J. Mol. Microbiol. Biotechnol. (1999) 1(2): 257-279.

The Major Facilitator Superfamily

Milton H. Saier, Jr.[1]*, J.Thomas Beatty[2], Andre Goffeau[3], Kevin T. Harley, Wilbert H.M. Heijne, Su-Chi Huang, Donald L. Jack, Peter S. Jähn, Katharine Lew[1], Jia Liu[4], Stephanie S. Pao, Ian T. Paulsen[1], Tsai-Tien Tseng[1], and Pritbir S. Virk[1]

[1]Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA
[2]Department of Microbiology and Immunology, The University of British Columbia, Vancouver, BC V6T 1Z3, Canada
[3]Unité de Biochimie Physiologique, Université Catholique de Louvain, Place Croix du Sud 2-20, B-1348 Louvain-La-Neuve, Belgium
[4]Infectious Diseases Department, Parke-Davis Pharmaceutical Research, Ann Arbor, MI 48106-1047, USA

Abstract

In 1998 we updated earlier descriptions of the largest family of secondary transport carriers found in living organisms, the major facilitator superfamily (MFS). Seventeen families of transport proteins were shown to comprise this superfamily. We here report expansion of the MFS to include 29 established families as well as five probable families. Structural, functional, and mechanistic features of the constituent permeases are described, and each newly identified family is shown to exhibit specificity for a single class of substrates. Phylogenetic analyses define the evolutionary relationships of the members of each family to each other, and multiple alignments allow definition of family-specific signature sequences as well as all well-conserved sequence motifs. The work described serves to update previous publications and allows extrapolation of structural, functional and mechanistic information obtained with any one member of the superfamily to other members with limitations determined by the degrees of sequence divergence.

Introduction

In 1998 the status of one of the two largest superfamilies of transmembrane solute transporters, the major facilitator superfamily, MFS, was reviewed and evaluated (Pao *et al.*, 1998). At that time, 17 families within this superfamily were recognized based on phylogenetic data, and each phylogenetic family in general included functionally characterized members that were specific for a single type of small molecule. Thus, three families (families 1, 5 and 7) were specific for sugars; two families (2 and 3) were specific for drugs; family 4 members transported organophosphates; family 6 permeases transported metabolites such as Krebs cycle intermediates; three families proved to be responsible for transport of inorganic anions (nitrate/nitrite, family 8; phosphate, family 9; and cyanate, family 17); family 10 proteins transported nucleosides; and five families included members that transported various monocarboxylic acids. These five families included proteins that transported (1) oxalate/formate (family 11), (2) sialate, lactate and pyruvate (family 12), (3) a wide variety of monocarboxylic acids (family 13), (4) an even wider range of organic anions plus inorganic phosphate (family 14), and (5) aromatic acids (family 15). One family recognized in 1998 was referred to as the unknown major facilitator (UMF) family (family 16) because no member of this phylogenetically distinct family had been functionally characterized (Goffeau *et al.*, 1997; Pao *et al.*, 1998). Recently, a member of this family has been shown to transport an iron-hydroxamate siderophore complex (Lesuisse *et al.*, 1998), and as its synthesis is controlled by iron availability, transport of this substrate seems to be its true physiological function. We have therefore renamed the UMF family the siderophore-iron-transporter (SIT) family in accordance with the designation by Lesuisse *et al.* (1998) of the newly characterized gene, *SIT1* (see below).

Statistical analyses conducted on established protein members of the MFS and members of a large family of peptide transporters known as the POT or PTR family (Paulsen and Skurray, 1994; Steiner *et al.*, 1995) revealed a possible distant phylogenetic relationship between members of the MFS. Our more recent PSI-BLAST results have confirmed and extended the suggestion that the POT family is indeed likely to be a divergent constituent family of the MFS.

In the present communication, we report expansion of the MFS from 17 to 29 established families and provide evidence that five additional families (including the POT family) are distantly related constituents of the MFS (see Table 1). Each of these novel families will be systematically described. Multiple alignments of the members of each family allow derivation of family-specific signature sequences; phylogenetic trees define the evolutionary relationships of the members of each family to each other, and hydropathy, similarity and amphipathicity plots provide information about structural features of these porters, thereby allowing interfamilial structural comparisons. We also analyze each of the 34 established and putative MFS families for characteristic sequence motifs, thus providing a firm basis for interfamilial motif comparisons. The results suggest that the importance of the MFS was underestimated in earlier analyses. This superfamily includes a larger percentage of the secondary carriers found in nature than was previously appreciated. Moreover, MFS carriers transport a much broader range of structurally divergent molecules than was realized. The results reveal that a major fraction of the secondary carriers found in nature are evolutionary related, and therefore probably similar in structure and mechanism of action.

Received July 10, 1999; revised August 30, 1999; accepted September 16, 1999. *For correspondence. Email saier@ucsd.edu; Tel. (858) 534-4084; Fax. (858) 534-7108.

Table 1. The Major Facilitator Superfamily (MFS) (TC #2.1)

TC #	Family Name and Abbreviation
2.1.1	The Sugar Porter (SP) Family
2.1.2	The Drug:H^+ Antiporter-1 (12 Spanner) (DHA1) Family
2.1.3	The Drug:H^+ Antiporter-2 (14 Spanner) DHA2) Family
2.1.4	The Organophosphate:P Antiporter (OPA) Family
2.1.5	The Oligosaccharide:H^+ Symporter (OHS) Family
2.1.6	The Metabolite:H^+ Symporter (MHS) Family
2.1.7	The Fucose:H^+ Symporter (FHS) Family
2.1.8	The Nitrate/Nitrite Porter (NNP) Family
2.1.9	The Phosphate:H^+ Symporter (PHS) Family
2.1.10	The Nucleoside:H^+ Symporter (NHS) Family
2.1.11	The Oxalate:Formate Antiporter (OFA) Family
2.1.12	The Sialate:H^+ Symporter (SHS) Family
2.1.13	The Monocarboxylate Porter (MCP) Family
2.1.14	The Anion:Cation Symporter (ACS) Family
2.1.15	The Aromatic Acid:H^+ Symporter (AAHS) Family
2.1.16	The Siderophore-Iron Transporter (SIT) Family
2.1.17	The Cyanate Permease (CP) Family
2.1.18	The Polyol Permease (PP) Family
2.1.19	The Organic Cation Transporter (OCT) Family
2.1.20	The Sugar Efflux Transporter (SET) Family
2.1.21	The Drug:H^+ Antiporter-3 (12 Spanner) (DHA3) Family
2.1.22	The Vesicular Neurotransmitter Transporter (VNT) Family
2.1.23	The Conjugated Bile Salt Transporter (BST) Family
2.1.24	The Unknown Major Facilitator-1 (UMF1) Family
2.1.25	The Peptide-Acetyl-Coenzyme A Transporter (PAT) Family
2.1.26	The Unknown Major Facilitator 2 (UMF2) Family
2.1.27	The Phenyl Propionate Permease (PPP) Family
2.1.28	The Unknown Major Facilitator-3 (UMF3) Family
2.1.29	The Unknown Major Facilitator-4 (UMF4) Family
2.2	The Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family
2.17	The Proton-dependent Oligopeptide Transporter (POT) Family
2.60	The Organo Anion Transporter (OAT) Family
2.71	The Folate-Biopterin Transporter (FBT) Family
97.7	The Putative Bacteriochlorophyll Delivery (BCD) Family

The SIT Family (TC #2.1.16)

Earlier phylogenetic studies of Goffeau *et al.* (1997) revealed the existence of a novel MFS family, and because no member of this family was functionally characterized, Pao *et al.* (1998) referred to it as the "unknown major facilitator"(UMF) family. All members of the UMF family were from the yeast, *Saccharomyces cerevisiae* (Goffeau *et al.*, 1997), and in March 1999, PSI-BLAST searches revealed that all sequenced members of this family are still from yeast species. The genomes of both *S. cerevisiae* and *Schizosaccharomyces pombe* encode multiple paralogues of this family (unpublished observations). In view of the large amount of eukaryotic and prokaryotic genome sequence information now available, it is reasonable to suggest that this family evolved from another primordial MFS family in yeast for a specialized function.

A recent report has provided a functional description of one of the established members of this UMF family (Lesuisse *et al.*, 1998). This protein, the product of the Yel065w gene of *S. cerevisiae* (Goffeau *et al.*, 1997) catalyzes the uptake of a hydroxamate siderophore-iron complex. The protein was designated the ferroxamine B permease. The *SIT1* (siderophore-iron transport-1) structural gene proved to be regulated by iron availability, and a *SIT1* null mutation eliminated uptake of iron-ferroxamine B. Uptake of this compound was competitively inhibited by another related iron complex, iron-ferricrocin. However, the latter compound was transported in an energy dependent process in the *SIT1* null mutant. These observations thus led to the conclusions that (1) the Sit1 permease exhibits a high degree of specificity for a restricted group of hydroxamate-siderophore-iron complexes, (2) other permeases in *S. cerevisiae* must transport other related compounds, and (3) in view of the induction properties of *SIT1* gene expression, the transport of iron siderophores is probably the true physiological function of the Sit1 protein. It seems likely that the other putative siderophore-iron transporter, recognized as the transporter of iron-ferricrocin, is a paralogue of Sit1.

The results reported by Lesuisse *et al.* are of particular interest because it has long been known that fungal siderophores, usually hydroxamates, as well as bacterial hydroxamate siderophores, can be used for iron acquisition by *S. cerevisiae* even though this yeast species does not synthesize siderophores (Lesuisse and Labbe, 1989; see Helm and Winkelmann, 1994 for a review). In view of these important observations, we have renamed the UMF family the siderophore-iron-transporter (SIT) family (TC #2.1.16) in accordance with the *SIT1* gene designation suggested by Lesuisse *et al.* (1998).

The Polyol Permease (PP) Family (TC #2.1.18)

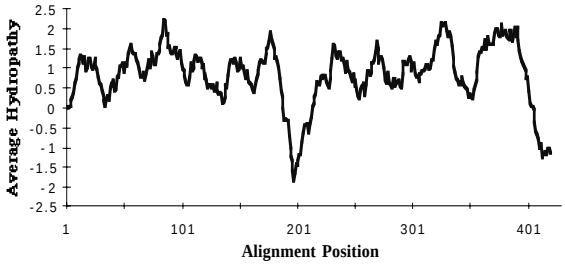
In our previous publication (Pao *et al.*, 1998) we identified 17 families of the MFS. An addendum added in proof described an eighteenth family that was recognized after publication of the molecular genetic and functional analyses

A

```
                  * *    **       ** *** *   **       *
Orf1 Bsu   (15)   GIPSHMVWGYIGVVIFMVGDGLEQGWLSPFLVDHGL
Orf2 Bsu   (6)    GIPKRLAWGFLGVVLFMMGDGLEQGWLSPFLIENGL
RbtT Kpn   (10)   GLPLNLIWGYVAIAVFMTGDGFELAFLSHYIKALGF
DalT Kpn   (10)   GLPLNLLWGYIAIAVFMTGDGFELAFLSHYIKALGF

Consensus         GLPLNL-WGYI----FMTGDG-E---LS---KALG-
```

B



C

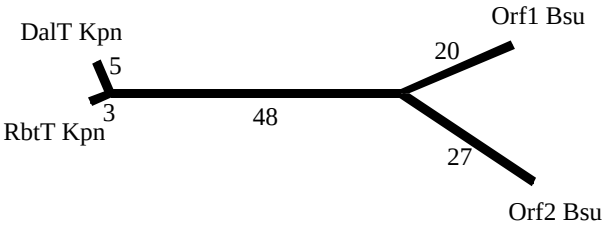


Figure 1. Partial multiple alignment (A), average hydropathy plot (B) and phylogenetic tree (C) for the polyol permease (PP) family (TC #2.1.18). The complete multiple alignment from which the partial alignment shown in A was derived using the TREE program of Feng and Doolittle (1990) was used to derive the average hydropathy plot shown in B as well as the phylogenetic tree shown in C. In A, fully conserved residues are presented in bold print with an asterisk above them. The first residue shown is presented in parentheses following the protein abbreviation (see Table 2). The consensus sequence indicates those residues that are present in the majority of the sequences. In B, a sliding window of 21 residues was used, with the hydropathy values of Kyte and Doolittle (1982). In C, branch length, presented in arbitrary units, is approximately proportional to phylogenetic distance.

described by Huel *et al.* (1997). These workers identified 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 amino acyl residues long, respectively, both with 12 putative TMSs. We conducted phylogenetic analyses of these two polyol permeases and found that they, together with two uncharacterized proteins encoded within the *Bacillus subtilis* genome, comprise a novel MFS family which we have termed the polyol permease (PP) family (family 18) (Table 2). The proteins of the PP family exhibit an approximation to the MFS-specific sequence motif between TMSs 2 and 3 of GVVAEIIGPRKTM (Pao *et al.*, 1998), 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 the *E. coli* KgtP protein gave a comparison score of 10.5 standard deviations for a segment of 107 residues (21% identity, 49% similarity, 0 gaps) (data not shown). This value is sufficient to establish that the proteins of the PP family are members of the MFS. The proteins of the PP family also exhibit recognizable sequence similarity to members of several other MFS permease families.

By hybrid protein construction, Heuel *et al.* (1997) demonstrated that the substrate specificities and kinetic properties for transport of DalT and RbtT are determined by the amino-terminal halves of the proteins. It is interesting to note that residues involved in sugar binding to the *E. coli* lactose permease (LacY; TC #2.1.5.1) have been found in both the amino-terminal half of this protein and the C-terminal half (Collins *et al.*, 1989; Matos *et al.*, 1994; see Varela and Wilson, 1996 for a review). The C-terminal half has been postulated to function in proton transport (Varela and Wilson, 1996; Venkatesan and Kaback, 1998).

A multiple alignment of the proteins listed in Table 2 was constructed. The four proteins of the PP family proved to be highly conserved with few gaps in the multiple alignment and many fully conserved residues. An example of a well-conserved portion of this complete multiple alignment is provided in Figure 1A. The following two signature sequences proved to be specific to the PP family:

1. G (L I V) P X (R H N) (L I V M A)$_2$ W G (F Y) (L I V) (G A) (L I V) (L I V A) (L I V) F M

and

2. G D G (L I V F) E X (G A) (F W) L S X (F Y) (L I V) X_3 G

(X is any residue; residues in parentheses represent alternative possibilities at a single position).

Figure 1B presents an average hydropathy plot of the complete multiple alignment for the 4 members of the PP family. It can be seen that the first six peaks of hydrophobicity are roughly equidistant from each other, as are the second six peaks. However, these two halves of the proteins are separated by a hydrophilic loop of substantial length. The two halves of the proteins proved to be

Table 2. The Polyol Permease (PP) Family (TC #2.1.18)

Abbreviation	Description	Organism	Length (amino acids)	Database & Accession No.
DalT Kpn	D-arabitol transporter	*Klebsiella pneumoniae*	425	gbAF045245
RbtT Kpn	Ribitol transporter	*Klebsiella pneumoniae*	427	gbAF045244
Orf1 Bsu	Sigma B transcribed gene	*Bacillus subtilis*	434	gbX93081
Orf2 Bsu	Putative transporter	*Bacillus subtilis*	414	gbAF027868

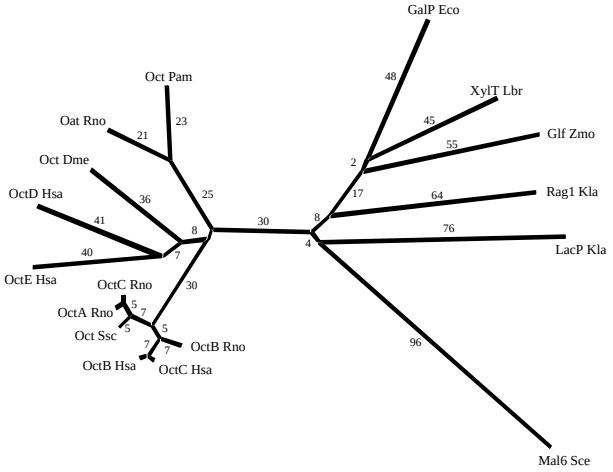


Figure 2. Phylogenetic tree for representative members of the OCT and SP families of the MFS. A well conserved segment of the complete multiple alignment (340-370 residues of the aligned proteins) was used to construct the phylogenetic tree using the TREE program of Feng and Doolittle (1990). Protein abbreviations are as indicated in Table 3.

