MICROBIOLOGICAL REVIEWS, Mar. 1994, p. 71–93
0146-0749/94/$04.00+0

Vol. 58, No. 1

Computer-Aided Analyses of Transport Protein Sequences: Gleaning Evidence concerning Function, Structure, Biogenesis, and Evolution

MILTON H. SAIER, JR.*

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

INTRODUCTION

"Unless all existence is a medium of revelation, no particular revelation is possible."

William Temple

Computer-aided sequence analysis has become an essential molecular biological tool for the estimation of the structures, functions, and evolutionary relationships of proteins. As DNA sequencing technology becomes perfected and the number of available gene-derived protein sequences increases, the utility of computer-aided approaches expands exponentially with the amount of data available. Sequence analyses can provide guides for experimentalists, leading them to the most direct and fruitful empirical approaches. Hundreds of competent scientists are currently devoting full-time efforts to the analyses of DNA and protein sequences, in an attempt to extract the information that is becoming available as a result of the efforts of individuals as well as of genome-sequencing projects.

Sequences of (unidentified) open reading frames (40) may be published without adequate searches of the data bases for homologous sequences, leading to a lack of recognition of functional, structural, and/or evolutionary aspects of the protein products of the sequenced genes (10, 76, 77, 118, 152, 154, 215, 238). In other cases, erroneous conclusions regarding the functions, origins, and cellular locations of the products of sequenced genes have resulted from inadequate analyses (103, 133, 151, 208). In still other cases, sequencing errors have been identified and subsequently corrected, merely as a result of computer analyses (26, 63, 109, 145).

In this review I summarize selected information concerning the structures, functions, biogenesis, and evolution of transport proteins which has been forthcoming largely as a result of the utility of programs currently available to the scientific commu-

* Phone: (619) 534-4084. Fax: (619) 534-7108. Electronic mail address: MSAIER@UCSD.EDU.

nity. I make no attempt to be exhaustive and freely select examples from our own published work. Because of the tremendous number of relevant publications, secondary rather than earlier primary references are frequently cited when appropriate. I hope that this review will help to reveal the incredible potential of computer-aided approaches so that an ever-increasing number of molecular biologists will come to utilize them.

PRINCIPLES AND TECHNIQUES OF COMPUTER-AIDED PROTEIN ANALYSES

The principles underlying the study of protein evolution and many of the computer-aided methods for quantifying the relatedness of any two proteins have been described in detail by Doolittle (40, 41, 43). A primary goal of the protein evolutionist is to reconstruct past events leading to the structures of current proteins, from which the relationships of present-day proteins that share a common ancestry can be defined (245, 248). A single sequence can evolve into different sequences either by speciation and divergence (such genes are called orthologous) or by gene duplication within a single organism followed by divergence (such genes are called paralogous). The latter event can give rise to proteins of different function or to isoforms of proteins having the same function. Because the rates of change of many proteins are very low, ancient events can be inferred by examining protein families. The sequence of a slowly evolving prokaryotic enzyme may, for example, exhibit as much as 40 to 50% identity with the corresponding enzyme from a eukaryote, even though eukaryotes diverged from prokaryotes over 1.5 billion years ago! High degrees of sequence similarity found in evolutionarily distant organisms, or low degrees of sequence similarity found in closely related organisms, may be due to horizontal transmission of genetic information (65, 144), even between prokaryotes and eukaryotes (71), and this information can probably be passed between these two kingdoms in both directions (see reference 15 and discussion therein). Homologous genes in a single organism that are acquired laterally (horizontally) are said to be xenologous. It is important to realize that different proteins evolve at very different rates and that different degrees of sequence similarity observed for two proteins will in general reflect both the time and the rate of divergence of their two structural genes (35, 101, 122).

The first task in defining the ancestry of a group of proteins is always to identify all sequenced members of a particular family by screening available data bases. Families or clusters of closely related proteins are first identified and then interrelated if possible. Signature sequences of established families can be used to identify newly sequenced members of these families (9). When several families of proteins are shown to share a common ancestry, we say that they make up a superfamily. Since they share a common origin, the members of a superfamily are said to be homologous. Proteins are either homologous or not homologous; there are no degrees of homology, although there are ranges of similarity or identity (40).

A binary comparison score [expressed in standard deviations (SDs)] can be calculated from the comparison of two amino acid sequences by using any of several programs (ALIGN [34], Los Alamos [81], and RDF2 [135]). The degree of similarity between the two sequences is compared with a large number of random shuffles of these two sequences (thus eliminating discrepancies due to unusual amino acid compositions) to establish significance. When two sequences give a comparison score of 3 SDs or less, there is little or no evidence for homology. However, if they give a comparison score of 6 SDs, the probability (P) that the degree of similarity exhibited by these two sequences arose by chance is about 10^{-9}. This suggests that such sequences probably arose from a common ancestor by divergent evolution, but this degree of sequence similarity could conceivably have arisen by a convergent evolutionary process, particularly if the sequences compared are short. Thus, homology is not established. On the other hand, when a comparison score is high, i.e., $\geq$9 SDs ($P \leq 10^{-19}$), the degree of similarity is considered to be too great to have arisen either by chance or by a convergent evolutionary process, so the two sequences are considered to be homologous (40). Methods for assessing the statistical significance and reliability of particular molecular sequence features have been presented (see, for example, references 4, 39, 56, and 84). It should be noted that the term "convergence" has been used loosely by different investigators to refer to the independent evolution of a similar sequence (sequence convergence), a similar specificity or catalytic function (functional convergence), or a similar topology or structural scaffold (structural convergence) (16).

Several questions can be asked to further substantiate the conclusion of common descent. (i) Do the two proteins (or protein domains) share a common function? (ii) Are the sequences compared derived from comparable portions of the proteins or protein domains? (iii) Are the two proteins or protein domains of about the same lengths? (iv) Do they have similar topologies or three-dimensional structures? If the answers to these questions are yes, one gains additional confidence in the conclusions derived solely from the computer-based statistical analyses described above. Simultaneous convergence of sequence, structure, and function is considered to be an exceptionally improbable evolutionary event.

Once homology has been established for a group of proteins or protein domains, the sequences are optimally aligned, yielding a single multiple alignment. Relative evolutionary distances and phylogenetic positions can be determined by using available computer programs. Results obtained with two distinct programs, based on different assumptions, i.e., the TREE and PAPA programs (44, 57), can be used to evaluate the reliability of the results obtained for any one program (215, 222). The construction of phylogenetic trees allows one to see the relatedness of members of a group of homologous proteins at a glance.

Other programs are available for (i) plotting average degrees of similarity as a function of residue position for a set of homologous proteins; (ii) locating signature sequences or motifs that are common to all members of a particular family but absent from other proteins; (iii) deriving primordial (ancestral) sequences from current sequences; (iv) predicting membrane protein topologies; (v) averaging hydropathy and amphipathicity plots; (vi) detecting internal duplications, deletions, and insertions within coevolving proteins; (vii) estimating regions of relative flexibility; (viii) estimating the probability that a sequence will exhibit a particular secondary structural feature (i.e., α-helix, β-strand, or β-turn), (ix) averaging secondary-structure and flexibility predictions for a group of homologous proteins, thereby increasing reliability; and (x) averaging comparison scores, similarity scores, or percent identities for estimation of relative degrees of evolutionary divergence for coevolving homologous proteins or protein domains or for approximating the relative ages of independently evolving protein families. The reader is referred to volume 183 of *Methods in Enzymology* for detailed consideration of specific computer programs concerned with the analysis of protein and nucleic acid sequences (42).

TABLE 1. Occurrence and properties of various transport protein families discussed in this review

Protein family	Family abbreviation used in this review	Multi-component	Occurrence in:		No. of sequenced proteins	Signature sequence available	Phylogenetic tree(s) constructed	References
			Prokaryotes	Eukaryotes				
Channel-mediated diffusion								
Outer membrane porins	Porin	−	+	+	>30	−	+	79
Voltage-sensitive ion channels	VIC	+ or −	−	+	>200	+	+	75, 95a, 214
Major intrinsic protein family	MIP	−	+	+	>20	+	+	155
ATP-driven transport								
Cation-translocating P-type ATPases	P type	+	+	+	>70	+	+	50
H^+ and Na^+ translocating F-type ATPases	F type	+	+	+	>30	−	+	14b
H^+ and Na^+ translocating V-type ATPases	V type	+	+	+	>10	−	+	14b
Arsenical resistance transporters	A type	+	+	+	>5	−	−	164
ATP binding cassette-type transporters	ABC type	+	+	+	>100[a]	−	+	52, 158, 208, 209
Electron flow-driven transport								
H^+-translocating electron carriers		+	+	+	Several families	−	+	55, 235
H^+-translocating transhydrogenases	Transhydrogenase	+ or −	+	+	>10	−	+ and −	68
Carrier-mediated transport (uni-, sym-, and antiport)								
Mitochondrial carrier family	MCF	−	−	+	>40	−	+	95
Major facilitator superfamily	MFS	−	+	+	>70	−	+	109
Amino acid/polyamine/choline family	APC	−	+	+	>20	+	+	147
Sodium:solute symporter family	SSF	−	+	+	>10	+	+?	156
Sodium:neurotransmitter symporter family	SNF	−	−	+	>20	+	+	156
Resistance/nodulation/cell division family	RND	−	+	−	4	+	+	172
Group translocation								
Phosphotransferase system permeases	PTS	+	+	−	>20; 3[b]	−	+	146, 178
Light-driven transport								
Bacteriorhodopsin family	BR	−	+	−	>10	+	+	95a
Reaction center family	RC	+	+	+	>10	−	−	249a

[a] Several ABC families exist, based on integral membrane protein sequence analyses.
[b] There are two PTS permease families, a major and a minor family.

FAMILIES AND SUPERFAMILIES OF TRANSPORT PROTEINS

The major families of transport proteins and the abbreviations used in this review for these families are listed in Table 1, which also indicates the approximate sizes of these families, some of their properties, the known occurrence of members of these families in the prokaryotic and eukaryotic kingdoms, and the availability of published sequence and phylogenetic information (168–170).

Outer membrane bacterial porin proteins, which transport many solutes nonselectively, have a structurally related counterpart in the anion channels of the outer mitochondrial membranes of eukaryotes (13, 14, 79, 125). The bacterial homotrimeric, 16-stranded, antiparallel β-barrel porins are structurally unlike any of the other transport proteins considered in this article (32, 241). Although all known porins consist primarily or exclusively of β-structure, it is not yet clear that all of these proteins belong to a single family (66, 203).

Five major types of solute-transporting ATPases (P type, F type, V type, A type, and ABC type) are found ubiquitously in nature. The P-type ATPases make up one superfamily, the F and V types together make up a second, A types make up a small but distinct family, and the ABC types probably form a huge and diverse superfamily, although this cannot be established solely on the basis of the sequences of their integral membrane constituents. P-type ATPases exhibit no significant sequence similarity with other proteins, and the same is true of the integral membrane constituents of F- and V-type ATPases (14a, 50, 62, 89). By contrast, ATP-binding proteins of the A- and ABC-type permeases have nontransport homologs (74, 164).

Mitochondrial enzymes and electron carriers are sometimes proton transport proteins. Among these are the membrane-bound transhydrogenases, the NADH dehydrogenase complexes, the ubiquinol:cytochrome *c* reductases, and the cytochrome oxidases (6, 68, 187, 216). These multisubunit protein complexes have prokaryotic counterparts, which, although usually less complex, can be shown to be homologous. Only in certain instances are their phylogenetic relationships well defined (55, 235). These proteins will not be considered further.

TABLE 2. Proposed structural features of integral membrane constituents of selected transport protein types[a]

Protein family	Probable no. of transmembrane spanners/unit[b]	No. of units/permease[b]	No. of transmembrane polypeptide chains	No. of internal repeats/polypeptide chain[c]	No. of spanners/repeat sequence	Proposed no. of tandem repeats/permease
Voltage-sensitive K^+ channels	6[d]	4	4	1	6	4
Voltage-sensitive Ca^{2+} channels	6[d]	4	1 (or 2)	4 (or 2)	6	4 (or 2)
Voltage-sensitive Na^+ channels	6[d]	4	1	4	6	1
MIP-type channels	6	2 or 4[e]	2 or 4[e]	2	3	2
Uniporters, symporters, antiporters (MFS, APC, SSS, and SNS families)	6	2 (8)[f]	1 (4)[f]	1	6	2
MCF	6	2	2	3	2	3
PTS permeases	6 or 8[g]	2	2	1	6 or 8	1
ABC-type systems	6	2	1 or 2[c]	1 or 2[c]	6	1 or 2
RND systems	4 or 6[h]	2	1	2	4 or 6	2
Bacteriorhodopsins	7	1	—[i]	2	3.5	2

[a] In all of the permease classes tabulated, the amino and carboxy termini of the proteins, as well as the loop regions between the six transmembrane segment domains or units, are believed to be on the cytoplasmic side of the membrane.

[b] Unit corresponds to the common transmembrane spanner unit of structure, presumed to correspond to a membrane-embedded protein domain (see text for explanatory remarks).

[c] The transmembrane constituent(s) of ABC-type permeases may be homo- or heterodimeric. They can exist as two distinct polypeptide chains or can be fused into a single chain (74, 153). In some of these permeases, five spanners for the integral constituents have been reported while a single additional spanner is reported for the ABC proteins of the same system (note: 5 + 1 = 6 [88]). In some cases when two units are fused together in a single polypeptide chain, the degree of divergence of their sequences suggests that the two halves of these proteins arose by fusion of two dissimilar genes rather than by tandem intragenic duplication (153).

[d] Six transmembrane α-helical spanners per unit are proposed, but the unit polypeptide chain may dip into the membrane at an additional position (117a).

[e] The MIP of the mammalian lens has been reported to be a homotetramer, whereas the glycerol facilitator of *E. coli* has been reported to be a dimer (2, 226).

[f] Few of these permeases have been examined with respect to their oligomeric state. However, only one exception to the proposal that these permeases are monomeric has been reported. This exceptional protein is the intestinal Na^+:glucose symporter, which appears to be tetrameric.

[g] Six transmembrane spanners have been reported for the mannitol permease (204), but analyses of the glucose permease clearly suggest eight (24). Considerations discussed by Buhr and Erni (24) lead to the possibility that most sequenced PTS permeases have eight spanners.

[h] Topological studies have not been carried out with these proteins, and the hydropathy plots are ambiguous (180).

[i] Bacteriorhodopsin forms a two-dimensional lattice in the membrane.

Other channel-type transporters include the ligand-gated and voltage-gated ion channels of the nerve and muscle cells of higher animals, members of the presumed pore-type major intrinsic protein (MIP) family and possibly the bacteriorhodopsins and reaction center protein complexes (Table 1). These transmembrane channels clearly have α-helical secondary structural elements called spanners; six of these spanners are believed to penetrate the membrane and make up the basic structural unit of the voltage-gated ion channels and the channel proteins of the MIP family (Table 2). The same structural feature characterizes many carriers or presumed carriers which may be prokaryote specific (the phosphotransferase system [PTS] and resistance-nodulation-cell division [RND] permeases), eukaryote specific (the mitochondrial carrier family [MCF] and sodium:neurotransmitter symporter [SNF] porters), or ubiquitous (the major facilitator superfamily [MFS]; ATP-binding cassette [ABC], amino acid-polyamine-choline [APC], and sodium:solute symporter [SSF] families) (Tables 1 and 2). Alternative topological arrangements have been suggested in a few instances (88, 220).

GUIDES TO TRANSPORT PROTEIN FUNCTION

Sequence Similarities Indicative of Common, Similar, or Overlapping Function or Specificity

In this section, examples will be selected which reveal that transport proteins showing a high degree of sequence similarity throughout their lengths usually serve the same function whereas related proteins with more divergent sequences are more likely to serve dissimilar but related functions. Proteins serving the same function in evolutionarily divergent organisms are usually (but not always) more closely related to each other than they are to proteins serving dissimilar functions from the same organism (see, for example, references 16, 95, 109, and 155). These observations suggest either that most of the principal systems found in nature existed before speciation or that systems of differing specificities diverged more rapidly from each other than did those which retained the same function.

