubtedly orthologs. The MopB protein of *Burkholderia cepacia*, specific for 4-methyl-*O*-phthalate, is on a branch by itself. Orf3 Shy clusters loosely with dicarboxylate permeases and therefore may exhibit specificity for such a compound. The phylogenetic tree does not provide clues to the functions of the remaining two unidentified proteins. However, these two proteins (YhjE Eco and HI0418 Hin) appear to be located at a phylogenetic distance from each other consistent with their being orthologs. Several MHS homologs of unknown function listed in Table 6 are not represented in Fig. 6 because the sequences were deposited in the databases after completion of the phylogenetic studies reported.

FAMILY 7: FUCOSE-GALACTOSE-GLUCOSE:H^+ SYMPORTER (FGHS) FAMILY

The first sequenced member of the FGHS family to be characterized was the FucP fucose permease of *E. coli* (31), and a 12-TMS topological model for this permease has been presented (32). Subsequently, a galactose/glucose permease of *Brucella abortus* (20) and a glucose/mannose permease of *Ba-*

TABLE 8. Members of the NNP family (family 8)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
CY04C12 Mtu	CY04C12.22c (putative nitrite exporter)	B	*Mycobacterium tuberculosis*	395	Z81360	GB
CY3G12 Mtu	MTCY3G12.04 (putative nitrite transporter)	B	*Mycobacterium tuberculosis*	515	Z79702	GB
MG294 Mge	Unknown function	B	*Mycoplasma genitalium*	474	E64232	PIR
NarK Bsu	Nitrite exporter 1	B	*Bacillus subtilis*	395	P46907	SP
NarK Eco	Nitrite exporter 1	B	*Escherichia coli*	463	P10903	SP
NarK Sty	Putative nitrite exporter (fragment)	B	*Salmonella typhimurium*	130	P37593	SP
NarU Eco	Nitrite exporter 2	B	*Escherichia coli*	462	P37758	SP
NasA Bsu	Putative nitrate transporter	B	*Bacillus subtilis*	421	P42432	SP
Ynt1 Hpo	Nitrate transporter	Y	*Hansenula polymorpha*	508	Z69783	GB
CrnA Eni	Nitrate transporter	F	*Emericella nidulans*	507	P22152	SP
Bch1 Hvu	Putative nitrate transporter	Pl	*Hordeum vulgare*	507	U34198	GB
Bch2 Hvu	Putative nitrate transporter	Pl	*Hordeum vulgare*	509	U34290	GB
Nar3 Cre	Nitrate transporter	Pl	*Chlamydomonas reinhardtii*	547	S40142	PIR

[a] For abbreviations, see Table 3 footnotes.

A NNP

```
Bch1 Hvu    (137) L I G F S L A T F V S C Q Y W M S T M F N S K I I G T V N G L A A G W G N M G G G
Bch2 Hvu    (139) L I G F S L A T F V S C Q Y W M S T M F N S K I I G T V N G L A A G W G N M G G G
Nar3 Cre    (144) F I G L S L C M F V C C Q F W C G T M F N V K I V G T A N A I A A G W G N M G G G
CrnA Eni    (132) F I G I L G G T F V P C Q V W C T G F F D K S I V G T A N S L A A G L G N A G G G
Ynt1 Hpo    (131) F I S F L G S S F I C C S Q F C A V F F D N N I I G T A N A I S A G W G N A G G G
NarU Eco    (137) L C G F A G A N F A S S M G N I S F F F P K A K Q G S A L G I N G G L G N L G V S
NarK Eco    (139) L C G F A G A N F A S S M A N I S F F F P K Q K Q G G A L G L N G G L G N M G V S
CY3G12 Mtu  (154) L T G L G G G N F A S S M S N A N A F Y P H R L K G S A L G I A G G V G N L G V P
NarK Bsu    (104) F L G I G G A V F S I G V T S L P K Y Y P K E K H G V V N G I Y   G A G N I G T A
CY04C12 Mtu (106) F L G V A G T I F A V G I P F A N N W Y Q P A R R G F S T G V F   G M G M V G T A
NasA Bsu    (111) L L G V A G A S F A V A L P M A S R W Y P P H L Q G L A M G I A   G A G N S G T L
MG294 Mge   (126) I W G L W G I T   S T L I F W T P L W K L A S Q Q A T K E N Q A L G F G I Q G A A

Consensus       - - G - - G - - F - - - - - - - - - - F - - - - - G - - - G - A - G - G N - G - -
```

B

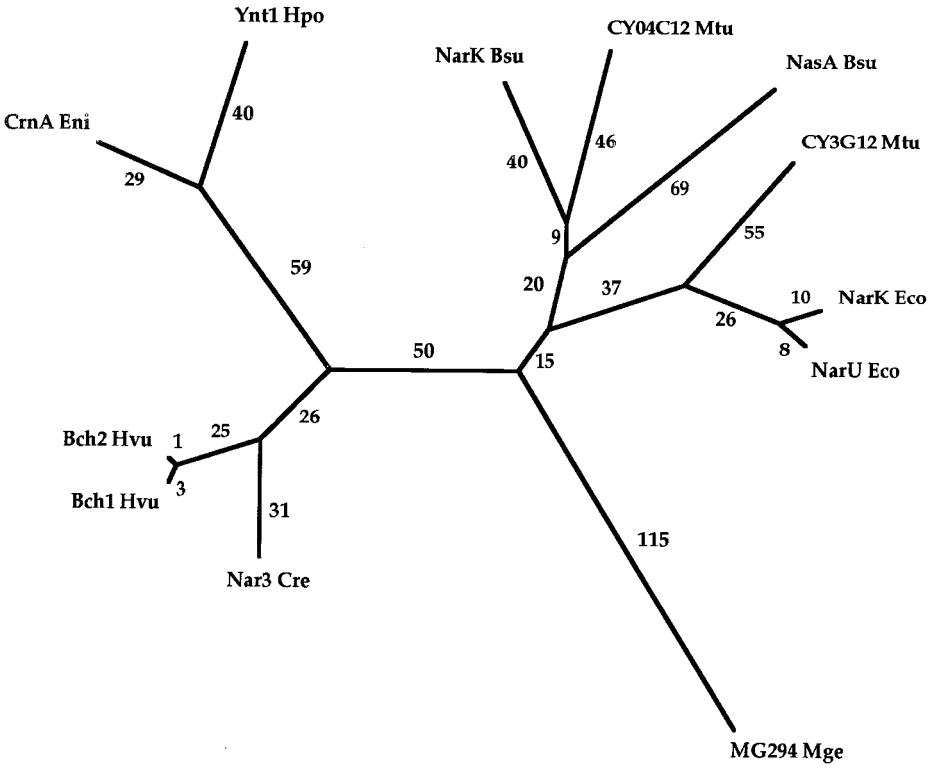

FIG. 8. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the NNP family (family 8).

TABLE 9. Members of the PHS family (family 9)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
Pho84 Sce	Inorganic phosphate transporter	Y	*Saccharomyces cerevisiae*	587	P25297	SP
YaeC Spo	Putative inorganic phosphate transporter	Y	*Schizosaccharomyces pombe*	559	Q09852	SP
YCR98c Sce	Probable metabolite transporter	Y	*Saccharomyces cerevisiae*	518	P25346	SP
GvPT Gve	Phosphate transporter	F	*Glomus versiforme*	521	U38650	GB
Pho-5+ Ncr	High-affinity phosphate transporter	F	*Neurospora crassa*	569	L36127	GB
AtPT1 Ath	Phosphate transporter	Pl	*Arabidopsis thaliana*	524	U62330	GB
AtPT2 Ath	Phosphate transporter	Pl	*Arabidopsis thaliana*	534	U62331	GB
PT1 Ath	Putative phosphate transporter	Pl	*Arabidopsis thaliana*	524	Y07681	GB
PT2 Ath	Putative phosphate transporter	Pl	*Arabidopsis thaliana*	524	Y07682	GB
PT1 Stu	Inorganic phosphate transporter 1	Pl	*Solanum tuberosum*	540	X98890	GB
PT2 Stu	Inorganic phosphate transporter 2	Pl	*Solanum tuberosum*	527	X98891	GB

[a] For abbreviations, see Table 3 footnotes.

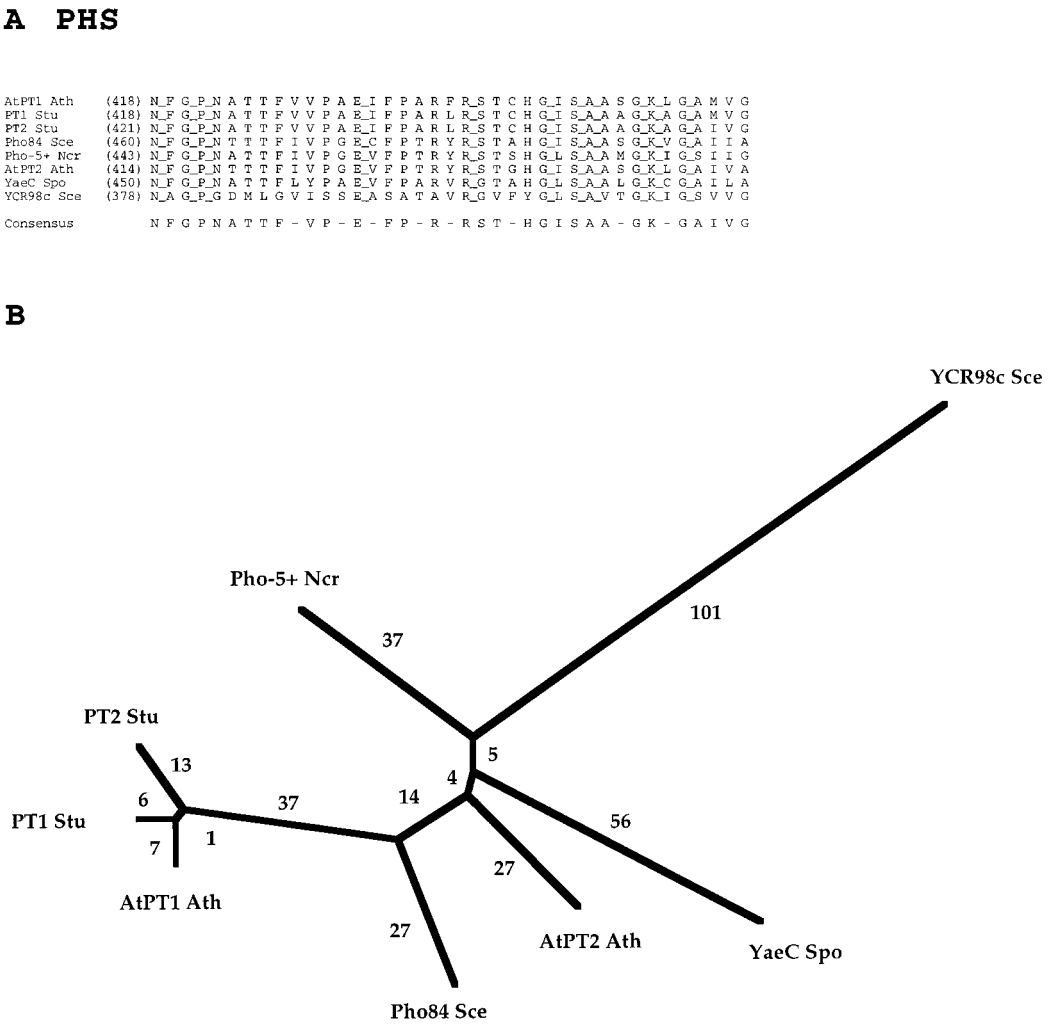

FIG. 9. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the PHS family (family 9).

cillus subtilis (61) were characterized and shown to be members of the FGHS family (Table 7). The *Bacillus* protein, like the *E. coli* FucP protein, is believed to be a sugar:proton symporter (61). Only four proteins are currently in the FGHS family.

In spite of the small size of the FGHS family, its members, all of which are derived from bacteria, exhibit a surprising degree of sequence diversion. Figure 7A shows the best-conserved portion of the complete multiple-sequence alignment of these sequences. Only 4 positions in the gap-free 32-position alignment shown are fully conserved, and a minority of the residue positions appear in the consensus sequence. The phylogenetic tree reveals that the *Haemophilus influenzae* protein clusters

TABLE 10. Members of the NHS family (family 10)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
NupG Eco	Nucleoside transporter	B	*Escherichia coli*	418	P09452	SP
XapB Eco	Xanthosine transporter	B	*Escherichia coli*	418	P45562	SP

[a] For abbreviations, see Table 3 footnotes.

NHS

```
NupG Eco (143) E L S H M Q L Y I G A A L S A I L V L F T L T L P H I P V A K Q Q A N Q S W T T L L G L D A F A L F K N K
XapB Eco (143) E L S S L Q L Y I A S G A S L L L S A Y A L T L P K I P V A E K K A T T S L A S K L G L D A F V L F K N P

Consensus      E L S - - Q L Y I - - - - S - - L - - - - L T L P - I P V A - - - A - - S - - - - L G L D A F - L F K N -
```

FIG. 10. Partial alignment of the sequences of the two currently sequenced proteins of the NHS family (family 10).

closely with FucP of *E. coli*, that the galactose/glucose permease of *B. abortus* is more distant, and that the hexose permease from *B. subtilis* is most divergent. These relative distances correlate with substrate specificity to the extent known since fucose is of the galacto configuration.

FAMILY 8: NITRATE-NITRITE PORTER (NNP) FAMILY

Thirteen proteins make up the current NNP family (Table 8). These proteins are derived from a variety of gram-negative and gram-positive bacteria as well as various eukaryotes including yeasts (Ynt1 Hpo), fungi (CrnA Eni), algae (Nar3 Cre), and higher plants (Bch1 Hvu and Bch2 Hvu). Irrespective of the organism, the nitrate permeases of the NNP family take up their substrate while the nitrite permeases apparently extrude theirs. Well-characterized members of the family are the NarK nitrite extrusion system involved in anaerobic nitrate-dependent respiration in *E. coli* (68) and the CrnA nitrate uptake permease of *Aspergillus nidulans* (81).

A portion of the NNP family protein multiple alignment is shown in Fig. 8A. Although this region is the best-conserved region in the complete multiple-sequence alignment, only three residues, all G's, are fully conserved. There are few gaps, and several residues are largely conserved (e.g., three additional G's, an F, and an N are conserved in all but one, two, or three of the proteins, respectively).