Table 3. Members of the Organic Cation Transporter (OCT; TC #2.1.19) and Sugar Porter (SP; TC #2.1.1) Families of the MFS Included in These Studies

Family	Abbreviation[1]	Name or Description	Source (Organism)	Length	Accession no.
Oct Family					
**	OctC Rno	Organic cation transporter protein 2	*Rattus norvegicus*	593	pirJC4884
**	OctA Rno	Organic cationic transporter	*Rattus norvegicus*	593	gbX98334
**	Oct Ssc	Apical organic cation transporter	*Sus scrofa*	554	gbY09400
**	OctB Rno	Organic cation transporter	*Rattus norvegicus*	556	pirI58089
**	OctC Hsa	Organic cationic transporter	*Homo sapiens*	554	gbX98332
**	OctB Hsa	Organic cation transporter 1	*Homo sapiens*	554	gbU77086
**	Oct Dme	Putative organic cation transporter	*Drosophila melanogaster*	548	gbY12400
**	Oat Rno	Renal organic anion transporter 1	*Rattus norvegicus*	551	gbAF008221
**	OctE Hsa	Polyspecific organic cation transporter	*Homo sapiens*	551	gbAB007448
**	Oct Pam	Renal organic anion transporter	*Pseudopleuronectes americanus*	562	gbZ97028
**	OctD Hsa	Organic cation transporter	*Homo sapiens*	456	gbAC002464
*	OctB Mmu	Organic cation transporter 2	*Mus musculus*	553	gbAJ006036
*	OctA Hsa	Organic cation transporter	*Homo sapiens*	555	gbX98333
*	OctD Rno	Organic cation transporter OCT1A	*Rattus norvegicus*	430	gbU76379
*	OctA Mmu	RST	*Mus musculus*	553	gbAB005451
*	OctF Hsa	Kidney organic cation transporter N2	*Homo sapiens*	557	gbAB015050
		Similar to the rat OCT1 transporter	*Mus musculus*	556	gbU38652
		Potential-sensitive polyspecific organic cation transporter	*Rattus norvegicus*	551	gbAF055286
		Organic cation transporter homolog	*Mus musculus*	545	gbU52842
		Similarity to rat organic cation transporter	*Caenorhabditis elegans*	576	gbZ83228
		Liver-specific transport protein	*Rattus norvegicus*	535	gbL27651
		Putative integral membrane transport protein	*Rattus norvegicus*	557	gbAJ001933
		Renal organic cation transporter	*Oryctolagus cuniculus*	554	gbAF015958
		Putative integral membrane transport protein	*Rattus norvegicus*	552	gbY09945
SP Family					
	XylT Lbr	Xylose/proton symporter	*Lactobacillus brevis*	457	gbAF045552
	Mal6 Sce	Maltose permease, Mal6T	*Saccharomyces cerevisiae*	614	spP15685
	Glf Zmo	Glucose facilitated diffusion protein	*Zymomonas mobilis*	473	spP21906
	Rag1 Kla	Low-affinity glucose transporter	*Kluyveromyces lactis*	567	spP18631
	GalP Eco	Galactose/proton symporter	*Escherichia coli*	464	spP37021
	LacP Kla	Lactose permease	*Kluyveromyces lactis*	587	spP07921

[1]Different Oct family paralogues from single species are distinguished by the letters "A,B,C,…." in chronological order according to the dates of submission to the database.
Proteins of the Oct family indicated with one asterisk (*) were used for the studies including only the Oct family members.
Proteins of the Oct family indicated with two asterisks (**) were used for all studies.
Proteins of the Oct family lacking an asterisk have not been functionally characterized and were not included in the reported studies.
Proteins of the SP family were used only for construction of the phylogenetic tree with Oct family members indicated with two asterisks.

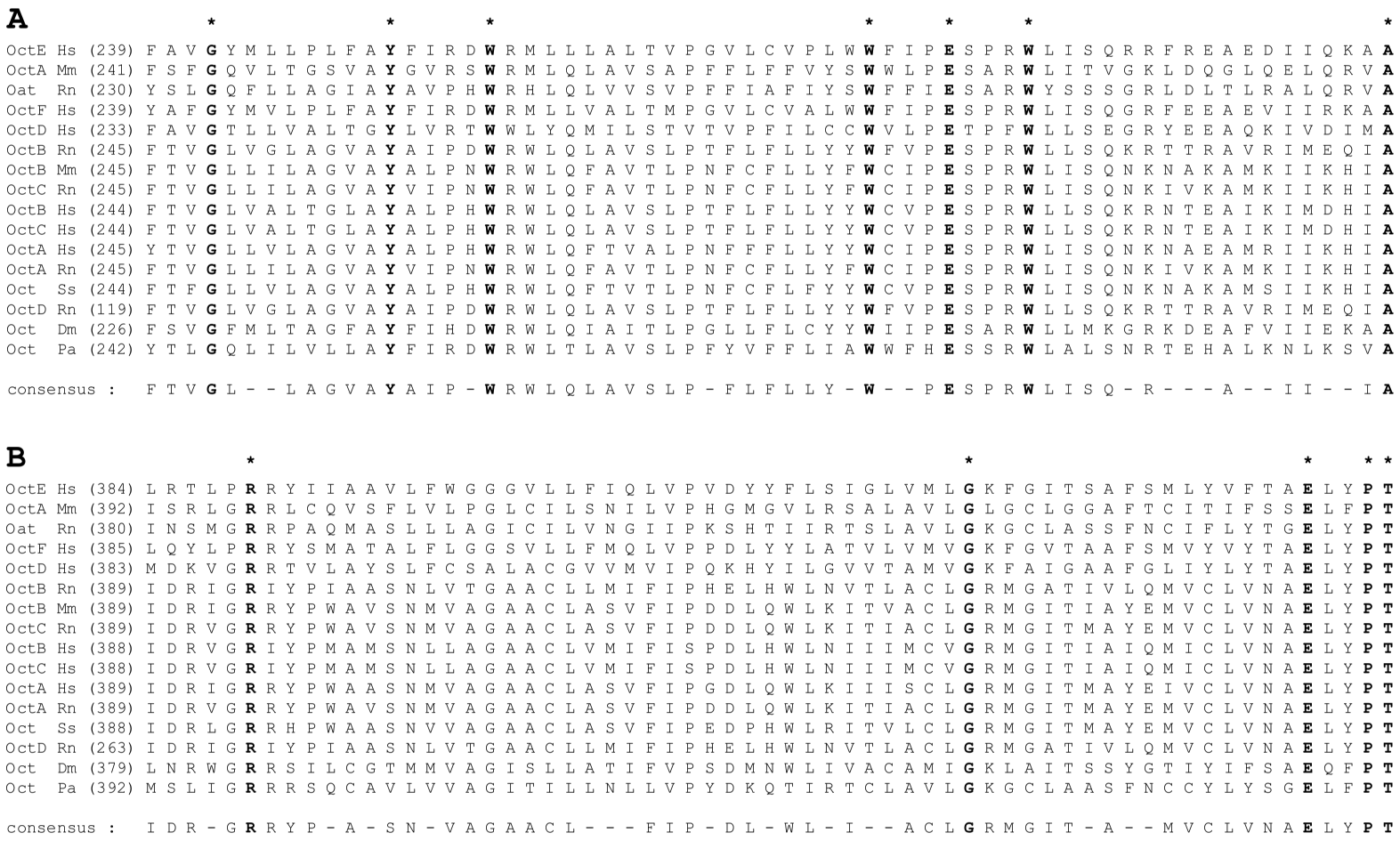


Figure 3. Two well conserved portions of the complete multiple alignment of functionally characterized members of the OCT family. The protein abbreviations are as indicated in Table 1. Fully conserved residues are indicated with asterisks and presented in bold print. The number of the first residue in each line is provided in parentheses following the protein abbreviation. The consensus sequence (consensus) (a majority of the residues at any one position conserved) is presented below the alignment.

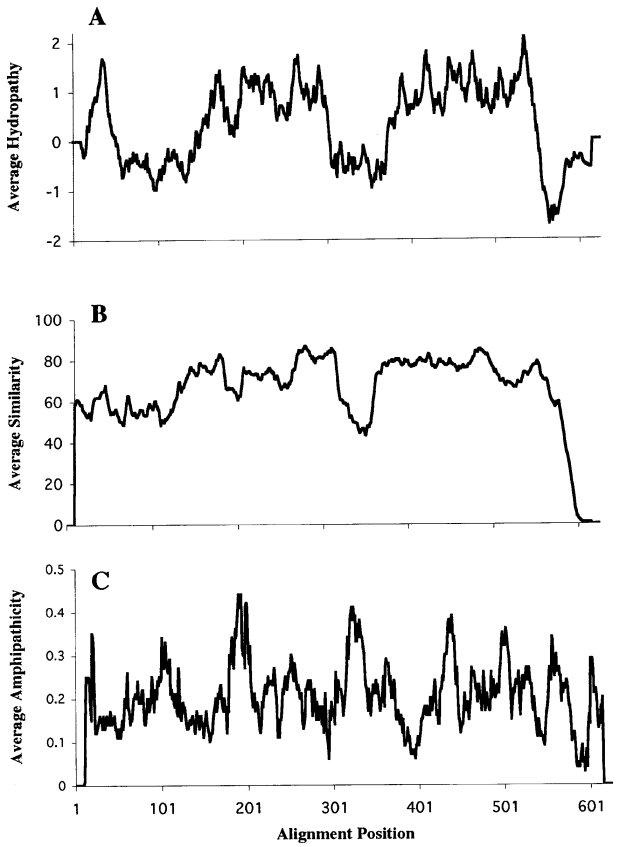


Figure 4. Average hydropathy (A), similarity (B) and amphipathicity (100° for α-helix) (C) plots for the fully aligned sequences of members of the OCT family as presented in Figure 3. A sliding window of 21 residues was used in all 3 plots. Hydropathy values for the individual amino acids were as calculated by Kyte and Doolittle (1982). The average amphipathicity program has been described (Le *et al.* 1999).

equally well conserved, but the peaks of hydrophobicity in general correlated with peaks of average similarity (not shown). An average amphipathicity plot with the angle per residue set at 100° for an α-helix revealed several peaks, the largest of which occurred at alignment positions just preceding and overlapping hydrophobic peaks 1 and 7. The C-terminal regions following putative TMS12 also proved to be strongly amphipathic. The results provide evidence that major regions of the proteins of the PP family are α-helical regardless of whether they are embedded in the membrane or surface localized.

The phylogenetic tree for the PP family (Figure 1C) shows clustering according to organism. Thus, the two *Klebsiella* proteins cluster tightly together as do the two *Bacillus* proteins. One can infer that extragenic duplications that gave rise to the pair of proteins in each organism occurred after Gram-negative bacteria diverged from Gram-positive bacteria.

The Organic Cation Transporter (OCT) Family (TC #2.1.19)

One of the 12 families described by Pao *et al.* (1998), the sugar porter (SP) family, was exceptionally large with 133 sequenced members. In contrast to most other MFS families, the SP family included members that were functionally diverse. While most members transported sugars, a few had been shown to transport organic cations and/or anions (see, for example, Gründemann *et al.*, 1994; Okuda *et al.*, 1996; Lopez-Nieto *et al.*, 1997; Kekuda *et al.*, 1998; see Koepsell (1998) for a current review). The latter proteins clustered distantly from the sugar porters on the SP family phylogenetic tree. These proteins catalyze uptake of cationic drugs such as tetramethyl ammonium, cimetidine, procainamide, quinidine and some endogenous metabolites such as N-methyl-nicotinamide. In view of these surprising observations, and because several additional such porters have since been characterized, the organic ion transporters were reexamined phylogenetically.

Table 3 lists established members of the OCT family as well as representative divergent members of the SP family used for the phylogenetic analyses. Only the proteins indicated with double asterisks were used for the analyses presented in Figure 2. As shown in Figure 2, all of the proteins known to function in organo-cation and anion transport clustered separately from representative transporters specific for sugars. This was observed regardless of the program used to construct the tree or dendogram (data not shown). Thus, while all of the organo-ion transport proteins showed greater sequence similarity to members of the SP family than to members of any one of the other MFS families, they clearly comprise a distinct family (or subfamily) both phylogenetically and functionally. We therefore have elected to designate this family the "organic cation transporter" (OCT) family (TC #2.1.19), named after the majority of the proteins which comprise this family. It is interesting to note that the single characterized organo-anion transporter, Oat Rno, clusters with an organo-cation transporter, Oct Pam (Figure 2) and transports both cations and anions (Koepsell, 1998).

All OCT family members indicated in Table 3 with either double or single asterisks were included in the analyses described below. Sixteen proteins, all from animals, plus several uncharacterized open reading frames, comprise the current OCT family. Two well conserved regions of the complete multiple alignment are presented in Figure 3. Both regions reveal a high degree of sequence identity, with seven and five fully conserved residues in the two portions shown, respectively. No gaps are present in these aligned sequences.

Two signature sequences were derived from the two well conserved regions shown in Figures 3A and B. These sequences are:

SS #1: [F Y W S A C] W [L I V F W C] [L I V F] X E [S T] [P A S] [R F] W [L Y] X_4 [R K]

SS #2: [L I V F Y] X_2 [L I V C] [C T Y F] [L I V] [V F Y] [S T N] [A S G] E X [Y F] P T [L I V F Y]

These two sequences retrieved only established members of the OCT family when screened against the SwissProt database, and they are therefore authentic signature sequences by this criterion.

Based on the complete multiple alignment including all of the proteins represented in Figure 3, average hydropathy, similarity and amphipathicity (100° for α-helix) were derived (see Figures 4A-C). The average hydropathy plot (Figure 4A) revealed the presence of one N-termi-

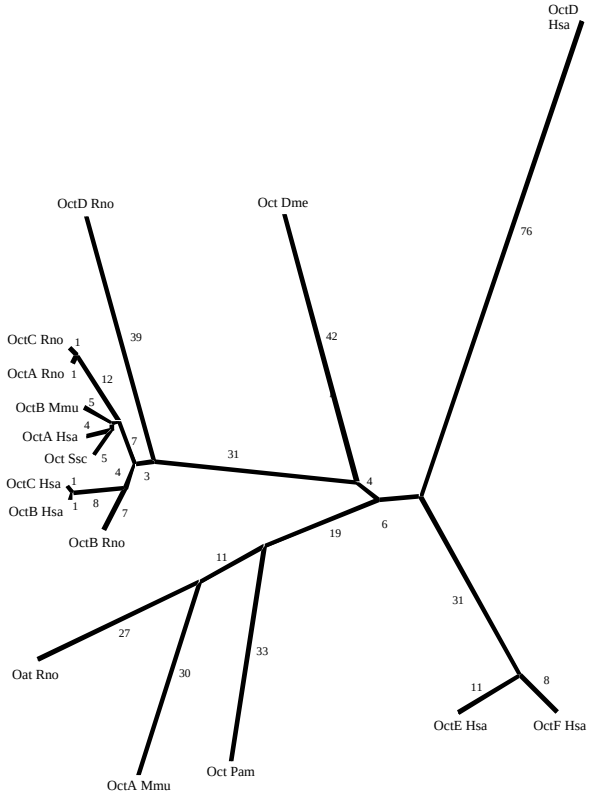


Figure 5. Phylogenetic tree of OCT family members, based on the complete multiple alignment of these proteins. The format of presentation and method of tree construction are as described in Figure 2.

nal hydrophobic segment of sufficient breadth and magnitude to span the membrane as an α-helix. Following an extended hydrophilic "loop" region, five additional peaks of hydrophobicity corresponding to five putative transmembrane spanning segments were observed. Following a second hydrophilic loop region, six additional putative transmembrane segments could be assigned. Noteworthy is the fact that the loop regions in general tend to be less well conserved than the transmembrane regions (Figure 4B). Further, the striking peaks of amphipathicity (Figure 4C) invariably correspond to hydrophilic inter-TMS regions. It can therefore be surmised, that not only the hydrophobic transmembrane regions, but also the hydrophilic "loop" regions occur largely as α-helices.

Figure 5 shows a phylogenetic tree where most of the currently recognized members of the OCT family are represented. The tree is based on the complete multiple alignment for these proteins, portions of which are shown in Figure 3. Many of these proteins, all from mammals, cluster tightly together suggesting that the paralogues within this cluster (3 from rats, and 3 from man) arose recently in evolutionary time by gene duplication events. Other paralogous members of the family are considerably more distant from each other and presumably arose as a result of much earlier gene duplication events. Examining the human paralogues, for example, revealed that OctA, B and C are similar in sequence, that OctE and F are similar to each other but very distant from all other human paralogues, and that OctD is the most distant human member of the family. The one organo-anion transporter represented (Oat Rno) clusters loosely with two cation transporters. This transporter is known to catalyze uptake of both cations and anions.

The Sugar Efflux Transporter (SET) Family (TC #2.1.20)

The proteins of the SET family are listed in Table 4. Five of the ten protein members are from *E. coli*, and three are from *Bacillus subtilis*. The other two are from *Mycobacterium tuberculosis* and *Yersinia pestis*. A homologue is also encoded within the *Deinococcus radiodurans* genome (not presented). The protein members of the SET family are distantly related to well characterized proteins from several different families within the MFS.

Three of the *E. coli* SET family proteins have been subjected to functional characterization (Liu *et al.*, 1999a,b). Two of these proteins have been shown to catalyze efflux of sugars and their derivatives. This fact provides the basis for the family name (SET). SetA (YabM) has been shown to catalyze efflux of isopropyl-thio-β-galactoside (IPTG), lactose and glucose. The efflux process was inhibited by a variety of other sugars such as aromatic α- and β-glucosides, aromatic α- and β-galactosides, cellobiose, maltose, α-methyl glucoside and L-glucose. The carrier thus apparently exhibits broad binding specificity. Additionally, sugar-containing amino glycoside antibiotics such as streptomycin and kanamycin were weakly expelled via this system as demonstrated using resistance tests (Liu *et al.*, 1999a,b). Sugars with five carbons or less proved to be poor inhibitors in the lactose transport assay.