Sugar permeases of the bacterial PTS. Detailed analyses of the sequences of a group of transmembrane sugar transport proteins, the Enzyme II complexes of the bacterial PTS, revealed that many of them fall into clusters of related proteins (80, 98, 99, 146, 162, 173, 176, 178, 179, 182). All of the proteins within this superfamily catalyze sugar transport, presumably by a single mechanism. Clustering patterns consistently correlate with sugar specificity, even though the sequenced proteins may be derived from a variety of evolutionarily divergent organisms. Thus, all sequenced permeases specific for sucrose (an α-glucoside), trehalose (another α-glucoside), and aromatic β-glucosides form a cluster; all sequenced glucose and *N*-acetylglucosamine permeases make up a second cluster; the fructose and mannitol permeases form a third major branch of the phylogenetic tree; and the lactose (a β-galactoside) and cellobiose (a β-glucoside) permeases cluster together, forming the fourth of the four major PTS sugar permease clusters (Fig. 1A) (91a, 146, 157, 178, 206). Moreover, within each cluster, the permeases specific for a particular sugar (i.e., fructose) are, in general, closer to each other than to those specific for another sugar (i.e., mannitol).

Interestingly, a second, minor family of PTS permeases was identified that could not be shown to be evolutionarily related to the major family. This family consists of three sequenced Enzyme II complexes, each phosphorylated by phospho-HPr and specific for a set of related sugars. One such permease in *Escherichia coli* exhibits specificity for aldo- and ketohexoses of the gluco-, manno-, and fructo- configurations with little specificity for the hydroxyl groups or carbonyl functions at carbons 1 and 2 (48, 49, 159). The second, found in *Bacillus subtilis*, is specific for fructose (110), whereas the third, present

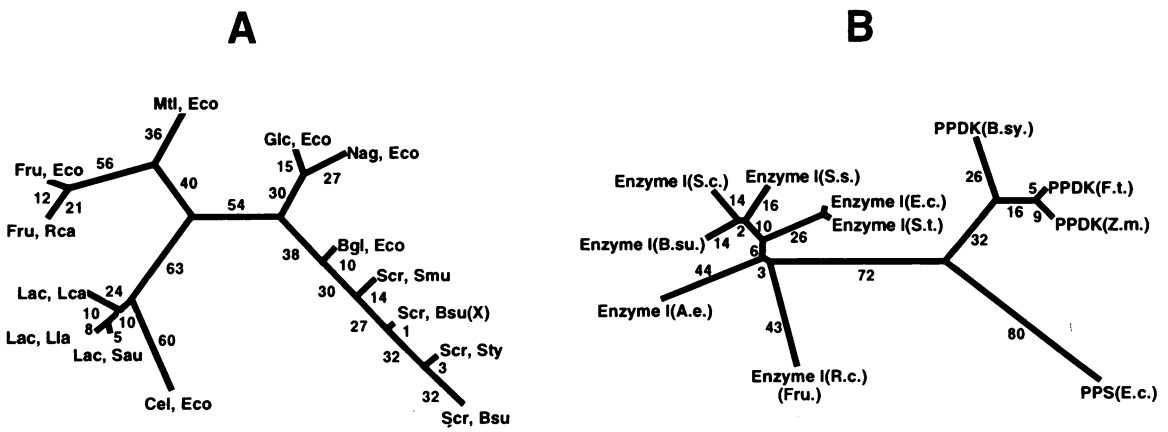

FIG. 1. Phylogenetic trees of selected proteins and protein domains of the bacterial PTS. (A) Tree for the integral membrane constituents of the PTS. The tree was based on the most similar portions of the Enzyme IIC sequences. These protein domains bear the sugar recognition sites of the PTS permeases. Abbreviations for the sugars are as follows: Glc, glucose; Nag, *N*-acetylglucosamine; Bgl, β-glucosides; Scr, sucrose; Mtl, mannitol; Fru, fructose; Lac, lactose; Cel, cellobiose. Reproduced from reference 178 with permission of the publisher. (B) Phylogenetic tree of sequenced Enzymes I of the PTS and their enzymic homologs, PEP synthase (PPS) and pyruvate:phosphate dikinases (PPDK). Reproduced from reference 148 with permission of the publisher. For species abbreviations and details of tree construction for these and subsequent figures, see the original publications.

in *Klebsiella pneumoniae*, is specific for sorbose, a fructose analog, but fructose is also transported by this system (97a). These three multisubunit protein complexes, which are closely related in sequence, all transport fructose and therefore exhibit overlapping sugar specificities.

Two other PTS Enzyme II complexes in *E. coli*, those specific for glucitol (250) and galactitol (97a), show domain sizes similar to those of members of the major family of PTS permeases. However, they lack sufficient sequence similarity to the latter proteins or to each other to allow postulation of a common evolutionary origin (97a, 250–252). These observations suggest that the PTS transport proteins, referred to as the Enzymes IIC (see below), may be members of a very ancient family with relatedness to other ancient families such as the major facilitator superfamily (152, 246, 247).

Energy-coupling proteins of the PTS. The PTS permeases discussed above are energized by a protein phosphorylation mechanism involving a phosphorelay (138). The first protein to be phosphorylated in this sequence is called Enzyme I. It is a protein of about 600 amino acids which is phosphorylated on the N-3 position of a histidyl residue at the expense of phosphoenolpyruvate (PEP). Enzyme I phosphate then transfers its phosphoryl moiety to a small (~85-residue), heat-stable phosphocarrier protein called HPr. HPr is phosphorylated on the N-1 position of a histidyl residue. This second, general, energy-coupling phosphorylated protein then phosphorylates the sugar-specific Enzyme II complexes (first the IIA domain or protein, then the IIB domain or protein) in preparation for sugar transport, catalyzed by the IIC domain or protein.

All Enzymes I and all HPrs have been found to make up two families of closely related proteins, respectively, regardless of the bacterium of origin (148, 188, 246, 247, 249). Moreover, the Enzymes I were found to be homologous to two other enzymes which are phosphorylated on the N-3 position of histidyl residues at the expense of PEP, pyruvate:phosphate dikinase, and PEP synthase (148, 249). These enzymes possess sizes, subunit compositions, and mechanisms of action similar to those of the Enzymes I (167). As clear indications of their related but divergent functions, the Enzymes I make up one cluster, whereas the pyruvate:phosphate dikinases and the one sequenced PEP synthase form two distant branches, on the phylogenetic tree constructed for these homologous proteins (Fig. 1B). As expected on the basis of function, the pyruvate: phosphate dikinases and the PEP synthase are more closely related to each other than they are to the Enzymes I.

Periplasmic receptors of ABC-type uptake permeases. Bacteria utilize extracellular solute-binding receptors to trigger chemoreception as well as solute uptake via the ABC-type transport systems (5, 74, 134, 172, 201). Among the well-characterized periplasmic solute-binding protein-dependent permeases are those specific for maltose, galactose, histidine, and oligopeptides (208). The periplasmic receptors which act in conjunction with the ABC-type solute uptake permeases make up several families of demonstrably homologous proteins. The proteins within each family bind solutes that are usually but not always related in structure.

Four examples will be provided (Fig. 2) (208). The family of receptors which includes the sequenced maltose (family 1)-binding proteins also includes members that are specific for multiple sugars, glycerol-3-phosphate, and iron (Fig. 2A). In this case, the receptor specificities are diverse. The family of receptors which includes the galactose-binding protein (family 2) includes homologous proteins which bind ribose, arabinose, and multiple sugars (Fig. 2B). The latter family includes several cytoplasmic repressor proteins, most but not all of which bind sugars (123, 223, 240). The histidine receptor, a member of family 3, is homologous to other bacterial receptors specific for glutamine, basic amino acids, and two derivatives of basic amino acids, octopine and nopaline (Fig. 2C). This family includes a single sequenced catalytic protein, cyclohexadienyl dehydratase of *Pseudomonas aeruginosa* (209), and the ligand-binding domains of two groups of eukaryotic glutamate receptors of nerve cells (120, 130). All of these proteins bind polar amino acids, and thus their binding specificities are similar. Finally, the fourth receptor family to be considered includes a coherent group of binding proteins specific for various di- and oligopeptides (Fig. 2D). Surprisingly, this family includes an exceptional member which is specific for nickel (121, 208). This Ni^{2+}-specific receptor represents a clear exception to the rule that closely related proteins have similar specificities since the nickel-binding protein is as similar to the various peptide-binding proteins as the latter are to each other. Newly discov-

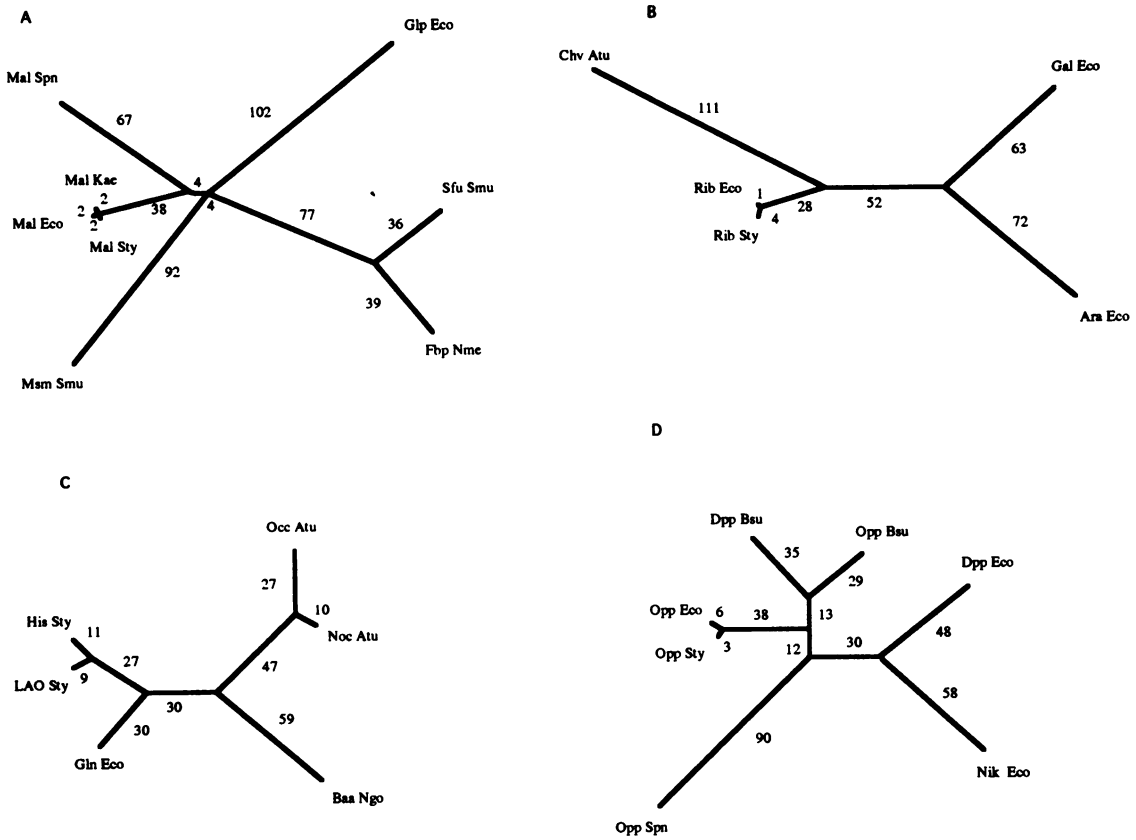

FIG. 2. Phylogenetic trees for four families of extracytoplasmic receptors of ABC-type transport systems. (A) Family 1 sugar-binding receptors specific for maltose (Mal), multiple sugars (Msm), α-glycerol phosphate (Glp), and iron (Sfu and Fbp). (B) Family 2 sugar-binding receptors specific for galactose (Gal), arabinose (Ara), ribose (Rib), and multiple sugars (Chv). Homologous ligand-binding domains of bacterial transcriptional regulators are not included (223). (C) Family 3 polar amino acid-binding receptors specific for histidine (His); lysine, arginine, and ornithine (LAO); glutamine (Gln); basic amino acids (Baa); octopine (Occ); and nopaline (Noc). A homologous bacterial enzyme and the ligand-binding domains of eukaryotic glutamate-specific neurotransmitter receptors are not included. (D) Family 4 dipeptide (Dpp)-, oligopeptide (Opp)-, and nickel (Nik)-specific receptors. Reproduced from reference 208 with permission; species abbreviations are provided in reference 208.

ered members of this family are (i) a virulence factor in *Salmonella typhimurium* (Fig. 2D) (134a, 208) and (ii) peptide sex pheromone receptors encoded on conjugative plasmids pAD1 and pCF1D of *Enterococcus faecalis* (31, 165, 211).

On the basis of three-dimensional structural analyses (82, 129, 141, 192, 200, 234), it is probable that all or most of the various families of extracytoplasmic bacterial receptor proteins share a common evolutionary origin. Evolutionary divergence therefore correlates to a remarkable degree with binding specificity for a restricted group of sugars, amino acids, peptides, or anions. The only clear exceptions to this generalization are the α-glycerol phosphate- and the iron-binding proteins of family 1 (Fig. 2A) and the nickel-binding protein of family 4 (Fig. 2D). The relationship of the metal-binding sites to the sugar- or peptide-binding site is not clear, particularly since the galactose-binding protein is known to possess a metal-binding site distant from the sugar-binding site (233). The Ca^{2+}-binding site in the galactose receptor is apparently of little functional significance because Ca^{2+} binding yields only local structural changes in the protein and does not appreciably influence sugar binding. Sugar binding, on the other hand, induces global conformational changes that activate the protein for interaction with the transport and chemoreception apparatuses in the membrane (105, 106). Similar global conformational changes presumably occur when the iron- and nickel-binding proteins bind their ligands, thereby initiating transport of these metallic ions.

ABC-type export permeases. ABC-type export permeases of bacteria and eukaryotes (5, 52, 74) lack the extracytoplasmic receptors considered above, but they each possess one or two integral membrane "channel" proteins as well as one or two cytoplasmic ATP-binding proteins (or domains) which hydrolyze ATP to energize transport. Among the substrates which are exported from the bacterial cytoplasm to the extracellular medium via these systems are proteins, peptides, capsular polysaccharides, modified oligosaccharides, and drugs (52, 153). The integral membrane constituents of the ABC-type carbohydrate and drug exporters are closely related to each other but not to those that export proteins and peptides or those that catalyze solute import (153, 224). Thus, as for the binding-protein constituents of the uptake systems, the degree of sequence similarity of the integral membrane proteins predictably correlates with function and specificity.

Phylogenetic analyses of the highly conserved ABC domains of numerous bacterial ABC-type exporters have recently been presented (52, 153). Results of the analyses of these energy-coupling domains proved to contrast with those of the integral membrane transport constituents noted above. The phylogenetic relatedness of the ABC domains correlates only secondarily with the substrate specificities of the transporters. Their

primary division is related to the nature of their disposition with the integral membrane domains. Thus, they fall into two major phylogenetic groups: those in which the soluble, ATP-binding domain is covalently linked to the integral membrane transport constituent as a single polypeptide chain (group A) and those in which these two structural constituents are present as distinct polypeptide chains (group B). ABC domains of the bacterial import systems generally fall into the latter category while those of the eukaryotic exporters fall into the former group (52).

On the basis of sequence similarities of the ABC domains as well as domain association states, the bacterial exporters fall into both categories. Bacterial export permeases of phylogenetic group A are generally specific for peptides and proteins, whereas those of phylogenetic group B transport capsular polysaccharides, oligosaccharides, and drugs (52, 153). These observations reveal a general, but not absolute, correlation of phylogenetic grouping with substrate specificity. It seems surprising that domain association has proven to be a strongly conserved characteristic since other types of domain shuffling have occurred repeatedly during the evolution of these and certain other classes of multidomain permeases (see below). However, it appears that the ABC domains did not always evolve in parallel with the integral membrane constituents of these permeases (74, 153).

Solute uniporters, symporters, and antiporters. Numerous transport proteins function to accumulate or extrude specific solutes by using a chemiosmotic form of energy (94). In these cases a solute is obligatorily cotransported (symported) with a charged species or countertransported (antiported) against a different but structurally related compound. The latter species flows down its electrochemical gradient in an energetically favorable process, thereby driving the second solute against a concentration gradient in an energetically unfavorable process. Symporters and antiporters are homologous to uniporters which catalyze simple facilitation of their solutes without energy expenditure or accumulation of the solute against a concentration gradient (73, 108). This fact suggests that the processes of uniport, antiport, and symport are very similar from a mechanistic standpoint. We refer to all such transport proteins as facilitators.