The phylogenetic tree for the NNP family (Fig. 8B) reveals that all of the eukaryotic proteins cluster together (on the left) as do the prokaryotic proteins (on the right). Further, within the eukaryotic cluster, the fungal proteins comprise one cluster while the plant proteins comprise another. Within the prokaryotic cluster, the two (putative) nitrite extrusion systems of *E. coli* (NarK and NarU) cluster tightly together. These paralogs probably arose by a recent gene duplication event. All other prokaryotic full-length proteins represented are from gram-positive bacteria. The NasA and NarK proteins of *B. subtilis* cluster loosely together, even though they are believed to catalyze nitrate uptake and nitrite efflux, respectively. The shorter phylogenetic distance of the *M. tuberculosis* protein (CY04C12) from NarK Bsu suggests that these two proteins may have the same or similar functions. Further, the other *M. tuberculosis* protein, CY3G12, resembles the two *E. coli* permeases, NarK and NarU (Fig. 8B). By contrast, the MG294 protein of *M. genitalium* is distant from all other members of the NNP family. Proteins from the mycoplasmas often exhibit greater distances from other gram-positive bacterial homologs than do other homologs from the latter bacteria (unpublished observations). The tree thus does not provide a clue to the function of this protein.

FAMILY 9: PHOSPHATE:H^+ SYMPORTER (PHS) FAMILY

The current PHS family is unusual in that it includes members from yeasts, fungi and plants but none from bacteria, animals, and other eukaryotes (Table 9). As a family within the MFS, it is presumed to be of ancient origin, and therefore one would expect members of the family to be found in bacteria (see Conclusions and Perspectives, below). Two well-characterized members of the PHS family are the Pho84 inorganic phosphate transporter of *S. cerevisiae* (9) and the GvPT phosphate transporter of *Glomus versiforme* (33).

The 11 members of the PHS family are fairly uniform in size (518 to 587 residues) and exhibit a striking degree of sequence similarity (Fig. 9A). Only eight proteins are included in Fig. 9 because of the near identity of the remaining three sequences to at least one protein that was included. Thus, in the 39-residue segment of the multiple-sequence alignment presented for the eight divergent proteins, 11 positions exhibit full conservation. Four of these residues are glycines, one is a proline, and one is alanine, all of which probably play structural roles. The remaining fully conserved residues (N, E, R, S, and K) are polar and therefore may function in substrate or proton binding or in catalysis of transport.

The phylogenetic tree for eight members of the PHS family is shown in Fig. 9B. Three of the four plant proteins cluster tightly together, and the short branch lengths separating them indicate that these proteins differ only slightly in sequence. This fact is also revealed by the partial multiple-sequence alignment shown in Fig. 9A. The yeast and fungal proteins cluster loosely together with the fourth plant protein. The occurrence of distant homologs in both the plant and fungal kingdoms suggests that both kingdoms possess isoforms that diverged from each other before plants diverged from fungi.

TABLE 11. Members of the OFA family (family 11)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
OxlT Ofo	Oxalate:formate antiporter	B	*Oxalobacter formigenes*	418	U40075	GB
YhjX Eco	Hypothetical 43-kDa protein	B	*Escherichia coli*	402	P37662	SP
c01003 Sso	Unknown	Ar	*Sulfolobus solfataricus*	430	Y08256	GB
Y38K Tte	Hypothetical 38-kDa protein	Ar	*Thermoproteus tenax*	373	P05715	SP
F10G7 Cel	tRNA-Leu in exon 5	An	*Caenorhabditis elegans*	470	U40029	GB

[a] For abbreviations, see Table 3 footnotes.

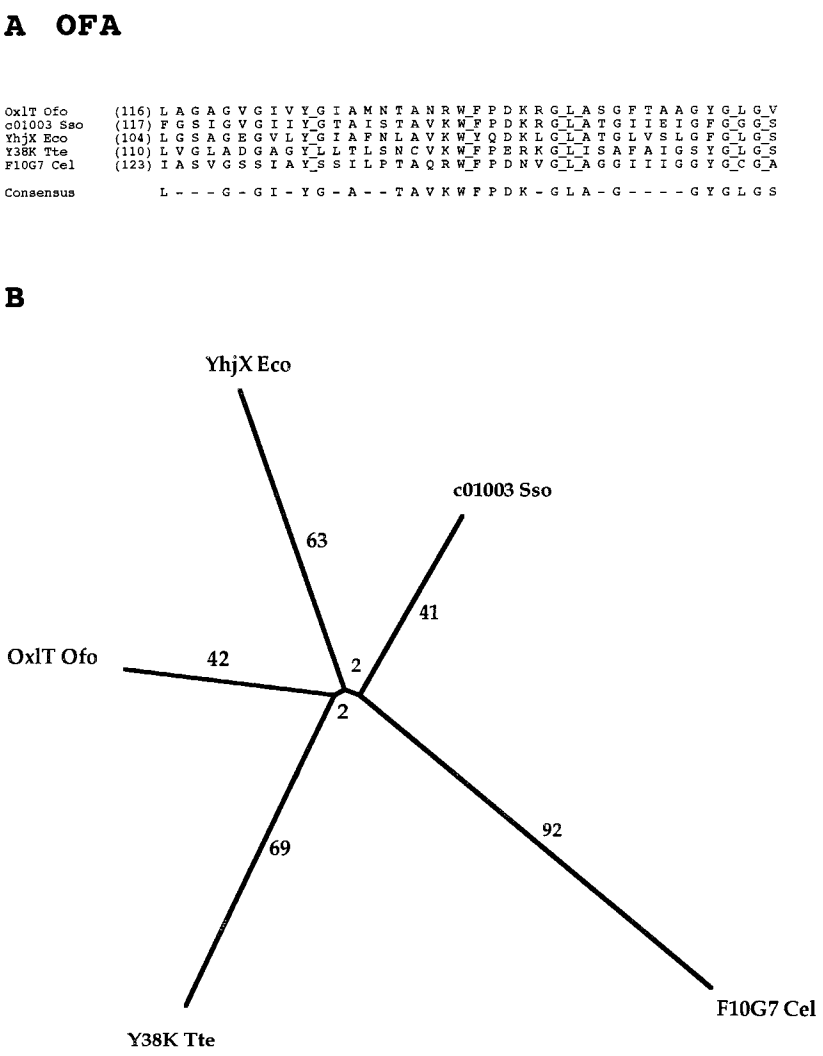

FIG. 11. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the OFA family (family 11).

FAMILY 10: NUCLEOSIDE:H$^+$ SYMPORTER (NHS) FAMILY

The NHS family currently has only two sequenced members, and both proteins are from *E. coli* (Table 10). One is a general nucleoside:proton symporter, NupG, and the other is a xanthosine permease. NupG has been examined structurally and is the better characterized of the two proteins (85).

As shown in Fig. 10, these two proteins are very similar, exhibiting close to 50% identity in the region shown. They evidently arose by a gene duplication event that occurred relatively recently in evolutionary time.

FAMILY 11: OXALATE:FORMATE ANTIPORTER (OFA) FAMILY

Five sequenced proteins comprise the OFA family, but only one of these proteins has been functionally characterized (Ta-

TABLE 12. Members of the SHS family (family 12)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
NanT Eco	Sialic acid transporter	B	*Escherichia coli*	496	P41036	SP
NanT Hin	Putative sialic acid transporter	B	*Haemophilus influenzae*	407	C64167	PIR
YjhB Eco	Putative sialic acid transporter	B	*Escherichia coli*	425	P39352	SP

[a] For abbreviations, see Table 3 footnotes.

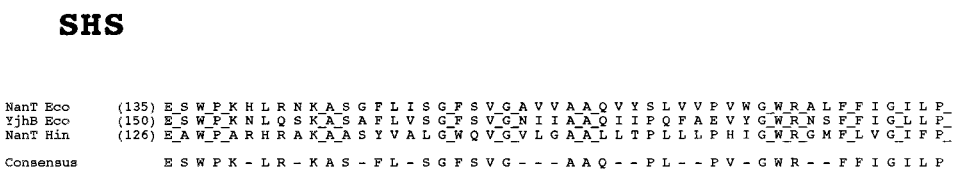

FIG. 12. Partial alignment of the sequences of the three currently sequenced proteins of the SHS family (family 12).

ble 11). This protein, the oxalate:formate antiporter from *Oxalobacter formigenes*, provides the basis for naming the OFA family (1, 4). The protein has been purified, reconstituted in an artificial membrane system, and studied structurally (24, 69). As is apparent from Table 11, members of the OFA family are widely distributed in nature, being present in the bacterial, archaeal, and eukaryotic kingdoms.

One would expect that a family that derives its members from all three kingdoms of life would exhibit little sequence similarity. However, as shown in Fig. 11A, 6 residues in the 40-residue segment shown are fully conserved and a majority of residues are present in the consensus sequence. The phylogenetic tree reveals that all five members of the OFA family are nearly equally distantly related to each other. This may be due in part to the fact that most of the organisms from which these proteins are derived are distantly related. However, one cannot draw conclusions about whether these proteins are orthologs of the same function. The answer to this dilemma will require direct experimentation.

FAMILY 12: SIALATE:H^+ SYMPORTER (SHS) FAMILY

Only three currently sequenced proteins comprise the SHS family, and only one of these, the NanT sialic acid permease of *E. coli*, is functionally characterized (Table 12) (52). *E. coli* possesses two SHS family paralogs, and *Haemophilus influenzae* possesses one homolog. Sequence comparisons (Fig. 12) and phylogenetic analyses (data not shown) reveal that the two *E. coli* paralogs are more closely related to each other than either of these proteins is related to the protein from *Haemophilus influenzae*, even though *H. influenzae* is closely related to *E. coli*. The function of this last protein can therefore not be surmised.

FAMILY 13: MONOCARBOXYLATE PORTER (MCP) FAMILY

Thirteen proteins comprise the MCP family, and all are from eukaryotes (Table 13). Most of these proteins are derived from various animal sources including three from *C. elegans*. However, *S. cerevisiae* possesses four paralogs. Only mammalian members of the MCP family have been functionally characterized. These permeases appear to be energized by proton symport (80). Monocarboxylates transported by these permeases include lactate, pyruvate, and mevalonate (25). Topological studies leading to a 12-TMS model have been reported (65).

The portion of the complete multiple-sequence alignment shown in Fig. 13A reveals that within the 53-residue segment presented, only 4 residues are fully conserved, and all are probably of structural significance (two G's, one F, and one A). In fact, except for the KRR motif and two serines, all of the residues that appear in the consensus sequence are structural or hydrophobic.

The phylogenetic tree for the MCP family is shown in Fig. 13B. All of the functionally characterized mammalian monocarboxylate transporters cluster tightly together, and even the outlying proteins from higher animals (RemP Gga and XpcT Hsa) are loosely associated with this cluster. The three *C. elegans* paralogs comprise a second diverse cluster. The four *S. cerevisiae* paralogs branch distantly from the animal proteins, and the yeast paralogs comprise two distinct clusters. Because of the extensive sequence diversion of the functionally uncharacterized proteins, we anticipate that they will prove to exhibit different transport functions.

FAMILY 14: ANION:CATION SYMPORTER (ACS) FAMILY

The ACS family is a large family with 40 currently sequenced members (Table 14). One of the members of this family was previously recognized to be a member of the MFS (67). This protein is the rabbit inorganic phosphate:Na^+ cotransporter (84). Several mammalian proteins of this specificity have now been characterized. All of the recognized substrates of the ACS family permeases are either organic or inorganic anions. Among the organic anions transported are glucarate, hexuronates, phthalate, allantoate, and probably tartrate (13, 66).

Proteins of the ACS family are widely distributed in nature. They are found in both gram-negative and gram-positive bacteria and in both the animal and fungal eukaryotic kingdoms. Oddly, no plant member has been sequenced, and none of the sequenced ACS proteins is from an archaeon.

Several organisms possess multiple ACS family paralogs. Thus, *B. subtilis* and *Rattus norvegicus* each have at least 2, *E. coli* has 5, *S. cerevisiae* has 7, and *C. elegans* has at least 15. Considering that the *C. elegans* genome was only about half sequenced when these analyses were conducted, one can anticipate that this one organism will prove to have nearly 30 paralogs within this one family of the MFS!

A portion of the complete multiple-sequence alignment for the proteins of the ACS family is shown in Fig. 14A. No residue

TABLE 13. Members of the MCP family (family 13)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
Ykl221w Sce	Putative monocarboxylate transporter	Y	*Saccharomyces cerevisiae*	473	P36032	SP
ORF23 Sce	pid:e183206 (putative monocarboxylate transporter)	Y	*Saccharomyces cerevisiae*	673	Z46843	GB
Yol119c Sce	Orf Yol119c (putative monocarboxylate transporter)	Y	*Saccharomyces cerevisiae*	501	Z74861	GB
Yor306c Sce	Orf Yor306c (putative monocarboxylate transporter)	Y	*Saccharomyces cerevisiae*	521	Z75214	GB
C49F8 Cel	C49F8.2 (putative monocarboxylate transporter)	An	*Caenorhabditis elegans*	807	Z70206	GB
K05B2 Cel	K05B2.5 (putative monocarboxylate transporter)	An	*Caenorhabditis elegans*	808	U29379	GB
Mct Clo	Monocarboxylate transporter 1	An	*Cricetulus longicaudatus*	494	Q03064	SP
Mct Hsa	Monocarboxylate transporter 2	An	*Homo sapiens*	500	P53985	SP
Mct2 Mau	Monocarboxylate transporter	An	*Mesocricetus auratus*	484	L31957	GB
Mct2 Rno	Monocarboxylate transporter	An	*Rattus norvegicus*	489	U62316	GB
RemP Gga	Putative monocarboxylate transporter	An	*Gallus gallus*	450	U15685	GB
T02G5 Cel	T02G5.12 (putative monocarboxylate transporter)	An	*Caenorhabditis elegans*	556	U41105	GB
XpcT Hsa	X-linked PEST-containing transporter	An	*Homo sapiens*	613	P36021	SP

[a] For abbreviations, see Table 3 footnotes.