SetB (YeiO) similarly catalyzes efflux of glucose and lactose, but IPTG and galactose were not transported. SetC

Table 4. Proteins of the Sugar Efflux Transporter (SET) Family of the MFS (TC #2.1.20)

Abbreviation	Name or Description	Organism	Length	Accession no.
YicK Eco	Hypothetical 43.5 KD protein in *selC-A* intergenic region	*Escherichia coli*	394	spP31436
YeiO Eco	Hypothetical 42.7 KD protein *in fruB-spr* intergenic region	*Escherichia coli*	393	spP33026
YabM Eco	Hypothetical 42.7 KD protein in *tbpA-leuD* intergenic region	*Escherichia coli*	392	spP31675
YceL Eco	Hypothetical 44.4 KD protein in *grxB-rimJ* intergenic region	*Escherichia coli*	402	spP77042
Orf Mtu	Hypothetical protein Rv0849	*Mycobacterium tuberculosis*	419	gbAL02204
YqjV Bsu	Hypothetical 44.7 KD protein in *glnQ-ansR* intergenic region	*Bacillus subtilis*	410	spP54559
YdeE Eco	Hypothetical 42.7 KD protein in *marB-dcP* intergenic region	*Escherichia coli*	395	spP31126
Orf2 Bsu	Similarity to tetracycline resistance protein from *E. coli* pBR322 plasmid	*Bacillus subtilis*	397	gbAF008220
Orf1 Bsu	Similarity to hypothetical protein YqjV from *B. subtilis*	*Bacillus subtilis*	401	gbY14081
Orf Ype	Open reading frame fragment	*Yersinia pestis*	304	—

A

```
                      *                                       *
Orf1 Bsu   (29) TGAMMGPFMVLYLHEQLNGSIMMPMLIISLQPFADIFLTLAAGRVTDRLG RRTAIL
Orf2 Bsu   (18) GASFLWPLNTIYIHNHLGKSLTVAGLVLMLNSGASVAGNLCGGFLFDKIG GFKSIM
YdeE Eco   (23) RGATL PFMTIYLSRQYSLSVDLIGYAMTIALTIGVVFSLGFGILADKFD KKRYML
Orf  Ype   (21) AGALQAPTLSLFLSTELKVRPLWVGLFYTVNAIAGITVSFLLAKRSDLGGDRRKLIL
YabM Eco   (28) AGALQAPTLSLFLSREVGAQPFWIGLFYTVNAIAGIGVSLWLAKRSDSQGDRRKLII
Orf  Mtu   (27) GFYMLMPYLADYLAGPLGLAAWAVGLVMGVRNFSQQGMFFVGGTLADRFG YKPLII
YeiO Eco   (30) AGALQTPTLSIFLTDEVHARPAMVGFFFTGSAVIGILVSQFLAGRSDKRGDRKSLIV
YicK Eco   (29) AGALQTPTLSIFLADELKARPIMVGFFFTGSAIMGILVSQFLARHSDKQGDRKLLIL
YqiV Bsu   (23) ATSMSIPFLAIYLTAVQGASASYAGLVIAASSSVGILASFYGGYISDKFG RKNMML
YceL Eco   (25) GFFVVFPLISIRFVDQMGWAAVMVGIALGLRQFIQQGLGIFGGAIADRFG AKPMIV

consensus :    AGALQ-P-LSIYL--LG------VGL--T-----GI--S---G--SDK-G-RK-LIL
```

B

```
                 *              *
Orf1 Bsu (141) F AVINAIYSTGLTAGPLVG
Orf2 Bsu (129) F NAIYVAQNAGVAVGSALG
YdeE Eco (134) F SINYTMLNIGWTIGPPLG
Orf  Ype (135) FSSIMRAQLSLAWVIGPPLS
YabM Eco (142) FSSVMRAQLSLAWVIGPPLA
Orf  Mtu (138) F AMFNVFYQSGILLGPLVG
YeiO Eco (144) FSSFLRAQVSLAWVIGPPLA
YicK Eco (143) FSTFLRAQISLAWVIGPPLA
YqiV Bsu (135) F NLRYAAINIGVVFGPVLG
YceL Eco (136) FFSLLMMQDSAGAVIGALLG

consensus :    F-S---AQ-S-GWVIGPPLG
```

Figure 6. Two relatively well conserved portions of the complete multiple alignment for the ten recognized proteins of the sugar efflux transporter (SET) family of the MFS. The protein abbreviations are as presented in Table 4. The methods and conventions of figure presentation are as described in the legend to Figure 3.

(YicK) did not expel any sugar tested including glucose, galactose, lactose or IPTG. Further, streptomycin and kanamycin were not substrates of either SetB or SetC. These results suggest that two closely related *E. coli* paralogues, but not a third, exhibit differing but overlapping specificities for sugars and their derivatives. The substrates of SetC have yet to be identified. A proton antiport mechanism has been inferred for all three *E. coli* SET family paralogues (Liu *et al.*, 1999a,b).

Figure 6 shows two fairly well conserved portions of the complete multiple alignment of the ten SET family proteins. Two residues are fully conserved in each of these gap-free regions, and from the regions shown in A and B, two signature sequences that retrieved only established SET family proteins from the SwissProt database were derived. These sequences are:

SS #1: P [L I V F Y T] [L I V M N] [S T A V] X_2 [L I V F] X_7 [L I V P A] X_2 [L I V P A] [M G] [L I V F Y] [L I V F A] [L I V F Y M] [S T A G M] [L I V A G] X_3 [L I V S A M] X_2 [L I V F T A G] [L I V F G A M] X_3 [L I V F A G] [A G] X_2 [T A S F] D

SS #2: [S F] [S A T N] [L I V F A M] X_5 [S N Q] [L I V S A T] [G A] [L I V W A] [L I V T A] [L I V A F] G [P A S] [L I V P A] [L I V] [G A S]

The average hydropathy plot, based on the complete multiple alignment from which the two alignments shown in Figure 6 was selected, is shown in Figure 7A. Twelve clear peaks of hydrophobicity are observed, and uniquely, they fall into six sets of two closely positioned peaks. Assuming a topology analogous to the 12 TMS proteins of the MFS, with both the N- and C-termini facing the cytoplasm, the results suggest that all periplasmic inter-TMS loops are short while all cytoplasmic inter-TMS loops are

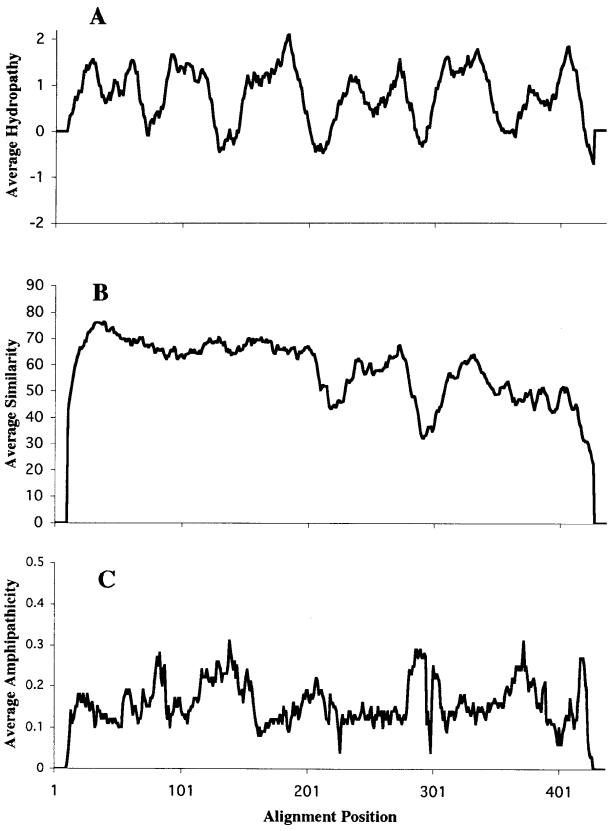


Figure 7. Average hydropathy (A), similarity (B), and amphipathicity (for α-helix) (C) plots for the proteins of the SET family. In all cases, the multiple alignment was generated using the TREE program (see Figure 6), and a sliding window of 21 residues was used.

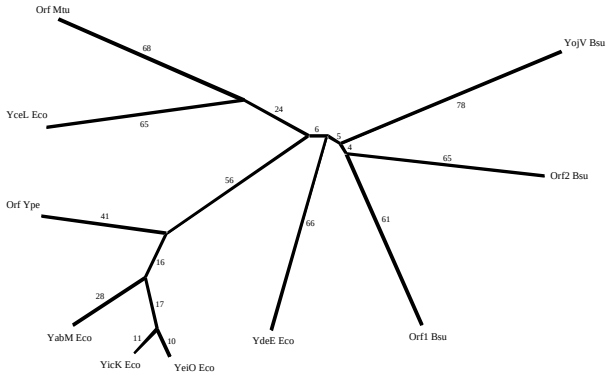


Figure 8. Phylogenetic tree for the proteins of the SET family (see legend to Figure 2 for format of presentation).

longer. The average similarity plot (Figure 7B) shows that, as for many other MFS families, the N-terminal domain is better conserved than the C-terminal domain. The least sequence similarity, reflecting multiple gaps in the aligned sequences, is found between putative TMSs 6 and 7 (the central loop) as well as between putative TMSs 8 and 9.

Figure 7C, showing the average amphipathicity plot with the angle set at 100° per residue as for an α-helix, revealed that the major peaks of amphipathicity occur in putative cytoplasmic loops 2-3, 4-5, 6-7 and 8-9, just preceding and overlapping putative TMS 11, and just following TMS 12. Most of these regions of strong amphipathicity occur within, or overlapping and immediately adjacent to the five cytoplasmic loops. This fact suggests that the cytoplasmic loops are present in large measure in α-helical configurations. No evidence concerning the secondary structures of the short external loops was obtained from these analyses.

The phylogenetic tree for the SET family proteins is shown in Figure 8. The three functionally examined *E. coli* proteins, SetA (YabM), SetB (YicK) and SetC (YeiO), are closely related paralogues. The sequence fragment from *Yersinia pestis* (Table 4) is also closely related to these proteins, suggesting a similar function. All other protein members of the family are distant from these four proteins and from each other, suggesting divergent functions. Although we would tentatively suggest that these proteins could function in the efflux of hydrophilic molecules, the phylogenetic distances between them renders even such a suggestion highly speculative.

The Drug:H$^+$ Antiporter-3 (DHA3) Family (TC #2.1.21)

The DHA3 family is a diverse, moderately sized family, several members of which exhibit limited sequence similarity with established members of the MFS and with the phylogenetically related GPH (TC #2.2) family (see below). All of the functionally characterized DHA3 transporters efflux drugs, probably by a proton antiport mechanism (Table 5). These proteins include the MefA macrolide resistance determinant of *Streptococcus pyogenes* (Clancy *et al.*, 1996), also found in *S. pneumoniae* and *Lactococcus lactis* (Table 5; Perreten *et al.*, 1997). MefA expels 14-membered macrolides such as erythromycin and oleandomycin as well as 15-membered macrolides such as azithromycin, but not 16-membered macrolides such as spiromycin and tylosin (Clancy *et al.*, 1996). Another characterized drug efflux pump is the Cmr multidrug resistance protein of *Corynebacterium glutamicum* which confers resistance to erythromycin, tetracycline, puromycin and bleomycin (Table 5; Jäger *et al.*, 1997). Others include the TetV tetracycline resistance determinant of *Mycobacterium smegmatis* (De Rossi *et al.*, 1998) and the Tap multidrug resistance efflux pump of *M. fortuitum* (Ainsa *et al.*, 1998). No description of the putative Ni^{2+} resistance protein of *Synechocystis*, NiR, mentioned in the database entry for this protein (see Table 5), is available. It can be anticipated that most if not all members of the DHA3 family will prove to be drug efflux pumps.

It is interesting to note that most (but not all) of the members of the DHA3 family are from Gram-positive bacteria. Thus, two of the proteins are from Gram-negative eubacteria, two are from cyanobacteria, one is from an archaeon, and 14 are from Gram-positive bacteria. None of the members of the DHA3 family is as yet from a

Table 5. The Drug:H$^+$ Antiporter-3 (DHA3) Family (TC #2.1.21)

Abbreviation	Description	Organism	Size (no. residues)	Database and Accession no.
YkuC Bsu	YkuC protein	*Bacillus subtilis*	430	gbZ99111
MefA Spy	Macrolide-efflux protein, MefA	*Streptococcus pyogenes*	405	gbU70055
MefA Spn	Macrolide-efflux determinant	*Streptococcus pneumoniae*	405	gbU83667
MefA Lla	Macrolide-efflux protein	*Lactococcus lactis*	418	gbX92946
Orf Bsu	Similar to multidrug resistance protein	*Bacillus subtilis*	417	gbZ99108
Orf Pho	403aa long hypothetical protein	*Pyrococcus horikoshii*	403	gbAB009504
Orf1 Mtu	Hypothetical 43.3 KD protein CY50.24	*Mycobacterium tubeculosis*	419	spQ11060
Tap Mfo	Tap protein	*Mycobacterium fortuitum*	409	gbAJ000283
YbdA Eco	Hypothetical membrane protein p43	*Escherichia coli*	416	spP24077
Orf1 Sco	Transmembrane protein	*Streptomyces coelicolor*	431	gbAL023496
Orf Msm	Putative transporter	*Mycobacterium smegmatis*	412	gbU46844
TetV Msm	Tetracycline-resistance determinant TetV	*Mycobacterium smegmatis*	419	gbAF030344
Orf1 Ssp	Hypothetical protein	*Synechocystis sp.*	465	gbD90915
Orf2 Mtu	Hypothetical 45.9 Kd protein CY10H4.37C	*Mycobacterium tubeculosis*	441	spP71607
Orf Axy	nreB	*Alicaligenes xylosoxidans*	474	gbL31491
Cmr Cgl	Multidrug resistance protein	*Corynebacterium glutamicum*	459	gbU43535
Orf2 Ssp	Hypothetical protein	*Synechocystis sp.*	427	gbD90899
NiR Ssp	Nickel resistance	*Synechocystis sp.*	445	gbD64005
Orf2 Sco	Putative integral membrane protein	*Streptomyces coelicolor*	630	gbAL023496

```
                              *
NiR  Ssp      (445)  GAIADRYDRKQMMVITHLARLGIVCLFPGV
Cmr  Cgl      (459)  GTVVDHNRKKSVMLFSSVTTLVFYCLSALV
Orf2 Ssp      (427)  GILTDYFSHKKLLIVSDIGSA  VCTFSVG
MefA Spy      (405)  GVLVDRHDRKKIMIGADLIIAAAGSVLTIV
Orf1 Mtu      (419)  GTAVDYFGRRRVSMVADALSGAAVAGVPLV
MefA Spn      (405)  GVLVDRHDRKKIMIGADLIIAAAGAVLAIV
Orf  Msm      (412)  GVLADRYSKRTILLWTALGGMLPALVLGVL
Tap  Mfo      (409)  GAAVDYLGRRRVSMISDLLSALSVAAVPVL
MefA Lla      (418)  GPFIDRINKKFLLISYDAVVAVIALGLFIY
Orf1 Ssp      (465)  GVYVDRWQKKQVLVVTNFCRGILILLLPFL
TetV Msm      (419)  GITADRINQRTIIIAVEVVNFVTVAVISAL
Orf  Bsu      (417)  GLLADRFDRKTIMFLSEIGRALTVISCVYV
YkuC Bsu      (430)  GVVPDRFDRKKVAENCDWIRAGLTVVLFFT
YbdA Eco      (416)  GVLADRYERKKVILLARGTCGIGFIGLCLN
Orf1 Sco      (431)  GALADAVDRRRVIVLTEAGLGLLAAVLLVN
Orf2 Mtu      (441)  GALMDRWDRRWVLVGANTGRLALIAGVGTI
Orf  Pho      (403)  GVIGDRYNRKHLMVGFDLARGVLLFLIIAL
Orf  Axy      (474)  GAYANRLPRRAFLVAMDLIRAAVAISLPFV
Orf2 Sco      (630)  GVLADRYPPRSVMRWASAVRLPLVAAMCAL

Consensus            G-L-DR--RK-V----D---------L---
```

Figure 9. Partial multiple alignment of the 19 members of the DHA3 family. Protein abbreviations are as presented in Table 5. Methods and conventions of presentation are as described in the legend to Figure 3.

eukaryote. The uniformity of the DHA3 family protein sizes is noteworthy. Thus, except for one protein from *Streptomyces coelicolor*, all proteins are in the size range of 405-474 (Table 5).