The largest group of facilitators yet to be characterized currently has over 70 sequenced members and is referred to as the MFS (Table 1) (64, 109). Six families make up this superfamily, each only distantly related to the others (see Fig. 3A for the phylogenetic relationships of five of these families). Members of the first family are specific for drugs, and all such transport systems catalyze efflux of toxic substances. These porters, of bacterial or lower eukaryotic origins, are specific for quinolone, tetracycline, antiseptics, methylenomycin A, aminotriazole, chloramphenicol, and multiple drugs. The drugs may be co- or countertransported with or against various monovalent or divalent ions. The second family is ubiquitous and includes members that exhibit specificities toward sugars or organic anions such as glucose, galactose, arabinose, lactose, xylose, maltose, and quinate. These porters may function by uniport, proton symport, and/or solute:solute antiport. The remaining four families of the MFS function with a distinct class of compounds. Family 3 acts on Krebs cycle intermediates (citrate and α-ketoglutarate) and uses proton symport to accumulate the solute. Family 4 acts upon phosphorylated compounds (phosphoglycerates, hexose phosphates, and glycerol-3-phosphate) and uses phosphate antiport as the mechanism to effect accumulation of the organic phosphate ester. Family 5 is specific for oligosaccharides (raffinose, lactose, and sucrose) and uses proton symport to accumulate the solute.

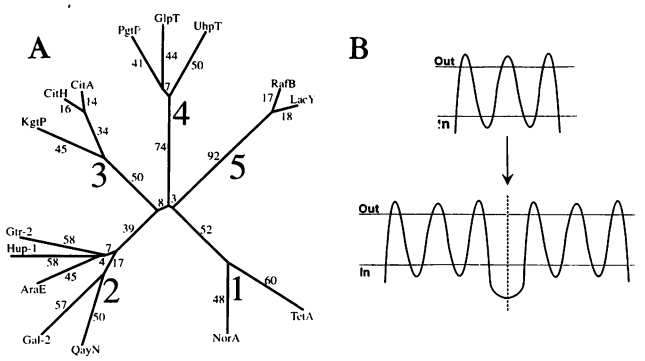

FIG. 3. (A) Phylogenetic tree of selected proteins of the MFS. Five of the six clusters discussed in the text (see the section on Solute Uniporters, Symporters, and Antiporters) are shown. These exhibit specificity for drugs (family 1), simple sugars and quinate (family 2), Krebs cycle intermediates (α-ketoglutarate and citrate) (family 3), phosphorylated organic esters (family 4), and oligosaccharides (raffinose, sucrose, and lactose) (family 5). Reproduced from reference 109 with permission of the publisher. Abbreviations and methods are presented in reference 109. (B) Proposed pathway for the evolution of genes encoding members of ubiquitous facilitator families such as the MFS, the APC family, and the SSF. It is possible that all three of these families share a common 6 + 6 (12)-spanner-encoding primordial gene, but this possibility cannot currently be established from sequence comparison data. It is possible (but not proven) that this primordial 12-spanner-encoding gene arose over 3.5 billion years ago by tandem duplication of a 6-spanner-encoding gene. The degrees of sequence similarity between the presumed repeat units in these permeases are insufficient to establish homology (64, 109).

Finally, family 6, not included in the phylogenetic tree shown in Fig. 3, includes a P_i porter of mammals that utilizes Na^+ symport to accumulate its substrate (94). It is therefore apparent that within this family of transport proteins, solute specificity and transport mode correlate well with phylogenetic grouping.

A distinct family of facilitators accumulates amino acids, polyamines, or choline (the APC family) (see reference 147 and references therein), utilizing either a proton symport mechanism, a solute:solute antiport mechanism, or both, depending on physiological conditions and the carrier under consideration. The proteins of this family resemble MFS porters in possessing two structural units, each consisting of six putative transmembrane α-helices and connected by a cytoplasmic loop. This structural feature is shared by facilitators that catalyze solute:Na^+ rather than solute:H^+ symport. With just one exception, the mammalian sodium:P_i symporter noted above, these proteins fall into several distinct and independent families (156). One of these families, the ubiquitous SSF (152, 156), is very diverse with respect to the solutes transported. These solutes include hexoses such as glucose and fructose in mammals, proline in bacteria, various neutral amino acids in mammals, pantothenate (a vitamin) in bacteria, adenosine (a nucleoside) in mammals, and myoinositol (a cyclic polyol) in mammals. Another such family, the SNF, transports neurotransmitters, small hormones, and osmolites, together with Na^+, and is found only in eukaryotes (30, 94, 219). Still others are found only in bacteria and are specific for an individual compound or a group of closely related compounds. The availability of additional related transport protein sequences may eventually allow the interconnection of families of facilitator proteins currently believed to be distinct.

P-type ATPases. A recent study (50) has revealed that all

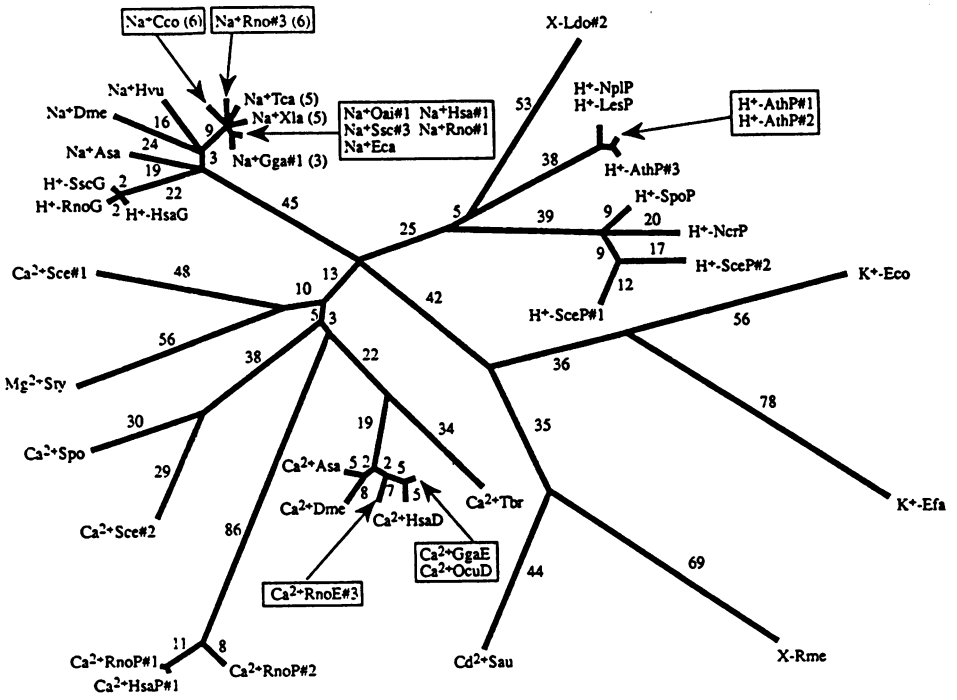

FIG. 4. Phylogenetic tree of 44 P-type, cation-transporting ATPases. The N-terminal, hydrophobic domains of these proteins were used for tree construction (see reference 50 for details of tree construction and species abbreviations). The four major clusters are specific for Ca^{2+} and Mg^{2+} (lower left), K^+ and H^+ (all from animal sources; upper left), H^+ (all from plants, fungi, and lower eukaryotes; upper right), and various mono- and divalent cations (all from bacteria except the Menkes disease Cu^{2+}-transporting ATPase of humans [not shown]; lower right). Note that the protein indicated K^+-Efa is now known to transport cupric ions (Cu^{2+}). Reproduced from reference 50 with permission of the publisher.

sequenced ATPases which form a phosphoryl intermediate (P-type ATPases) make up a single superfamily with four major subfamilies. These enzymes are all believed to utilize similar mechanisms to couple cation transport to ATP hydrolysis. Phylogenetic tree construction revealed that the P-type ATPase superfamily is subdivided largely in accordance with the cation specificities of the individual transport systems (Fig. 4). One cluster of these proteins (lower left cluster in Fig. 4) encompasses all of the Ca^{2+}-transporting enzymes found in plants, animals, and lower eukaryotes, as well as the bacterial Mg^{2+}-transporting ATPases. A second cluster (Fig. 4, upper left) is specific for Na^+ and K^+ or H^+ and K^+ and is thus far restricted to higher eukaryotes. A third group of P-type ATPases (Fig. 4, upper right) transports protons in yeasts, fungi, and lower eukaryotes. Some unsequenced bacterial ATPases specific for Ca^{2+}, Na^+, and H^+ may prove to cluster with these eukaryotic proteins (69, 70, 90, 93). However, the Na^+-transporting ATPase of the gram-positive eubacterium, *Enterococcus hirae*, has recently been shown to be of the V type (207).

The fourth cluster of P-type ATPases (Fig. 4, lower right) is diverse in its cation specificities and direction of transport. These porters are largely prokaryotic. One prokaryotic member of this group can pump K^+ into *E. coli* cells. Two others catalyze Cu^{2+} uptake and extrusion, respectively, in *Enterococcus hirae*, and a fourth catalyzes Cd^{2+} efflux in *Staphylococcus aureus* (50, 127). The one known eukaryotic member of this cluster, the Cu^{2+} efflux ATPase of humans, is apparently defective in Menkes disease (232). The human and the two bacterial cupric ion-transporting ATPases are more closely related to each other than they are to the other members of this subfamily, which exhibit differing cation specificities (unpublished results). Thus, whereas cluster 4 proteins are diverse in specificity, they also are widely divergent in sequence. As more of these proteins are diverse in specificity, they also are widely divergent in sequence. As more of these proteins are sequenced, a clearer correlation of sequence relatedness with cation specificity is likely to emerge. The four subfamilies of P-type ATPases presumably represent variations on a single mechanistic theme with different sets of related properties and cation specificities.

Use of Sequence Comparisons To Identify Conserved Regions of Presumed Structural or Functional Significance

Numerous studies have shown that the degree of conservation of a particular region or a specific residue in a family of homologous proteins correlates with its functional or structural importance. Conserved glycyl, prolyl, and hydrophobic residues usually have structural significance whereas hydrophilic or charged residues more often have catalytic significance. Examples of these principles have been noted for the P-type ATPases (50), in which regions of catalytic significance (i.e., the ATP-binding site, the aspartyl phosphorylation site) are very strongly conserved (see the next section). In the Enzymes I, the HPr proteins, and the Enzyme II complexes of the bacterial PTS, the most highly conserved regions surround the active site histidinyl and cysteinyl residues which become phosphorylated (112, 146, 148). Similarly, in ABC-type permease systems, the ATP-binding proteins are more highly conserved than are the integral membrane proteins or the periplasmic solute-binding receptors, and the sequences which make up the ATP-binding sites are more strongly conserved than other regions of the former proteins (74, 153, 164, 208).

Thus, regions exhibiting striking degrees of conservation almost always prove to be regions of specific functional significance.

Sequence Motifs as an Indication of Function

The presence of a specific motif may prove useful in identifying a functional type. Serrano (194) and Higgins (74) have summarized evidence for two widely distributed ATP-binding motifs (the Walker motifs A and B [236]), occurring in numerous transport proteins. The consensus sequences for these two motifs are (K/R) $(Hy)_{4-6}$G[XGXX or XXGX]GK and (K/R)$X_{2-3}GX_{2-3}(Hy)_5$D (ambiguous residues in parentheses; X represents any residue, and Hy represents a hydrophobic residue; alternative sequences in brackets). Transport proteins containing these sequences include (i) the α- and β-subunits of F_1-ATPases of bacteria, mitochondria, and chloroplasts; (ii) evolutionarily related V-type ATPases; and (iii) the energizing subunits of the ABC-type solute transporters. Many other proteins contain just sequence A, although additional variations on the sequence B theme may become recognized as more sequence analyses are conducted.

P-type ATPases possess an entirely different ATP-binding motif (50). This strongly conserved sequence and the sequence surrounding the aspartyl phosphorylation site in these cation-transporting ATPases are the most conserved sequences in these enzymes. The ATP-binding consensus motif is TGDG VNDHyPAL, whereas the aspartyl phosphorylation consensus motif is HyCSDKTGTHyT (Hy represents a hydrophobic residue and underlined residues are fully conserved in all currently sequenced P-type ATPases). These sequences not only allow identification of functional regions in these proteins but also serve as signature sequences for the identification of P-type ATPases (9, 50).

Although different families of transport proteins can be identified with the use of signature sequences, now available for several of the major transport protein families (Table 1), few additional functional motifs are found in these proteins. The presence of these other motifs in functionally uncharacterized sequences may provide evidence against a transport function (21, 33, 67, 145, 149, 223, 240).

GUIDES TO INTEGRAL MEMBRANE TRANSPORT PROTEIN STRUCTURE

Known and Postulated Three-Dimensional Structures of Membrane Proteins

Three-dimensional structural analyses by X-ray crystallography and by two-dimensional protein lattice analysis with electron microscopy have revealed two principal structural motifs which occur in transmembrane proteins: a modified β-barrel structure, exemplified by bacterial outer membrane porins (32, 241), and a structure consisting largely of several parallel membrane-spanning α-helices (spanners), established for the photosynthetic reaction centers localized to the inner membranes of gram-negative photosynthetic bacteria (38, 131), as well as for bacteriorhodopsin found in halophilic archaebacteria (128). A "hybrid" structure consisting of both α-helical and β-strand transmembrane elements has been proposed for the acetylcholine receptor (220), but the low-resolution (9 Å [0.9 nm]) structure obtained is insufficient to establish this suggestion.

All of the proteins to be discussed here are believed to consist primarily of transmembrane α-helical structural elements. They are all found in the cytoplasmic membranes of bacteria and eukaryotes or in the membranes of eukaryotic organelles. Significantly, units of six spanners, occurring most frequently in duplicate or quadruplicate, typify the structures of many of these transport proteins (Table 2) (126, 171, 177). For example, several members of the MFS have been shown to possess 12 spanners (see below). It is important to note, however, that families of transport proteins postulated or known to have fewer or more spanners per unit (from 3 to 8 and possibly up to 14) have been sequenced and topologically characterized (24, 97, 128, 130, 199, 237). Available computer methods for predicting the subcellular localizations of soluble and transmembrane proteins as well as their topologies and conformations have been presented and evaluated (18, 27, 51, 78, 119, 186, 191, 197, 230, 231).

Topological Predictions for Sequenced Integral Membrane Transport Proteins

The three-dimensional structures of the transport proteins to be discussed in this section have not yet been elucidated. Because crystallization of these proteins is proving to be an arduous task which has met with repeated failure in the past, alternative approaches must be contemplated. Use of primary-sequence data to estimate the topologies of transmembrane proteins, coupled with molecular genetic and biochemical approaches, particularly involving the construction and characterization of transport protein-reporter gene fusions, has proven illuminating. Thus, the lactose:H^+ symporter, the α-ketoglutarate:H^+ symporter, a tetracycline:H^+ symporter, and the hexose-phosphate:phosphate antiporter, all homologous *E. coli* proteins of the MFS (93), have been shown to have a common 6 + 6 or 12 spanner structure (3, 25, 45, 103, 193).

Topological predictions based exclusively on hydropathy analyses (96) of the sequences of these same proteins have provided a surprising degree of agreement with the empirical results (i.e., see reference 64). Recently, von Heijne (231) and Sipos and von Heijne (197) have further enhanced the reliability of these methods by superimposing the "positive inside" rule onto traditional hydropathy analyses. This rule states that strongly basic residues (arginine and lysine) occur with much higher frequency in cytoplasmic protein loops connecting transmembrane spanners than in extracellular loops. Although application of this method to selected bacterial and eukaryotic membrane proteins, including several members of the MFS, was shown to correctly predict their topologies, relatively poor predictive capacity was observed for PTS and MIP family permeases, in which the number of spanners was overestimated, as well as for MCF proteins, in which the number of spanners was underestimated (213a). In the MCF proteins, underestimation was due to the hydrophilic nature of some of the transmembrane spanners.

Examination of several families of transport proteins has revealed that, in general, transmembrane helices are more strongly conserved than loop regions joining these spanners and that these loops are frequently of variable length, suggesting that insertional and deletional mutations were permissively introduced during evolution. Such events suggest either a relative lack of importance for structure or function or a specialization of function for the different protein members of the family. Very clear examples of this type of behavior are provided by the proteins of the mitochondrial carrier family (95). All six putative transmembrane helices are strongly conserved with few gaps in the multiple alignment, and these spanners can be clearly defined by virtue of the sequence similarities observed for the three internal, two-spanner repeats (see below). Almost all gaps in the aligned sequences

occur between these spanners. The MIP family of channel proteins presents a similar picture. The six putative spanners are highly conserved, whereas poor conservation and numerous gaps are observed for the loop regions connecting these spanners (155).