A MCP

```
Mct Hsa      (92)  A S F C N T V Q Q L Y V C I G V I G G L G L A F N L N P A L T M I G K Y F Y K R R P L A N G L A M A G S P V
Mct Clo      (92)  A S F C N T V Q E L Y L C I G V I G G L G L A F N L N P A L T M I G K Y F Y K K R P L A N G L A M A G S P V
Mct2 Rno     (98)  A S F S S S V I E L Y L T V G F I G G L G L A F N L Q P A L T I I G K Y F Y R K R P L A N G F A M A G S P V
Mct2 Mau     (93)  A S F S N S V L E L Y L T I G F I G G L G L A F N L Q P A L T I I G K Y F Y R R R P M A N G L A M A G S P V
RemP Gga      (4)  A S F T T N I I E L Y L T A G V L T G L G M A L N F Q P S L I M L G T Y F D K R R P L A N G L A A A G S P V
XpcT Hsa    (251)  S S F T S S L S L R Y F T Y G I L F G C G C S F A F Q P S L V I L G H Y F Q R R L G L A N G V V S A G S S I
K05B2 Cel   (162)  A P A S P N I Y V F M L I Y G V M G G L G F G M I Y L P A I V V V G F Y F D S K R A M A T G I S V A G S G V
T02G5 Cel    (92)  S C F A T E I W H F V I S V G V I M G I G F G L V Y C P A I V I V T M Y F E S K R S L A T G I A V A G A G V
C49F8 Cel   (181)  A M F C S H I F F F M I S F G L G C G V G M S F I Y N A A I V I V T Y Y F E K R R G L A T S F A V S G T G V
YOR306c Sce (181)  T A N S T K Y W H F I L S F A I V C G F G N G I V L S P L V S V P A H Y F F K R R G T A L A M A T I G G S V
YOL119c Sce (171)  L A N C K S V W Q F I L A F S V C S G L G T G I L M T P L I G T V A T W F L K R R G I A T S I S T M G G S I
Orf23 Sce   (186)  A S F T T K L W Q L Y V T Q G F M V G C S I S L I F V P A T T V L P G W F L K K R A V A M G V S L L G T G A
YKL221w Sce (115)  A A F S V T L W E I Y L T Q G V L I G F G L A F I F I P S V T L I P L W F R N K R S L A S G I G T A G S G L

Consensus          A S F - - - - - - - Y L - - G V - - G L G - - - - - - P A - - - - G - Y F - K R R - L A - G - A - A G S - V
```

B

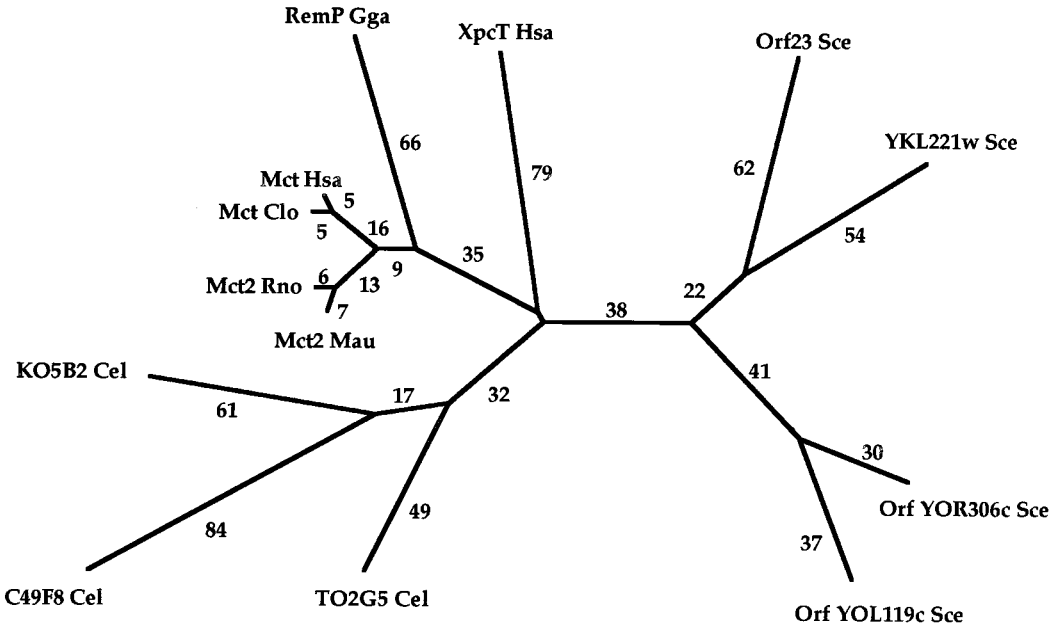

FIG. 13. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the MCP family (family 13).

TABLE 14. Members of the ACS family (family 14)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
ExuT Eco	Hexuronate transporter	B	*Escherichia coli*	472	P42609	SP
YcbE Bsu	Putative glucarate transporter	B	*Bacillus subtilis*	455	P42237	SP
YcbE Ppu	Putative glucarate transporter	B	*Pseudomonas putidas*	456	P42205	SP
HpaX Eco	4-Hydroxyphenylacetate transporter	B	*Escherichia coli*	458	Z37980	GB
Orf f450 Eco	Putative glucarate transporter	B	*Escherichia coli*	450	U29581	GB
Pht1 Ppu	Phthalate transporter	B	*Pseudomonas putida*	451	Q05181	SP
TtuB Avi	Tartrate transporter	B	*Agrobacterium vitis*	433	U25634	GB
TtuB1 Avi	Tartrate transporter	B	*Agrobacterium vitis*	449	U32375	GB
YhaU Eco	Putative glucarate transporter	B	*Escherichia coli*	444	P42613	SP
YidT Eco	Galactonate transporter	B	*Escherichia coli*	445	P31457	SP
YybO Bsu	Glucarate transporter	B	*Bacillus subtilis*	435	P37489	SP
C11D3.18c Spo	C11D3.18c (putative Na^+/phosphate cotransporter)	Y	*Schizosaccharomyces pombe*	498	Z68166	GB
Dal5 Sce	Allantoate permease	Y	*Saccharomyces cerevisiae*	543	P15365	SP
L0578 Sce	Unknown function	Y	*Saccharomyces cerevisiae*	531	S50965	PIR
Yal0670 Sce	Unknown function	Y	*Saccharomyces cerevisiae*	593	P39709	SP
Ycr28c Sce	Unknown function	Y	*Saccharomyces cerevisiae*	512	P25621	SP
Ygr260w Sce	Unknown function	Y	*Saccharomyces cerevisiae*	534	Z73044	GB
Yil166c Sce	Unknown function	Y	*Saccharomyces cerevisiae*	542	P40445	SP
Ylr004c Sce	Orf Ylr004c (function unknown)	Y	*Saccharomyces cerevisiae*	523	Z73176	GB
C18D1.2 Cel	C18D1.2 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	493	Z48543	GB
C35A5 Cel	C35A5.3 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	596	Z71185	GB
D1046.4 Cel	D1046.4 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	583	Z68160	GB
F09A5.1 Cel	F09A5.1 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	411	Z69788	GB
F41C3.2 Cel	F41C3.2 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	467	U23521	GB
K08C7.1 Cel	K08C7.1 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	498	Z70286	GB
M117 Cel	M117.1 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	552	Z73910	GB
Na/Pi-1 Ocu	Na^+/phosphate cotransporter, renal	An	*Oryctolagus cuniculus*	465	A56410	PIR
Na/Pi-4 Hsa	Na^+/phosphate cotransporter, renal	An	*Homo sapiens*	465	D28532	GB
Npt1 Mmu	Na^+/phosphate cotransporter 1	An	*Mus musculus*	465	X77241	GB
NTP1 Hsa	Na^+/phosphate cotransporter	An	*Homo sapiens*	467	A48916	PIR
RNaPi-1 Rno	Na^+/phosphate cotransporter, hepatic	An	*Rattus norvegicus*	465	U28504	GB
T07A5 Cel	T07A5.3 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	385	Z48055	GB
T09B9 Cel	T09B9.2 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	516	Z47070	GB
T27D12.1 Cel	T27D12.1 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	493	Z70037	GB
Ykh4 Cel	C02C2.4 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	568	P34272	SP
Yld2 Cel	C38C10.2 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	472	Q03567	SP
ZK512 Cel	Putative Na^+/phosphate cotransporter	An	*Caenorhabditis elegans*	576	P34644	SP
ZK54 Cel	ZK54.1 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	380	U58737	GB
ZK652.10 Cel	ZK652.10 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	420	S44900	PIR
ZK682 Cel	YK61E12.5 (putative Na^+/phosphate cotransporter)	An	*Caenorhabditis elegans*	506	U41110	GB

[a] For abbreviations, see Table 3 footnotes.

is fully conserved at any one position. However, the aligned sequences are essentially gap free except for an incompletely sequenced region of one protein from *C. elegans* and another *C. elegans* protein which exhibits a single-residue insertion not found in the other proteins. Particularly worthy of note are the following residues found in the consensus sequence. The G is conserved in all but one protein; the W is conserved in all of the top 31 proteins, and the ER motif is conserved in many of the proteins. It is clear that the proteins have been correctly aligned in spite of very significant sequence divergence.

The phylogenetic tree for the ACS family is shown in Fig. 14B. Distinct clustering of the many proteins represented is apparent. Most striking is the fact that all bacterial proteins comprise one cluster, all the yeast proteins comprise a second, and the animal proteins comprise two additional diverse clusters. Proteins of known and similar specificities cluster tightly together (e.g., the mammalian inorganic phosphate:Na^+ cotransporters or the glucarate transporters of gram-negative and gram-positive bacteria). Proteins specific for different but structurally related substrates (e.g., phthalate and 4-hydroxyphenylacetate, or glucarate and hexuronate) are found within the same cluster but distantly, while those of very dissimilar substrate specificities do not cluster at all. This tendency of permeases of similar specificities to "flock" together has been noted before (71, 72) and provides a basis for assigning tentative functions to several of the uncharacterized members of the ACS family.

FAMILY 15: AROMATIC ACID:H^+ SYMPORTER (AAHS) FAMILY

Table 15 presents the seven members of the AAHS family. These proteins are derived exclusively from gram-negative bacteria, and they all transport aromatic acids. They exhibit a relatively narrow range of sizes (418 to 460 residues).

The part of the complete multiple-sequence alignment reproduced in Fig. 15A for the seven members of the AAHS family reveals full conservation in 6 of the 49 positions shown, while 26 positions appear in the consensus sequence. Four of the fully conserved residues are glycines, of likely structural

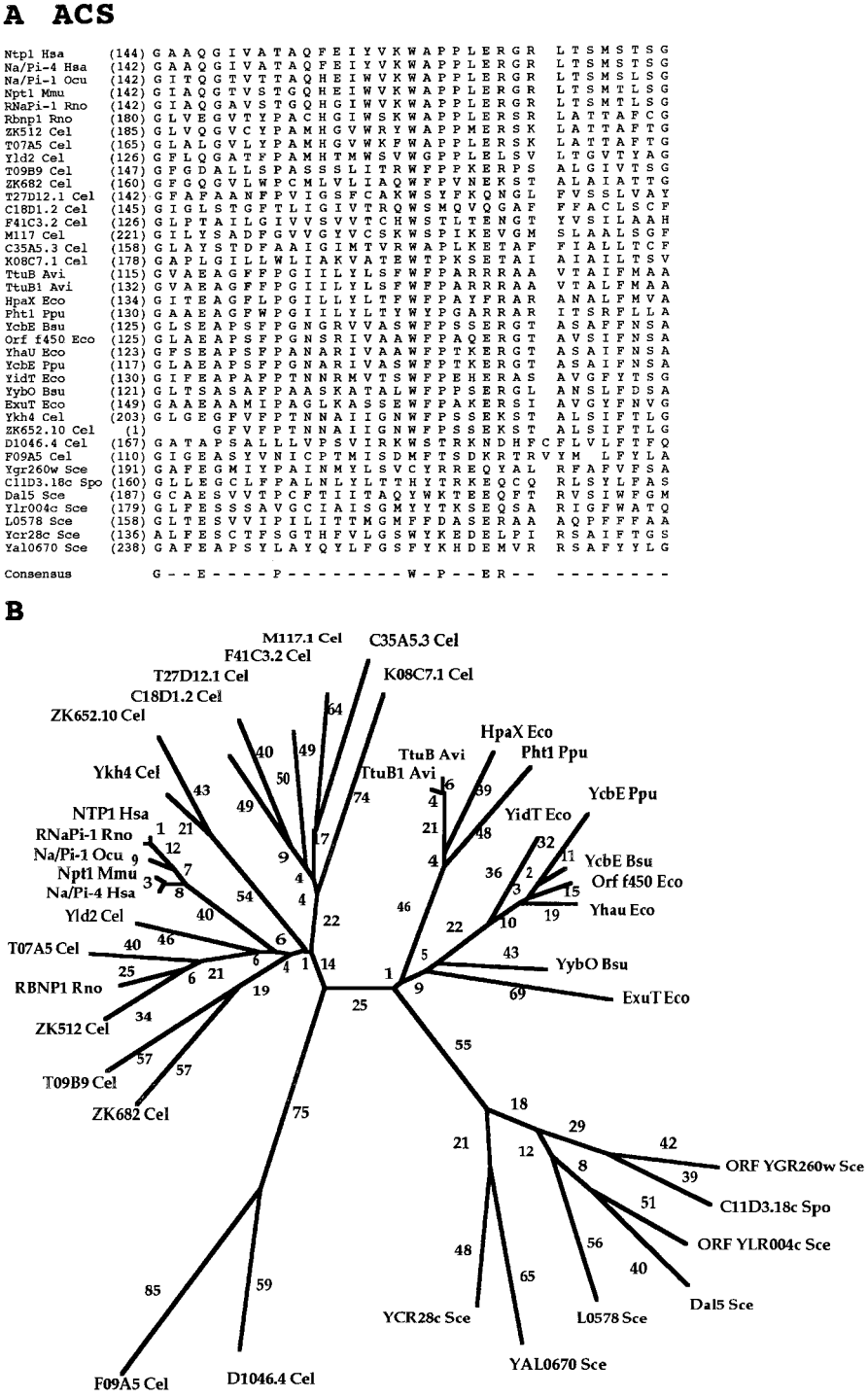

FIG. 14. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the ACS family (family 14).

significance. The remaining two fully conserved residues are S and D. The fully conserved D and the adjacent R/K residues are part of a motif seen in part in the consensus sequence as follows: LADRFGRKR. This region appears to comprise a hydrophilic loop separating two very hydrophobic TMSs (TMS2 and TMS3 [see Table 19]).

The phylogenetic tree of the seven proteins of the AAHS family (Fig. 15B) reveals that the two PcaK proteins of the same specificity cluster together, but the other proteins do not cluster. Based on this tree, we doubt that the *E. coli* protein is a 3-hydroxyphenylpropionate transporter like HppK.