Figure 9 presents a relatively well conserved portion of the multiple alignment that includes the sequences of the 19 identified members of the DHA3 family. A single glycyl residue is completely conserved, but several residues are largely conserved (see Figure 9). The DHA3 family signature sequence derived from this region of the alignment is:

G - (L I V T A P) - (L I V F Y A T) - (L I V T A P G M) - (D N) - (Y R H A) -X_2- (R K H P Q) - (K R) - (R K H Q S T A W F) - (L I V M F) - (L I V M S A) - (L I V M F E R) -X_2- (D N E H S A R) - (L I V F A W G T) - (A T G C L I V) -X- (L I V GA M F)

Figure 10A shows the average hydropathy plot for the identified DHA3 family members. Twelve peaks of hydrophobicity correspond to twelve putative transmembrane segments (TMSs). The average similarity plot (Figure 10B) reveals that for each peak of hydrophobicity there is a peak of similarity. This fact shows that the transmembrane segments are better conserved that the inter-TMS loops. As shown in Figure 10C, regions of strong amphipathicity are usually found between TMSs. However, the putative TMS at position 100 is both well conserved and amphipathic.

The phylogenetic tree for the DHA3 family is shown in Figure 11. Most protein members of the family are distant from each other. However, MefA Spy and MefA Spn cluster tightly together, and these proteins cluster loosely with MefA Lla. These three proteins may well be orthologues as also suggested by their biochemical designations and available functional data. Similarly, Tap Mfo and Orf1 Mtu are very similar in sequence suggesting that these two myobacterial proteins are orthologues. All other proteins in the DHA3 family are distantly related to these proteins as well as to each other.

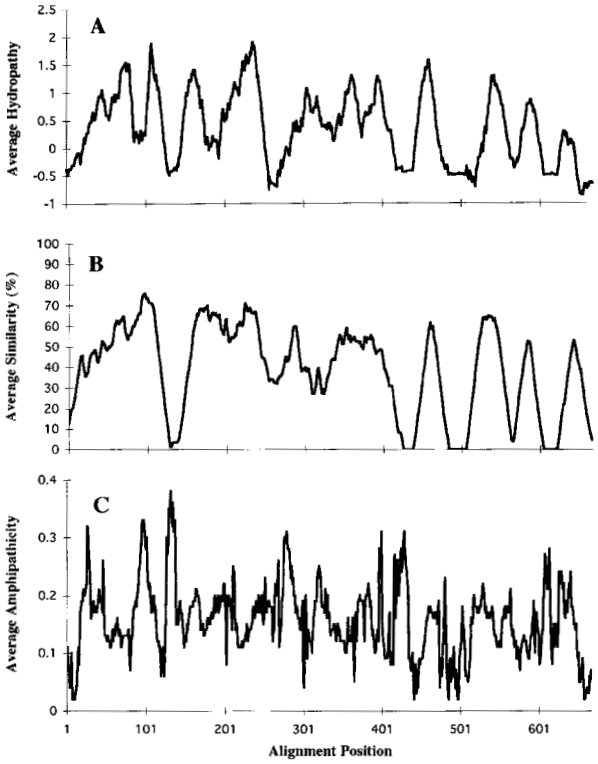


Figure 10. Average hydropathy (A), similarity (B) and amphipathicity (for α-helix, C) plots for the proteins of the DHA3 family. Plots are based on the multiple alignment generated with the TREE program using a sliding window of 21 residues.

The Vesicular Neurotransmitter (VNT) Family (TC #2.1.22)

In our earlier analysis of the MFS (Pao *et al.*, 1998), we included the few vesicular neurotransmitter transporters that were at that time sequences in the sugar porter (SP) family (TC #2.1.1) because of their close phylogenetic association. With more members available for analysis, it is now clear that these proteins comprise their own cluster or family which, however, is more closely related to the SP family than to other MFS families. We have consequently assigned these proteins to a separate family.

Sequenced members of the VNT family are presented in Table 6. The better characterized members of the VNT family are synaptic vesicle proteins from mammals, the electric eel and insects (Bajjalieh *et al.*, 1992, 1993; Gingrich *et al.*, 1992; Bindra *et al.*, 1993; Janz *et al.*, 1998; Nagase *et al.*, 1998; Wang and Fallon, 1998). These proteins constitute a novel family of 12 putative TMS proteins of about 700 amino acyl residues.

Seven members of the VNT family are listed in Table 6. However, three of these proteins (Orf Bta, KIAA Hsa, Sv2A Rno) are nearly identical in sequence. Similarly, KIAB Hsa and Sv2B Rno are nearly identical in sequence. A phylogenetic tree of the 4 dissimilar proteins revealed that KIAA Hsa, KIAB Hsa and Sv2 Dom are about equally distant from each other while Sv2 Aal is only distantly related to these three proteins (Figure 12).

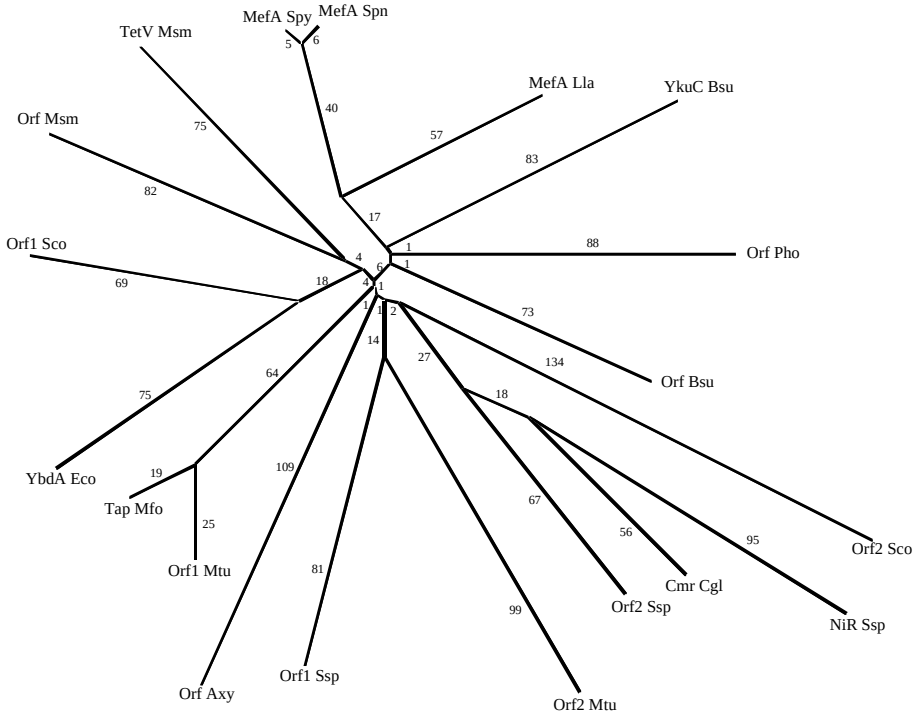


Figure 11. Phylogenetic tree for the DHA3 family (see Figure 2 legend for methods and format of presentation).

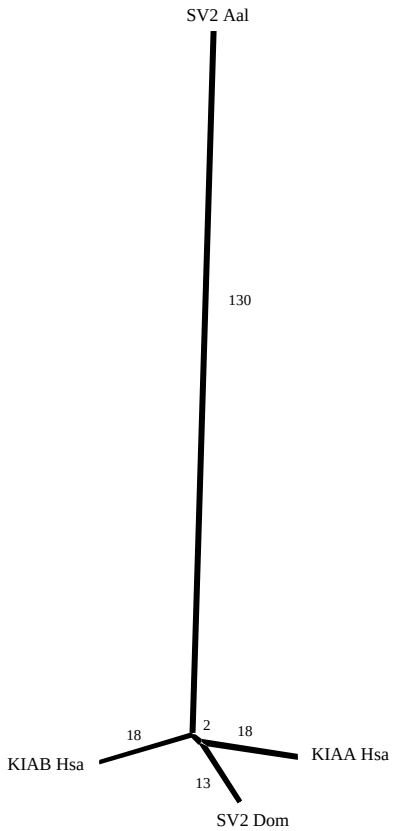


Figure 12. Phylogenetic tree for the vesicular neurotransmitter (VNT) family of the MFS.

The (Putative) Conjugated Bile Salt Transporter (BST) Family (TC #2.1.23)

A single fully sequenced protein, Bsh, and a fragment of a second protein, Orf, both from *Lactobacillus johnsonii*, constitute the BST family (Table 6). When produced in *E. coli*, the fully sequenced protein produced a strain with a three-fold increase in the uptake rate for taurocholic acid (Elkins and Savage, 1998). Cholate was apparently not transported leading to the suggestion that the transporter is specific for conjugated bile salts. The protein is 451 amino acids in length and exhibits 12 putative TMSs. The two homologous ORFs proved to be about 80% identical in the region of the 200 residue fragment that corresponded to the C-terminus of the Bsh protein. Because of the small size of the family, no further analyses are reported.

The Unknown Major Facilitator-1 (UMF1) Family (TC #2.1.24)

Only three proteins comprise the UMF1 family (Table 6). Two of these proteins are from two different yeast species, and one is from the bacterium, *Bacillus subtilis*. The two yeast proteins exhibit extensive sequence similarity throughout their lengths, are of the same size and are predicted to possess 12 TMSs. They exhibit sufficient sequence similarity with an uncharacterized protein, YxiO from *B. subtilis,* to establish that these three proteins are homologous and belong to a single family. YxiO is a 428 residue protein exhibiting 12 putative TMSs (Table 6). With a single iteration, the PSI-BLAST program revealed motif

Table 6. Protein Members of Small, Newly Discovered Families Within the MFS

Abbreviation	Description in Database	Size	Organism	Accession #
The Vesicular Neurotransmitter Transporter (VNT) Family (TC #2.1.22)				
Sv2 Aal	Synaptic vesicle protein	401 aa	*Aedes albopictus*	gbAF049228
Orf Bta	Transporter-like protein	742 aa	*Bos taurus*	gbQ29397
Sv2 Dom	Transmembrane transporter	724 aa	*Discopyge ommata*	gbQ90406
KIAA Hsa	KIAA 0736 protein	742 aa	*Homo sapiens*	gbAB018279
KIAB Hsa	KIAA 0735 protein	683 aa	*Homo sapiens*	gbAB018278
Sv2B Rno	Synaptic vesicle protein	683 aa	*Rattus norvegicus*	pirS34961
Sv2A Rno	Synaptic vesicle protein	742 aa	*Rattus norvegicus*	spQ02563
The Conjugated Bile Salt Transporter (BST) Family (TC #2.1.23)				
Bsh Ljo	Putative conjugated bile salt transporter	451 aa	*Lactobacillus johnsonii*	gbAF054971
Orf Ljo	Putative conjugated bile salt transporter (fragment)	279 aa	*Lactobacillus johnsonii*	gbAF054971
The Unknown Major Facilitator-1 (UMF1) Family (TC #2.1.24)				
YxiO Bsu	Hypothetical 47.3 kd protein in WAPA-LICT intergenic region	428 aa	*Bacillus subtilis*	spP42306
Orf Sce	Hypothetical 58.8 kd protein in GLK1-SRO9 intergenic region	528 aa	*Saccharomyces cerevisiae*	spP25568
Orf Spo	Hypothetical 58.6 kd protein in C2G11.13 in chromosome 1	529 aa	*Schizosaccharomyces pombe*	spQ09812
The Unknown Major Facilitator-2 (UMF2) Family (TC #2.1.26)				
YfkF Bsu	YfkF protein	391 aa	*Bacillus subtilis*	gbD83967
YcaD Eco	Hypothetical 41.4 kd Protein in DMSC-PFLA Intergenic Region	382 aa	*Escherichia coli*	spP21503
The Phenyl Propionate Permease (PPP) Family (TC#2.1.27)				
HcaT Eco	Putative Phenyl Propionate Uptake Permease	379 aa	*Escherichia coli*	spQ47142
YfhS Hin	Hypothetical Protein HI0308	388 aa	*Haemophilus influenzae*	spP44629
The Unknown Major Facilitator-3 (UMF3) Family (TC #2.1.28)				
Orf Hsa	C-receptor	555 aa	*Homo sapiens*	AF118637.1
Orf1 Cel	Weak similarity to *Bacillus* and *Pseudomonas* probable glucarate transporters (GI: 709999 and PIR:S27616)	623 aa	*Caenorhabditis elegans*	AF002196.1
YT45 Cel	Hypothetical 55.1 kd protein B0416.5 in chromosome X	507 aa	*Caenorhabditis elegans*	spQ11073
Orf2 Cel	C05G5.1	456 aa	*Caenorhabditis elegans*	gbZ70203.1
Orf3 Cel	CELC42C1	544 aa	*Caenorhabditis elegans*	gbAF043695.1
Orf4 Cel	Predicted using Genefinder	487 aa	*Caenorhabditis elegans*	gbCAB07317.1 Z92825
Orf5 Cel	Predicted using Genefinder	467 aa	*Caenorhabditis elegans*	CAB07312 Z92825
The Unknown Major Facilitator-4 (UMF4) Family (TC #2.1.29)				
Orf Ape	Hypothetical protein	369 aa	*Aeropyrum pernix*	AP000064
Orf1 Afu	Conserved hypothetical protein	388 aa	*Archaeoglobus fulgidus*	AE000946
Orf2 Afu	Conserved hypothetical protein (AF2103 and AF2102)	147 & 217	*Archaeoglobus fulgidus*	AE000958.1

similarity of the two yeast UMF1 proteins with established members of the GPH family (TC #2.2) which is distantly related to the MFS (see below), as well as with members of the DHA1 family of the MFS (TC #2.1.2). The functionally uncharacterized *B. subtilis* YxiO protein is distantly related to DHA1 family members. Thus, the UMF1 proteins comprise a novel family in the MFS.

An average hydropathy plot (not shown) was in agreement with a 12 TMS topology in a 4 + 2 + 6 arrangement. Thus, putative inter-TMS cytoplasmic loops 4-5 and 6-7 as well as extracytoplasmic loop 1-2 and the N- and C-termini of these proteins are the largest strongly hydrophilic portions of these proteins. Unlike most MFS families, a greater degree of sequence similarity was observed in the second halves of these proteins than for the first halves. Strongly amphipathic regions included the N- and C-termini as well as loops 1-2, 4-5, 6-7 and 10-11. Thus, all of the large hydrophilic portions of these proteins may be present in α-helical configuration. There is no indication as to the functions of these proteins, although of the various members of the MFS, they exhibit greatest sequence and motif similarity to sugar and drug transporters as noted above.

The Peptide-Acetyl-CoA Transporter (PAT) Family (TC #2.1.25)

Two members of the PAT family have been functionally characterized, but the precise biochemical functions of these proteins are not certain. One of these proteins is the putative Acetyl-Coenzyme A transporter found in the endoplasmic reticular and Golgi membranes of man (Kanamori *et al.*, 1997). It is homologous to proteins in *Caenorhabditis elegans*, *Saccharomyces cerevisiae* and

Table 7. Sequenced Members of the Peptide-Acetyl-CoA Transporter (PAT) Family (TC #2.1.25)

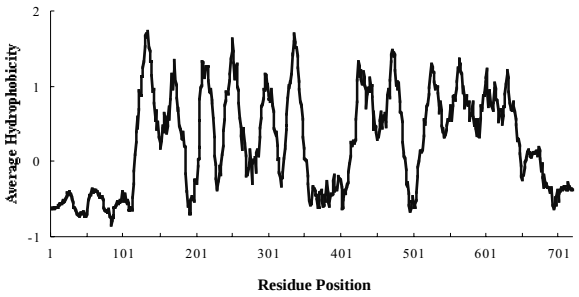
Abbreviation	Name or Database Description	Organism	Size (No. residues)	Database and Accession No.
AmpG Eco	Signal transducer encoded by *ampG*	*Escherichia coli*	491	spP36670
OrfX Ngo	Functionally uncharacterized OrfX	*Neisseria gonorrhoeae*	427	gbU82701
Orf3 Hin	Functionally uncharacterized protein HI0350 (Orf3)	*Haemophilus influenzae*	425	spP24326
AmpG1 Rpr	AmpG protein (AmpG1)	*Rickettsia prowazekii*	452	gbAJ235271
AmpG2 Rpr	AmpG protein (AmpG2)	*Rickettsia prowazekii*	408	gbAJ235272
AmpG3 Rpr	AmpG protein (AmpG3)	*Rickettsia prowazekii*	421	gbAJ235273
YbtX Ype	YbtX functionally uncharacterized protein	*Yersinia pestis*	455	gbAF091251
AcCoAT Hsa	Acetyl-coenzyme A:CoA antiporter	*Homo sapiens*	549	gbD88152
Orf Sce	Hypothetical 63KD protein (YBR220c)	*Saccharomyces cerevisiae*	560	spP38318
Orf Cel	Functionally uncharacterized Orf	*Caenorhabditis elegans*	538	gbZ50859

A

```
                                                *
Orf3   Hin (41)   KHLSIELIGAVTGVMLPYGLKFLWAPLLD
OrfX   Ngo (44)   EQVDLKSIGLMALIGLPFTWKFLWSPLMD
AmpG   Eco (41)   ENIDLKTIGFFSLVGQAYVFKFLWSPLMD
AmpG1  Rpr (43)   KDIALQTIGMLSFITLPYSINFLLAPVFD
AmpG3  Rpr (34)   AKYTTDIIGAISLAAFPYCLKVIWSPFID
AmpG2  Rpr (41)   SDFDKITIGLFGLVNFIHIFKFLWGPLLE
YbtX   Ype (75)   AGGSLALAGATTLFMLPWALKFIWAPWIE
Orf    Cel(160)   KHVSYGSQAIFSFAYWPFSLKLLWAPIVD
AcCoAT Has(101)   KNVSYTDQAFFSFVFWPFSLKLLWAPLVD
Orf    Sce (47)   KETSFTSLGIFSMATYPYSLKIIWSPIVD

Consensus         -------IG--S----P--LKFLW-P--D
```

B



C

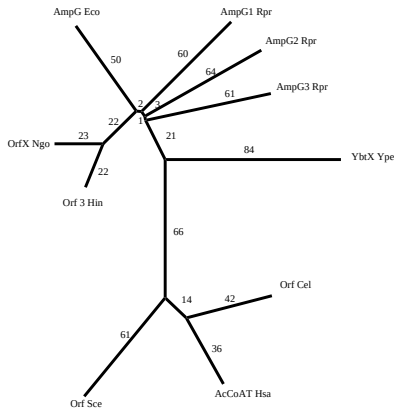


Figure 13. Partial multiple alignment (A), average hydropathy plot (B), and phylogenetic tree (C) for the peptide/acetyl CoA transporter (PAT) family.

several Gram-negative bacteria. The other of these proteins, the homologous *E. coli* AmpG protein, probably brings into the cell peptides, including cell wall degradative peptides and glycopeptides, which act as inducers of β-lactamase synthesis (Lindquist *et al.*, 1993; Jacobs *et al.*, 1994; Park *et al.*, 1998). In *Haemophilus influenzae*, the gene encoding a PAT family homologue is found in a gene cluster concerned with lipopolysaccharide synthesis. A homologue from *Neisseria gonorrhoeae* has also been sequenced. These proteins are of 425-632 amino acyl residues in length and exhibit 12 putative transmembrane α-helical spanners (TMSs). The mechanism of energy coupling is not absolutely established, but the topology of these proteins and their established inclusion in the MFS suggest that they are secondary carriers. The acetyl-CoA transporter is expected to function by Acetyl-CoA:CoA antiport while the AmpG protein is most likely energized by substrate:H^+ symport.