Strongly hydrophobic stretches of 18 or more residues do not always indicate the presence of a transmembrane helical segment. Thus, the exopolyphosphatase and the guanosine pentaphosphatase of *E. coli*, two strikingly similar proteins, both possess hydrophobic regions, 18 amino acyl residues long, which are located about 270 residues from the N termini of these proteins (150). By analogy with glycerol kinase, believed to be homologous to the phosphatases, the region in question is predicted to serve as an interior α-helix connecting two globular domains of each protein by hydrophobic forces. This example illustrates the danger in interpreting hydropathy plots indiscriminately in terms of transmembrane topology.

Conserved glycyl, prolyl, and hydrophobic residues which do not have side chains that can function in catalysis are generally believed to have structural significance. Glycyl and prolyl residues, for example, normally occur with low frequency in α-helices and β-strands of soluble proteins, but they occur with high frequency in β-turns, where the peptide chain changes direction (167). In spite of these observations with soluble proteins, careful surveys of membrane proteins have surprisingly revealed that prolyl and glycyl residues are often buried in transmembrane helices of catalytic or functional membrane proteins but not in transmembrane regions of proteins which merely serve to anchor the protein to the membrane (19, 20, 36). It has been suggested that the lipid environment stabilizes a folded structure in which prolines in transmembrane helices introduce slight changes in direction or "kinks" in the helix axis without disrupting the helix as a whole (37, 254).

In both bacteriorhodopsin of *Halobacterium halobium* (116) and the lactose permease of *E. coli* (160), systematic replacement of prolines and other structural residues in the transmembrane helices has been achieved. In most cases, complete loss of transport function does not result from any one substitution, although reduced activity is frequently observed. Moreover, in the ion channel-forming peptide, pardaxin, replacement of the putative transmembrane proline at position 13 with L-alanine or D-proline impaired channel conductivity and blocked voltage activation (139, 195). It therefore appears that conserved transmembrane residues in functionally active integral membrane proteins determine the detailed topology of the helix and its orientation to other helices in the protein. Helix distortion presumably allows for optimal function of the protein, possibly by allowing for movement of the protein helices in the membrane and hence of the substrate across the membrane.

Examination of strongly conserved residues in the mitochondrial carriers proves interesting in this regard. The three odd-numbered spanners were found to be more hydrophobic than the three even-numbered spanners, and the former contained strongly conserved glycyl and prolyl residues, presumably allowing flexibility or helix deformation or both (95). Each of the even-numbered spanners was preceded by a conserved glycyl residue, which can effectively participate in a β-turn (167), and positions were found in the homologous transmembrane helices which showed conservation of hydrophilicity. Five of the six helices were followed by the largely conserved sequence, (D/E)-Hy(K/R) (- represents any residue, Hy represents hydrophobic, D and E are acidic, and K and R are basic) (92, 95). On the basis of the relative degrees of conservation and the nature of the residues conserved, it was possible to predict that residues in the odd-numbered helices are more important for flexibility and structural maintenance of the channels whereas residues in the even-numbered helices may line the aqueous pores and provide specificity.

Delimitation of Protein Domains Based on Interdomain Linker Flexibility

The study of protein structure in essence boils down to the study of the constituent domain structures and their interactions (28, 29). As noted above in sections dealing with the bacterial PTS, this enzyme system normally consists of the two general energy coupling proteins, Enzyme I and HPr, as well as three sugar-specific proteins or protein domains termed IIA, IIB, and IIC (178). When these protein domains are covalently linked to one another in single polypeptide chains, they are usually connected by recognizable flexible linkers.

Two types of such linkers, termed AP (alanine-proline-rich) linkers (142) and Q (glutamine-rich) linkers (243), have been found repeatedly in PTS proteins. Wu et al. (249) sequenced a multiphosphoryl transfer protein of the fructose-specific PTS of *Rhodobacter capsulatus* in which a single large polypeptide chain was found to encode the structural and functional equivalents of Enzyme I (C terminal), HPr (central) and Enzyme IIA^{Fru} (N terminal). Connecting these three catalytic domains were AP-rich linkers, not only of similar length and amino acid composition but also of similar sequence. Both linkers began with the sequence GAAAP-AA. At the DNA level, these linker regions were found to form imperfect inverted repeats (palindromes) of greater than 90% G+C content. Similar linkers were found in five other sequenced PTS proteins as well as in numerous other proteins of prokaryotic, eukaryotic, and viral origins (249). In virtually all cases that had been examined, these AP-rich regions served as flexible segments connecting relatively rigid protein domains.

The mannose Enzyme II complex of the *E. coli* PTS is perhaps best characterized in this regard. In the Enzyme IIAB protein of this complex, the AP-rich linker could be drastically increased or decreased in length, and it could even be cleaved, without appreciable loss or gain of function (48). This fact explains why frequent insertions and deletions have occurred during the evolution of these regions, giving rise to gaps in the interdomain regions of the multiple alignments of homologous multidomain proteins.

The genetic origin of flexible linkers is an interesting topic (see references 246 and 249 for discussions of the possibilities). Because AP-rich linkers are encoded by G+C-rich DNA sequences which, at least in some instances, form stable stem-loop structures, they may have arisen as autonomous genetic elements which could insert between cistrons to fuse protein domains into longer polypeptide chains. Further, because of their unusual nucleotide sequences and compositions, they may serve as hot spots for recombination. Computer analyses of the gene segments encoding large numbers of these linkers might be revealing.

Q-linkers most commonly connect the IIA domains of PTS Enzymes II with the IIB and IIC domains (205). In addition to glutamine, the presence of other strongly hydrophilic residues such as arginine and glutamate, as well as of semipolar residues such as serine and proline, characterizes these linkers. Among the PTS permeases containing Q-linkers are those specific for glucose in *Bacillus subtilis*, sucrose in *Streptococcus mutans*, β-glucosides in *E. coli*, and *N*-acetylglucosamine in *E. coli*. Additionally, Q-linkers connect IIA^{Glc}-like domains to the lactose permeases of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* (136, 137). These lactose permeases function by H^+ symport, and the IIA^{Glc} domains, linked to the C-

termini of these permeases, can be phosphorylated by the PTS and presumably play a regulatory role rather than a catalytic role in sugar transport (205).

When a recognizable AP- or Q-linker is not identifiable, the presence of other types of flexible linkers such as highly charged linkers (60) can sometimes be found by using appropriate computer programs. Their identification is useful not only for defining the boundaries of discrete domains but also for estimating the types and extent of motion these domains may undergo, relative to each other (18, 28, 29, 47, 143).

Targeting Sequences and Subcellular Localization

In eukaryotic cells, a variety of targeting sequences may be included in proteins to determine their subcellular localization. C-terminal targeting sequences are common in proteins that are exported from the bacterial cytoplasm via ABC-type exporters (52). Cleavable, N-terminal hydrophobic leaders are found in proteins targeted to the eukaryotic endoplasmic reticulum or the gram-negative bacterial periplasm, whereas amphipathic leaders target nucleus-encoded proteins of eukaryotes to mitochondria and chloroplasts (61, 161, 181, 227, 228) and prokaryotic PTS permeases to the cytoplasmic membrane (183, 184, 253). Identification of hydrophobic and amphipathic leader sequences and other types of organelle-specific targeting sequences can lead to correct postulation of the subcellular localization of a protein and consequently to a clearer understanding of its function.

N-terminal hydrophobic leaders are characteristically 15 to 30 residues in length. They consist of a positively charged N terminus, a central hydrophobic region, and a C-terminal region, predominating in polar residues. Usually within this third region is a cleavage site for proteolytic removal of the targeting signal following export. Many such signal sequences have been tabulated, and analyses of their characteristics have been published (59, 229, 239). Because there is more than one signal peptidase in most organisms, there is more than one cleavage site consensus sequence. Most proteins exported to the periplasm or outer membrane of *E. coli* are cleaved by the major signal peptidase, but lipoprotein signal peptidase recognizes a different sequence (see references 172, 181, and 208 for reviews).

Amphipathic sequences in any region of a protein consisting of β-strands or helices of any dimension can be identified by using appropriate computer programs (7, 47, 143). On the basis of computer analyses of almost 2,000 proteins, it was suggested that N-terminal, amphipathic, α-helical sequences function primarily in macromolecular recognition (174). N-terminal, amphipathic leaders of PTS permeases, having strongly polar residues on one side of an α-helix and strongly nonpolar residues on the other side, have been shown to play a role in the insertion of these proteins into the membrane. Further, they can initiate secretion of alkaline phosphatase into the periplasm when these leaders replace the natural hydrophobic leader of alkaline phosphatase (253). Like mitochondrial targeting sequences, they exhibit a high proportion of hydrophobic and basic residues but a low proportion of helix breakers (proline, glycine, and cysteine). Unlike mitochondrial targeting sequences, they show an increased proportion of acidic residues but a decreased proportion of hydroxy residues. Further, although the proportions of hydrophobic and basic residues are elevated in both the prokaryotic and eukaryotic leaders, they differ in content. Isoleucine and lysine predominate in bacterial leaders whereas leucine and arginine predominate in mitochondrial targeting sequences (181). Although these differences are not understood from a functional standpoint, they clearly must reflect specific structural or functional requirements of the two systems.

Specific mutations in the amphipathic leader of one of the PTS permeases, the mannitol Enzyme II, were found to block insertion of the protein into the membrane. Moreover, when the same mutations were introduced into exported *mtlA-phoA* fusion proteins, export was blocked (253). Recently it has been found that peptides corresponding in sequence to the natural PTS permease leaders readily incorporate into planar phospholipid bilayers with a preferential alignment of the α-helix long axis parallel to the membrane surface (210). When the sequence of the mannitol leader peptide was changed to correspond to those of the nonfunctional mutant leaders, the peptides retained their helical character in aqueous solution but became disordered in lipid bilayers. Further, these peptides showed an enhanced ability to disorder the lipid molecules within the bilayer (210). It therefore seems that helical association of the peptide with the lipid bilayer may be a prerequisite to proper association with the secretory machinery.

RECONSTRUCTING PATHWAYS OF TRANSPORT PROTEIN EVOLUTION

With the sequences of literally hundreds of transport proteins now available, it has become possible to delineate the major families of transport proteins, to define their structural similarities and differences, and to gain some evidence regarding the times, places, and pathways of their evolution. The available evidence suggests that several families of transport proteins evolved independently of each other, although others, distantly related by descent, together make up large superfamilies. In other cases, indirect evidence suggests that certain families of transport systems exhibit the characteristic of homology, although this cannot be established from sequence analyses alone owing to large degrees of sequence divergence. Finally, many transport proteins for which sequence data are available do not fall into any of the well-defined transporter families. Some of these may represent patriarchs of new families that are yet to be recognized, but others share a sufficient number of characteristics with members of established families to suggest a common origin. Three-dimensional analyses will ultimately be required to help define the evolutionary relationships of some of these distantly related proteins. The categorization of transport proteins into their various families, the establishment of superfamilies, and reconstruction of the pathways that led to their appearance are the primary topics of this section.

Identifying Genetic Rearrangements during Evolution

Some transport protein families have evolved with the apparent lack of genetic rearrangements, so that members of the family exhibit continuous sequence similarity from their N termini to their C termini. An apparent lack of gene splicing, fusion, and transposition during the evolution of the MFS (109), the SSF (156), and the APC family (147) has been noted. Other, more recently appearing families, the MCF and the MIP family, include members which similarly have not undergone obvious genetic rearrangements during their evolution.

Multiple domains are found within the PTS- and ABC-type permeases, and, in contrast to the facilitators mentioned above, these permease systems have undergone extensive intragenic splicing, fusion, duplication, deletion, and shuffling during their evolution (5, 52, 74, 146, 178, 247). These genetic rearrangement events have almost always occurred between

rather than within genetic elements coding for discrete domains. Additionally, "foreign" domains have sometimes become associated with these permeases as is observed for the fructose PTS of *Salmonella typhimurium* (249) and the cystic fibrosis transport regulator of the ABC-type transporter family (74). Although the pairing and splicing of certain domains in these permeases have occurred frequently, those of others, particularly in the ABC family, have apparently occurred as rare evolutionary events (52).

Examining the transport proteins of the PTS, the three essential domains of the Enzyme II complexes, IIA, IIB, and IIC, can theoretically occur within a single polypeptide chain in any one of six orders. Only two such orders are actually found: CBA and BCA (N terminal to C terminal). For example, in *E. coli*, the *N*-acetylglucosamine Enzyme II has its domains in the order C, B, A, but the homologous β-glucoside permease has its domains in the order B, C, A. The fructose protein has two B domains in the order B′, B, C, and it functions with a IIA protein that is linked to other domains that function to phosphorylate it. It is most closely related to the mannitol Enzyme II, which has the domain order C, B, A. It is therefore clear that the intragenic shuffling events, resulting in altered domain order in the protein products, have occurred more than once during the evolution of the PTS permeases but in a nonrandom fashion (146, 178).

Additionally, the various domains may be lost or be present as distinct polypeptide chains. The sucrose Enzymes IIBC of *E. coli* and *B. subtilis* have no sucrose-specific A domain at all, presumably because the IIA^{Scr} structural genes have been either lost or silenced. They function with the glucose IIA. By contrast, in *Streptococcus mutans*, IIA^{Scr} is found covalently linked to $IIBC^{Scr}$. The trehalose Enzyme II similarly functions with IIA^{Glc} (91a). In the cellobiose Enzyme II complex of *E. coli*, each of the three domains (IIA, IIB, and IIC) forms a distinct protein (151, 178).

The PTS in most bacteria consists of Enzyme I and HPr, two general energy-coupling proteins of the PTS, as well as the Enzyme II complexes. In enteric bacteria the fructose operon encodes a three-domain protein, the FruBMH protein, which has a IIA^{Fru} domain (the B domain) linked to the M (modulator) domain, which lacks homology with other PTS permease domains, linked in turn to an HPr-like (H) domain termed FPr (178, 249). Further, in *Rhodobacter capsulatus*, the IIA^{Fru} domain, HPr, and Enzyme I are covalently linked in a single polypeptide chain as noted above. Newly discovered genes encoding PTS proteins, sequenced as a result of the *E. coli* genome-sequencing project, have revealed additional variations on the known themes (156a). It is therefore clear that domain fusion, splicing, duplication, deletion, and shuffling have occurred repeatedly at the genetic level during evolution of the system.

Domain organization of ABC transporters has recently been reviewed (52, 74), and the interested reader is referred to these articles for more detailed discussions of the various documented domain combinations known to occur in this superfamily of permeases. Of major evolutionary interest is the fact that when phylogenetic trees of the ABC domains of these permeases are constructed, it is found that those covalently linked to the transmembrane transport domains (all sequenced eukaryotic proteins and many bacterial export systems) form one major cluster whereas those which are not covalently linked (all sequenced bacterial uptake systems as well as the remaining bacterial export systems) make up the other. This observation suggests (but does not establish) that a single fusion or splicing event occurred early during the evolution of this family of proteins and that the fusional state of these primordial proteins was faithfully transmitted through evolutionary history to the present day. Other domain splicing and fusion events (e.g., between two ABC domains and between the two integral membrane domains) occurred more recently and more often during the evolutionary process (52, 74).

Parallel versus Independent Evolution of Domains within Multidomain Proteins

When proteins of the PTS are analyzed, it is found that, in general, the different domains (A, B, and C) of homologous Enzyme II complexes evolved in parallel from a single primordial complex even though intragenic rearrangements have given rise to domain shuffling (146, 178). Thus, phylogenetic trees of the Enzymes II show similar patterns regardless of the domain analyzed. This need not necessarily be the case.

The solute-binding receptors that function with bacterial ABC-type uptake systems are synthesized with an N-terminal signal sequence that targets the newly synthesized protein to the extracytoplasmic milieu. These receptors fall into several families of demonstrably homologous proteins as discussed in the section Periplasmic Receptors of ABC-Type Uptake Permeases, above. One of these families of receptors, specific for monosaccharides such as galactose, ribose, and xylose (Fig. 2B), includes a large number of homologous proteins which function in the cytoplasm as transcriptional repressors (192, 223, 240). The latter proteins include the lactose repressor, LacI, the galactose repressor, GalR, and the fructose repressor, FruR, all of *E. coli*. Another family of extracytoplasmic receptors, specific for aliphatic amino acids such as leucine, isoleucine, and valine, also includes a cytoplasmic repressor protein, the AmiC protein, which binds aliphatic amides and thereby regulates transcription of an aliphatic amidase in *P. aeruginosa* (209, 242).