Two additional proteins, the BenE benzoate:H^+ symporter of *Acinetobacter calcoaceticus* and an *E. coli* protein, exhibit 36% identity and 61% similarity to each other. These two proteins show limited sequence similarity to members of the AAHS family, but the degree of similarity is insufficient to establish homology. The BenE family is therefore considered a distinct family (TC 2.46) (72b).

TABLE 15. Members of the AAHS family (family 15)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
HppK Rgl	Putative 3-hydroxyphenyl propionate transporter	B	*Rhodococcus globerulus*	453	U89712	GB
MucK Aca	*cis*,*cis*-Muconate transporter	B	*Acinetobacter calcoaceticus*	413	U87258	GB
PcaK Aca	Putative 4-hydroxybenzoate transporter	B	*Acinetobacter calcoaceticus*	421	L05770	GB
Min6 Eco	Unknown function	B	*Escherichia coli*	418	U73857	GB
PcaK Ppu	4-Hydroxybenzoate transporter	B	*Pseudomonas putida*	448	U10895	GB
TfdK Reu	Putative 2,4-dichlorophenoxyacetate transporter	B	*Ralstonia eutropha*	460	U16782	GB
BenK Aca	Benzoate transporter	B	*Acinetobacter calcoaceticus*	466	AF009224	GB

[a] For abbreviations, see Table 3 footnotes.

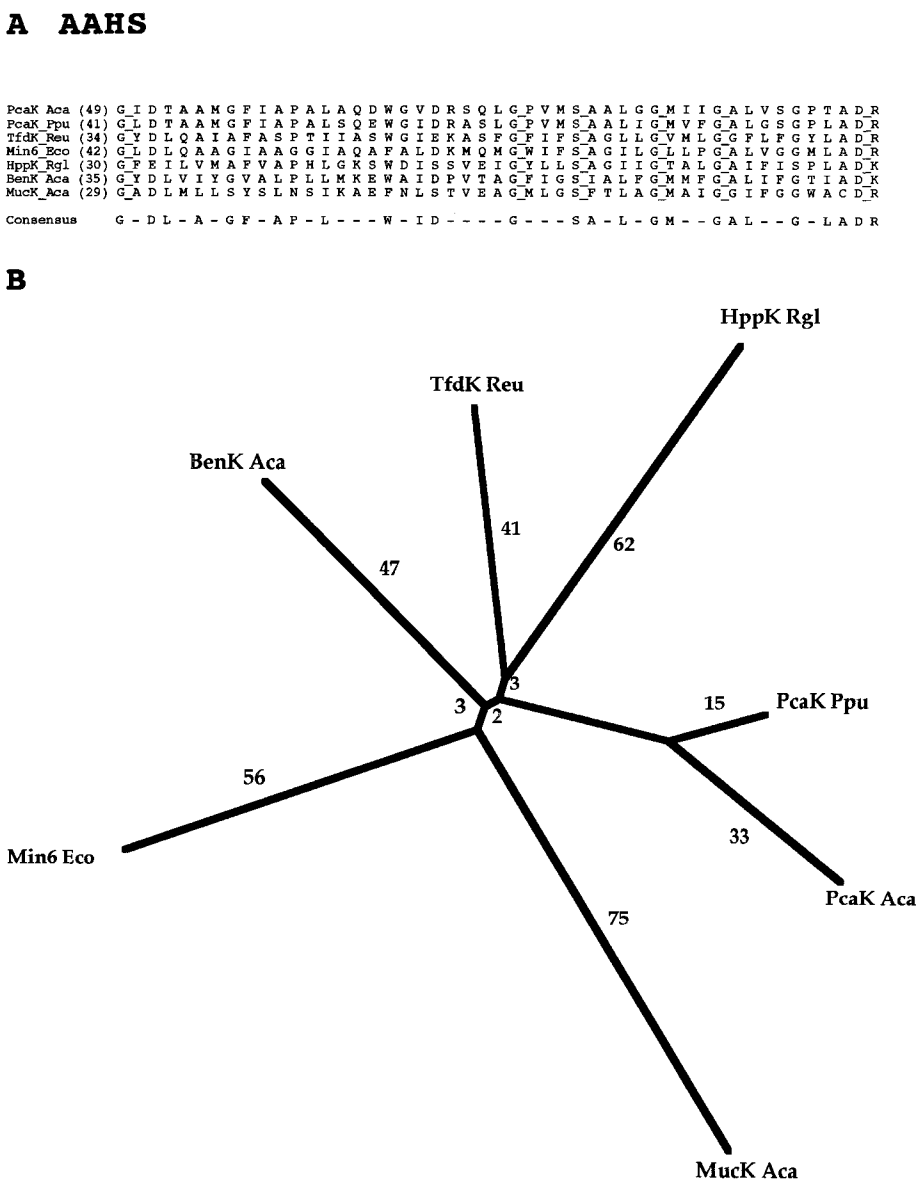

FIG. 15. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the AAHS family (family 15).

FAMILY 16: UNKNOWN MAJOR FACILITATOR (UMF) FAMILY

While screening the complete *S. cerevisiae* genome, Goffeau et al. (26) recently identified a new family within the MFS. The *S. cerevisiae* genome was found to encode a total of 28 proteins that either cluster with the drug resistance families 2 and 3 of the MFS or were members of a new cluster (cluster III in reference 26). Six *S. cerevisiae* proteins comprise the new family, and because none of the encoded proteins has been functionally characterized, we have designated this cluster the un-

TABLE 16. Members of the UMF family (family 16)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
Ycl070c Sce	Unknown function	Y	*Saccharomyces cerevisiae*	631	P25595	SP
Yel065w Sce	Unknown function	Y	*Saccharomyces cerevisiae*	628	P39980	SP
Yhl040c Sce	Unknown function	Y	*Saccharomyces cerevisiae*	627	P38731	SP
Yhl047c Sce	Unknown function	Y	*Saccharomyces cerevisiae*	637	P38724	SP
Ykr065w Sce	Unknown function	Y	*Saccharomyces cerevisiae*	615	P36173	SP
Yol158c Sce	Unknown function	Y	*Saccharomyces cerevisiae*	606	Z74900	GB

[a] For abbreviations, see Table 3 footnotes.

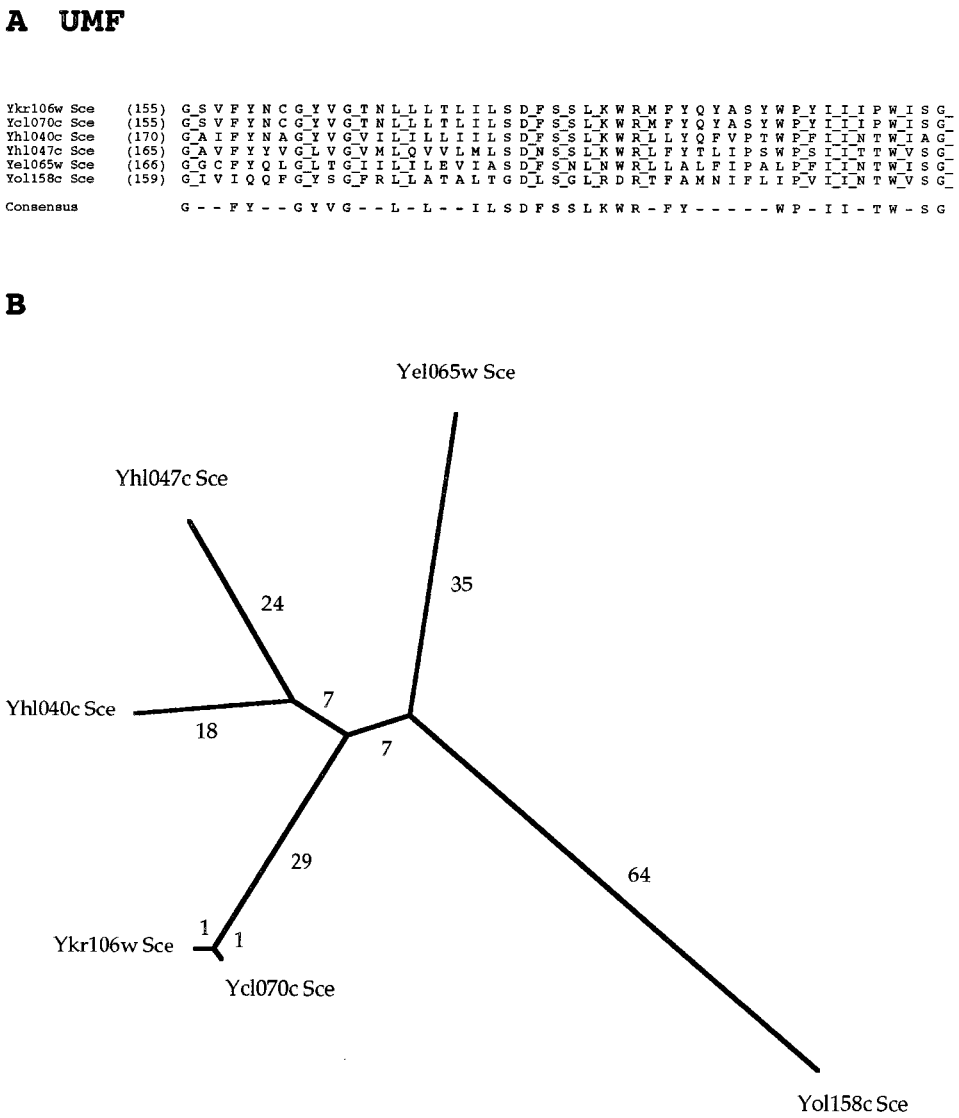

FIG. 16. Partial multiple-sequence alignment (A) and phylogenetic tree (B) for the six currently sequenced proteins of the UMF family (family 16).

known major facilitator (UMF) family (Table 16). Like family 2 drug resistance permeases (DHA14; Table 1), the members of this new family possess 14 putative TMSs. These novel yeast proteins are fairly uniform in size (606 to 637 residues). Surprisingly, no sequenced close homologs are found in any other organism.

A partial multiple-sequence alignment and the phylogenetic tree for the UMF family are shown in Fig. 16A and B, respectively. Figure 16A reveals the high degrees of sequence similarity observed for these six proteins. Two of the UMF paralogs (Ykr106w and Yc1070c) are extremely close in sequence and must have arisen from a very recent gene duplication event. Two other UMF paralogs (Yh1047c and Yh1040c) cluster loosely together and presumably resulted from an earlier gene duplication event. The remaining two paralogs (Yel065w and Yol158c) are the most divergent members of the family.

TABLE 17. Members of the CP family (family 17)

Abbreviation	Description	Phylum[a]	Organism	Size (no. of residues)	Accession no.	Database[a]
CynX Eco	Cyanate transporter	B	*Escherichia coli*	393	U73857	GB
Orf 393 Eco	Putative cyanate transporter	B	*Escherichia coli*	393	D90822	GB
YycB Bsu	Unknown function	B	*Bacillus subtilis*	402	P37482	SP

[a] For abbreviations, see Table 3 footnotes.

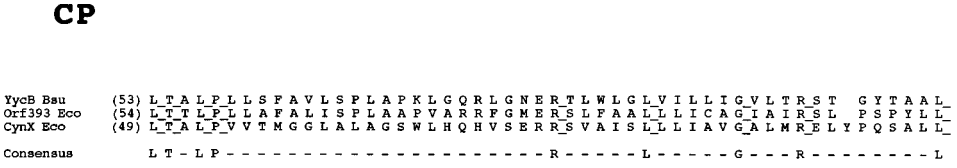

FIG. 17. Partial multiple-sequence alignment for the three members of the CP family (family 17).

FAMILY 17: CYANATE PERMEASE (CP) FAMILY

Three currently sequenced proteins are found within the cyanate permease family, two from *E. coli* and one from *B. subtilis* (Table 17). One of these proteins, CynX, is encoded within the cyanate degradative operon *cynTSX* of *E. coli* (79). CynX is 393 amino acyl residues long and possesses 11 or 12 putative TMSs (Fig. 17). The other members of the family are an *E. coli* protein of 393 amino acids, demonstrably homologous to the tartrate permease of *Agrobacterium vitis* (Table 14), and a *B. subtilis* protein of 402 amino acids (Table 17).

POSSIBLE INCLUSION OF THE PROTON-DEPENDENT OLIGOPEPTIDE TRANSPORTER (POT) FAMILY IN THE MFS

The POT family has been described by two different laboratories. In 1994, Paulsen and Skurray (62) recognized eight sequenced or partially sequenced members of this family and named it the proton- (or proton motive force [pmf])-dependent oligopeptide transporter (POT) family. Subsequently, Steiner et al. (78) renamed the family the peptide transport (PTR) family. Neither group noticed similarities of the mem-

TABLE 18. Signature sequences for 16 of the 17 families of the MFS as well as for the related POT family