Table 7 presents the currently sequenced members of the PAT family. Members are derived from bacteria, yeast and animals. The prokaryotic proteins are smaller than the eukaryotic proteins by about 100 amino acyl residues (408-491 residues versus 538-560 residues). As noted above, the two functionally characterized proteins, AmpG of *E. coli* and AcCoAT of man probably transport cell wall peptides and Acetyl-Coenzyme A, respectively. Since Acetyl-CoA contains several secondary amide (peptide-like) bonds, the inclusion of a substrate such as Acetyl-CoA in a family of peptide transporters is not entirely surprising. *Rickettsia prowazekii* encodes 3 AmpG-like paralogues within its small (1.1 Mbp) genome (Andersson *et al.*, 1998) although other bacteria (*E. coli* and *H. influenzae* and the two sequenced eukaryotic genomes, *S. cerevisiae* and *C. elegans*, all with much larger genomes, only encode one. Most of the twelve bacteria for which fully sequenced genomes are available, and all of the four archaea with sequenced genomes do not encode a recognizable PAT family member.

A partial multiple alignment of the ten PAT family members is shown in Figure 13A. Only one residue in this alignment is fully conserved, but at several positions, substitutions are strictly conservative. A signature sequence derived from this portion of the complete alignment is as follows:

(G A) (L I V F M T A)$_2$ (S A G T) (L I V M F A G) X_3 (P A I) (F Y H W) X (L I V F W) (K N) (L I V F) (L I V) (W L) (G A S) P (L I V F W) (L I V F M) (D E)

The average hydropathy plot, based on the complete multiple alignment for the ten PAT family members, presented in Figure 13B, reveals 12 peaks, presumably corresponding to 12 TMSs in a 6 + 6 arrangement. PAT family permeases therefore exhibit the expected MFS topology.

The phylogenetic tree for the PAT family is shown in Figure 13C. While the orthologues from *H. influenzae* and *N. gonorrhoeae* cluster together, all other bacterial proteins are relatively distant from each other. Thus, the *E. coli* and *H. influenzae* proteins are too distantly related to be orthologues, and the three *R. prowazekii* paralogues are equidistant from each other and the *E. coli* homologue. The *R. prowazekii* paralogues presumably arose by gene duplication events that occurred a long time ago, possibly about the time when the α- (*R. prowazekii*) and γ- (*E. coli*) proteobacteria diverged from each other.

The three eukaryotic protein members of the PAT family are found on a branch distant from the prokaryotic proteins, and the branching patterns and relative phylogenetic distances are roughly consistent with the possibility that these three proteins in man, worm and yeast are orthologues. They may all be Acetyl-Coenzyme A:Coenzyme A antiporters found in the endoplasmic reticular membrane of the eukaryotic cell as has been shown for the human protein.

The Unknown Major Facilitator-2 (UMF2) Family (TC #2.1.26)

The UMF2 family consists of just two bacterial proteins (Table 6). One is the YcaD protein of *E. coli*, and the other is the YfkF protein of *Bacillus subtilis*. These two proteins, of 12 putative TMSs, are of unknown function. They show greatest sequence similarity to the cis, cis-muconate transporter, MucK of *Actinobacter* (TC #2.1.15.4) with lower sequence similarity to members of the sugar porter family (TC #2.1.1).

The Phenyl Propionate Permease (PPP) Family (TC #2.1.27)

The PPP family consists of a single poorly characterized protein which probably functions as a phenyl propionate permease in *E. coli* (Diaz *et al.*, 1998). A homologue is present in *Haemophilus influenzae* (Table 6). These proteins are of about 380 residues and exhibit 12 putative TMSs. The transport function of the *E. coli* protein was deduced from the nature of the 3-phenyl propionate catabolic operon, several of the encoded constituents of which were characterized functionally.

The Unknown Major Facilitator-3 (UMF3) Family (TC #2.1.28)

The UMF3 family consists of one human and six *C. elegans* proteins. The human protein is the cell surface receptor (c-receptor) for anemia-inducing feline leukemia virus subgroup C (Tailor *et al.*, 1999). Its transport substrate is unknown. Similarly, none of the *C. elegans* proteins are functionally characterized. These proteins are of 456-623 residues and exhibit the expected 12 TMSs.

The Unknown Major Facilitator-4 (UMF4) Family (TC #2.1.29)

The UMF4 family consists of three archaeal proteins, two from *Archaeoglobus fulgidus* and one from *Aeropyrum pernix*. The two full length proteins are of 369 and 388 residues and exhibit 12 putative TMSs. The third protein is reported as two distinct Orfs in *A. fulgidus*, probably due to a sequencing error. These proteins are functionally uncharacterized.

The Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family (TC #2.2)

The GPH family was first described in 1994 (Reizer *et al.*, 1994), but in 1996, Poolman *et al.* comprehensively reviewed the extensive literature concerning the cation and sugar selectivity determinants for this family (Poolman *et al.*, 1996). This family of permeases includes the well-characterized melibiose:Na^+ symporters of *E. coli*, *Salmonella typhimurium* and *Klebsiella pneumoniae* which can use Na^+, Li^+ and H^+ as the cotransported cation as well as the lactose permease of *Streptococcus thermophilus* which functions by sugar:H^+ symport. Mutants were described in which the cation and/or sugar substrate specificities of the permeases were altered, or in which sugar transport was uncoupled from cation cotransport. Most of the mutations proved to occur in the N-terminal halves of the permease proteins, particularly in or near the putative amphipathic transmembrane helices (TMS) 2 and 4 although some occurred in the inter-TMS loop 10-11 in the second halves of these proteins. Subsequently, Wilson and Wilson (1998) described compensatory double mutations that led them to propose that helices 4 and 11 are in close proximity and may comprise part of the active site.

A dendogram of the most studied protein members of the GPH family that were then available revealed three clusters, first the lactose/raffinose permeases of Gram-positive bacteria, second, the melibiose permeases of enteric Gram-negative bacteria, and third, all remaining proteins (glucuronide and xyloside transporters) from both Gram-negative and Gram-positive bacteria. Naderi and Saier (1996) subsequently provided evidence that the well-characterized and physiologically important sucrose:H^+ symporters of plants are distant members to this family, and additional computational analyses revealed that sequence similarity with various established members of the MFS could be observed. In fact, PSI-BLAST results clearly suggest that the GPH family exhibits conserved motifs in common with MFS proteins, and we therefore consider it highly likely that these two families of permeases share a common origin. Because of the extensive sequence and phylogenetic analyses reported by Poolman *et al.* (1996), no further analyses will be reported here. Poolman *et al.* (1996) believed that members of the GPH family transport pentoses, and they therefore designated the family the galactoside-pentose-hexuronide family. However, in a recent report, the substrate specificity of XylP, the isoprimeverose permease of *Lactobacillus plantarum*, was clarified (Chaillou *et al.*, 1998). This protein was shown to be highly specific for isoprimeverose, an α-xyloside, and the parental sugar, D-xylose, was not transported. Thus, contrary to the suggestion of Poolman *et al.* (1996) these permeases do not transport free pentoses, and the correct name of the family is the galactoside-pentoside-hexuronide family.

The Proton-Dependent Oligopeptide Transporter (POT) Family (TC #2.17)

Proteins of the POT family (Paulsen and Skurray, 1994) (also called the PTR [peptide transport] family) (Steiner *et al.*, 1995) consist of proteins from animals, plants, yeast and both Gram-negative and Gram-positive bacteria. Several of these organisms possess multiple POT family paralogues. The proteins are of about 450-600 amino acyl residues in length with the eukaryotic proteins in general being longer than the bacterial proteins. They exhibit 12 putative or established transmembrane α-helical spanners. Some members of the POT family exhibit limited sequence similarity to protein members of the major facilitator superfamily (MFS; TC #2.1) (comparison scores of up to 8 standard deviations for segments in excess of 60 residues in length). Thus the POT family is probably a family within the MFS (Pao *et al.*, 1998; Saier *et al.*, 1999).

While most members of the POT family catalyze peptide transport, one is a nitrate permease and one can transport histidine as well as peptides. Some of the peptide transporters can also transport antibiotics. These proton symporters thus transport a wide range of compounds.

The phylogeny of the POT family has recently been published (Saier *et al.*, 1999), and consequently detailed analyses will not be reported here. However, the proteins of the POT family proved to cluster into four easily distinguishable clusters. Cluster 1 contained all bacterial proteins, cluster 2 contained all animal proteins, cluster 3 contained all yeast proteins plus one plant protein, and cluster 4 contained all remaining plant proteins. These facts suggest that POT family members have diverged from a common ancestor primarily due to speciation and late gene duplication events. The reader is referred to Saier *et al.* (1999) as well as our web site for more detailed information about this family as well as references to the primary literature.

The Organoanion Transporter (OAT) Family (TC #2.60)

PSI-BLAST results with a single iteration suggested that the OAT family represents a distant familial constituent of the MFS. Table 8 provides the current protein members of this family. Proteins of the OAT family catalyze the Na^+-independent facilitated transport of organic anions such as bromosulfobromophthalein and prostaglandins as well as conjugated and unconjugated bile acids (taurocholate and cholate, respectively) (Hakes and Berezney, 1991; Jacquemin *et al.*, 1994; Kanai *et al.*, 1995; Hagenbuch, 1997; Abe *et al.*, 1998; Chan *et al.*, 1998; Schuster, 1998). These transporters are found exclusively in animals. Some exhibit a high degree of tissue specificity. For example, the rat OAT is found at high levels in liver and kidney, and at lower levels in other tissues. These proteins consist of 643-809 amino acyl residues with one exception (Table 8) and possess 10-12 putative α-helical transmembrane spanners. They may catalyze electrogenic anion uniport or anion exchange.

Figures 14A and B present two portions of the multiple alignment of the OAT family. The first region corresponds to putative TMS6 and represents the most conserved region within the complete multiple alignment (see Figure 15B). From this region a signature sequence for the OAT family was derived as follows:

D X_2 (W F) (L I V) G (A M C) W W (L I V F) (G S) (F L) (L I V) (L I V A) (S A C F) (G A S).

The second region shown in Figure 14B corresponds to an unusual hydrophilic, cysteine-rich region that occurs between putative TMSs 9 and 10 (see Figure 15A). Because this loop is predicted to be localized to the extracellular milieu, and is therefore in an oxidizing environment, one can predict that the conserved cysteine residues are oxidized primarily to cystine residues. Thus, this extracellular domain undoubtedly contains disulfide bridges. Within this loop region ten fully conserved cysteine residues plus one nearly conserved cysteine residue are found. Six of these cysteine residues are portrayed in Figure 14B. One can therefore suggest that this extracellular domain of about 120 residues is extensively cross-linked by disulfide bonds. We suggest that this region serves as an extracellular receptor domain as has been demonstrated for the cysteine-rich extracellular domains of epithelial Na^+ channel (ENaC) family members (TC #1.2) (Le and Saier, 1996). This suggestion implies that the OAT transporters may be regulated by extracellular molecules or stimuli.

The average hydropathy and similarity plots for the OAT family are shown in Figures 15A and B, respectively. It can be seen that 12 hydrophobic peaks are observed in

Table 8. Proteins of the Organo Anion Transporter (OAT) Family (TC #2.60)

Abbreviation	Name and Description	Organism	# Residues	Accession #
Pgt Rno	Prostaglandin transporter (PGT) matrin F/G	*Rattus norvegicus*	643	spQ00910
Pgt Hsa	Prostaglandin transporter (PGT)	*Homo sapiens*	643	spQ29259
Orf1 Hsa	KIAA0880 protein	*Homo sapiens*	709	gbAB020687
OatP Rno	Sodium-independent organic anion transporter	*Rattus norvegicus*	670	spP46720
OatP Hsa	Sodium-independent organic anion transporter	*Homo sapiens*	670	spP46721
OatB Rno	Sodium-independent organic anion transporter	*Rattus norvegicus*	661	spO35913
Orf2 Rno	Organic anion transporter 3	*Rattus norvegicus*	670	gbAF041105
OatK1 Rno	Organo anion transporter K1	*Rattus norvegicus*	669	gbAB020687
Orf3 Rno	Similarity to rat prostaglandin transporter	*Rattus norvegicus*	674	gbZ81016
Orf4 Cel	Predicted using genefinder	*Caenorhabditis elegans*	690	gbAL032660
Orf5 Cel	Coded for by C. elegans cDNA	*Caenorhabditis elegans*	1451	gbU39993
Orf6 Cel	CDNA EST EMBL:D68039	*Caenorhabditis elegans*	544	gbAL021475
Orf7 Cel	Similar to zinc-finger DNA-binding protein	*Caenorhabditis elegans*	655	gbU40415
Orf8 Cel	Similar to matrin F/G	*Caenorhabditis elegans*	809	gbU40953

A

```
                          *    * **
Pgt    Rno   (245) NLSPGDPRWIGAWWLGLLISSGFLIVTSLPFFFFP
Pgt    Hsa   (245) NLVPGDPRWIGAWWLGLLISSALLVLTSFPFFFFP
Orf1   Hsa   (264) SLTIKDPRWVGAWWLGFLIAAGAVALAAIPYFFFP
OatP   Rno   (233) TITPSDTRWVGAWWIGFLVCAGVNILTSIPFFFLP
Orf2   Rno   (233) TITPTDTRWVGAWWIGFLICAGVNILSSIPFFFFP
OatB   Rno   (232) TITPTDTRWVGAWWIGFLVCAGVNILTSFPFFFFP
OatP   Hsa   (233) IITPTDTRWVGAWWFGFLICAGVNVLTAIPFFFLP
OatK1  Rno   (233) TITPTDIRWVGAWWIGFLVCAGVNILISIPFFFFP
Orf4   Cel   (252) PMERSDPRWVGAWWVGFIISSISALMIAFPILAFA
Orf3   Rno   (259) HIGTHDEHWIGAWWLGFLVCGSAYLILAVPFFFFP
Orf8   Cel   (323) SSGETDPTWVGAWWLSFIAASFVGFVAVLPLASLP
Orf7   Cel   (251) IDNSADPRFIGMWWIGFVVCGFVALFTAFPLIMFP
Orf6   Cel   (300) GLTPLDPMWIGCWWLGFLIFGTLLFGPSLVLYFFP

Consensus          --TP-DPRWVGAWW-GFLIC-GV--L-S-PFFFFP
```

B

```
                          *   * *                          * ***
Pgt    Rno   (444) CRRDCSCPDSFFHPVCG  DNG  VEYVSPCHAGC
Pgt    Hsa   (444) CRRDCSCPDSIFHPVCG  DNG  IEYLSPCHAGC
Orf1   Hsa   (489) CMEACSCPLDGFNPVCD  PST RVEYITPCHAGC
OatP   Rno   (439) CNTRCSCSTNTWDPVCG  DNG  VAYMSACLAGC
Orf2   Rno   (439) CNRGCSCSTNSWDPVCG  DNG  LAYMSACLAGC
OatB   Rno   (438) CNTRCNCSTNTWDPVCG  DNG  LAYMSACLAGC
OatP   Hsa   (439) CNVDCNCPSKIWDPVCG  NNG  LSYLSACLAGC
OatK1  Rno   (439) CNTRCSCLTKTWDPVCG  DNG  LAYMSACLAGC
Orf4   Cel   (453) CNADCHCKME WNPVCD  RNT GHMYYSACHAGC
Orf3   Rno   (470) CLEYCNCETVLKFDGVS  YNG  QNFYSPCHAGC
Orf8   Cel   (600) CNKQCTCDPSEYRPVCAELDDGRQFTYYSPCYAGC
Orf7   Cel   (445) CSENCHC DSFFNPVCS  EDS KLTFLSPCHAGC
Orf7   Cel   (512) CRDDCMCEQTPLYPVCD  VSG  SAYYSPCHAGC

Consensus          CN--C-C------PVCG--DNG----Y-SPCHAGC
```

Figure 14. Two partial multiple alignments (A and B) of the proteins of the organo anion transporter (OAT) family.