Examination of these proteins reveals that all of the transcriptional regulatory proteins lack the N-terminal periplasmic targeting sequence and, instead, possess an N-terminal DNA-binding domain with a typical helix-turn-helix motif. The N-terminal signal sequences of the receptors show no detectable similarity to the N-terminal DNA-binding domains of the repressors. It is therefore probable that domain swapping occurred in the evolution of these proteins. The sugar-binding domains of both the external receptors and the cytoplasmic repressors are evolutionarily related by common descent, but the N-terminal domains of these proteins are not. Their common functional feature is ligand binding, even though the biological functions of these two classes of evolutionarily related proteins are entirely different. In this case domain swapping allowed the evolution of proteins of differing function with the use of common structural elements.

Among the neurotransmitter-binding receptors of higher organisms are two types of glutamate receptors (75). There are the ionotropic (ion-conducting, channel-type) receptors and the metabotropic (G-protein-coupled) receptors (120, 130). In addition to the transmembrane channel and GTP-binding protein-interacting domains, both classes of these proteins possess putative, glutamate-binding domains localized to the external surface of the cell. The glutamate-binding domains are homologous to the bacterial polar amino acid-binding receptors, although the integral membrane constituents of the neurotransmitter receptors are dissimilar in sequence and structure to the transmembrane domains of ABC-type transporters (120, 130). It seems highly likely that gene fusion during evolution was responsible for the appearance of homologous solute-binding domains within these two dissimilar types of neurotransmitter receptors.

Intragenic Duplication as an Indicator of Independent Evolution

Extensive evidence now suggests that several families of transport proteins that are constructed from similar structural units of six semiparallel transmembrane spanners nevertheless arose independently of each other at different times in evolutionary history. The ABC superfamily and the major facilitator superfamily probably arose first, over 3.5 billion years ago; the MIP family possibly arose second, about 2.5 billion years ago; the RND family arose in its present form soon thereafter (possibly about 2.0 billion years ago) by a distinct route; the MCF arose next, most probably about 1.5 billion years ago, after the advent of eukaryotes; and the current forms of voltage-sensitive Na^+ and Ca^{2+} channels probably arose most recently, maybe even less than 1 billion years ago, even though primitive ion channels of related sequence clearly existed before this time. All of these transport protein families arose by distinguishable tandem intragenic duplication events that led to superficially similar structural units. Below, these transport protein families will be discussed in the order in which they were presumed to have appeared.

MFS, ABC-type permeases, and arsenical resistance pumps. Carrier-type transport systems that allow the transmembrane passage of solutes by facilitation (uniport), countertransport (antiport), or ion cotransport (symport) are believed to function by similar mechanisms (166). This belief is supported by recent sequence analyses which have shown that sugar facilitators of eukaryotes are homologous to symporters and antiporters of bacteria (64, 73, 108, 109, but see reference 58). In fact, this large superfamily of proteins extends from enteric bacteria and cyanobacteria to the lower eukaryotes such as yeasts, algae, and protozoans and on up to higher plants and animals including humans. These facilitators are specific for simple sugars, oligosaccharides, organic acids, organophosphate esters, and drugs (Fig. 3A) (109). They consist of protein subunits each consisting of two units of six transmembrane helical segments separated by a cytoplasmically localized loop of variable length (Table 2). These 12 (6 + 6) spanner proteins may have been derived from a common ancestral protein of half this size, although this possibility cannot be established from sequence comparisons (Fig. 3B) (see references 3, 64, and 193) for discussions of these possibilities and references thereto). This family of transmembrane solute facilitators, the MFS, is a very old family which evidently arose in prokaryotes and probably dates back more than 3.5 billion years (64, 109).

ABC-type permeases possess integral membrane subunits which can either be homodimeric or heterodimeric, in which instances each subunit contains six transmembrane spanners, or it can be "pseudodimeric," in which case a single polypeptide chain forms two units of six spanners with a central, cytoplasmic loop connecting them (52, 74, 88, 153). The last-mentioned "pseudodimeric" integral membrane proteins superficially resemble facilitators of the MFS, but an evolutionary connection has not been established. Since at least some MFS permeases can be cleaved in the central loop without loss of activity (160), a striking structural parallel between the MFS proteins and the ABC-type family of permeases can be drawn.

Relevant to this fact are recent developments concerning the arsenical resistance (A-type) ATPases encoded by various bacterial plasmids (163, 196). These efflux pumps render bacteria resistant to arsenite, antimonite, and tellurite (85, 86, 218). The topology of the integral membrane ArsB protein of *E. coli* resembles that of the iron-hydroxamate ABC-type permease of *E. coli* and the MFS permeases in having 12 spanners (244). Insufficient sequence similarity is observed, however, to allow the postulation of homology. A second dimeric subunit of the *E. coli* arsenical resistance pump, ArsA, serves as the ATP-binding constituent of the permease which couples ATP hydrolysis to oxyanion efflux (85–87). This subunit superficially resembles the ABC domains of ABC-type permeases, although a degree of sequence similarity sufficient to establish homology is not observed.

In two sequenced *ars* operons present on staphylococcal plasmids, *arsB* genes which encode proteins that exhibit 58% identity to the *E. coli* ArsB protein are found. Surprisingly, these two operons do not encode ArsA proteins, which normally provide the ATP-dependent energy-coupling function (196). Recent analyses have suggested that efflux of arsenite via the staphylococcal ArsB-dependent, ArsA-independent transport system can be driven by the membrane potential (22, 164). In this respect, the system resembles the MFS permeases. However, when the *E. coli arsA* gene is coexpressed in *trans* with the staphylococcal *arsB* gene, increased resistance to arsenite is observed relative to cells bearing only the *arsB* gene. This increased degree of resistance is apparently due to efficient ATP-driven arsenite efflux as observed for the *E. coli* arsenical resistance pump. The system thus functionally resembles a typical ABC-type export permease when ArsA is available but an MFS permease when ArsA is not present (52, 164). These exciting results suggest a functional (and thus possibly an evolutionary) connection between the integral membrane constituents of the arsenical resistance pump, the structurally similar ABC-type ATP-driven permeases, and the MFS-type proton motive force-driven permeases.

MIP family. Sequence analyses have revealed a family of intrinsic membrane "channel" proteins from plants, animals, yeasts, and bacteria (Fig. 5A) (10, 72, 132, 155). These proteins include the MIP from the lens fiber cell of the mammalian eye, after which this family was named. Functional and structural aspects of these channel proteins have recently been summarized (46, 155). Five of them possess recognized transport functions: (i) MIP probably transports salt, particularly Na^+; (ii) ChIP, the channel intrinsic protein of human erythrocytes and of kidney epithelia, transports water and is apparently the long-sought aqua channel of these cells and tissues; (iii) TIP, a group of tonoplast intrinsic proteins of plants, also serves as a water channel, the various isoforms being present in diverse plant tissues and organelles: (iv) NOD, nodulin-26 of soybeans, may transport dicarboxylates across the peribacteroid membrane in a process which allows communication between the host plant cell and symbiotic, nitrogen-fixing *Rhizobium* species; and (v) GLP, the glycerol facilitator of *E. coli*, *Bacillus subtilis*, *Streptomyces coelicolor*, and possibly *Lactococcus lactis*, allows nonspecific passive diffusion of glycerol and other straight-chain carbon compounds across the bacterial cytoplasmic membrane (72, 111, 113, 155, 221). The invertebrate protein Big Brain (BIB) is an essential neurogenic protein in *Drosophila melanogaster* and apparently has a human analog (1), but its transport specificity has not been determined. The yeast protein FPS-1, which functions as a suppressor of a genetic growth deficit on fermentable sugars, is also not characterized biochemically. Most of these proteins are similar in size, have six putative membrane-spanning domains per subunit, and may be tetrameric (MIP) or dimeric (GLP) (Table 2) (2, 198, 225, 226).

Sequence comparisons surprisingly revealed that the putative six transmembrane helical segments in each of these proteins arose by intragenic duplication of a primordial gene encoding just three transmembrane helical segments (Fig. 5B) (132, 155). The two halves of these proteins thus have opposite

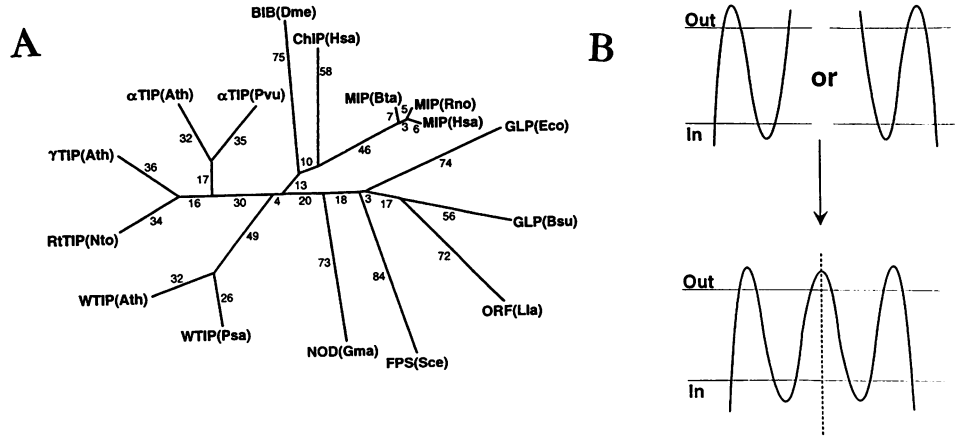

FIG. 5. (A) Phylogenetic tree of the ubiquitous MIP family of channel proteins. All plant proteins (NOD and TIP) are located in one cluster (lower left). All proteins from animal sources (BIB, ChIP, and MIP) form another cluster. The single yeast protein (FPS) forms a distinct branch (lower center). Finally, the bacterial glycerol facilitators (GLP) make up a fourth diffuse cluster (right). Reproduced from reference 155, with permission of the publisher; abbreviations and methods of tree construction are described in reference 155. (B) Proposed pathway for the evolution of genes encoding protein members of the MIP family. The primordial gene is believed to have encoded a single three-spanner unit. This genetic element tandemly duplicated to give a gene encoding a single six-spanner unit. The duplication event is estimated to have occurred about 2.5 billion years ago in prokaryotes. Vertical transmission to eukaryotes followed by duplication and divergence gave rise to most sequenced members of the family. Homology between repeat segments has been established for some of the proteins (155).

orientation in the membrane. On the basis of the degree of similarity of the two repeat units in these proteins, this duplication event, giving rise to the MIP precursor protein, is estimated to have occurred about 2.5 billion years ago in prokaryotes. The common prokaryotic MIP family precursor gene, which presumably encoded a glycerol facilitator, was then transmitted vertically to eukaryotes, in which it underwent extensive duplication and divergence so that the encoded proteins gained a diversity of functions, subcellular locations, and substrate specificities (46, 155). Statistical analyses suggested that duplication and divergence of the primordial MIP family gene in eukaryotes to give the various eukaryotic proteins of differing specificities occurred long after the intragenic duplication event which gave rise to the *glpF* primordial gene in prokaryotes (155). Gene duplication in the various eukaryotic kingdoms accounts for the grouping of proteins in Fig. 5A according to organismal phylogeny.

RND family. A novel family of bacterial proteins which functions generally in heavy metal ion (Co^{2+}, Ni^{2+}, Cd^{2+} and Zn^{2+}) resistance in (i) *Alcaligenes eutrophus* (102, 124), (ii) nodulation of alfalfa by *Rhizobium meliloti* (8), and (iii) cell division as well as resistance to toxic substances (antibiotics, detergents, and crystal violet) in *E. coli* (91) (the RND family of presumed transporters [Table 1; Fig. 6] [180]) has recently been identified. The substrates of some but not all protein members of this family have been identified. We have provided evidence suggesting that these presumed transport systems catalyze solute (or macromolecular) efflux (180). They consist of a long N-terminal hydrophilic stretch of about 300 amino acyl residues; a hydrophobic stretch of about 200 residues, which, on the basis of hydropathy analyses, appears to contain four to six transmembrane spanners; and a repeat sequence of 500 residues with a hydrophilic 300-residue region again preceding a C-terminal hydrophobic region of four to six spanners. These proteins thus consist of about 1,000 residues forming alternating hydrophilic and hydrophobic regions.

Quantitative sequence comparisons of the first and second halves of these large proteins established the suggestion that they arose by a tandem intragenic duplication event (Fig. 6) (180). The first 500 residues thus share a common origin with the second 500 residues. Although the comparison score for the two halves of each of these proteins varies substantially depending on the protein examined, one of these proteins, the cobalt, zinc, cadmium resistance pump (CzcA protein) of *A. eutrophus*, exhibits a score of 14 SDs when the two halves are compared (180). This high score indicates that the duplication event must have occurred relatively recently during evolutionary history, perhaps about 2 billion years ago, long after the MFS was established and probably after the MIP family was established as well.

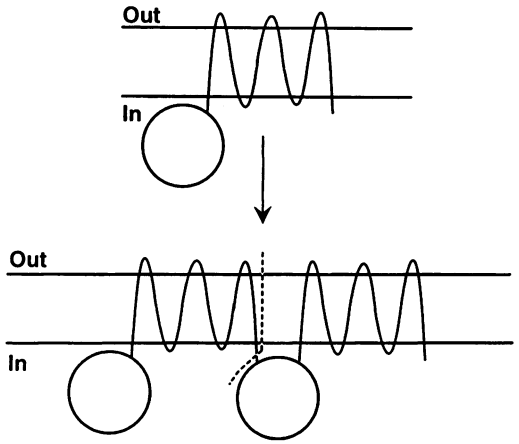

FIG. 6. Proposed pathway for the evolution of genes encoding members of the bacterial metal RND family of transporters. The primordial gene is believed to have encoded an N-terminal, 300-residue, hydrophilic domain localized to the cytoplasmic side of the membrane (large circles) and a C-terminal, 200-residue, hydrophobic, transmembrane domain of four to six spanners. It tandemly duplicated to give a gene of twice this size, encoding a 1,000-residue protein. This duplication event may have occurred about 2.0 billion years ago. The homology of the two repeats in one such protein (CzcA) has been established (180).

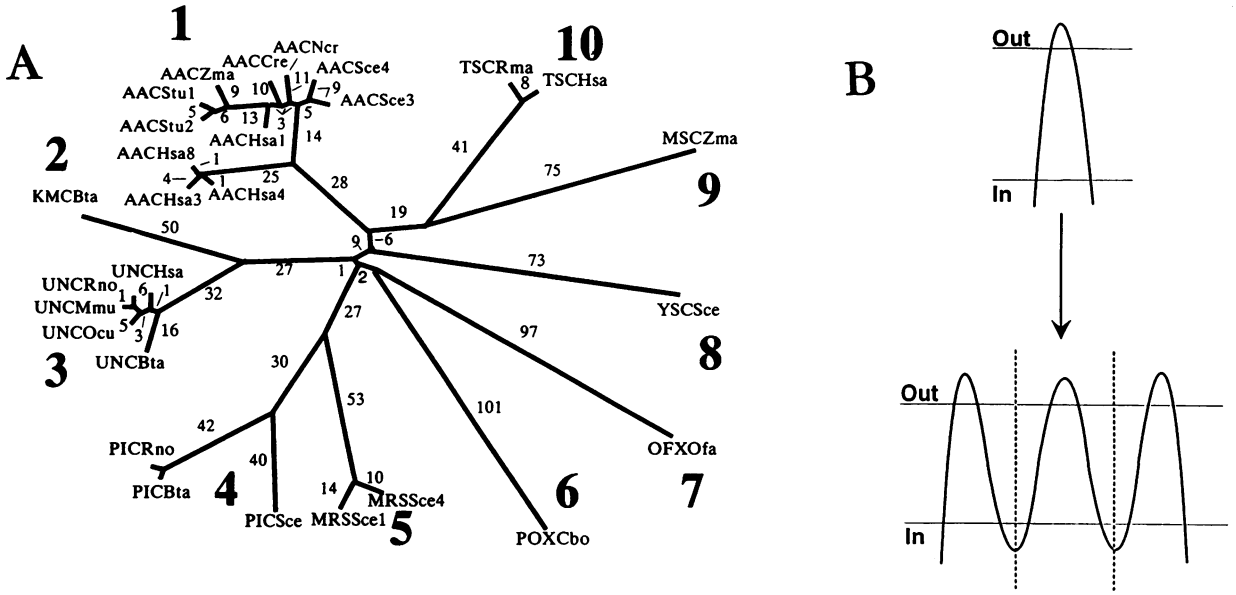

FIG. 7. (A) Phylogenetic tree for the proteins of the MCF. These proteins include the ATP-ADP exchangers (AAC; cluster 1), the α-ketoglutarate–malate exchanger (BMC; cluster 2), the uncoupling protein of brown adipose tissue (UNC; cluster 3), and the P_i carriers (PIC; cluster 4), all of eukaryotic mitochondria. Other proteins have not been functionally identified. Reproduced from reference 95 with permission of the publisher; abbreviations and methods are presented in reference 195. (B) Proposed pathway for the evolution of genes encoding members of the MCF. The primordial gene is believed to have encoded a single two-spanner unit. It tandemly triplicated to give a gene encoding a single six-spanner unit. On the basis of the degree of sequence similarity between the two-spanner repeats, this event may have occurred about 1.5 billion years ago in eukaryotes. The degree of sequence similarity observed between two-spanner repeat units is sufficient to establish homology (95).