Family		Signature sequence
No.	Name	
1	SP[a]	[LIVFYW]-x-[SAGTN]-E-[LIVFMTN]-x_5-[RS]-[GRKYA]-x_7-[LIVMFST]-x-[LIVWF]-x_2-G-x-[LIVFM]-[LIVGTFY]
2	DHA14[b]	W-[RPSEGH]-[WYSA]-[LIVCA]-[FL]-[LIVWYF]-[LIVA]-[NAPY]-[LIVMS]-[PI]-[LIVMATF]-$x_{1,3}$-[LIVANGS]-[LIVMFCAP]-[LIVAFG]-[LIMAGY]
3	DHA12[b]	[TINFYKAS]-[GYLIVTAS]-x_2-[LIVFAP]-x_2-[LIFSMWP]-[LIVFWS]-x-[GAM]-x_2-[ANSTV]-[DANR]-x_2-[GEP]-[RYK]-[KRHTPW]-x-[LIVMPS]-[LIVMY]-x_2-[GATSCF]-x_2-[LIVGATF]-x_2-[LIVCAT]-x_2-[LIVFMA]-[LIVMGPSA]-x-[LIVMCAG]-[LIVCATF]-[NVPATSG]
4	OPA	R-[GI]-x_3-[GAST]-[LIVFA]-W-[NS]-x-[SACN]-[HAG]-[NS]-[LIVT]-G-[GN]-[GAI]
5	OHS	D-[KR]-L-G-L-[RK]-K-x-L-[LIV]-W-x_2-[ST]-x_2-L-[LIV]-x_2-[AG]-P-F
6	MHS	P-[LIV]-G-[GAS]-x-[LIVF]-[LIVFM]-G-x_3-D-[RK]-x-G-R
7	FGHS	A-[GAP]-x_4-[RK]-x-S-x_3-[GT]-[LIVT]$_3$-[GA]-L-x-[LIV]
8	NNP	[FYK]-x_3-[ILQRK]-x-[GA]-x-[VASK]-x-[GASN]-[LIVFQ]-$x_{1,2}$-G-x-G-[NIM]-x-G-[GVTA]
9	PHS	N-x-G-P-x_5-[LIV]-[LIVY]-[PS]-[AGS]-E-x-[FS]-[PA]-[AT]-x_2-R-[SG]-x_3-G-[LIV]-S-A-x_2-G-K-x-G
10	NHS	L-G-L-D-A-F-x-L-F-K-N
11	OFA	Y-x_6-[AC]-x-[KR]-W-[FY]-[PQ]-[DE]-x_2-G-L-x-[STG]-[GA]-x_4-[GS]-[FY]-G-x-G
12	SHS	E-x-W-P-x_5-K-A-x_6-G-x_2-V-G-x_3-A-A
13	MCP	G-[LIVCF]-[GS]-x-[ASG]-x_4-[PA]-x-[LIVT]-x_5-[YW]-F-x_2-[RKL]$_2$-x-[LIVTM]-A-x-[GAS]-x_4-G
14	ACS[c]	
15	UMF	G-[LIVM]-[LIV]-L-[LIVFM]-[TG]-x_4-[LIVC]-[LIV]-L-[LIV]-P
16	AAHS	S-[FIA]-[GAT]-[LIV]-x-G-x_3-G-[AG]-[LIVF]-x_2-[GS]-x-[ALIT]-[AC]-D-[RK]
17	CP	R-[TS]-[LV]-x_3-L-[LV]-[LI]$_2$-x_2-G-x_3-R
18	POT	[LIVT]-[LIVGAS]-[LIVFCA]-[GA]-[TASNY]-G-[GLM]-[LIVMF]-[KRN]-[APS]-[SCN]-[LIVP]-[LIVSA]-X-[LIVMF]-[LIVMAG]

[a] The "signature sequence" of the SP family (family 1) is based on those representative proteins depicted in Fig. 3. It is therefore not a true signature sequence.
[b] The signature sequences of the DHA14 and DHA12 families are based on the multiple-sequence alignments published by Paulsen et al. (60).
[c] No family-specific signature sequence could be derived for the ACS family.

TABLE 19. Common sequence motifs localized between TMS2 and TMS3 in the 17 families of the MFS[a]

Family		Predominant residue at position:												
No.	Name	1 G	2 [RKPATY]	3 L	4 [GAS]	5 [DN]	6 [RK]	7 [FY]	8 G	9 R	10 [RK]	11 [RKP]	12 [LIVGST]	13 [LIM]
1	SP	G	X	L	G	N	R	F	G	R	R	X	X	L
2	DHA14	G	R	L	A	D	R	F	G	R	K	R	X	L
3	DHA12	G	X	L	S	D	R	F	G	R	R	P	V	L
4	OPA	G	T	L	G	D	H	Y	N	P	R	R	X	L
5	OHS	G	L	L	S	D	K	L	G	L	R	K	H	L
6	MHS	G	A	Y	G	D	R	Y	G	R	K	K	G	L
7	FGHS	K	R	Y	S	Y	K	A	G	I	X	X	G	L
8	NNP	G	P	L	T	D	R	F	G	P	R	X	X	X
9	PHS	V	A	F	I	D	T	I	G	R	K	P	I	Q
10	NHS[b]	G	I	I	A	V	Q	M	A	R	R	T	C	I
11	OFA	N	K	W	F	D	K	R	G	L	A	S	G	I
12	SHS	G	A	M	A	D	K	Y	G	R	K	P	V	M
13	MCP	S	X	L	V	N	K	Y	G	S	R	P	V	M
14	ACS	G	Y	L	L	D	K	K	G	A	K	K	V	I
15	UMF	A	R	L	S	D	I	F	G	R	L	X	L	F
16	AAHS	G	P	L	A	D	R	F	G	R	K	R	V	L
17	CP	X	X	L	X	Q	R	X	G	X	E	R	S	L
Σ[c]		11	11	11	11	14	13	11	15	9	13	11	11	14

[a] The residue indicated is the one present in the largest number of proteins in the indicated family at the position specified. In some cases two similar residues (e.g., S and T, or R and K) were present with the same frequency. One of these residues was then arbitrarily selected. An X indicates that no single residue predominated at that position.
[b] Only one of the two members of this family was used to provide the sequence shown. The other protein exhibited gaps in the region of this motif.
[c] The numbers of families that exhibit the residue(s) indicated in the "consensus" sequence at the top of the table are indicated.

bers of this permease family to members of the MFS.

The current POT family includes 34 proteins derived from bacteria, yeast, fungi, plants and animals (74a). In addition to transporting peptides of two to four residues, some have been shown to transport individual amino acids. One plant protein transports nitrate and chlorate, and another transports nitrite.

The POT family (TC 2.17) was found to exhibit sequence similarity to selected members of MFS families 1 (SP) and 2 (DHA14) (maximal optimized comparison score of 8 SD for an aligned segment of 70 amino acyl residues). This value is sufficient to strongly suggest that the POT family is a member of the MFS. However, the shortness of the segments compared

TABLE 20. Motif characterization exemplified by proteins of the PHS family (family 9)[a]

Sequence name[b]	Residues compared[c]	Motif diagram[d]	Motif B[e]	Motif D[e]
PT1 Stu (a)	1–283	26-[A]--52--[B]-42-[C]--63--[D]-38	DKMGRKKVY	MPETARYTALV
PT2 Ath (a)	1–278	27-[A]--52--[B]-41-[C]--63--[D]-33	DKLGRKKVY	MPETARYTALV
AtPT2 Ath (a)	1–269	36-[A]--45--[B]-42-[C]--53--[D]-31	DRLGRKRMY	IPETPRYTMDV
GuPT Gve (a)	1–282	36-[A]--45--[B]-42-[C]--53--[D]-44	DRLGRKRMY	IPETPRYTMDV
PT2 Stu (a)	1–283	29-[A]--53--[B]-42-[C]--63--[D]-34	AKMGRKKVY	MPETARYTALV
AtPT1 Ath (a)	1–281	27-[A]--52--[B]-42-[C]--63--[D]-35	DKLGRKKVY	MPETARYTALV
Pho84 Sce (a)	1–317	68-[A]--45--[B]-42-[C]--63--[D]-37	DIVGRKRIY	IPESPRYQLDV
Pho-5+ Ncr (a)	1–303	53-[A]--97---------[C]--63--[D]-37	DKMWRTVIG	IPETPRYTFDV
PT1 Stu (b)	284–540	20-[A′]-54--[B]-41-[C′]-57--[D]-27	DRIGRFAIQ	VPESKGKSLEE
PT2 Ath (b)	279–523	23-[A′]-54--[B]-41-[C′]-57--[D]-12	DTIGRFKIQ	VPEPKGKSLEE
PT2 Stu (b)	284–527	23-[A′]-54--[B]-41-[C′]-57--[D]-11	DIIGRFAIQ	VPESKGKSLEE
Pho84 Sce (b)	318–587	35-[A′]-51--[B]-37-[C′]-54--[D]-35	DIIGRKPIQ	IPETKRKTLEE
AtPT1 Ath (b)	282–534	22-[A′]-54--[B]-41-[C′]-57--[D]-21	DVIGRFAIQ	VPESKGKSLEE
GuPT Gve (b)	283–521	15-[A′]-57--[B]-40-[C′]-48--[D]-21	DSWGRKPIQ	IPETKGLSLEE
AtPT2 Ath (b)	270–521	28-[A′]-57--[B]-40-[C′]-48--[D]-21	DSWGRKPIQ	IPETKGLSLEE
Pho-5+ Ncr (b)	304–569	30-[A′]-53--[B]-37-[C′]-53--[D]-35	DTVGRKPIQ	IPETKRKTLEE
		Multilevel consensus sequence	DKIGRKKIQ	IPETKRYTLEE
			L FP Y	V AGKSADV
				M P L

[a] The MEME and MAST programs (5) were used for motif identification.
[b] Abbreviation of the protein (Table 9) and an indication of whether the N-terminal half (a) or the C-terminal half (b) was examined.
[c] Actual residues in each protein segment analyzed.
[d] Spacing (in numbers of amino acyl residues) between the various motifs detected. These common motifs were as follows: motif A, AGMGFFTDAYDLF; motif A′, SWFLLDIAFY; motif B, DKWGRKKIY; motif C, DYPLSATIMSEYANKKWRGAMMAAVFAMQ; motif C′, NFGPNATTFIVPGEIFPTRYRSTCHGIS; motif D, MPETKRYTLEE.
[e] Motifs B and D as they occur in each of the protein segments examined. The multilevel motif consensus sequences for motifs B and D are provided below the motif B and D sequences, respectively.

TABLE 21. Systematic classification of MFS permeases

TC no.	Name or description[a]	Source	Organism	Accession no.	Database[b]
2.1.1	SP family (family 1)				
2.1.1.1	Galactose:H^+ symporter	Bacteria	GalP of *Escherichia coli*	P37021	SP
2.1.1.2	Arabinose:H^+ symporter	Bacteria	AraE of *Escherichia coli*	P09830	SP
2.1.1.3	Xylose:H^+ symporter	Bacteria	XylE of *Escherichia coli*	P09098	SP
2.1.1.4	Glucose uniporter	Bacteria	Glf of *Zymomonas mobilis*	P21906	SP
2.1.1.5	Hexose uniporter	Yeasts	HxtO of *Saccharomyces cerevisiae*	P43581	SP
2.1.1.6	Galactose uniporter	Yeasts	Gal2 of *Saccharomyces cerevisiae*	P13181	SP
2.1.1.7	Quinate:H^+ symporter	Fungi	Qay of *Neurospora crassa*	P11636	SP
2.1.1.8	Myoinositol:H^+ symporter	Yeasts	ITR1 of *Saccharomyces cerevisiae*	P30605	SP
2.1.1.9	Lactose:H^+ symporter	Yeasts	LacP of *Kluyveromyces lactis*	P07921	SP
2.1.1.10	Maltose:H^+ symporter	Yeasts	MAL6 of *Saccharomyces cerevisiae*	P15685	SP
2.1.1.11	α-Glucoside:H^+ symporter	Yeasts	αGlc permease of *Saccharomyces cerevisiae*	S59368	PIR
2.1.1.12	Glucose uniporter	Animals	Gtr3 of *Rattus norvegicus*	Q07647	SP
2.1.1.13	Fructose uniporter	Animals	Ftr of *Homo sapiens*	U11843	GB
2.1.1.14	Hexose:H^+ symporter	Plants	Hup1 of *Chlorella kessleri*	P15686	SP
2.1.1.15	Synaptic vesicle neurotransmitter transporter	Animals	SYV2 of *Rattus norvegicus*	Q02563	SP
2.1.1.16	Organic cation transporter	Animals	Oca of *Rattus norvegicus*	X78855	GB
2.1.1.17	Glucose transporter	Protists	Th2A of *Trypanosoma brucei*	Q06222	SP
2.1.2	DHA14 family (family 2)				
2.1.2.1	Actinorhordin	Gram-positive bacteria	ActVa of *Streptomyces coelicolor*	X58833	GB
2.1.2.2	Cephamycin	Gram-positive bacteria	CmcT of *Nocardia lactamdurans*	Q04733	SP
2.1.2.3	Lincomycin	Gram-positive bacteria	LmrA of *Streptomyces lincolnensis*	X59926	GB
2.1.2.4	Methylenomycin	Gram-positive bacteria	MmrB of *Bacillus subtilis*	Q00538	SP
2.1.2.5	Puromycin	Gram-positive bacteria	Pur8 of *Streptomyces lipmanii*	X76855	GB
2.1.2.6	Tetracenomycin	Gram-positive bacteria	TemA of *Streptomyces glaucescens*	M80674	GB
2.1.2.7	Bicyclomycin, sulfathiazole, etc.	Gram-negative bacteria	Bcr of *Escherichia coli*	JN0659	PIR
2.1.2.8	Fluoroquinolones, acriflavin, chloramphenicol, ethidium bromide, etc.	Gram-positive bacteria	Blt of *Bacillus subtilis*	L32599	GB
2.1.2.9	Hydrophobic uncouplers (e.g., CCCP)	Gram-negative bacteria	EmrD of *Escherichia coli*	P31442	SP
2.1.2.10	Daunomycin, ethidium bromide	Gram-positive bacteria	LmrP of *Lactococcus lactis*	X89779	GB
2.1.2.11	Benomyl, cycloheximide, methotrexate, etc.	Yeasts	CaMDR1 of *Candida albicans*	P28873	SP
2.1.2.12	Doxorubicin, ethidium bromide, rhodamine-6-G	Mammals	VMAT1 of *Rattus norvegicus*	M97380	GB
2.1.2.13	Amiloride	Yeasts	Car1 of *Schizosaccharomyces cerevisiae*	P33532	SP
2.1.2.14	Cycloheximide	Yeasts	CyhR of *Candida maltosa*	P32071	SP
2.1.2.15	Chloramphenicol	Bacteria	CmlA of *Pseudomonas aeruginosa*	P32482	SP
2.1.2.16	Tetracycline	Bacteria	TetA of *Escherichia coli*	X00006	GB
2.1.2.17	Acetylcholine	Animals	Unc17 of *Caenorhabditis elegans*	P34711	SP
2.1.3	DHA12 family (family 3)				
2.1.3.1	Aminotriazole, 4-nitroquinoline-*N*-oxide	Yeasts	Atr1 of *Saccharomyces cerevisiae*	Z49210	GB
2.1.3.2	CCCP, nalidixic acid, organomercurials, etc.	Gram-negative bacteria	EmrB of *Escherichia coli*	P27304	SP
2.1.3.3	Acriflavin, ethidium bromide, fluoroquinolones	Gram-positive bacteria	LfrA of *Mycobacterium smegmatis*	U40487	GB
2.1.3.4	Mono- and divalent organocations	Gram-positive bacteria	QacA of *Staphylococcus aureus*	X56628	GB
2.1.3.5	Pristinamycin I and II, rifamycin	Gram-positive bacteria	Ptr of *Streptomyces pristinaespiralis*	X84072	GB
2.1.3.6	Tetracycline	Bacteria	TetK of *Staphylococcus aureus*	M16217	GB
2.1.4	OPA family (family 4)				
2.1.4.1	Hexose-phosphate:P_i antiporter	Bacteria	UhpC of *Escherichia coli*	P09836	SP
2.1.4.2	Phosphoglycerate:P_i antiporter	Bacteria	PgtP of *Salmonella typhimurium*	P12681	SP
2.1.4.3	Glycerol-phosphate:P_i antiporter	Bacteria	GlpT of *Escherichia coli*	P08194	SP
2.1.5	OHS family (family 5)				
2.1.5.1	Lactose:H^+ symporter	Bacteria	LacY of *Escherichia coli*	P02920	SP
2.1.5.2	Raffinose:H^+ symporter	Bacteria	RafB of *Escherichia coli*	P16552	SP
2.1.5.3	Sucrose:H^+ symporter	Bacteria	CscB of *Escherichia coli*	P30000	SP
2.1.6	MHS family (family 6)				
2.1.6.1	Citrate:H^+ symporter	Bacteria	Cit of *Escherichia coli*	P07860	SP
2.1.6.2	α-Ketoglutarate:H^+ symporter	Bacteria	KgtP of *Escherichia coli*	P17448	SP
2.1.6.3	Dicarboxylate:H^+ symporter	Bacteria	PcaT of *Pseudomonas putida*	U48776	GB
2.1.6.4	(Proline/betaine):(H^+/Na^+) symporter	Bacteria	ProP of *Escherichia coli*	P30848	SP
2.1.6.5	4-Methyl-*o*-phthalate:H^+ symporter	Bacteria	MopB of *Burkholderia cepacia*	U29532	GB
2.1.7	FHS family (family 7)				
2.1.7.1	L-Fucose:H^+ symporter	Bacteria	FucP of *Escherichia coli*	P11551	SP
2.1.7.2	Glucose/galactose permease	Bacteria	Ggp of *Brucella abortus*	U43785	GB
2.1.8	NNP family (family 8)				
2.1.8.1	Nitrite extrusion permease	Bacteria	NarK of *Escherichia coli*	P10903	SP
2.1.8.2	Nitrate uptake permease	Bacteria	NasA of *Bacillus subtilis*	P42432	SP