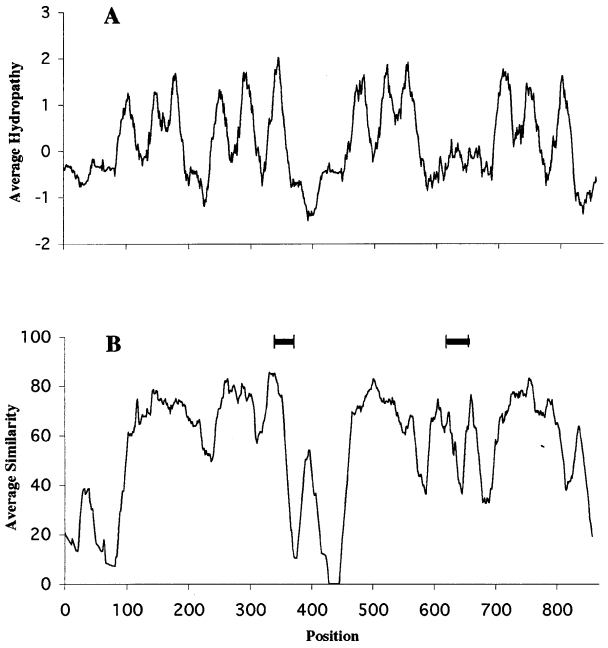


Figure 15. Average hydropathy (A) and similarity (B) plots for the proteins of the organo anion transporter (OAT) family. The bars in Figure B are the regions of the complete multiple alignment shown in Figure 14.

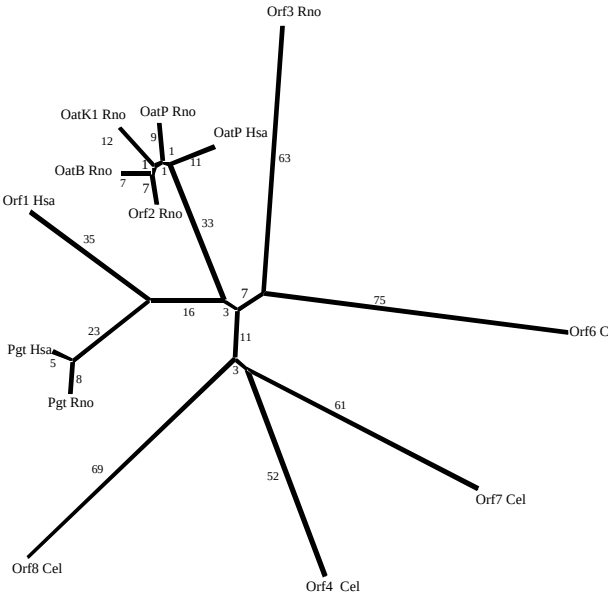


Figure 16. Phylogenetic tree for the proteins of the organo anion transporter (OAT) family.

Figure 15A, all of which are well conserved, as shown in Figure 15B. The two regions of the multiple alignment represented in Figures14A and B are shown by the dark bars in Figure 15B. In the latter figure, it can be seen that the N- and C-termini as well as the central loop separating TMS6 from TMS7 are poorly conserved as is often observed for eukaryotic members of the MFS. However, the putative extracellular receptor domain separating putative TMSs 9 and 10 includes regions that are well conserved. This fact further suggests that this region is of functional significance.

The phylogenetic tree for the proteins of the OAT family is reproduced in Figure 16. There are eight major branches, four represented by proteins from *C. elegans*, and four including proteins derived exclusively from mammals. Six of these mammalian proteins are derived from the rat, and they fall into three distinct clusters. The three human homologues similarly fall into three distinct clusters. The close Pgt orthologues undoubtedly serve the same function of prostaglandin transport in rats and humans, respectively. The cluster of five rat and one human organo anion transporters undoubtedly serve very similar biochemical functions. No function can be predicted for the dissimilar Orf3 of *R. norvegicus* or the distant *C. elegans* homologues (Figure 16).

The Folate-Biopterin Transporter (FBT) Family (TC #2.71)

PSI-BLAST searches suggested that the FBT family, with members characterized in protozoa (Gottesdiener, 1994; Moore and Beverley, 1996), is a distant constituent of the MFS. Protein members of the FBT family are listed in Table 9. These proteins are from plants and cyanobacteria as well as protozoa. While the protozoan proteins are reported to be large (627-704 amino acyl residues), the plant and cyanobacterial proteins are much smaller (408-494 residues).

Table 9. Sequenced Members of the Folate-Biopterin Transporter (FBT) Family (TC #2.71)

Abbreviation	Name or Database Description	Organism	Size (No. residues)	Database and Accession No.
BT1 Ldo	BT1 biopterin/folate (not methotrexate) transporter	*Leishmania donovani*	627	gbL38571
FT1 Ldo	FT1 folate/methotrexate (not biopterin) transporter	*Leishmania donovani*	704	No Acc. #
Orf Lin	Integral membrane protein	*Leishmania infantum*	627	gbL25643
BT1 Lme	Biopterin transporter	*Leishmania mexicana*	631	gbAF078929
FT1 Tbr	FT1 (ESAG10) folate/biopterin transporter	*Trypanosoma brucei*	686 (*597)	pirS33475
Orf1 Ath	Similar to Synechocystis integral membrane protein	*Arabidopsis thaliana*	431	gbAC002376
Orf2 Ath	Functionally uncharacterized Orf	*Arabidopsis thaliana*	408	gbAC002332
Orf3 Ath	Putative integral membrane protein	*Arabidopsis thaliana*	429	gbAC006223
Orf Sco	Functionally uncharacterized Orf	*Synechococcus PCC7942*	453	gbAF055873
Orf Scy	Integral membrane protein	*Synechocystis PCC6803*	494	gbD64002

An extended portion of the complete multiple alignment of the FBT family members is shown in Figure 17A. Several residues are fully conserved and many more residues appear in the consensus sequence. The signature sequence for the FBT family is:

(S A C) X (L I V M) (A C) P X G X E (S A G) X (L I V) (F Y T) (A S) (L I V F T) (L M) (A M) (S G)

The average hydropathy plot (Figure 17B) indicates that the FBT family proteins exhibit 12 or 13 putative TMSs.

The phylogenetic tree (Figure 17C) shows clustering generally in accordance with the phylogenies of the organisms. Thus the proteins from protozoa cluster together as do the plant proteins and the cyanobacterial proteins. One exception is the plant protein (Orf3 Ath) which clusters with the cyanobacterial proteins. It would be predicted on this basis to be a chloroplast protein.

The Putative Bacteriochlorophyll Delivery (BCD) Family (TC #97.7)

Table 10 presents the seven currently sequenced members of the putative BCD family. All of these proteins are of about the same size, and, as shown below, exhibit similar topological features. The function of none of these proteins is established. Use of the PSI-BLAST program clearly suggested that these proteins are distant members of the MFS. The suggestion that some of them are pigment synthases (Table 10) is likely to be in error. Several of these proteins have been shown to be essential for normal photosynthetic activity (Youvan *et al.*, 1984; Zsebo and Hearst, 1984; Tichy *et al.*, 1989; Gibson *et al.*, 1992).

Although none of the members of the BCD family is functionally characterized, the topology of two of them (PucC Rca (LeBlanc and Beatty, 1996) and YpuM Rca [recently renamed LhaA] (Young and Beatty, 1998)) have been experimentally determined. As expected for members of the MFS, they exhibit a 12 TMS topology with both the N- and C-termini facing the cytoplasm. LhaA has recently been speculated to be a bacteriochlorophyll "delivery" export permease (Young and Beatty, 1998), and this proposal provides the basis for naming the BCD family.

Figure 18 shows two well conserved portions of the complete multiple alignment for the BCD family proteins. These two regions correspond to the ends of TMSs 1 and 7 as well as the loop regions between TMSs 1 and 2, and 7 and 8, respectively. Limited sequence similarity between these two aligned segments can be detected, and this similarity presumably reflects the ancient gene duplication event which is believed to have given rise to all members of the MFS (see Pao *et al.*, 1998 for discussion of the published evidence).

Two signature sequences were derived from the two well conserved portions of the complete multiple alignment shown in Figure 18. These sequences are:

SS #1: L N R [L I V] [M L] [L I V] X E L X [L I V]

SS #2: [D E] [L I V A] [L I V] L E P [Y F] [G A] G

When these sequences were screened against the SwissProt database, they retrieved only established members of the BCD family, thus showing that by this criterion, they are authentic signature sequences for this family.

Two of the seven members of the BCD family have been shown experimentally to exhibit a 12 TMS topology. This fact is in agreement with the average hydropathy plot shown in Figure 19A. Thus, 6 peaks of hydrophobicity, followed by a large hydrophilic "loop" region is then followed by 6 additional hydrophobic peaks. Each of these peaks is of sufficient magnitude and length to span the membrane as an α-helix. As noted for several other MFS families such as the SET family (TC #2.1.20) (see above), the average

Table 10. Members of the Putative Bacteriochlorophyll Delivery (BCD) Family (TC #97.7)

Abbreviation	Name or Description	Organism	Length	Accession no.
PucC Rsu	Putative regulatory protein PucC	*Rhodovulum sulfidophilum*	454	spP95656
PucC Rsp	PucC protein	*Rhodobacter sphaeroides*	459	spQ02443
Orf Rru	Hypothetical protein GF115	*Rhodospirillum rubrum*	480	pirB61213
PucC Rca	PucC protein	*Rhodobacter capsulatus*	461	spP23462
YpuM Rca	Hypothetical 50.4 KD protein	*Rhodobacter capsulatus*	477	spP26176
Bch2 Rca	Bacteriochlorophyll synthase	*Rhodobacter capsulatus*	428	spP26171
Orf Ssp	Bacteriochlorophyll synthase	*Synechocystis sp.*	484	gbD90910

A

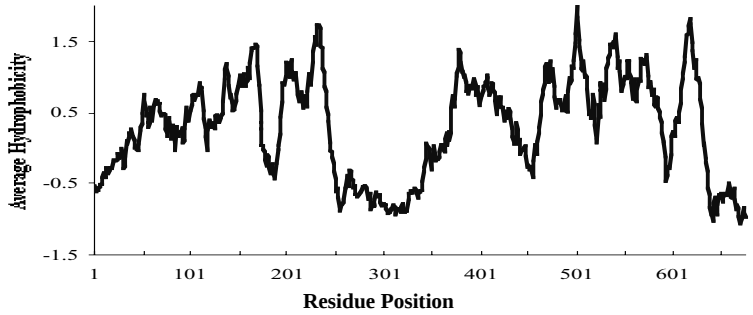
```
                                *                       *
BT1   Ldo (440) VFTHLFPHHSYRFVMGLSAVLLPAASMFDLLILKRWNLVIGIPDHAMYILGDAI
BT1   Lme (444) LFTHLFPNYSYRLVMGLSAVLLPAASMFDVVILKRWNLAIGIPDHAMYIFGDAI
FT1   Ldo (485) LFNFLFAKHGYRLTFIVTTIMQVLAALFDIIMVKRWNLYIGIPDHAMYIWGDAV
FT1   Tbr (410) LFRYVFSKRSYRLTFIVTTLIEIVSSIFDIIIVERWNRPY VSDHVVFVLGDQI
Orf-1 Ath (227) VYDRYWKKLPMRALIHIVQLLYAFSLLFDYILVKQINLAFGIS NTAFVLCFSS
Orf-2 Ath (245) VYDRYLKTLPMRPLIHIIQLLYGLSILLDYILVKQINLGFGIS NEVYVLCFSS
Orf   Sco (289) IFQRFLRGVPIRRIFGWMIVVTTLLGLTSLILVTHLNRSWGISDQ WFSLGDSL
Orf   Scy (297) LYQRFLKTLPFRVIMGWSTVISSLLGLTTLILITHANRAMGIDDH WFSLGDSI
Orf-3 Ath (270) LYNGFLKTVPLRKIFLVTTIFGTGLGMTQVILVSGFNRQLGISDE WFAIGDSL

Consensus       LF-------P-R--------------LFD-ILVK--NL--GISDH--F-LGDS-

                          *         * * *                       *
BT1   Ldo       IYEVCDMLLNMPMMMLMCRIAPRGSESMVFALLASIYHLGTSTSSAIGYLLMET
BT1   Lme       IYEVCNMLLNMPMMMLMCRIAPRGSESMVFALLASIYHLGTSTSSAIGYLLMET
FT1   Ldo       VGEIVYMLGFMPQIVLLSRLCPRGSESVVYALMAGFARLGRTTAASLGAILLEY
FT1   Tbr       IHQVCYMMHFMPTVILISRLCPSGYESAVYSVLAGCAHFGRSVSNTLGWLLMEY
Orf-1 Ath       VAEILAQFKILPFSVLLANMCPGGCEGSITSFLASTLCLSSVVSGFTGVGMANM
Orf-2 Ath       LAEILAQFKILPFAVRLASMCPQGCEGSVTSFLASTLCLSQIVSAFLGVGLANL
Orf   Sco       ILTVAGQLSFMPVLILAARLCPSGIEATLFALLMSVLNLAHFGSVELGALLTHW
Orf   Scy       ILTVTGQIAFMPVLVLAARLCPPGIEATLFALLMSVMNLAGVLSFEVGSLLTHW
Orf-3 Ath       ILTVLAQASFMPVLVLAARLCPEGMEATLFATLMSISNGGSVLGGLMGAGLTQA

Consensus       I-EV--Q--FMP---L-ARLCP-G-E--VFALLAS---LG---S---G-LL---
```

B



C

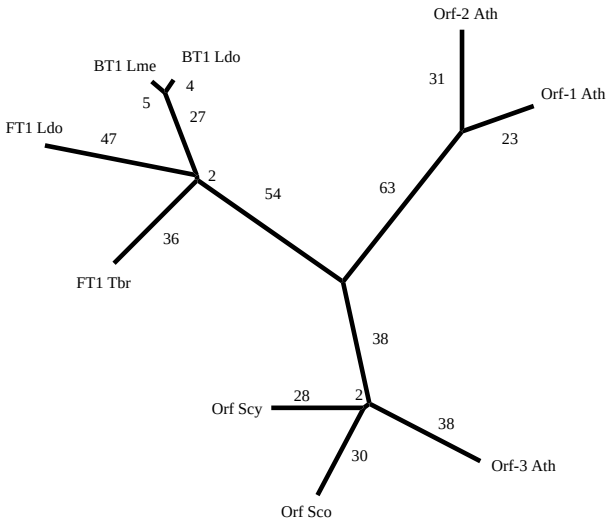


Figure 17. An extended portion of the complete multiple alignment (A), average hydropathy plot (B), and phylogenetic tree (C) for the proteins of the folate-biopterin transporter (FBT) family.