Each of several members of the RND family of exporters functions in conjunction with a cytoplasmic membrane protein of about 400 amino acyl residues. The latter protein possesses an N-terminal membrane anchor but otherwise extends from the cytoplasmic membrane into the periplasm of the gram-negative bacterial cell, where it may interact with the outer membrane (190). Homologous proteins function in conjunction with transporters of the MFS, which export drugs, as well as with transporters of the ABC family, which export proteins (54, 104). All transporters which function with these proteins appear to catalyze efflux of fairly large substrates, substances that are too large to pass through the porin-type channels in the outer membrane (180). As a result of these observations and the nature of conserved residues in these proteins, it has been proposed that they function to allow the transporters to export their substrates across both the inner and outer membranes rather than across just the inner membrane (180). Possibly this family of homologous proteins functions to allow localized fusion of the inner and outer membranes, thereby creating a single rather than a double barrier to transport, just in the vicinity of the transport system. Alternatively, they may allow creation of a *trans*-outer membrane pore which is functionally connected to the inner membrane permease.

MCF. In mitochondria of eukaryotes, there exists a number of functionally distinct carriers which include an ATP-ADP exchanger, a phosphate porter, an α-ketoglutarate–malate exchanger, a tricarboxylate carrier, and the uncoupling protein of mammalian brown adipose tissue. Sequence comparisons have shown that these porters, as well as six clusters of functionally uncharacterized (or poorly characterized) proteins on the MCF phylogenetic tree, are homologous (Fig. 7A) (83, 92, 95, 95a, 115, 127). Each subunit of each such transport system apparently spans the membrane 6 times, and the dimeric forms of these proteins are presumed to be the structural and functional equivalents of the monomeric forms of carriers that span the membrane 12 times (Table 2). Despite their superficial resemblance to other transport protein families, they nevertheless form a distinct evolutionarily unrelated family.

The MCF clearly arose by triplication of a small primordial gene that encoded a protein having only two transmembrane helical segments (Fig. 7B) (92, 95), and this triplication event is believed to have occurred after the appearance of eukaryotes. It may have arisen specifically for eukaryotic organellar function because no member of this family has yet been identified in bacteria or in the cytoplasmic membranes of eukaryotes.

The degree of similarity between the three internal repeats in MCF proteins is much greater than that between the two three-spanner internal repeats in the MIP family proteins or the very distantly related six-spanner presumed internal repeats in the MFS proteins. This fact suggests that this triplication event occurred relatively recently in evolutionary time, possibly about 1.5 billion years ago, after the development of the eukaryotic cell. The fact that members of the MCF have been found only in eukaryotic organelles but are exclusively nucleus encoded further substantiates this possibility. Statistical analyses suggest either that duplication and divergence of the six-spanner-encoding primordial gene occurred shortly after the triplication event which gave rise to the primordial gene or that this triplication event occurred more than once from the same primordial two-spanner-encoding gene at about the same time as did divergence to give the various known subclusters (95). Limited sequence similarity between odd- and even-numbered spanners suggests that the two-spanner-encoding primordial gene may have arisen from a very small one-spanner-encoding gene by duplication, but the observed

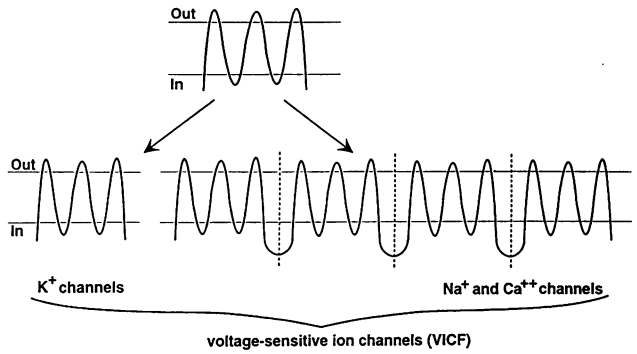

FIG. 8. Proposed pathway for the evolution of the K^+, Ca^{2+}, and Na^+ channels of the voltage-sensitive ion channel (VIC) family. The primordial gene is believed to have encoded a single six-spanner unit, analogous to that found in current-day K^+ channels, and it duplicated twice to give genes encoding four such units in the Ca^{2+} and Na^+ channels. On the basis of the degree of sequence similarity among the four repeat segments, these two duplication events may have occurred about 1 billion years ago in eukaryotic cells. In confirmation of this notion, some Ca^{2+} channels have only two repeat segments (95a). In the four repeat unit channels, the first duplication event is believed to have occurred substantially before the second. The degree of sequence similarity between repeat segments is in excess of that required to establish homology.

sequence similarity could have arisen by a convergent evolutionary process.

Voltage-sensitive ion channels. Nerve and muscle cells of higher eukaryotes possess homologous proteins that form either ligand-gated or voltage-sensitive ion channels (11, 75, 100). The latter group includes cation channels specific for Na^+, Ca^{2+}, or K^+ (12, 114, 117, 202, 217). Voltage-sensitive ion channels are believed to exist in plants, in lower single-celled eukaryotes such as yeasts and *Paramecium* species, and in eubacteria such as *E. coli* (185, 189, 212, 213). However, the genes encoding microbial channel proteins have not yet been sequenced, and their relationships to the well-defined ion channels of animal nerve and muscle cells are therefore not clear.

Proteins which form the voltage-sensitive channels that facilitate Na^+ or Ca^{2+} transport in nerves and muscle tissues consist of single polypeptide chains of about 2,000 amino acyl residues. Each one of most of these channel proteins contains four homologous internal repeats (I to IV) of about 300 residues, which are strikingly similar in sequence (percent sequence identity commonly of about 50%). Each of the four repeated domains contains six presumed α-helical transmembrane segments (Table 2). There is no evidence for repeat sequences within any of these internal repeats, suggesting that these proteins evolved independently of the MIP channel proteins and the mitochondrial carriers. Intragenic quadruplication events (actually two duplication events) (75) must have given rise to the large polypeptide chain of each such channel protein, which is about four times larger than the presumptive primordial protein (Fig. 8) (75, 95b).

The voltage-sensitive K^+ channels are sufficiently similar in sequence to the Na^+ and Ca^{2+} channels to establish homology, but they differ from the latter proteins in being homotetramers in which each subunit corresponds to one of the four internal repeats of the Na^+ or Ca^{2+} channels (Table 2) (75, 114, 117, 202). The degree of sequence identity between the K^+ channels and either the Na^+ or the Ca^{2+} channels is frequently only about 20%, showing that K^+ channels diverged from Na^+ and Ca^{2+} channels long before the two tandem duplication events that gave rise to the large polypeptide chains encompassing the four repeat units referred to above. Possibly the primordial Na^+ and Ca^{2+} channels were once homotetramers resembling current-day K^+ channels. It is reasonable to presume that monomeric proteins were the evolutionary precursors of these tetramers and that oligomerization arose to allow cooperativity or sensitivity to regulatory agents. Intragenic tandem quadruplication may subsequently have occurred to allow fine control over channel function. It has been suggested that Na^+ channels arose from primitive Ca^{2+} channels, thus explaining the striking sequence similarity of all sequenced Na^+ channels, their high degree of sequence similarity with Ca^{2+} channels, and the fact that a few sequenced Ca^{2+} channels apparently have only two rather than four repeat units (75, 95b).

RELATIVE DEGREES OF CONSERVATION IN DIFFERENT PORTIONS OF TRANSPORT PROTEINS

Several studies have attempted to evaluate the relative degrees of conservation found within different portions of the proteins which make up the various families of transport proteins described in the previous sections. Thus, when the two-domain periplasmic solute-binding receptors of the ABC transporters are analyzed, it is found that for all eight characterized families of these proteins, the N-terminal domains are more strongly conserved than the C-terminal domains (208). Similarly, the N-terminal repeat of the three repeat segments in the MCF is most strongly conserved (95), as is the N-terminal repeat of the two repeat segments in the MIP family (155), at least when dissimilar members of these families are examined. Interestingly, when closely related members (e.g., members of a single cluster of the MIP family or the MCF exhibiting the same or similar specificities) were examined, this generalization did not always hold.

The same has been noted when proteins in different clusters of the phylogenetic tree for the MFS were compared (109). The most striking regions of sequence similarity were always found in the N-terminal halves of these proteins, and well-conserved sequence motifs found in both the first and second halves of these proteins were more strongly conserved in the N-terminal domains (109). However, as for the MIP family and the MCF, within specific clusters of the phylogenetic trees of the MFS this preferential conservation of the N-terminal segments was not always observed. Thus, the C-terminal domains may preferentially serve cluster-specific functions whereas the N-terminal domains preferentially serve family-specific functions. Exactly what these functions are remains to be determined. It is possible that the N-terminal domains are more important for energy coupling, membrane insertion, and/or proper structural maintenance, whereas C-terminal domains are more important for specificity determination. Relevant to this point, the N-terminal part of the melibiose permease of *E. coli* has been implicated in Na^+ coupling to sugar transport (140, 255). The melibiose:Na^+ or melibiose:H^+ symporters of enteric bacteria are not demonstrably homologous to members of the MFS, but they are homologous to H^+:lactose symporters of gram-positive bacteria (136, 137), and they exhibit structural and mechanistic features as well as sequence motifs similar to those of MFS permeases (see references 17, 23, 156, and 160 for comparisons of the melibiose and lactose permeases of *E. coli*).

Examination of some nontransport protein families has also revealed greater conservation of N-terminal domains than of C-terminal domains. For example, proteins of the fimbrial chaperone family 2, whose constituent members are large

proteins of over 800 amino acyl residues, exhibit greatest conservation in their N-terminal domains (222). It may be that for these proteins as well as for the periplasmic receptors and transport proteins discussed above, the N-terminal regions of these families exhibit greatest sequence similarity because they share a requisite functional, structural, or biogenic feature that can be provided only by the N-terminal domain(s).

The general feature of N-terminal conservation noted above for several classes of transport, receptor and chaperone proteins is not always observed. For example, in the small RND family of presumed porters (Tables 1 and 2; see above), the C-terminal domains are more strongly conserved than are the N-terminal domains (180). In this case, the C-terminal, transmembrane, hydrophobic regions are most strongly conserved.

CONCLUSIONS

Transport proteins translocate their solutes either by channel-type mechanisms or by carrier-type mechanisms, and a close relationship between these two kinetically distinguishable types of transport mechanisms has been postulated (126, 168, 169). A preformed channel may be a structural requirement for carrier-mediated transport. Introduction of specific residues in the channel, allowing for stereospecific solute recognition as well as conformational vacillation between two alternative states, with the solute-binding site alternating between orientations facing inward versus outward, may be all that is required to convert a channel into a carrier (126).

The unit hydrophobic domain of the constituent members of many families of transport proteins apparently exhibits a characteristic pattern of six tightly clustered transmembrane helical segments with both the amino and carboxy termini localized to the cytoplasmic surface of the membrane (Table 2). This distinctive structural feature, as well as recognition of the fact that this pattern evolved independently many times during evolutionary history to give transport proteins of the channel or carrier type, leads to the possibility that such a structural motif is particularly well suited for the formation of transmembrane solute channels. Three-dimensional analysis is required to reveal the precise reasons for the suitability of six spanner units for the construction of transport proteins.

Although several distinct (and currently unrelated) ancient families of transporters, such as the MFS, the APC family, the SSF, the ABC family, and most PTS permeases, may eventually prove to be related by common descent despite their sequence divergence, it is clear that others are not. Thus, voltage-sensitive ion channels of nerve and muscle cells, the MCF, the RND family, the MIP family, and the MFS clearly arose independently of each other, at different times in evolutionary history, in different cell types, using different routes. It is therefore suggested that a three-dimensional transmembrane structure consisting of six parallel α-helices is particularly well suited to transmembrane channel formation. Such constraints could account both for the retention of this structural motif during evolutionary divergence of homologous transport proteins and for apparent structural convergence of evolutionarily unrelated transport proteins to give distinct families of proteins possessing similar topologies.

Functional and structural as well as evolutionary aspects of transport proteins have been revealed by computer analyses that would have been all but impossible without them. First, it is clear that substrate specificity within any one family (or superfamily) of porters correlates to a remarkable degree with sequence conservation. This generalization appears valid for functionally diverse proteins with a range of specificities from simple ions, sugars, amino acids, and drugs to oligosaccharides and peptides and on up to complex carbohydrates, proteins, and lipophilic substances. The nature of the protein (be it a cytoplasmic, integral membrane, or extracellular constituent of the cell) seems to be immaterial to the conclusion that the degree of sequence relatedness generally correlates with the degree of functional similarity, particularly regarding solute-binding specificity.

Second, functionally and structurally important regions in proteins can be identified on the basis of sequence comparisons for members of homologous families, provided that these functional and structural features are common to the different members of the family. These regions can be identified on the basis of sequence conservation. The nature of the residues conserved within a specific region of a family of proteins may provide hints about whether a region of conservation plays a structural or a functional role, and specific, currently recognized motifs, if present, can provide clear indications of function. In most cases, however, site-specific mutagenic and biochemical analyses are required to establish the specific function involved.

Third, topological and secondary-structure predictions can be made with a remarkable degree of reliability solely on the basis of computer-aided methods. Thus, α-structure, β-structure, β-turns, α-helical transmembrane spanners, cytoplasmic sidedness, etc., can be predicted with ever-increasing reliability, particularly when a set of homologous proteins is analyzed. The utility of these methods is of particular importance since the three-dimensional structures of very few transport proteins (and none of the ones which form the focus of our discussion here) have been elucidated.

Fourth, identification of AP-rich, Q-rich, and highly charged linkers, as well as other regions of polypeptide flexibility, serves to delimit independently folding polypeptide domains in multidomain proteins. When several homologous proteins are available, the occurrence of gaps in the multiple alignment of these proteins further substantiates predictions regarding the presence of flexible linkers. Use of homology with other (possibly shuffled) multidomain proteins provides a confirming but independent computer-aided method to define structural domains.

Finally, primary sequence determines subcellular localization in both prokaryotes and eukaryotes. N-terminal leader sequences of hydrophobic and amphipathic character have been discussed in this review, but C-terminal and internal targeting motifs are also important, particularly in eukaryotes. Thus, the identification of specific types of targeting sequences and their cleavage sites (or the absence of these sequences) provides useful information which indirectly relates both to structure and function.

We have seen that computer applications have greatly enhanced our understanding of the structural, functional, biogenic, and evolutionary aspects of transport protein families. Elucidation of the detailed three-dimensional structures of transmembrane transport proteins and thus of their common and divergent structural features will ultimately be required to fully define these relationships. In the meantime, further application of currently available computer programs, refinement of these programs, and development of entirely new programs will undoubtedly allow us to draw an increasingly detailed picture of these remarkably varied proteins.