Continued on following page

TABLE 21—*Continued*

TC no.	Name or description[a]	Source	Organism	Accession no.	Database[b]
2.1.9	PHS family (family 9)				
2.1.9.1	P_i uptake permease	Yeasts	Ph84 of *Saccharomyces cerevisiae*	P25297	SP
2.1.9.2	P_i uptake permease	Fungi	Pho-5 of *Neurospora crassa*	L36127	GB
2.1.9.3	P_i uptake permease	Plants	PT1 of *Solanum tuberosum*	X98890	GB
2.1.10	NHS family (family 10)				
2.1.10.1	Nucleoside permease	Bacteria	NupG of *Escherichia coli*	P09452	SP
2.1.10.2	Xanthosine permease	Bacteria	XapB of *Escherichia coli*	P45562	SP
2.1.11	OFA family (family 11)				
2.1.11.1	The oxalate:formate antiporter	Bacteria	OxlT of *Escherichia coli*	U40075	GB
2.1.12	SHS family (family 12)				
2.1.12.1	The sialic acid permease	Bacteria	NanT of *Escherichia coli*	P41036	SP
2.1.13	MCP family (family 13)				
2.1.13.1	The monocarboxylate (lactate, pyruvate, mevalonate) uptake/efflux permease	Animals, yeasts, fungi	Mot-1 of *Homo sapiens*	A55568	PIR
2.1.14	ACS family (family 14)				
2.1.14.1	Glucarate permease	Bacteria	GudT of *Bacillus subtilis*	P42237	SP
2.1.14.2	Hexuronate permease	Bacteria	ExuT of *Escherichia coli*	P42609	SP
2.1.14.3	Putative tartrate permease	Bacteria	TtuB of *Agrobacterium vitis*	U32375	GB
2.1.14.4	Allantoate permease	Yeasts	Dal5 of *Saccharomyces cerevisiae*	P15365	SP
2.1.14.5	Phthalate permease	Bacteria	Pht1 of *Pseudomonas putida*	Q05181	SP
2.1.14.6	Na:P_i symporter	Animals	Npt1 of *Mus musculus*	X77241	GB
2.1.15	AAS family (family 15)				
2.1.15.1	4-Hydroxybenzoate (protocatachuate) permease	Bacteria	PcaK of *Pseudomonas putida*	U10895	GB
2.1.15.2	3-Hydroxyphenyl propionate permease	Bacteria	MhpT of *Escherichia coli*	X97543	GB
2.1.15.3	2,4-Dichlorophenoxy-acetate permease	Bacteria	TfdK of *Ralstonia eutropha*	U16782	GB
2.1.16	UMF family (family 16)				
2.1.16.1	Permease of unknown specificity	Yeasts	Ykr106w of *Saccharomyces cerevisiae*	P36173	SP
2.1.17	CP family (family 17)				
2.1.17.1	Cyanate transport system	Bacteria	CynX of *Escherichia coli*	P17583	SP

[a] CCCP, carbonyl cyanide *m*-chlorophenylhydrazone; P_i, inorganic phosphate.
[b] See Table 3, footnote *b*.

and the fact that only a few currently sequenced MFS proteins exhibit significant similarity to POT family members lead us to retain the POT family as a distinct family, at least at present. A signature sequence for the POT family is provided in Table 18.

MFS FAMILY-SPECIFIC SIGNATURE SEQUENCES

Using the portions of the multiple-sequence alignments shown in Fig. 3 to 17, we attempted to derive valid signature sequences for all of the 17 constituent families of the MFS. We were successful except in the case of the very diverse family 14 (ACS) and the large family 1 (SP), for which a complete multiple-sequence alignment that includes all 133 members is not available. A potential signature sequence for family 1 (SP) which encompasses the proteins included in the alignment shown in Fig. 3A was, however, derived. This and other MFS family-specific signature sequences are presented in Table 18. They should be useful in identifying and characterizing new members of these MFS families as they become sequenced.

MFS-SPECIFIC SEQUENCE MOTIF

In 1990, Henderson and Maiden noted a sequence motif common to several of the MFS permeases then available for study (36). This five-residue motif (RXGRR) occurred between TMS2 and TMS3. The motif was suggested to form a β-turn linking the adjacent transmembrane helices, and the cationic residues were suggested to interact with negative charges in membrane lipid head groups. Subsequent studies revealed that this motif could be identified even in MFS members that exhibit extensive sequence divergence (30). A similar but less well conserved motif was found in the second half of the protein, between TMS8 and TMS9.

Jessen-Marshall et al. (39) examined the functional significance of an expanded form of this motif [GX_3(D/E)(R/K)XG[X](R/K)(R/K)], an 11-residue motif, by using the lactose permease of *E. coli* as a representative model system. Comparable studies on the same motif in the Tn*10*-encoded Tet carrier of *E. coli* have also been reported (87–89). The work on this metal-tetracycline:H^+ antiporter of *E. coli* suggested that the negative charge at position 5, at least one of the basic residues, and both glycines at positions 1 and 8 are important for function. Jessen-Marshall et al. determined the phenotypes of 28 mutants with site-specific mutations in LacY in which this motif was altered. Their studies led to the conclusions that small side chain volume at specific positions and high β-turn propensity are probably of structural importance within this motif. Additionally, the acidic residue (D) at position 5 in the above-cited motif proved to be important for transport activity, in agreement with the results of Yamaguchi et al. (87–89). Although replacement of the glycine at position 8, or of any one of the basic residues within the motif, frequently had little or no pronounced effect on transport, any

two such replacements resulted in reduced transport rates. Taken together, the results suggested that the motif is important for both structural and functional aspects of the lactose permease. It was suggested that the motif may be important in promoting global conformational changes of the permease that accompany transport.

Since this motif appears to be one of the most strongly conserved features of previously characterized MFS proteins, we were interested to determine if it could be identified between TMS2 and TMS3 in all or most of the MFS families characterized here. The complete multiple-sequence alignments of the proteins which comprise the 17 MFS families were therefore examined to attempt an identification of comparable motifs between TMS2 and TMS3. In these analyses, the single residue which was present in the largest number of the proteins which comprise the family was recorded, even when that residue was not present in a majority of the proteins. When no residue predominated, an X was recorded at that position. The alignment used for the SP family (family 1) included the representative members examined in Fig. 3, but for all the other families, all (or almost all) of the protein members were studied. The results are recorded in Table 19. Thirteen positions were included in the analysis, and in all cases the residues recorded represent those that occur sequentially (without gaps or insertions). In one case, the NHS family (family 10), just one of the two members of the family was used to derive the motif since gaps occurred in this region of the other family member.

The consensus motif for this 13-residue sequence is G-[RK PATY]-L-[GAS]-[DN]-[RK]-[FY]-G-R-[RK]-[RKP]-[LIVGST]-[LIM]. Within this motif, four positions are specified by a single amino acid (G-1, L-3, G-8, and R-9), four positions are specified by a pair of closely related residues (DN-5, RK-6, FY-7, and RK-10), three positions are specified by three possible residues (GAS-4, RKP-11, and LIM-13), and two positions are more degenerate (RKPATY-2 and LIVGST-12). At the bottom of Table 19, the number of families that conform to the consensus signature sequence at each of the 13 positions is indicated. The G at position 8 is in 15 of the 17 families, DN at position 5 and LIM at position 13 are found in 14 of the families, RK at positions 6 and 10 are found in 13 of the families, and R at position 9 is found in 9 of the families. The seven other positions coincidentally show 11 families that conform to the expected residue or set of residues. For single-residue conservation, the order is as follows: G-8 (15) > D-5 (12) > G-1 (11) and L-3 (11) > R-9 (9) and L-13 (9) > R-6 (7), F-7 (7) and R-10 (7), where the numbers in parentheses indicate the number of families. Thus, the degree of conservation is striking.

EVIDENCE FOR REPEAT SEQUENCES IN MFS PROTEINS

Henderson and coworkers have noted the presence of shared sequence motifs in the first and second halves of MFS sugar porters (family 1) (30, 36). Rubin et al. (70) presented convincing statistical evidence that tetracycline efflux proteins of the MFS (family 3) exhibit sufficient sequence similarity in their first and second halves to establish homology on statistical grounds. Using the MEME and MAST programs (5), we have found further evidence that proteins within several of the MFS families consist of duplicated six-TMS units. The evidence will be presented here for one of these families, the PHS family (family 9).

The output from the MEME and MAST programs for the two halves of eight members of the PHS family is presented in Table 20. The protein abbreviations recorded in Table 9, as well as the residues that comprise the two halves of these proteins, are given. The programs identified six motifs (motifs A, A′, B, C, C′, and D). Motifs A, B, C, and D were found in the first halves of these proteins, while motifs A′, B, C′, and D were identified in the second halves of these proteins. The spacing between motifs A and B was the same as that between motifs A′ and B, that between motifs B and C was the same as that between motifs B and C′, and that between motifs C and D was the same as that between motifs C′ and D for the first and second halves of these proteins, respectively (Table 20). Motifs A and A′ as well as motifs C and C′ were found to exhibit minimal sequence similarity, even though they occurred in corresponding parts of the two halves of these proteins. However, motifs B and D proved to be essentially the same in both halves of these proteins (Table 20). Consensus motif B is DKIGRKKIQ, while consensus motif D is IPETKRYTLEE. The occurrence of these motifs has been well documented in previous publications concerned with sugar porters (family 1) (30). Table 20 presents the alignments of these two motifs in the eight proteins examined for both the first and second halves of these proteins. The degree of sequence conservation is striking.

When regions of the two halves of these proteins were compared with the GAP program (16), comparison scores of 8 to 10 SD were often observed. For example, PT1 Stu of family 9 (PHS; Table 9) exhibited a comparison score of 10 SD with 26.4% identity when residues 68 to 121 were compared with residues 344 to 397. Sometimes heterologous protein comparisons also gave high comparison scores. Thus, when residues 29 to 94 of LacY Eco of family 5 (OHS; Table 5) were compared with residues 281 to 346 of ProP Eco of family 6 (MHS; Table 6), a comparison score of 13 SD (34.8% identity) was obtained. These values clearly demonstrate a significant degree of sequence similarity that is strongly suggestive of homology. In agreement with Rubin et al. (70), we therefore conclude that proteins of the MFS arose by an internal gene duplication event which probably occurred prior to divergence of the MFS families.

CLASSIFICATION OF MFS PERMEASES

We have recently formed a "transport commission" (TC) with the primary goal of classifying all transmembrane transport systems. These systems are classified according to four criteria as follows: W, transport mode and energy-coupling mechanism; X, family or superfamily; Y, subfamily or family, respectively; Z, substrate specificity. Every transporter family or superfamily thus has a two-digit TC number (W.X). The TC number for the MFS is 2.1 (Table 21). Each of the 17 families within the MFS has been assigned a three-digit TC number (W.X.Y). Thus, the SP family has the TC number 2.1.1; the 14-TMS drug:H^+ antiporter (DHA14) family has the TC number 2.1.2, and the CP family has the TC number 2.1.17 (Table 21). Because we were not able to satisfactorily establish homology of POT family permeases to established MFS proteins, the POT family is not included within the MFS and has its own TC number (TC 2.17).

Each permease of dissimilar function within each MFS family is given its own four-digit TC number (W.X.Y.Z), but more than one ortholog of the same function in different organisms, more than one paralog of the same function in a single organism, and functionally uncharacterized permeases are not included if they are within the same MFS family. Thus, all orthologous galactose:H^+ symporters in the SP family have the TC number 2.1.1.1, while all arabinose:H^+ symporters in the

SP family have the TC number 2.1.1.2 (Table 21). The information provided in the TC tables includes the TC number for a particular permease, the name of that permease, its biological sources, and an example of such a permease. The accession number of the permease chosen to exemplify a particular entry is provided, allowing easy access to its sequence. A BLAST search (2) should allow the identification of other orthologs. A more complete description of the TC classification system, its rationale, and its benefits are described elsewhere (73).