A

```
                    **   *   *    *    ***    **
Orf Rru      (37)  RLSLFQVTVGMAGVLLTGTLNRVMIVELGVPT
Orf Ssp      (22)  RLGLFQMGLGIMSLLTLGVLNRVLIDELAVLP
PucC Rca     (36)  RLSLFQITVGMTLTLLAGTLNRVMIVELAVPA
YpuM Rca     (33)  RLSLFQVSVGMAQVLLLGTLNRVMILELGVPA
PucC Rsu     (37)  RLSMFQVSVGMAMVLLVGTLNRVMIVELEVPA
PucC Rsp     (33)  RLSLFQVAVGMAIVLLVGTLNRVMIVELKVPA
Bch2 Rca     (10)  RLGLVQLCIGAVVVLTTSTLNRLMVVELALPA

Consensus          RLSLFQV-VGMA-VLL-GTLNRVMIVELAVPA
```

B

```
                       ***  *         ** *      *
Orfr Ru      (280) DILLEPYGGEILHLSVGATTMLTAMMATGTLV
Orf Ssp      (293) DAVLEPYGGEVFNLCISETTQLNAFFGMGTLL
PucC Rca     (282) DVLLEPYGGQALHLTVGETTKLTALFALGTLA
YpuM Rca     (276) DVLLEPYGGQVLGLKVGQTTWLTAGWAFGALV
PucC Rsu     (284) DVLLEPFGGQVLDMSVAATTKLTAAVAGGTLV
PucC Rsp     (277) DVILEPYGGEVLSMTVAETTRLTATFAGGGLV
Bch2 Rca     (241) ELILEPYAGLVFGFTAGETTKLSGMQNGGVFF

Consenus           DVLLEPYGG-VL-LTVGETTKLTAMFAGGTLV
```

Figure 18. Two well conserved portions of the complete multiple alignment for the seven sequenced proteins of the bacteriochlorophyll-delivery (BCD) family. Numbers in parentheses following the protein abbreviation give the first residue in each line. Abbreviations of the proteins are presented in Table 10. Fully conserved residues are indicated by asterisks and are presented in bold print. The consensus sequence (consensus) (4 of 7 residues conserved) is presented at the bottom of the alignment.

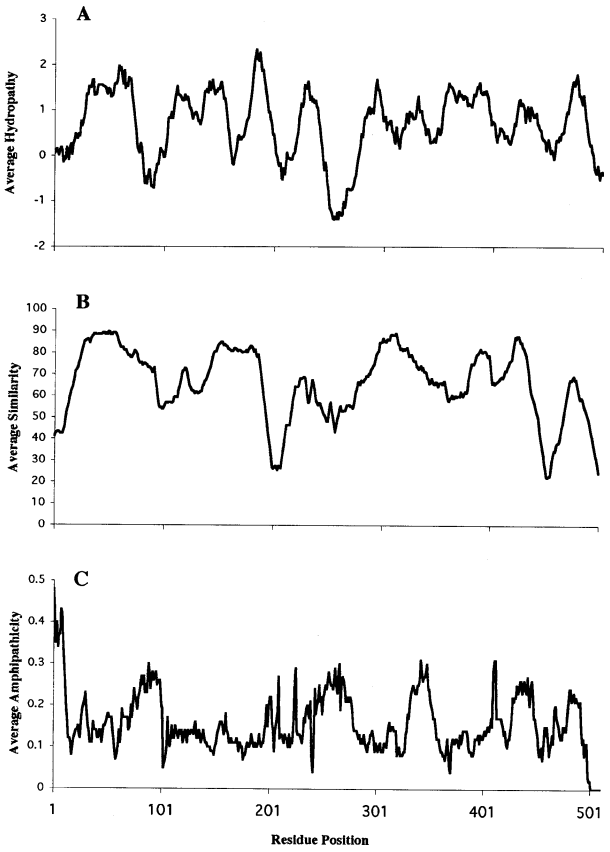


Figure 19. Average hydropathy (A), similarity (B), and amphipathicity (100° for α-helix; C) for the proteins of the bacteriochlorophyll-delivery (BCD) family.

hydropathy plot shown in Figure 19A suggests that the cytoplasmic loops connecting TMSs are in general longer than the extracytoplasmic loops. Thus, while peaks 1 and 2, 3 and 4, 7 and 8, and 9 and 10 are close to each other, peaks 5 and 6, and 11 and 12 are not.

The average similarity plot (Figure 19B) reveals that TMSs 1 and 2 as well as the connecting loop region, and the homologous TMSs 7 and 8 as well as their loop region are the best conserved portions of these proteins. However, the loop region between TMSs 4 and 5 as well as TMS 5 is also well conserved. In this connection it is interesting to note that the homologous TMS 11 is also well conserved, but that the loop region between TMSs 10 and 11 is not as well conserved as that between TMSs 4 and 5. This fact is consistent with our earlier observation, confirmed by the profile shown in Figure 19B, showing that the first halves of MFS proteins are generally better conserved that the second halves (Marger and Saier, 1993).

The average amphipathicity plot shown in Figure 19C reveals that all major peaks of amphipathicity (when plotted for an α-helix) occur before, in between, or following the 12 TMSs. Strikingly, the large peaks between TMSs 2 and 3 and TMSs 8 and 9 occur in corresponding positions of the two halves of the proteins. The large peak of amphipathicity observed at the beginning of the alignment (Figure 19C) is poorly conserved, and, in fact, was observed for only one member of the family.

The phylogenetic tree for the BCD family proteins is shown in Figure 20. The three PucC proteins are closely related as are the YpuM Rca and Orf Rru, suggesting that these two clusters each consist of orthologues serving the same function. The last two proteins (Orf Sce and Bch2 Rca) are distant members of the family. No correspondence of function can be proposed for these proteins.

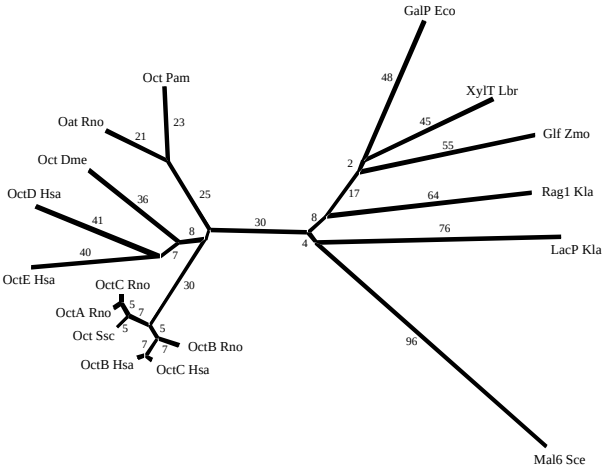


Figure 20. Phylogenetic tree for the seven members of the bacteriochlorophyll-delivery (BCD) family.

A Subfamily of Vesicular Monoamine Transporters (VMAT) Within the Drug:H^+ Antiporter-1 (DHA1) Family of the MFS (TC #2.1.2)

Within the DHA1 family of drug exporters is a number of transporters that are capable of transporting either drugs or neurotransmitters (Table 11). Phylogenetic analyses reported below have shown that these proteins form two distinct clusters within the DHA1 family (Paulsen *et al.*, 1996). They are all derived from animals and may be localized to neurotransmitter-containing vesicles. However, because many of these transporters have been shown to transport drugs, and therefore exhibit overlapping specificities with other members of this family, we have retained this group of proteins within the DHA1 family. We thus classify this subfamily the VMAT subfamily of the DHA1 family of the MFS.

Two regions of the VMAT family are exceptionally well conserved as shown in Figures 21A and B. These two regions encompass putative TMSs 5-6 and TMSs 10-11 (see Figure 22C). From these two partial alignments, two VMAT subfamily-specific signature sequences were derived. They are:

(A) L X_2 V X_2 A X L L D N M L X_2 V X V P I X P

(B) L V D X R X_2 S V Y G S X Y A I A D

The average hydropathy plot for the VMAT family is shown in Figure 22A. Twelve peaks of hydrophobicity are apparent, and these presumably correspond to TMSs 1-12 as expected for most members of the MFS. The average similarity plot (not shown) revealed that the second halves of these proteins are better conserved than the first halves, and that within each of these two halves, the last two TMSs are best conserved. Gaps in the multiple alignment are in part responsible for the unusually large spacing between putative TMS1 and 2. The C-termini of these proteins are also poorly conserved.

An average amphipathicity plot (100° as for an α-helix; not shown) revealed that the largest peak of amphipathicity preceded TMS1, but several additional smaller peaks were present, particularly between TMSs 2 and 3, 3 and 4, 6 and 7, 8 and 9 and following TMS12. Thus, it can be suggested that much of these proteins, both the transmembrane regions and the inter-TMS loop regions, assume α-helical configurations.

The phylogenetic tree for the VMAT family (Figure 22B) reveals three clusters of mammalian paralogues, one of which includes a more distant homologue from *Torpedo marmorata*. The *C. elegans* and *D. melanogaster* proteins are distant to all of the mammalian proteins. However, the clustering of these proteins into two major groups suggests two functional types. The VMAT cluster is probably concerned with monoamine transport while the Unc cluster is concerned with acetylcholine transport. Thus, we suggest that the functionally uncharacterized proteins from *C. elegans* and *D. melanogaster* are acetylcholine transporters.

Conclusions and Perspectives

The present study of the major facilitator superfamily (MFS), the largest superfamily of secondary carriers found in nature, reveals that it is substantially larger and more diverse than was recognized in 1998. The MFS includes 29 established and five additional probable families as compared with only 17 families recognized in 1998 (Pao *et al.*, 1998). If one considers the "extended" superfamily , including the five distantly related families (see bottom of Table 1), there is a 2x increase in family representation. In addition to the compounds that were then recognized as substrates of MFS permeases, we now know that one family within the MFS (SIT) can take up iron siderophores in yeast, that a second family (SET) can efflux sugars in bacteria, and that two families, the VNT and VMAT families, most closely related to the SP family and part of the DHA1 family, respectively, function in neurotransmitter transport. Two eukaryotic families within the extended MFS, the OCT and OAT families, are concerned with transport of organo cations and anions including a variety of drugs and toxic substances. Two families within the extended MFS (PAT and POT) transport peptides, and both of these families include members that transport a range of compounds in addition to peptides. Thus, PAT family members probably transport acetyl-CoA, coenzyme A, and glycopeptides in addition to peptides, while POT family members transport nitrate, chlorate, an amino acid (histidine) and various antibiotics in addition to peptides. Both bacterial and eukaryotic MFS permeases, belonging to different families, transport conjugated bile salts. Vitamins and their precursors are also likely substrates of a distant MFS family (the FBT family).

Table 11. Sequenced Vesicular Monoamine Transporter (VMAT) Subfamily of the Drug:H^+ Antiporter-1 (DHA1) Family (TC #2.1.2)

Abbreviation	Protein Description	Size	Organism	Accession #
Vmat2 Bta	Bovine synaptic vesicular monoamine transporter	517 aa	*Bos taurus*	spQ27963
Orf Cel	Similar to synaptic vesicle amine transporter	319 aa	*Caenorhabditis elegans*	gbU41508
Unc17 Cel	Vesicular acetylcholine transporter	532 aa	*Caenorhabditis elegans*	spP34711
Unc17 Dme	Vesicular acetylcholine transporter	578 aa	*Drosophila melanogaster*	gbAF030197
Vmat1 Hsa	Human chromaffin granule monoamine transporter	525 aa	*Homo sapiens*	spP54219
Vmat2 Hsa	Human synaptic vesicular monoamine transporter	514 aa	*Homo sapiens*	spQ05940
Unc17 Hsa	Human vesicular acetylcholine transporter	532 aa	*Homo sapiens*	pirI38658
Unc17 Mmu	Vesicular acetylcholine transporter	530 aa	*Mus musculus*	gbAF019045
Vmat1 Rno	Rat vesicular chromaffin granule monoamine transporter	521 aa	*Rattus norvegicus*	spQ01818
Sv2 Rno	Synaptic vesicle monoamine transport protein	515 aa	*Rattus norvegicus*	pirB43319
Vmat2 Rno	Vesicular monoamine transport protein	515 aa	*Rattus norvegicus*	spQ01827
Unc17 Rno	Vesicular acetylcholine transporter	530 aa	*Rattus norvegicus*	gbU09838
Unc17 Toc	Vesicular acetylcholine transporter	511 aa	*Torpedo ocellata*	pirS43686

A

```
Vmat2 Hsa (23) LFIVFLALLLDNMLLTVVVPIIP
Vmat2 Bta (23) LFIVFLALLLDNMLLTVVVPIIP
Vmat2 Rno (23) LFIVFLALLLDNMLLTVVVPIIP
Vmat1 Hsa (24) LVVVFVALLLDNMLFTVVVPIVP
Vmat1 Rno (24) LVVVFVALLLDNMLLTVVVPIVP
Unc17 Hsa (36) LVIVCVALLLDNMLYMVIVPIVP
Unc17 Mmu (36) LVIVCVALLLDNMLYMVIVPIVP
Unc17 Toc (39) LVIVCIAMLLDNMLYMVIVPIIP
Unc17 Dme (35) LVIVSIALLLDNMLYMVIVPIIP
Unc17 Cel (34) LVIVSIALLLDNMLYMVIVPIIP
```

B

```
Vmat2 Hsa (409) LVDLRHVSVYGSVYAIADV
Vmat2 Bta (412) LVDLRHVSVYGSVYAIADV
Vmat2 Rno (410) LVDLRHVSVYGSVYAIADV
Vmat1 Hsa (417) LVDLRHTSVYGSVYAIADV
Vmat1 Rno (414) LVDLRHTSVYGSVYAIADV
Unc17 Hsa (408) LVDVRHVSVYGSVYAIADI
Unc17 Mmu (408) LVDVRHVSVYGSVYAIADI
Unc17 Toc (389) LVDIRYVSVYGSVYAIADI
Unc17 Dme (382) LVDVRYVSVYGSIYAIADI
Unc17 Cel (387) LVDTRHVSVYGSVYAIADI
```

Figure 21. Two well conserved regions of the complete multiple alignment (A and B) for the vesicular monoamine transporter (VMAT) subfamily of the DHA1 family of the MFS.

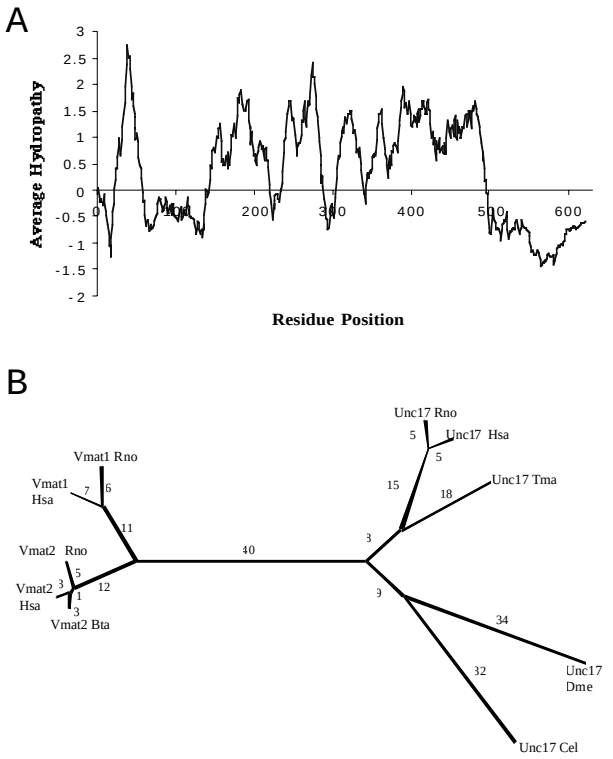


Figure 22. An average hydropathy plot (A) and a phylogenetic tree (B) for the vesicular monoamine transporter (VMAT) subfamily of the DHA1 family of the MFS.

Finally, four of the novel MFS families (UMF1-4) are not functionally defined. Consequently, we can anticipate that the range of substrates transported by MFS permeases will continue to expand. As more genomes become sequenced and published, all currently recognized families will undoubtedly expand in size and functional diversity, and new families will be discovered. We can predict that the currently recognized UMF families as well as novel, yet-to-be-discovered families will exclusively transport small to medium sized molecules. This prediction is based on the fact that no member of the MFS has yet been shown to transport a macromolecule (*i.e.*, a protein, a complex carbohydrate, a nucleic acid or a lipid), and none has been shown to transport an inorganic cation as its primary substrate. We anticipate that MFS permeases are not capable of accommodating macromolecular substrates, due to architectural restrictions, but we recognize no reason why they should not be able to transport inorganic cations such as K^+, Mg^{2+}, Mn^{2+}, Ca^{2+}, Fe^{3+}, etc.

Most of the 34 MFS families described here function primarily in solute uptake. However, five of these families (DHA1-3, SET and BCD) expel their solutes. In all five cases, a proton antiport mechanism is probable. We further predict that several of the UMF families and additional yet-to-be-discovered MFS families will prove to function in efflux, particularly in prokaryotic organisms where facilitated diffusion is rare and active transport is the rule. With the exception of drug efflux pumps, past experimentation has focused primarily on uptake systems. We anticipate that many novel families of permeases, both within the MFS and outside of this superfamily, will prove to function with outwardly directed polarity.