ACKNOWLEDGMENTS

I acknowledge the invaluable research efforts of my coworkers Michael P. Eddy, Matthew J. Fagan, Jeffrey Kuan, Michael D. Marger, Gerald M. Pao, Aiala Reizer, Jonathan Reizer, Friederich Titgemeyer, Christine Trandinh, and Mieke Van Rosmalen. Together they have

made a unique contribution to our understanding of the evolutionary origins of complex transport systems. Mary Beth Hiller provided expert assistance in the preparation of the manuscript.

The work was supported by USPHS grants 5RO1AI21702 and 2 RO1AI14176 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

1. **Adams, M. D., M. Dubnick, A. R. Kerlavage, R. Moreno, J. M. Kelley, T. R. Utterback, J. W. Nagle, C. Fields, and J. C. Venter.** 1992. Sequence identification of 2,375 human brain genes. Nature (London) **355**:632–634.
2. **Aerts, T., J.-Z. Xia, H. Slegers, J. de Block, and J. Clauwaert.** 1990. Hydrodynamic characterization of the major intrinsic protein from the bovine lens fiber membranes. J. Biol. Chem. **265**:8675–8680.
3. **Allard, J. D., and K. P. Bertrand.** 1992. Membrane topology of the pBR322 tetracycline resistance protein. TetA-PhoA gene fusions and implications for the mechanism of TetA membrane insertion. J. Biol. Chem. **267**:17809–17819.
4. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
5. **Ames, G. F.-L., C. S. Mimura, and V. Shyamala.** 1990. Bacterial periplasmic permeases belong to a family of transport proteins operating from *E. coli* to human: traffic ATPases. FEMS Microbiol. Rev. **75**:429–446.
6. **Anraku, Y.** 1988. Bacterial electron transport chains. Annu. Rev. Biochem. **57**:101–132.
7. **Auger, I. E.** 1993. Computational techniques to predict amphipathic helical segments, p. 7–19. *In* R. M. Epand (ed.), The amphipathic helix. CRC Press, Inc., Boca Raton, Fla.
8. **Baev, N., G. Endre, G. Petrovics, Z. Banfalvi, and A. Kondorosi.** 1991. Six nodulation genes of *nod* box locus 4 in *Rhizobium meliloti* are involved in nodulation signal production: *nodM* codes for D-glucosamine synthetase. Mol. Gen. Genet. **228**:113–124.
9. **Bairoch, A.** 1992. PROSITE—a dictionary of sites and patterns in proteins. Nucleic Acids Res. **20**(Suppl.):S2013–S2018.
10. **Baker, M. E., and M. H. Saier, Jr.** 1990. A common ancestor for bovine lens fiber major intrinsic protein, soybean nodulin-26 protein, and *Escherichia coli* glycerol facilitator. Cell **60**:185–186.
11. **Barnard, E. A.** 1992. Receptor classes and the transmitter-gated ion channels. Trends Biochem. Sci. **17**:368–374.
12. **Begenisich, T.** 1987. Molecular properties of ion permeation through sodium channels. Annu. Rev. Biophys. Biophys. Chem. **16**:247–263.
13. **Benz, R.** 1985. Porin from bacterial and mitochondrial outer membranes. Crit. Rev. Biochem. **19**:145–190.
14. **Blachly-Dyson, E., S. Peng, M. Colombini, and M. Forte.** 1990. Selectivity changes in site-directed mutants of the VDAC ion channel: structural implications. Science **247**:1233–1236.

14a. **Blair, A., L. K. Ngo, J. Park, and M. H. Saier, Jr.** Unpublished data.

14b. **Blair, A., L. Tran, J. Park, and M. H. Saier, Jr.** Unpublished results.

15. **Bork, P., and R. F. Doolittle.** 1992. Proposed acquisition of an animal protein domain by bacteria. Proc. Natl. Acad. Sci. USA **89**:8990–8994.
16. **Bork, P., C. Sander, and A. Valencia.** 1993. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. Protein Sci. **2**:31–40.
17. **Botfield, M. C., D. M. Wilson, and T. H. Wilson.** 1990. The melibiose carrier of *Escherichia coli*. Res. Microbiol. **141**:328–331.
18. **Bowie, J. U., J. F. Reidhaar-Olson, W. A. Lim, and R. T. Sauer.** 1990. Deciphering the message in protein sequences: tolerance to amino acid substitutions. Science **247**:1306–1310.
19. **Brandl, C. J., and C. M. Deber.** 1986. Hypothesis about the function of membrane-buried proline residues in transport proteins. Proc. Natl. Acad. Sci. USA **83**:917–921.
20. **Brandl, C. J., R. B. Deber, L. C. Hsu, G. A. Woolley, X. K. Young, and C. M. Deber.** 1988. Evidence for similar function of transmembrane segments in receptor and membrane-anchored proteins. Biopolymers **27**:1171–1182.
21. **Brennan, R. G.** 1991. Interactions of the helix-turn-helix binding domain. Curr. Opin. Struct. Biol. **1**:80–88.
22. **Bröer, S., G. Ji, A. Bröer, and S. Silver.** 1993. Arsenic efflux governed by the arsenic resistance determinant of *Staphylococcus aureus* plasmid pI258. J. Bacteriol. **175**:3480–3485.
23. **Brooker, R. J.** 1990. The lactose permease of *Escherichia coli*. Res. Microbiol. **141**:309–315.
24. **Buhr, A., and B. Erni.** 1993. Membrane topology of the glucose transporter of *Escherichia coli*. J. Biol. Chem. **268**:11599–11603.
25. **Calamia, J., and C. Manoil.** 1990. lac permease of *Escherichia coli* topology and sequence elements promoting membrane insertion. Proc. Natl. Acad. Sci. USA **87**:4937–4941.
26. **Carne, A., and J. E. Walker.** 1983. Amino acid sequence of ovine 6-phosphogluconate dehydrogenase. J. Biol. Chem. **258**:12895–12906.
27. **Chelvanayagam, G., Z. Reich, R. Bringas, and P. Argos.** 1992. Prediction of protein folding pathways. J. Mol. Biol. **227**:901–916.
28. **Chen, Y., W. J. Fairbrother, and P. E. Wright.** 1993. Three-dimensional structures of the central regulatory proteins of the bacterial phosphotransferase system, HPr and Enzyme IIA^{glc}. J. Cell. Biochem. **51**:75–82.
29. **Chen, Y., J. Reizer, M. H. Saier, Jr., W. J. Fairbrother, and P. E. Wright.** 1993. Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, HPr and IIA^{glc}. Biochemistry **32**:32–37.
30. **Clark, J. A., and S. G. Amara.** 1993. Amino acid neurotransmitter transporters: structure, function, and molecular diversity. BioEssays **15**:323–332.
31. **Clewell, D. B.** 1993. Bacterial sex pheromone-induced plasmid transfer. Cell **73**:9–12.
32. **Cowan, S. W., T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Pauptit, J. N. Jansonius, and J. P. Rosenbusch.** 1992. Crystal structures explain functional properties of two *E. coli* porins. Nature (London) **358**:727–733.
33. **Davis, T., M. Yamada, M. Elgort, and M. H. Saier, Jr.** 1988. Nucleotide sequence of the mannitol (*mtl*) operon in *Escherichia coli*. Mol. Microbiol. **2**:405–412.
34. **Dayhoff, M. O., W. C. Barker, and L. T. Hunt.** 1983. Establishing homologies in protein sequences. Methods Enzymol. **91**:524–545.
35. **Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt.** 1978. A model of evolutionary change in proteins, p. 345–352. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, Suppl. 3. National Biomedical Research Foundation, Silver Spring, Md.
36. **Deber, C. M., C. J. Brandl, R. B. Deber, L. C. Hsu, and X. K. Young.** 1986. Amino acid composition of the membrane and aqueous domains of integral membrane proteins. Arch. Biochem. Biophys. **251**:68–76.
37. **Deber, C. M., M. Glibowicka, and G. A. Woolley.** 1990. Conformations of proline residues in membrane environments. Biopolymers **29**:149–157.
38. **Deisenhofer, J., and H. Michel.** 1991. Structures of bacterial photosynthetic reaction centers. Annu. Rev. Cell Biol. **7**:1–23.
39. **Devereux, J., P. Haeberli, and N. O. Smithies.** 1984. A comprehensive set of sequence analyses for the VAX. Nucleic Acids Res. **12**:387–395.
40. **Doolittle, R. F.** 1986. Of urfs and orfs: a primer on how to analyze derived amino acid sequences. University Science Books, Mill Valley, Calif.
41. **Doolittle, R. F.** 1989. Redundancies in protein sequences, p. 599–623. *In* G. D. Fasman (ed.), Prediction of protein structure and the principles of protein conformation. Plenum Publishing Corp., New York.
42. **Doolittle, R. F. (ed.).** 1990. Molecular evolution: computer analysis of protein and nucleic acid sequences. Methods Enzymol., vol. 183.
43. **Doolittle, R. F.** 1992. Reconstructing history with amino acid sequences. Protein Sci. **1**:191–200.
44. **Doolittle, R. F., and D.-F. Feng.** 1990. Nearest neighbor procedure for relating progressively aligned amino acid sequences. Methods Enzymol. **183**:659–669.
45. **Eckert, B., and C. F. Beck.** 1989. Topology of the transmembrane

Tn10-encoded tetracycline resistance protein within the inner-membrane of *Escherichia coli*. J. Biol. Chem. **64**:11663–11670.
46. **Ehring, G. R., G. A. Zampighi, and J. E. Hall.** 1993. Does MIP play a role in cell-cell communication? p. 143–152. *In* J. E. Hall, G. A. Zampighi, and R. M. Davis (ed.), Gap junctions, progress in cell research III. Elsevier Publishers, Amsterdam.
47. **Eisenberg, D.** 1984. Three-dimensional structure of membrane and surface proteins. Annu. Rev. Biochem. **53**:595–623.
48. **Erni, B., B. Zanolari, P. Graff, and H. P. Kocher.** 1989. Mannose permease of *Escherichia coli*: domain structure and function of the phosphorylating subunit. J. Biol. Chem. **264**:18733–18741.
49. **Erni, B., B. Zanolari, and H. P. Kocher.** 1987. The mannose permease of *Escherichia coli* consists of three different proteins. Amino acid sequence and function in sugar transport, sugar phosphorylation, and penetration of phage λ DNA. J. Biol. Chem. **262**:5238–5247.
50. **Fagan, M. J., and M. H. Saier, Jr.** 1994. P-type ATPases of eukaryotes and bacteria: sequence comparisons and construction of phylogenetic trees. J. Mol. Evol. **38**:57–99.
51. **Fasman, G. D., and W. A. Gilbert.** 1990. The prediction of transmembrane protein sequences and their conformation: an evaluation. Trends Biochem. Sci. **15**:89–92.
52. **Fath, M. J., and R. Kolter.** 1993. ABC-transporters: the bacterial exporters. Microbiol. Rev. **57**:995–1017.
53. **Fath, M. J., R. Skvirsky, L. Gilson, H. K. Mahanty, and R. Kolter.** 1992. The secretion of colicin V, p. 331–348. *In* R. James, F. Lazdunski and F. Pattus (ed.), Bacteriocins, microcins, and antibiotics. Springer-Verlag KG, Heidelberg, Germany.
54. **Fath, M. J., R. C. Skvirsky, and R. Kolter.** 1991. Functional complementation between bacterial MDR-like export systems: colicin V, alpha-hemolysin, and *Erwinia* protease. J. Bacteriol. **173**:7549–7556.
55. **Fearnley, I. M., and J. E. Walker.** 1992. Conservation of sequences of subunits of mitochondrial complex I and their relationships with other proteins. Biochim. Biophys. Acta **1140**:105–134.
56. **Felsenstein, J.** 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22**:521–565.
57. **Feng, D.-F., and R. F. Doolittle.** 1990. Progressive alignment and phylogenetic tree construction of protein sequences. Methods Enzymol. **183**:375–387.
58. **Flügge, U. I., K. Fischer, A. Gross, W. Sebald, F. Lottspeich, and C. Eckerskorn.** 1989. The triose phosphate-3-phosphoglycerate-phosphate translocator from spinach chloroplasts: nucleotide sequence of a full-length cDNA clone and import of the *in vitro* synthesized precursor protein into chloroplasts. EMBO J. **8**:39–46.
59. **Geller, B., H.-Y. Zhu, S. Cheng, A. Kuhn, and R. E. Dalbey.** 1993. Charged residues render Pro-OmpA potential dependent for initiation of membrane translocation. J. Biol. Chem. **268**:9442–9447.
60. **Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. **55**:303–315.
61. **Glover, L. A., and J. G. Lindsay.** 1992. Targeting proteins to mitochondria: a current overview. Biochem. J. **284**:609–620.
62. **Gogarten, J. P., T. Starke, H. Kibak, J. Fishmann, and L. Taiz.** 1992. Evolution and isoforms of V-ATPase subunits. J. Exp. Biol. **172**:137–147.
63. **Goldrick, D., G.-Q. Yu, S.-Q. Jiang, and J.-S. Hong.** 1988. Nucleotide sequence and transcription start point of the phosphoglycerate transporter gene of *Salmonella typhimurium*. J. Bacteriol. **170**:3421–3426.
64. **Griffith, J. K., M. E. Baker, D. A. Rouch, M. G. P. Page, R. A. Skurray, I. T. Paulsen, K. F. Chater, S. A. Baldwin, and P. J. F. Henderson.** 1992. Membrane transport proteins: implications of sequence comparisons. Curr. Opin. Cell Biol. **4**:684–695.
65. **Groisman, E. A., M. H. Saier, Jr., and H. Ochman.** 1992. Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome. EMBO J. **11**:1309–1316.
66. **Hancock, R. E. W., R. Siehnel, and N. Martin.** 1990. Outer membrane proteins of *Pseudomonas*. Mol. Microbiol. **4**:1069–1075.
67. **Harrison, S. C., and A. K. Aggarwal.** 1990. DNA recognition by proteins with the helix-turn-helix motif. Annu. Rev. Biochem. **59**:933–969.
68. **Hatefi, Y.** 1985. The mitochondrial electron transport and oxidative phosphorylation system. Annu. Rev. Biochem. **54**:1015–1069.
69. **Heefner, D. L., and F. M. Harold.** 1980. ATP-linked sodium transport in *Streptococcus faecalis*. J. Biol. Chem. **255**:11396–11402.
70. **Heefner, D. L., and F. M. Harold.** 1982. ATP-driven sodium pump in *Streptococcus faecalis*. Proc. Natl. Acad. Sci. USA **79**:2798–2802.
71. **Heinemann, J. A., and G. F. Sprague, Jr.** 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature (London) **340**:205–209.
72. **Heller, K. B., E. C. C. Lin, and T. H. Wilson.** 1980. Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. J. Bacteriol. **144**:274–278.
73. **Henderson, P. J. F., and M. C. J. Maiden.** 1990. Homologous sugar transport proteins in *Escherichia coli* and their relatives in both prokaryotes and eukaryotes. Philos. Trans. R. Soc. London Ser. B **326**:391–410.
74. **Higgins, C. F.** 1992. ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. **8**:67–113.
75. **Hille, B.** 1992. Ionic channels of excitable membranes, 2nd ed. Sinauer Associates, Sunderland, Mass.
76. **Isaki, L., R. Beers, and H. C. Wu.** 1990. Nucleotide sequence of the *Pseudomonas fluorescens* signal peptidase II gene (*lsp*) and flanking genes. J. Bacteriol. **172**:6512–6517.
77. **Jackowski, S., and J.-H. Alix.** 1990. Cloning, sequence, and expression of the pantothenate permease (*panF*) gene of *Escherichia coli*. J. Bacteriol. **172**:3842–3848.
78. **Jähnig, F.** 1990. Structure predictions of membrane proteins are not that bad. Trends Biochem. Sci. **15**:93–95.
79. **Jeanteur, D., J. H. Lakey, and F. Pattus.** 1991. The bacterial porin superfamily: sequence alignment and structure prediction. Mol. Microbiol. **5**:2153–2164.
80. **Jiang, W., L.-F. Wu, J. M. Tomich, M. H. Saier, Jr., and W. G. Niehaus.** 1990. Corrected sequence of the mannitol (*mtl*) operon in *Escherichia coli*. Mol. Microbiol. **4**:2003–2006.
81. **Kanehisa, M.** 1982. Los Alamos sequence analysis package for nucleic acids and proteins. Nucleic Acids Res. **10**:183–196.
82. **Kang, C.-H., W.-C. Shin, Y. Yamagata, S. Gokcen, G. F.-L. Ames, and S.-H. Kim.** 1991. Crystal structure of the lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium* at 2.7-Å resolution. J. Biol. Chem. **266**:23893–23899.
83. **Kaplan, R. S., J. A. Mayor, and D. O. Wood.** 1993. The mitochondrial tricarboxylate transport protein. cDNA cloning, primary structure, and comparison with other mitochondrial transport proteins. J. Biol. Chem. **268**:13682–13690.
84. **Karlin, S., and S. F. Altschul.** 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA **87**:2264–2268.
85. **Kaur, P., and B. P. Rosen.** 1992. Mutagenesis of the C-terminal nucleotide-binding site of an anion-translocating ATPase. J. Biol. Chem. **267**:19272–19277.
86. **Kaur, P., and B. P. Rosen.** 1992. Plasmid-encoded resistance to arsenic and antimony. Plasmid **27**:29–40.
87. **Kaur, P., and B. P. Rosen.** 1993. Complementation between nucleotide-binding domains in an anion-translocating ATPase. J. Bacteriol. **175**:351–357.
88. **Kerppola, R. E., and G. F.-L. Ames.** 1992. Topology of the hydrophobic membrane-bound components of the histidine periplasmic permease. Comparison with other members of the family. J. Biol. Chem. **267**:2329–2336.
89. **Kibak, H., L. Taiz, T. Starke, P. Bernasconi, and J. P. Gogarten.** 1992. Evolution of structure and function of V-ATPases. J. Bioenerg. Biomembr. **24**:415–424.
90. **Kinoshita, N., T. Unemoto, and H. Kobayashi.** 1984. Sodium-stimulated ATPase in *Streptococcus faecalis*. J. Bacteriol. **158**:844–848.
91. **Klein, J. R., B. Henrich, and R. Plapp.** 1991. Molecular analysis and nucleotide sequence of the *envCD* operon of *Escherichia coli*.