CONCLUSIONS AND PERSPECTIVES

In this computational analysis, we have attempted to identify all recognizable sequenced members of the MFS that have been deposited in the current databases. We have used sequence similarity as a basis for inclusion of proteins in the superfamily, with a cutoff point of 8 to 9 SD for the comparison score of a test sequence with that of an established member of the family (18, 71). Two families, the FGHS family (family 7) and the POT family (putative family 18), gave optimized comparison scores of 8 SD for segments of these permeases that are in excess of 100 and 60 residues, respectively (Table 2). These families are likely to be constituents of the MFS, but their assignments to the superfamily are more tenuous than for those of the other families. This is particularly true of the POT family, where the segment exhibiting sequence similarity to an established MFS member is short. All the other families included members that gave comparison scores with an established member of the superfamily of 10 SD or more. Furthermore, these scores were generated with large segments of the compared proteins, and usually the entirety of the sequences was compared. We are therefore confident that all the families listed in Table 1 (with the possible exceptions of the FGHS and POT families) are members of the MFS.

Sixteen to eighteen MFS families were identified and generally shown to be restricted to a specific type of substrate. Thus, functionally characterized members of families 1, 5, and 7 are almost without exception specific for sugars; characterized members of families 2 and 3 are without exception specific for drugs and other deleterious substances; and families 4, 6, 8, 9, 11 to 14, and 17 are specific for various classes of anionic compounds. Furthermore, the only nucleoside permeases in the MFS are found in family 10, and most of the aromatic acid permeases are found in family 15. These observations clearly show that substrate specificity is a well-conserved trait and that phylogenetic classification provides a limited but reliable guide to function.

Similar considerations can be applied to pump polarity. Thus, while members of families 1, 5, 7, and 8 can apparently function quite readily by one or more modes (e.g., uniport, symport with inwardly directed polarity, and/or antiport), families 2, 3, 4, and 11 apparently function with a high propensity for an antiport mechanism, and families 6, 9, 10, and 12 to 15 probably function with a high propensity for a cation symport mechanism. Clearly, these mechanistic differences must reflect structural and catalytic residue differences, regardless of whether they reflect qualitative or quantitative differences. The molecular bases for these differences should be subject to biochemical, biophysical, and molecular genetic analyses.

The phylogenetic tree in which representative members of each MFS family were included (Fig. 2) revealed that the major families diverged from each other long ago, possibly more than 2 billion years ago. Moreover, all currently recognized, topologically studied members of the MFS possess either 12 or 14 putative TMSs. We predict that the 14-TMS topology arose from the 12-TMS topology more than once during the evolution of the MFS. These considerations suggest that the MFS is one of the oldest protein families on Earth and that MFS proteins were present more than 3 billion years ago in their present form, probably with 12 TMSs. The age of the MFS, as well as currently unrecognized architectural features of these proteins, can account in part for its functional and sequence diversity (73).

During our analyses of transport protein families, we have noted that some families have diversified extensively while other families have not. Thus, while proteins of the MFS (TC 2.1) and the ABC superfamily (TC 3.1) have evolved to transport almost any substrate of biological interest, other ancient families have apparently not done so. For example, the ammonium transporter (Amt) family (TC 2.49), which includes the ammonium carrier of *Corynebacterium glutamicum* (77), consists of homologous proteins found ubiquitously in bacteria, archaea, and eukarya, but all functionally characterized members of the family transport NH_4^+ (73). Many other such examples exist. Why one ancient family has been capable of evolving functional diversity while another has not is not presently understood.

As noted above, the MFS transports a wide array of substrates with either inwardly or outwardly directed polarity or without polarity (Table 1). These substrates include sugars, drugs, an array of metabolites, amino acids, nucleosides, vitamins, and both inorganic and organic anions and cations. Such compounds are also transported via ABC-type permeases, but the latter family of permeases can also transport macromolecules such as proteins, complex carbohydrates, and intact phospholipids. Of the several hundred MFS carriers so far identified, none has yet been found to be capable of transporting a macromolecule. Some do transport oligosaccharides and peptides, but the sizes of these polymers appear to be limited to about 4 amino acyl or sugar residues. The basis for this fundamental functional difference between ABC and MFS permeases is not yet understood, but it presumably reflects the dimensions and flexible associations of the channel-forming α-helices. Perhaps ABC-type channels can "breathe" and expand or even dissociate into constituent helices whereas MFS channels cannot. Perhaps the channels of MFS permeases are bounded by the helices of a single domain whereas ABC permease channels are formed from dissociable domains. We suggest that very basic architectural features of these two classes of permeases differ and provide an explanation for their distinctive functional characteristics.

As summarized in Table 1, several MFS families are restricted to certain types of organisms. For example, families 6 (NHS) and 12 (SHS) are restricted to gram-negative bacteria while family 16 is restricted to yeasts. It is interesting to ask whether this restriction is a true reflection of the distribution of these proteins in nature or whether it merely reflects the need for more extensive sequence data. Examination of the data summarized in Table 1 reveals that within limits, the largest families are represented in the largest numbers of phyla while the smallest families are restricted to the smallest numbers of phyla. Thus, families 1, 3, 14, and 18 are represented in four or more of the indicated phyla; except for family 2, these are the largest MFS families. Similarly, families 5, 7, 9, 10, 12, 15, 16, and 17 are found in only one or two phyla; except for family 11, these are the smallest families. We therefore anticipate that most of the MFS families will eventually prove to be ubiquitous. This suggestion is consistent with our proposal that most of the individual MFS families diverged from each other more than 2 billion years ago, before eukarya and archaea diverged from bacteria. It should be noted, however, that we have tacitly

assumed that branch length is approximately proportional to phylogenetic distance. Branch length is, in fact, a true reflection of sequence divergence and therefore a reflection of both the time and rate of divergence. If evolutionary pressure for sequence divergence is variable over evolutionary time, a "phylogenetic" tree can be misleading. While we believe that the MFS tree is in general reflective of evolutionary time (i.e., that the rate of evolutionary divergence was fairly constant for the different families), exceptions may occur. Thus, family 16 (UMF) includes six closely related proteins, all from *S. cerevisiae*, and no member of this family has yet been found in another organism. Since none of these proteins has been functionally characterized, it is possible that they have evolved to serve a yeast-specific function and that the six paralogous members of the family arose relatively recently due to gene duplication events that occurred in yeasts. These proteins may serve a transport function that is specific to yeasts, a function perhaps involving mating, budding, or differentiation.

A final question arises, i.e., whether additional MFS families are likely to be found. Families with greater sequence divergence than those currently recognized (e.g., those with members that give less than 8 SD in comparison scores with members of known MFS families) and families not currently represented in the databases (e.g., those that will eventually prove to be the smallest MFS families) will undoubtedly be discovered as additional sequence data become available. However, when we screened the sequence motif presented at the top of Table 19 against the SwissProt database with 0, 1, 2, 3, or 4 mismatches, the only transport proteins that were retrieved were recognized members of the MFS already listed in Tables 3 to 17. This observation leads us to suggest that if additional MFS proteins are in the current databases, they have diverged in sequence to such an extent that they are unrecognizable on the basis of primary structure alone. Three-dimensional structural data on numerous secondary carriers, including both MFS and non-MFS permeases, may be required to resolve this question.

ACKNOWLEDGMENTS

We thank Tim Bailey, Charles Elkan, Jayna Ditty, André Goffeau, Caroline Harwood, Gary Kuan, Ellen Neidle, Jonathan Reizer, and Marek Sliwinski for valuable discussions. We also thank Mary Beth Hiller, Lyn Alkan, and Milda Simonaitis for assistance in the preparation of the manuscript.

Work in our laboratory was supported by USPHS grants 2RO1 AI14176 from the National Institute of Allergy and Infectious Diseases and 2RO1 GM55434 from the National Institute of General Medical Science (to M.H.S.). I.T.P. was supported by a C. J. Martin Fellowship from the National Health and Medical Council of Australia.

ADDENDUM IN PROOF

H. Huel, S. Turgut, K. Schmid, and J. W. Lengeler (J. Bacteriol. **179:**6014–6019, 1997) have recently reported the sequences of the D-arabinitol:H^+ and ribitol:H^+ symport permeases of *Klebsiella pneumoniae* (DalT and RbtT, respectively). These two proteins are 86% identical and are 425 and 427 aminoacyl residues long, respectively, both with 12 putative TMSs. We have conducted phylogenetic analyses of these two polyol permeases and have found that they, together with an uncharacterized protein encoded within the *Bacillus subtilis* genome, comprise a novel MFS family which we have termed the polyol permease (PP) family (family 18) (T.-T. Tseng and M. H. Saier, Jr., unpublished observations). The proteins of the PP family exhibit an approximation to the MFS-specific sequence motif between TMSs 2 and 3 (Table 19) of GVVAEIIG PRKTM, thus showing poor correspondence to the N-terminal half of this MFS-specific motif but excellent correspondence to the C-terminal half. Binary comparison of DalT with KgtP Eco (Table 6) gave a comparison score of 10.5 standard deviations for a segment of 107 residues (21% identity, 49% similarity, 0 gaps). This value is sufficient to establish that the three proteins of the PP family are members of the MFS (T.-T. Tseng and M. H. Saier, Jr., unpublished results). The three proteins of the PP family also exhibit recognizable sequence similarity to members of several other MFS permease families. The following two signature sequences proved to be specific to the PP family: Y(A/G)(L/I/V)RGX(A/G)YPLFXYSF(L/I/V)V and $GEX_2TLWXALXFX_3GG(L/I/V)_2AL$ (X is any residue). By hybrid protein construction, Heuel et al. demonstrated that the substrate specificities and kinetic properties for transport of DalT and RbtT are determined by the amino-terminal halves of the proteins. This result contrasts with those reported for the lactose permease of *E. coli* (LacY; TC 2.1.5.1) in which substrate specificity appears to be determined primarily by residues in the carboxy-terminal half of the protein.

In recent work, M. J. Whipp, H. Camakaris, and A. J. Pittard have cloned and analyzed the *shiA* gene, which encodes the shikimate transport system of *E. coli* K-12 (SwissProt accession no. P76350; 438 amino acids) (Gene, in press). This permease proved to be a member of the metabolite:H^+ symporter (MHS) family (family 6) of the MFS.

REFERENCES

1. **Abe, K., Z. S. Ruan, and P. C. Maloney.** 1996. Cloning, sequencing, and expression in *Escherichia coli* of OxlT, the oxalate:formate exchange protein of *Oxalobacter formigenes*. J. Biol. Chem. **271:**6789–6793.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Ambudkar, S. V., and P. C. Maloney.** 1984. Characterization of phosphate: hexose 6-phosphate antiport in membrane vesicles of *Streptococcus lactis*. J. Biol. Chem. **259:**12576–12585.
4. **Anantharam, V., M. J. Allison, and P. C. Maloney.** 1989. Oxalate:formate exchange: the basis for energy coupling in *Oxalobacter*. J. Biol. Chem. **264:**7244–7250.
5. **Bailey, T. L., and C. Elkan.** 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers, p. 28–36. *In* Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, Calif.
6. **Bairoch, A., P. Bucher, and K. Hofmann.** 1997. The PROSITE database, its status in 1997. Nucleic Acids Res. **25:**217–221.
7. **Baldwin, S. A.** 1993. Mammalian passive glucose transporters: members of an ubiquitous family of active and passive transport proteins. Biochim. Biophys. Acta **1154:**17–49.

7a. **Boles, E., and C. P. Hollenberg.** 1997. The molecular genetics of hexose transport in yeasts. FEMS. Microbiol. Rev. **21:**85–111.

8. **Brooker, R. J., and T. H. Wilson.** 1985. Isolation and nucleotide sequencing of lactose carrier mutants that transport maltose. Proc. Natl. Acad. Sci. USA **82:**3959–3963.
9. **Bun-Ya, M., M. Nishimura, S. Harashima, and Y. Oshima.** 1991. The PHO84 gene of *Saccharomyces cerevisiae* encodes an inorganic phosphate transporter. Mol. Cell. Biol. **11:**3229–3238.
10. **Calamia, J., and C. Manoil.** 1990. Lac permease of *Escherichia coli*: topology and sequence elements promoting membrane insertion. Proc. Natl. Acad. Sci. USA **87:**4937–4941.
11. **Collier, L. S., N. N. Nichols, and E. L. Neidle.** 1997. *benK* encodes a hydrophobic permease-like protein involved in benzoate degradation by *Acinetobacter* sp. strain ADP1. J. Bacteriol. **179:**5943–5946.
12. **Collins, J. C., S. F. Permuth, and R. J. Brooker.** 1989. Isolation and characterization of lactose permease mutants with an enhanced recognition of maltose and diminished recognition of cellobiose. J. Biol. Chem. **264:**14698–14703.
13. **Crouzet, P., and L. Otten.** 1995. Sequence and mutational analysis of a tartrate utilization operon from *Agrobacterium vitis*. J. Bacteriol. **177:**6518–6526.
14. **Dayhoff, M. O., W. C. Barker, and L. T. Hunt.** 1983. Establishing homologies in protein sequences. Methods Enzymol. **91:**524–545.
15. **Dean, M., and R. Allikmets.** 1995. Evolution of ATP-binding cassette transporter genes. Curr. Opin. Genet. Dev. **5:**779–785.
16. **Devereux, J., P. Hasberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programmes for the VAX. Nucleic Acids Res. **12:**387–395.
17. **Dimroth, P.** 1997. Primary sodium ion translocating enzymes. Biochim. Biophys. Acta **1318:**11–51.