If one includes the five distantly related MFS families (see bottom of Table 1), in what we have called the extended MFS, and analyzes completely sequenced genomes for MFS permeases, most organisms, both prokaryotes and eukaryotes, show a significant fraction of their secondary carriers as MFS permeases. Thus, based on the data published by Paulsen *et al.* (1998a,b), various organisms exhibit between 11 and 47% of their recognized secondary carriers as MFS permeases as follows: *Saccharomyces cerevisiae*, 47%; *Bacillus subtilis*, 44%; *Escherichia coli*, 42%; *Helicobacter pylori*, 23%; *Haemophilus influenzae*, 22%; *Mycoplasma genitalium*, 17%; *Synechocystis*, 17%; *Methanococcus jannaschii*, 11%. These values reveal that in general, large genome organisms have a greater ratio of MFS to total secondary permeases than small genome organisms, a fact that presumably reflects the need that all organisms have to maintain ionic homeostasis. Ionic homeostasis depends primarily on non-MFS permeases (Paulsen *et al.*, 1998a,b). Thus, large genome organisms exhibit the phenomenon of nutritional versatility, being able to use many exogenous nutrients for growth. This versatility arose in part by proliferation of MFS paralogues. By contrast, small genome organisms are generally restricted to a narrow range of organic nutrients for growth, and they consequently display a limited repertoire of MFS permeases. We expect that in eukaryotes, the MFS will generally prove to be much larger than any other superfamily of transport proteins. It will be interesting to determine if eukaryotic MFS permeases also have an increased degree of functional diversity relative to prokaryotes. Our preliminary results suggest that this may not be the case. The proliferation of eukaryotic MFS paralogues seems to reflect the need for elaborate temporal and spatial regulatory constraints; that in prokaryotes may instead have resulted from the need to adapt to a tremendous range of ecological niches.

Acknowledgements

We are grateful to Mary Beth Hiller and Milda Simonaitis for their assistance in the preparation of this manuscript. Work in our laboratory was supported by USPHS grants 5RO1 AI21702 from the National Institutes of Allergy and Infectious Diseases and 9RO1 GM55434 from the National Institute of General Medical Sciences, as well as by the M.H. Saier, Sr. memorial research fund.

References

Abe, T., Kakyo, M., Sakagami, H., Tokui, T., Nishio, T., Tanemoto, M., Nomura, H., Hebert, S.C., Matsuno, S., Kondo, H., and Yawo, H. 1998. Molecular characterization and tissue distribution of a new organic anion transporter subtype (oatp3) that transports thyroid hormones and taurocholate and comparison with oatp2. J. Biol. Chem. 273: 22395–22401.

Aínsa, J.A., Blokpoel, M.C.J., Otal, I., Young, D.B., De Smet, K.A.L., and Martín, C. 1998. Molecular cloning and characterization of Tap, a putative multidrug efflux pump present in *Mycobacterium fortuitum* and *Mycobacterium tuberculosis*. J. Bacteriol. 180: 5836–5843.

Alfonso, A., Grundahl, K., McManus, J.R., Asbury, J.M., and Rand, J.B. 1994. Alternative splicing leads to two cholinergic proteins in *Caenorhabditis elegans*. J. Mol. Biol. 241: 627–630.

Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., Näslund, A.K., Eriksson, A.-S., Winkler, H.H., and Kurland, C.G. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature, 396: 133–140.

Bajjalieh, S.M., Peterson, K., Shingal, R., and Scheller, R.H. 1992. SV2, a brain synaptic vesicle protein homologous to bacterial transporters. Science, 257: 1271–1273.

Bajjalieh, S.M., Peterson, K., Linial, M., and Scheller, R.H. 1993. Brain contains two forms of synaptic vesicle protein 2. Proc. Natl. Acad. Sci. USA, 90: 2150–2154.

Bindra, P.S., Knowles, R., and Buckley, K.M. 1993. Conservation of the amino acid sequence of SV2, a transmembrane transporter in synaptic vesicles and endocrine cells. Gene, 137: 299–302.

Chaillou, S., Postma, P.W., and Pouwels, P.H. 1998. Functional expression in *Lactobacillus plantarum* of *xylP* encoding the isoprimeverose transporter of *Lactobacillus pentosus*. J. Bacteriol. 180: 4011–4014.

Chan, B.S., Satriano, J.A., Pucci, M., and Schuster, V.L. 1998. Mechanism of prostaglandin E_2 transport across the plasma membrane of HeLa cells and *Xenopus* oocytes expressing the prostaglandin transporter "PGT". J. Biol. Chem. 273: 6689–6697.

Clancy, J., Petitpas, J., Dib-Hajj, F., Yuan, W., Cronan, M., Kamath, A.V., Bergeron, J., and Retsema, J.A. 1996. Molecular cloning and functional analysis of a novel macrolide-resistance determinant, *mefA*, from *Streptococcus pyogenes*. Mol. Microbiol. 22: 867–879.

Collins, J.C., Permuth, S.F., and Brooker, R.J. 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.

De Rossi, E., Blokpoel, M.C., Cantoni, R., Branzoni, M., Riccardi, G., Young, D.B., De Smet, K.A., and Ciferri, O. 1998. Molecular cloning and functional analysis of a novel tetracycline resistance determinant, *tet(V)*, from *Mycobacterium smegmatis*. Antimicrob. Agents Chemotherapy, 42: 1931–1937.

Díaz, E., Ferrández, A., and García, J.L. 1998. Characterization of the *hca* cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid in *Escherichia coli* K-12. J. Bacteriol. 180: 2915–2923.

Elkins, C.A., and Savage, D.C. 1998. Identification of genes encoding conjugated bile salt hydrolase and transport in *Lactobacillus johnsonii* 100-100. J. Bacteriol. 180: 4344–4349.

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

Gibson, L.C.D., McGlynn, P., Chaudhri, M., and Hunter, C.N. 1992. A putative anaerobic coproporphyrinogen III oxidase in *Rhodobacter sphaeroides*. II. Analysis of a region of the genome encoding *hemF* and the *puc* operon. Mol. Microbiol. 6: 3171–3186.

Gingrich, J.A., Andersen, P.H., Tiberi, M., el Mestikawy, S., Jorgensen, P.N., Fremeau, R.T., Jr., and Caron, M.G. 1992. Identification, characterization, and molecular cloning of a novel transporter-like protein localized to the central nervous system. FEBS Lett. 312: 115–122.

Goffeau, A., Park, J., Paulsen, I.T., Jonniaux. J.-L., Dinh, T., Mordant, P., and Saier, M.H., 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.

Gottesdiener, K.M. 1994. A new VSG expression site-associated gene (ESAG) in the promoter region of *Trypanosoma brucei* encodes a protein with ten potential transmembrane domains. Mol. Biochem. Parasitol. 63: 143–151.

Gründemann, D., Gorboulev, V., Gambaryan, S., Veyhl, M., and Koepsell, H. 1994. Drug excretion mediated by a new prototype of polyspecific transporter. Nature, 372: 549–552.

Hagenbuch, B. 1997. Molecular properties of hepatic uptake systems for bile acids and organic acids. J. Membr. Biol. 160: 1–8.

Hakes, D.J., and Berezney, R. 1991. Molecular cloning of matrix F/G: A DNA binding protein of the nuclear matrix that contains putative zinc finger motifs. Proc. Natl. Acad. Sci. USA, 88: 6186–6190.

Jacobs, C., Huang, L., Bartowsky, E., Normark, S., and Park, J.T. 1994. Bacterial cell wall recycling provides cystolic muropeptides as effectors for β-lactamase induction. EMBO J. 13: 4684–4694.

Jacquemin, E., Hagenbuch, B., Stieger, B., Wolkoff, A.W., and Meier, P.J. 1994. Expression cloning of a rat liver Na(+)-independent organic anion transporter. Proc. Natl. Acad. Sci. USA, 91: 133–137.

Jäger, W., Kalinowski, J., and Pühler, A. 1997. A *Corynebacterium glutamicum* gene conferring multidrug resistance in the heterologous host *Escherichia coli*. J. Bacteriol. 179: 2449–2451.

Janz, R., Hofmann, K., and Sudhof, T.C. 1998. SVOP, an evolutionarily conserved synaptic vesicle protein, suggests novel transport functions of synaptic vesicles. J. Neurosci. 18: 9269–9281.

Kanai, N., Lu, R., Satriano, J.A., Bao, Y., Wolkoff, A.W., and Schuster, V.L. 1995. Identification and characterization of a prostaglandin transporter. Science, 268: 866–869.

Kanamori, A., Nakayama, J., Fukuda, M.N., Stallcup, W.B., Sasaki, K., Fukuda, M., and Hirabayashi, Y. 1997. Expression cloning and characterization of a cDNA encoding a novel membrane protein required for the formation of *O*-acetylated ganglioside: A putative acetyl-CoA transporter. Proc. Natl. Acad. Sci. USA, 94: 2897–2902.

Kekuda, R., Prasad, P.D., Wu, X., Wang, H., Fei, Y.-J., Leibach, F.H., and Ganapathy, V. 1998. Cloning and functional characterization of a potential-sensitive polyspecific organic cation transporter (OCT3) most abundantly expressed in placenta. J. Biol. Chem. 273: 15971–15979.

Koepsell, E. 1998. Organic cation transporters in intestine, kidney, liver and brain. Annu. Rev. Physiol. 60: 243–266.

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

Le, T., and Saier, M.H., Jr. 1996. Phylogenetic characterization of the epithelial Na^+ channel (ENaC) family. Mol. Membr. Biol. 13: 149–157.

Le, T., Tseng, T.-T., and Saier, M.H., Jr. 1999. Flexible programs for the estimation of average amphipathicity of multiply aligned homologous proteins: Application to integral membrane transport proteins. Mol. Membr. Biol. 16: 173–179.

LeBlanc, H.N., and Beatty, J.T. 1996. Topological analysis of the *Rhodobacter capsulatus* PucC protein and effects of C-terminal deletions on light-harvesting Complex II. J. Bacteriol. 178: 4801–4806.

Lesuisse, E., Simon-Casteras, M., and Labbe, P. 1998. Siderophore-mediated iron uptake in *Saccharomyces cerevisiae*: The *SIT1* gene encodes a ferrioxamine B permease that belongs to the major facilitator superfamily. Microbiology, 144: 3455–3462.

Lindquist, S., Weston-Hafer, K., Schmidt, H., Pul, C., Korfmann, G., Erickson, J., Sanders, C., Martin, H.H., and Normark, S. 1993. AmpG, a single transducer in chromosomal β-lactamase induction. Mol. Microbiol. 9: 703–715.

Liu, J.Y., Miller, P.F., Gosink, M., and Olson, E. 1999a. The identification of a new family of sugar efflux pumps in *Escherichia coli*. Mol. Microbiol. 31: 1845–1851.

Liu, J.Y., Miller, P.F., Willard, J., and Olson, E.R.. 1999b. Functional and biochemical characterization of *Escherichia coli* sugar efflux transporters. J. Biol. Chem. 274: 22977–22984.

Liu, Y., Peter, D., Roghani, A., Schuldiner, S., Prive, G.G., Eisenberg, D., Brecha, N., and Edwards, R.H. 1992. A cDNA that suppresses MPP+ toxicity encodes a vesicular amine transporter. Cell, 70: 539–551.

Lopez-Nieto, C.E., You, G., Bush, K.T., Barros, E.J.G., Beier, D.R., and Nigam, S.K. 1997. Molecular cloning and characterization of NKT, a gene product related to the organic cation transporter family that is almost exclusively expressed in the kidney. J. Biol. Chem. 272: 6471–6478.

Marger, M.D., and Saier, M.H., Jr. 1993. A major superfamily of transmembrane facilitators catalyzing uniport, symport and antiport. Trends Biochem. Sci. 18: 13–20.

Matos, M.E., and Wilson, T.H. 1994. Characterization and sequencing of an uncoupled lactose carrier mutant of *Escherichia coli*. Biochem. Biophys. Res. Commun. 200: 268–274.

Moore, J., and Beverley, S.M. 1996. Pteridine transport and recurrent amplification of extrachromosomal DNAs in *Leishmania*. Woods Hole Mol. Parasitol. Meeting Abstracts.

Naderi, S., and Saier, M.H., Jr. 1996. Plant sucrose:H^+ symporters are homologous to the melibiose permease of *Escherichia coli*. Mol. Microbiol. 22: 389–391.

Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O. 1998. Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. DNA Res. 5: 277–286.

Okuda, M., Saito, H., Urakami, Y., Takano, M., and Inui, K. 1996. cDNA cloning and functional expression of a novel rat kidney organic cation transporter, OCT2. Biochem. Biophys. Res. Commun. 224: 500–507.

Pao, S.S., Paulsen, I.T., and Saier, M.H., Jr. 1998. Major facilitator superfamily. Microbiol. Mol. Biol. Rev. 62: 1–34.

Park, J.T., Raychaudhuri, D., Li, H., Normark, S., and Mengin-Lecreulx, D. 1998. MppA, a periplasmic binding protein essential for import of the bacterial cell wall peptide L-alanyl-γ-D-glutamyl-*meso*diaminopimelate. J. Bacteriol. 180: 1215–1223.

Paulsen, I.T., Brown, M.H., and Skurray, R.A. 1996. Proton-dependent multidrug efflux systems. Microbiol. Rev. 60: 575–608.

Paulsen, I.T., and Skurray, R.A. 1994. The POT family of transport proteins. TIBS, 18: 404.

Paulsen, I.T., Sliwinski, M.K., and Saier, M.H., Jr. 1998a. Microbial genome analyses: Global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. J. Mol. Biol. 277: 573–592.

Paulsen, I.T., Sliwinski, M.K., Nelissen, B., Goffeau, A., and Saier, M.H., Jr. 1998b. Unified inventory of established and putative transporters encoded within the complete genome of *Saccharomyces cerevisiae*. FEBS Lett. 430: 116–125.

Perreten, V., Schwarz, F., Cresta, L., Boeglin, M., Dasen, G., and Teuber, M. 1997. Antibiotic resistance spread in food (letter). Nature, 389: 801–802.

Poolman, B., Knol, J., van der Does, C., Henderson, P.J.F., Liang, W.-J., Leblanc, G., Pourcher, T., and Mus-Veteau, I. 1996. Cation and sugar selectivity determinants in a novel family of transport proteins. Mol. Microbiol. 19: 911–922.

Reizer, J., Reizer, A., and Saier, M.H., Jr. 1994. A functional superfamily of sodium/solute symporters. Biochim. Biophys. Acta, 1197: 133–166.

Saier, M.H., Jr., Eng., B.H., Fard, S, Garg, J., Haggerty, D.A., Hutchinson, W.J., Jack, D.L., Lai, E.C., Liu, H.J., Nusinew, D.P., Omar, A.M., Pao, S.S., Paulsen, I.T., Quan, J.A., Sliwinski, M., Tseng, T.-T., Wachi, S., and Young, G.B. 1999. Phylogenetic characterization of novel transport protein families revealed by genome analyses. Biochim. Biophys. Acta, 1422: 1–56.

Schuldiner, S., Shirvan, A., and Linial, M. 1995. Vesicular neurotransmitter transporters: From bacteria to humans. Physiol. Rev. 75: 369–392.

Schuster, V.L. 1998. Molecular mechanisms of prostaglandin transport. Annu. Rev. Physiol. 60: 221–242.

Sneath, P.H.A., and Sokal, R.R. 1973. Numerical taxonomy, the principles and practice of numerical classification. San Francisco: W.H. Freeman and Co.

Steiner, H.-Y., Naider, F., and Becker, J.M. 1995. The PTR family: A new group of peptide transporters. Mol. Microbiol. 16: 825–834.

Tailor, C.S., Willett, B.J., and Kabat, D. 1999. A putative cell surface receptor for anemia-inducing feline leukemia virus subgroup C is a member of a transporter superfamily. J. Virol. 73: 6500–6505.

Tichy, H.V., Oberlé, B., Stiehle, H., Schiltz, E., and Drews, G. 1989. Genes downstream from *pucB* and *pucA* are essential for formation of the B800-850 complex of *Rhodobacter capsulatus*. J. Bacteriol. 171: 4914–4922.

Varela, M.F., and Wilson, T.H. 1996. Molecular biology of the lactose carrier of *Escherichia coli*. Biochim. Biophys. Acta, 1276: 21–34.

Venkatesan, P., and Kaback, H.R. 1998. The substrate-binding site in the lactose permease of *Escherichia coli*. Proc. Natl. Acad. Sci. USA, 95: 9802–9807.

Wang, Z.H., and Fallon, A.M. 1998. The mosquito dihydrofolate reductase amplicon contains a truncated synaptic vesicle protein gene. Insect Mol. Biol. 7: 317–325.

Wilson, T.H., and Wilson, D.M. 1998. Evidence for a close association between helix IV and helix XI in the melibiose carrier of *Escherichia coli*. Biochim. Biophys. Acta, 1374: 77–82.

Young, C.S., and Beatty, J.T. 1998. A topological model of the *Rhodobacter capsulatus* light-harvesting I complex assembly protein LhaA (previously known as ORF1696). J. Bacteriol. 180: 4742–4745.

Young, G.B., Jack, D.L., Smith, D.W., and Saier, M.H., Jr. 1999. The amino acid/auxin:proton symport permease family. Biochim. Biophys. Acta, 1415: 306–322.

Youvan, D.C., Bylina, E.J., Alberti, M., Begusch, H., and Hearst, J.E. 1984. Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna, and flanking polypeptides from *R. capsulata*. Cell, 37: 949–957.

Zsebo, K.M., and Hearst, J.E. 1984. Genetic-physical mapping of a photosynthetic gene cluster from *R. capsulato*. Cell. 37: 937-947.