18. **Doolittle, R. F.** 1986. Of urfs and orfs: a primer on how to analyze derived amino acid sequences. University Science Books, Mill Valley, Calif.
19. **Essenberg, R. C.** 1987. The stimulation by salts of hexose phosphate uptake by *Escherichia coli*. Biochem. J. **243:**345–350.
20. **Essenberg, R. C., C. Candler, and S. K. Nida.** 1997. *Brucella abortus* strain 2308 putative glucose and galactose transporter gene: cloning and characterization. Microbiology **143:**1549–1555.
21. **Fath, M. J., and R. Kolter.** 1993. ABC transporters: bacterial exporters. Microbiol. Rev. **57:**995–1017.
22. **Feng, D.-F., and R. F. Doolittle.** 1990. Progressive alignment and phylogenetic tree construction of protein sequences. Methods Enzymol. **183:**375–387.
23. **Franco, P. J., and R. J. Brooker.** 1994. Functional roles of Glu-269 and Glu-325 within the lactose permease of *Escherichia coli*. J. Biol. Chem. **269:**7379–7386.
24. **Fu, D., and P. C. Maloney.** 1997. Evaluation of secondary structure of OxlT, the oxalate transporter of *Oxalobacter formigenes*, by circular dichroism spectroscopy. J. Biol. Chem. **272:**2129–2135.
25. **Garcia, C. K., J. L. Goldstein, R. K. Pathak, R. G. Anderson, and M. S. Brown.** 1994. Molecular characterization of a membrane transporter for lactate, pyruvate, and other monocarboxylates: implications for the Cori cycle. Cell **76:**865–873.
26. **Goffeau, A., J. Park, I. T. Paulsen, J.-L. Jonniaux, T. Dinh, P. Mordant, and M. H. Saier, Jr.** 1997. Multidrug-resistant transport proteins in yeast: complete inventory and phylogenetic characterization of yeast open reading frames within the major facilitator superfamily. Yeast **13:**43–54.
27. **Goswitz, V. C., and R. J. Brooker.** 1993. Isolation of lactose permease mutants which recognize arabinose. Membr. Biochem. **10:**61–70.
28. **Goswitz, V. C., and R. J. Brooker.** 1995. Structural features of the uniporter/symporter/anti-porter superfamily. Protein Sci. **4:**534–537.
29. **Gott, P., and W. Boos.** 1988. The transmembrane topology of the *sn*-glycerol-3-phosphate permease of *Escherichia coli* analyzed by *phoA* and *lacZ* protein fusions. Mol. Microbiol. **2:**655–663.
30. **Griffith, J. K., M. E. Baker, D. A. Rouch, M. G. P. Page, R. A. Skurray, I. T. Paulsen, K. F. Chater, S. A. Baldwin, and P. J. F. Henderson.** 1992. Membrane transport proteins: implications of sequence comparisons. Curr. Opin. Cell Biol. **4:**684–695.
31. **Gunn, F. J., C. G. Tate, and P. J. F. Henderson.** 1994. Identification of a novel sugar-H^+ symport protein, FucP, for transport of L-fucose into *Escherichia coli*. Mol. Microbiol. **12:**799–809.
32. **Gunn, F. J., C. G. Tate, C. E. Sansom, and P. J. F. Henderson.** 1995. Topological analyses of the L-fucose-H^+ symport protein, FucP, from *Escherichia coli*. Mol. Microbiol. **15:**771–783.
33. **Harrison, M. J., and M. L. van Buuren.** 1995. A phosphate transporter from the mycorrhizal fungus *Glomus versiforme*. Nature **378:**626–629.
34. **Harwood, C. S., N. N. Nichols, M.-K. Kim, J. L. Ditty, and R. E. Parales.** 1994. Identification of the *pcaRKF* gene cluster from *Pseudomonas putida*: involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. J. Bacteriol. **176:**6479–6488.
35. **Henderson, P. J. F.** 1991. Sugar transport proteins. Curr. Opin. Struct. Biol. **1:**590–601.
36. **Henderson, P. J. F., and M. C. J. Maiden.** 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.
37. **Higgins, C. F.** 1992. ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. **8:**67–113.
38. **Island, M. D., and R. J. Kadner.** 1993. Interplay between the membrane-associated UhpB and UhpC regulatory proteins. J. Bacteriol. **175:**5028–5034.
39. **Jessen-Marshall, A. E., N. J. Paul, and R. J. Brooker.** 1995. The conserved motif G*XXX* (D/E) (R/K)*X* G[*X*] (R/K) (R/K), in hydrophilic loop 2/3 of the lactose permease. J. Biol. Chem. **270:**16251–16257.
40. **Kaback, H. R.** 1997. A molecular mechanism for energy coupling in a membrane transport protein, the lactose permease of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **94:**5539–5543.
41. **Kaback, H. R., E. Bibi, and P. D. Roepe.** 1990. β-Galactoside transport in *E. coli*: a functional dissection of *lac* permease. Trends Biochem. Sci. **15:**309–314.
42. **Kaczorowski, G. J., and H. R. Kaback.** 1979. Mechanism of lactose translocation in membrane vesicles from *Escherichia coli*. 1. Effect of pH on efflux, exchange, and counterflow. Biochemistry **18:**3691–3697.
43. **King, S. C., and T. H. Wilson.** 1990. Identification of valine 177 as a mutation altering specificity for transport of sugars by the *Escherichia coli* lactose carrier. Enhanced specificity for sucrose and maltose. J. Biol. Chem. **265:**9638–9644.
44. **Kuan, G., E. Dassa, W. Saurin, M. Hofnung, and M. H. Saier, Jr.** 1995. Phylogenetic analyses of the ATP-binding constituents of bacterial extracytoplasmic receptor-dependent ABC-type nutrient uptake permeases. Res. Microbiol. **146:**271–278.
45. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.

45a. **Le, T., and M. H. Saier, Jr.** Unpublished programs.

46. **Maiden, M. C. J., E. O. Davis, S. A. Baldwin, D. C. M. Moore, and P. J. F. Henderson.** 1987. Mammalian and bacterial sugar transport proteins are homologous. Nature **325:**641–643.
47. **Maiden, M. C. J., M. C. Jones-Mortimer, and P. J. F. Hendersen.** 1988. The cloning, DNA sequence, and overexpression of the gene *araE* coding for arabinose-proton symport in *Escherichia coli* K12. J. Biol. Chem. **263:**8003–8010.
48. **Maloney, P. C.** 1990. Microbes and membrane biology. FEMS Microbiol. Rev. **87:**91–102.
49. **Maloney, P. C.** 1992. The molecular and cell biology of anion transport by bacteria. Bioessays **14:**757–762.
50. **Maloney, P. C., S. V. Ambudkar, V. Anantharam, L. A. Sonna, and A. Varadhachary.** 1990. Anion-exchange mechanisms in bacteria. Microbiol. Rev. **54:**1–17.
51. **Marger, M. D., and M. H. Saier, Jr.** 1993. A major superfamily of transmembrane facilitators catalyzing uniport, symport and antiport. Trends Biochem. Sci. **18:**13–20.
52. **Martinez, J., S. Steenbergen, and E. Vimr.** 1995. Derived structure of the putative sialic acid transporter from *Escherichia coli* predicts a novel sugar permease domain. J. Bacteriol. **177:**6005–6010.
53. **Mitchell, P.** 1967. Translocations through natural membranes. Adv. Enzymol. **29:**33–87.
54. **Mitchell, P.** 1967. Proton-translocation phosphorylation in mitochondria, chloroplasts and bacteria: natural fuel cells and solar cells. Fed. Proc. **26:**1370–1379.
55. **Neidle, E. L., C. Hartnett, L. N. Ornston, A. Bairoch, M. Rekik, and S. Harayama.** 1991. Nucleotide sequences of the *Acinetobacter calcoaceticus benABC* genes for benzoate 1,2-dioxygenase reveal evolutionary relationships among multicomponent oxygenases. J. Bacteriol. **173:**5385–5395.
56. **Olsen, S. G., and R. J. Brooker.** 1989. Analysis of the structural specificity of the lactose permease toward sugars. J. Biol. Chem. **264:**15982–15987.
57. **Olsen, S. G., K. M. Greene, and R. J. Brooker.** 1993. Lactose permease mutants which transport (malto-)oligosaccharides. J. Bacteriol. **175:**6269–6275.
58. **Olson, A. L., and J. E. Pessin.** 1996. Structure, function, and regulation of the mammalian facilitative glucose transporter gene family. Annu. Rev. Nutr. **16:**235–256.
59. **Paulsen, I. T., A. M. Beness, and M. H. Saier, Jr.** 1997. Computer-based analyses of the protein constituents of transport systems catalyzing export of complex carbohydrates in bacteria. Microbiology **143:**2685–2699.
60. **Paulsen, I. T., M. H. Brown, and R. A. Skurray.** 1996. Proton-dependent multidrug efflux pumps. Microbiol. Rev. **60:**575–608.
61. **Paulsen, I. T., S. Chauvaux, P. Choi, and M. H. Saier, Jr.** 1998. Characterization of glucose-specific catabolite repression-resistant mutants of *Bacillus subtilis*: identification of a novel hexose:H^+ symporter. J. Bacteriol. **180:**498–504.
62. **Paulsen, I. T., and R. A. Skurray.** 1994. The POT family of transport proteins. Trends Biochem. Sci. **18:**404.
63. **Paulsen, I. T., M. K. Sliwinski, and M. H. Saier, Jr.** Microbial genome analyses: Global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. J. Mol. Biol., in press.
64. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
65. **Poole, R. C., C. E. Sansom, and A. P. Halestrap.** 1996. Studies of the membrane topology of the rat erythrocyte H^+/lactate cotransporter (MCT1). Biochem. J. **320:**817–824.
66. **Rai, R., F. S. Genbauffe, and T. G. Cooper.** 1988. Structure and transcription of the allantoate permease gene (DAL5) from *Saccharomyces cerevisiae*. J. Bacteriol. **170:**266–271.
67. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1994. A functional superfamily of sodium/solute symporters. Biochim. Biophys. Acta **1197:**133–166.
68. **Rowe, J. J., T. Ubbink-Kok, D. Molenaar, W. N. Konings, and A. J. Driessen.** 1994. NarK is a nitrite-extrusion system involved in anaerobic nitrate respiration by *Escherichia coli*. Mol. Microbiol. **12:**579–586.
69. **Ruan, Z. S., V. Anantharam, I. T. Crawford, S. V. Ambudkar, S. Y. Rhee, M. J. Allison, and P. C. Maloney.** 1992. Identification, purification, and reconstitution of OxlT, the oxalate:formate antiport protein of *Oxalobacter formigenes*. J. Biol. Chem. **267:**10537–10543.
70. **Rubin, R. A., S. B. Levy, R. L. Heinrikson, and F. J. Kézdy.** 1990. Gene duplication in the evolution of the two complementing domains of Gram-negative bacterial tetracycline efflux proteins. Gene **87:**7–13.
71. **Saier, M. H., Jr.** 1994. Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution. Microbiol. Rev. **58:**71–93.
72. **Saier, M. H., Jr.** 1996. Phylogenetic approaches to the identification and characterization of protein families and superfamilies. Microb. Comp. Genomics **1:**129–150.

72a. **Saier, M. H., Jr.** Unpublished data.

72b. **Saier, M. H., Jr.** Unpublished data.

73. **Saier, M. H., Jr.** Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. Adv. Microb. Physiol., in press.
74. **Saier, M. H., Jr., I. T. Paulsen, M. K. Sliwinski, S. S. Pao, R. A. Skurray, and**

H. Nikaido. Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. FASEB J., in press.
74a. **Saier, M. H., Jr., et al.** Unpublished data.
75. **Seol, W., and A. J. Shatkin.** 1991. *Escherichia coli kgtP* encodes an α-ketoglutarate transporter. Proc. Natl. Acad. Sci. USA **88:**3802–3806.
76. **Seol, W., and A. J. Shatkin.** 1993. Membrane topology model of *Escherichia coli* α-ketoglutarate permease by PhoA fusion analysis. J. Bacteriol. **175:**565–567.
77. **Siewe, R. M., B. Weil, A. Burkovski, B. J. Eikmanns, M. Eikmanns, and R. Krämer.** 1996. Functional and genetic characterization of the (methyl)ammonium uptake carrier of *Corynebacterium glutamicum*. J. Biol. Chem. **271:**5398–5403.
78. **Steiner, H.-Y., F. Naider, and J. M. Becker.** 1995. The PTR family: a new group of peptide transporters. Mol. Microbiol. **16:**825–834.
79. **Sung, Y.-C., and J. A. Fuchs.** 1988. Characterization of the *cyn* operon in *Escherichia coli* K12. J. Biol. Chem. **263:**14769–14775.
80. **Tamai, I., H. Takanaga, H. Maeda, Y. Sai, T. Ogihara, H. Higashida, and A. Tsuji.** 1995. Participation of a proton-cotransporter, MCT1, in the intestinal transport of monocarboxylic acids. Biochem. Biophys. Res. Commun. **214:**482–489.
81. **Unkles, S. E., K. L. Hawker, C. Grieve, E. I. Campbell, P. Montague, and J. R. Kinghorn.** 1991. *crnA* encodes a nitrate transporter in *Aspergillus nidulans*. Proc. Natl. Acad. Sci. USA **88:**204–208 (Errata, **88:**4564, 1991, and **92:**3076, 1995.)
82. **Varela, M. F., and T. H. Wilson.** 1996. Molecular biology of the lactose carrier of *Escherichia coli*. Biochim. Biophys. Acta **1276:**21–34.
83. **von Heijne, G.** 1992. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J. Mol. Biol. **225:**487–494.
84. **Werner, A., M. L. Moore, N. Mantei, J. Biber, G. Semenza, and H. Murer.** 1991. Cloning and expression of cDNA for a Na/Pi cotransport system of kidney cortex. Proc. Natl. Acad. Sci. USA **88:**9608–9612.
85. **Westh Hansen, S. E., N. Jensen, and A. Munch-Petersen.** 1987. Studies on the sequence and structure of the *Escherichia coli* K-12 *nupG* gene, encoding a nucleoside-transport system. Eur. J. Biochem. **168:**385–391.
86. **Williams, P. A., and L. E. Shaw.** 1997. *mucK*, a gene in *Acinetobacter calcoaceticus* ADP1 (BD413), encodes the ability to grow an exogenous *cis*,*cis*-muconate as the sole carbon source. J. Bacteriol. **179:**5935–5942.
87. **Yamaguchi, A., T. Kimura, and T. Sawai.** 1994. Hot spots for sulfhydryl inactivation of Cys mutants in the widely conserved sequence motifs of the metal-tetracycline/H^+ antiporter of *Escherichia coli*. J. Biochem. **115:**958–964.
88. **Yamaguchi, A., N. Ono, T. Akasaka, T. Noumi, and T. Sawai.** 1990. Metal-tetracycline/H^+ antiporter of *Escherichia coli* encoded by a transposon, Tn10. The role of the conserved dipeptide, Ser^{65}-Asp^{66}, in tetracycline transport. J. Biol. Chem. **265:**15525–15530.
89. **Yamaguchi, A., Y. Someya, and T. Sawai.** 1992. Metal-tetracycline/H^+ antiporter of *Escherichia coli* encoded by a transposon, Tn10. The role of a conserved sequence motif, G*XXXX*R*X*GRR, in a putative cytoplasmic loop between helices 2 and 3. J. Biol. Chem. **267:**19155–19162.
90. **Yan, R.-T., and P. C. Maloney.** 1993. Identification of a residue in the translocation pathway of a membrane carrier. Cell **75:**37–44.
91. **Yan, R.-T., and P. C. Maloney.** 1995. Residues in the pathway through a membrane transporter. Proc. Natl. Acad. Sci. USA **92:**5973–5976.