Mol. Gen. Genet. **230:**230–240.
91a. **Klein, W., and W. Boos.** Personal communication.
92. **Klingenberg, M.** 1990. Mechanism and evolution of the uncoupling protein of brown adipose tissue. Trends Biochem. Sci. **15:**108–112.
93. **Kobayashi, H., J. V. Brunt, and F. M. Harold.** 1978. ATP-linked calcium transport in vesicles of *Streptococcus faecalis*. J. Biol. Chem. **253:**2085–2092.
94. **Konings, W. N., B. Poolman, and A. J. M. Driessen.** 1989. Bioenergetics and solute transport in lactococci. Crit. Rev. Microbiol. **16:**419–476.
95. **Kuan, J., and M. H. Saier, Jr.** 1993. The mitochondrial carrier family of transport proteins: structural, functional and evolutionary relationships. Crit. Rev. Biochem. Mol. Biol. **28:**209–233.
95a. **Kuan, J., and M. H. Saier, Jr.** 1993. Expansion of the mitochondrial carrier family. Res. Microbiol. **144:**671–672.
95b. **Kuan, J., and M. H. Saier, Jr.** Unpublished results.
96. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
97. **Landry, D., S. Sullivan, M. Nicolaides, C. Redhead, A. Edelman, M. Field, Q. Al-Awqati, and J. Edwards.** 1993. Molecular cloning and characterization of p64, a chloride channel protein from kidney microsomes. J. Biol. Chem. **268:**14948–14955.
97a. **Lengeler, J.** Personal communication.
98. **Lengeler, J. W.** 1990. Molecular analysis of the Enzyme II-complexes of the bacterial phosphotransferase system (PTS) as carbohydrate transport systems. Biochim. Biophys. Acta **1018:** 155–159.
99. **Lengeler, J. W., F. Titgemeyer, A. P. Vogler, and B. M. Wöhrl.** 1990. Structures and homologies of carbohydrate: phosphotransferase system (PTS) proteins. Philos. Trans. R. Soc. London B **326:**489–504.
100. **Lester, H. A.** 1992. The permeation pathway of neurotransmitter-gated ion channels. Annu. Rev. Biophys. Biomol. Struct. **21:**267–292.
101. **Li, W.-H., and D. Graur.** 1991. Fundamentals of molecular evolution. Sinauer Associates, Sunderland, Mass.
102. **Liesegang, H., K. Lemke, R. A. Siddiqui, and H.-G. Schlegel.** 1993. Characterization of the inducible nickel and cobalt resistance determinant cnr from pMOL28 of *Alcaligenes eutrophus* CH34. J. Bacteriol. **175:**767–778.
103. **Lloyd, A. D., and R. J. Kadner.** 1990. Topology of the *Escherichia coli uhpT* sugar-phosphate transporter analyzed by using Tn*phoA* fusions. J. Bacteriol. **172:**1688–1693.
104. **Lomovskaya, O., and K. Lewis.** 1992. Emr, an *Escherichia coli* locus for multidrug resistance. Proc. Natl. Acad. Sci. USA **89:**8938–8942.
105. **Luck, L. A., and J. J. Falke.** 1991. ^{19}F NMR studies of the D-galactose chemosensory receptor. 1. Sugar binding yields a global structural change. Biochemistry **30:**4248–4256.
106. **Luck, L. A., and J. J. Falke.** 1991. ^{19}F NMR studies of the D-galactose chemosensory receptor. 2. Ca(II) binding yields a local structural change. Biochemistry **30:**4257–4261.
107. **Luque, F., L. A. Mitchenall, M. Chapman, R. Christine, and R. N. Pau.** 1993. Characterization of genes involved in molybdenum transport in *Azotobacter vinelandii*. Mol. Microbiol. **7:**447–459.
108. **Maiden, M. C. J., E. O. Davis, S. A. Baldwin, D. C. M. Moore, and P. J. F. Henderson.** 1987. Mammalian and bacterial sugar transport proteins are homologous. Nature (London) **325:**641–643.
109. **Marger, M. D., and M. H. Saier, Jr.** 1993. A major superfamily of transmembrane facilitators catalyzing uniport, symport and antiport. Trends Biochem. Sci. **18:**13–20.
110. **Martin-Verstraete, I., M. Debarbouille, A. Klier, and G. Rapoport.** 1990. Levanase operon of *Bacillus subtilis* includes a fructose-specific phosphotransferase system regulating the expression of the operon. J. Mol. Biol. **214:**657–671.
111. **Maurel, C., J. Reizer, J. I. Schroeder, and M. J. Chrispeels.** 1993. The vacuolar membrane protein γ-TIP creates water specific channels in *Xenopus oocytes*. EMBO J. **12:**2241–2247.
112. **Meins, M., P. Jenö, D. Müller, W. J. Richter, J. P. Rosenbusch, and B. Erni.** 1993. Cysteine phosphorylation of the glucose transporter of *Escherichia coli*. J. Biol. Chem. **268:**11604–11609.
113. **Miao, G.-H., Z. Hong, and D. P. S. Verma.** 1992. Topology and phosphorylation of soybean nodulin-26, an intrinsic protein of the peribacteroid membrane. J. Cell Biol. **118:**481–490.
114. **Miller, R. J.** 1992. Voltage-sensitive Ca^{2+} channels. J. Biol. Chem. **267:**1403–1406.
115. **Mitchell, P.** 1979. Compartmentation and communication in living systems. Ligand conduction: a general catalytic principle in chemical, osmotic and chemiosmotic reaction systems. Eur. J. Biochem. **95:**1–20.
116. **Mogi, T., L. J. Stern, B. H. Chao, and H. G. Khorana.** 1989. Structure-function studies on bacteriorhodopsin. VIII. Substitutions of the membrane-embedded prolines 50, 91, and 186: the effects are determined by the substituting amino acids. J. Biol. Chem. **264:**14192–14196.
117. **Montal, M.** 1990. Molecular anatomy and molecular design of channel proteins. FASEB J. **4:**2623–2635.
117a. **Montal, M.** Personal communication.
118. **Muramatsu, S., and T. Mizuno.** 1989. Nucleotide sequence of the region encompassing the *glpKF* operon and its upstream region containing a bent DNA sequence of *Escherichia coli*. Nucleic Acids Res. **17:**4378.
119. **Nakai, K., and M. Kanehisa.** 1991. Expert system for predicting protein localization sites in Gram-negative bacteria. Proteins Struct. Funct. Genet. **11:**95–110.
120. **Nakanishi, N., N. A. Shneider, and R. Axel.** 1990. A family of glutamate receptor genes: evidence for the formation of heteromultimeric receptors with distinct channel properties. Neuron **5:**569–581.
121. **Navarro, C., L.-F. Wu, and M.-A. Mandrand-Berthelot.** 1993. The *nik* operon of *Escherichia coli* encodes a periplasmic binding-protein-dependent transport system for nickel. Mol. Microbiol. **9:**1181–1191.
122. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York.
123. **Nichols, J. C., N. K. Vyas, F. A. Quiocho, and K. S. Matthews.** 1993. Model of lactose repressor core based on alignment with sugar-binding proteins is concordant with genetic and chemical data. J. Biol. Chem. **268:**17602–17612.
124. **Nies, D. H., A. Nies, L. Chu, and S. Silver.** 1989. Expression and nucleotide sequence of a plasmid-determined divalent cation efflux system from *Alcaligenes eutrophus*. Proc. Natl. Acad. Sci. USA **86:**7351–7355.
125. **Nikaido, H.** 1992. Porins and specific channels of bacterial outer membranes. Mol. Microbiol. **6:**435–442.
126. **Nikaido, H., and M. H. Saier, Jr.** 1992. Transport proteins in bacteria: common themes in their design. Science **258:**936–942.
127. **Odermatt, A., H. Suter, R. Krapf, and M. Solioz.** 1993. Primary structure of two P-type ATPases involved in copper homeostasis in *Enterococcus hirae*. J. Biol. Chem. **268:**12775–12779.
128. **Oesterhelt, D., C. Bräuchle, and N. Hampp.** 1991. Bacteriorhodopsin: a biological material for information processing. Q. Rev. Biophys. **24:**425–478.
129. **Oh, B.-H., J. Pandit, C.-H. Kang, K. Nikaido, S. Gokcen, G. F.-L. Ames, and S.-H. Kim.** 1993. Three-dimensional structures of the periplasmic lysine-, arginine-, ornithine-binding protein with and without a ligand. J. Biol. Chem. **268:**11348–11355.
130. **O'Hara, P. J., P. O. Sheppard, H. Thøgersen, D. Venezie, B. A. Haldeman, V. McGrane, K. M. Houamed, C. Thomsen, T. L. Gilbert, and E. R. Mulvihill.** 1993. The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. Neuron **11:**41–52.
131. **Okamura, M. Y., and G. Feher.** 1992. Proton transfer in reaction centers from photosynthetic bacteria. Annu. Rev. Biochem. **61:**861–896.
132. **Pao, G. M., L.-F. Wu, K. D. Johnson, H. Höfte, M. J. Chrispeels, G. Sweet, N. N. Sandal, and M. H. Saier, Jr.** 1991. Evolution of the MIP family of integral membrane transport proteins. Mol. Microbiol. **5:**33–37.
133. **Parker, L. L., and B. G. Hall.** 1990. Characterization and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. Genetics **124:**455–471.
134. **Parkinson, J. S., and E. C. Kofoid.** 1992. Communication mod-

ules in bacterial signaling proteins. Annu. Rev. Genet. **26:**71–112.
134a. **Parralopez, C., M. T. Baer, and E. Groisman.** 1993. Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in *Salmonella typhimurium*. EMBO J. **12:**4053–4062.
135. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
136. **Poolman, B.** 1990. Precursor/product antiport in bacteria. Mol. Microbiol. **4:**1629–1636.
137. **Poolman, B., R. Modderman, and J. Reizer.** 1992. Lactose transport system of *Streptococcus thermophilus*. The role of histidine residues. J. Biol. Chem. **267:**9150–9157.
138. **Postma, P., J. Lengeler, and G. R. Jacobson.** 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. **57:**543–594.
139. **Pouny, Y., and Y. Shai.** 1992. Interaction of D-amino acid incorporated analogues of pardaxin with membranes. Biochemistry **31:**9482–9490.
140. **Pourcher, T., M.-L. Zani, and G. Leblanc.** 1993. Mutagenesis of acidic residues in putative membrane-spanning segments of the melibiose permease of *Escherichia coli*. I. Effect on Na^+-dependent transport and binding properties. J. Biol. Chem. **268:**3209–3215.
141. **Quiocho, F. A.** 1986. Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. Annu. Rev. Biochem. **55:**287–315.
142. **Radford, S. E., E. D. Laue, R. N. Perham, S. R. Martin, and E. Appella.** 1989. Conformational flexibility and folding of synthetic peptides representing an interdomain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. J. Biol. Chem. **264:**767–775.
143. **Rees, D. C., L. DeAntonio, and D. Eisenberg.** 1989. Hydrophobic organization of membrane proteins. Science **245:**510–513.
144. **Reeves, P.** 1993. Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. Trends Genet. **9:**17–22.
145. **Reizer, A., J. Deutscher, M. H. Saier, Jr., and J. Reizer.** 1991. Analysis of the gluconate (*gnt*) operon of *Bacillus subtilis*. Mol. Microbiol. **5:**1081–1089.
146. **Reizer, A., G. M. Pao, and M. H. Saier, Jr.** 1991. Evolutionary relationships among the permease proteins of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. Construction of phylogenetic trees and possible relatedness to proteins of eukaryotic mitochondria. J. Mol. Evol. **33:**179–193.
147. **Reizer, J., K. Finley, D. Kakuda, C. L. MacLeod, A. Reizer, and M. H. Saier, Jr.** 1993. Mammalian integral membrane receptors are homologous to facilitators and antiporters of yeast, fungi and eubacteria. Protein Sci. **2:**20–30.
148. **Reizer, J., C. Hoischen, A. Reizer, T. N. Phan, and M. H. Saier, Jr.** 1993. Sequence analyses and evolutionary relationships among the energy-coupling proteins Enzyme I and HPr of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. Protein Sci. **2:**506–521.
149. **Reizer, J., A. Reizer, A. Bairoch, and M. H. Saier, Jr.** 1993. A diverse transketolase family that includes the RecP protein of *Streptococcus pneumoniae*, a protein implicated in genetic recombination. Res. Microbiol. **144:**341–347.
150. **Reizer, J., A. Reizer, P. Bork, C. Sander, and M. H. Saier, Jr.** 1993. Exopolyphosphate phosphatase and guanosine pentaphosphate phosphatase belong to the sugar kinase/actin/hsp70 superfamily. Trends Biochem. Sci. **18:**247–248.
151. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1990. The cellobiose permease of *Escherichia coli* consists of three proteins and is homologous to the lactose permease of *Staphylococcus aureus*. Res. Microbiol. **141:**1061–1067.
152. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1990. The Na^+:pantothenate symporter (*PanF*) of *E. coli* is homologous to the Na^+:proline symporter (*PutP*) of *E. coli* and the Na^+:glucose symporters of mammals. Res. Microbiol. **141:**1069–1072.
153. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1992. A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. Protein Sci. **1:**1326–1332.
154. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1992. The putative Na^+/H^+ antiporter (NapA) of *Enterococcus hirae* is homologous to the putative K^+/H^+ antiporter (KefC) of *Escherichia coli*. Microbiol. Lett. **94:**161–164.
155. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1993. The MIP family of integral membrane channel proteins: sequence comparisons, evolutionary relationships, reconstructed pathway of evolution, and proposed functional differentiation of the two repeated halves of the proteins. Crit. Rev. Biochem. Mol. Biol. **28:**235–257.
156. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** A functional superfamily of sodium:solute symporters. Biochim. Biophys. Acta, in press.
156a. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** Unpublished results.
157. **Reizer, J., A. Reizer, M. H. Saier, Jr., and G. R. Jacobson.** 1992. A proposed link between nitrogen and carbon metabolism involving protein phosphorylation in bacteria. Protein Sci. **1:**722–726.
158. **Reizer, J., M. H. Saier, Jr., and A. Reizer.** 1992. Possible problems with the protein sequence comparison program FASTA. Trends Biochem. Sci. **17:**60.
159. **Rephaeli, A. W., and M. H. Saier, Jr.** 1980. Substrate specificity and kinetic characterization of sugar uptake and phosphorylation, catalyzed by the mannose Enzyme II of the phosphotransferase system in *Salmonella typhimurium*. J. Biol. Chem. **255:**8585–8591.
160. **Roepe, P. D., T. G. Consler, M. E. Menezes, and H. R. Kaback.** 1990. The lac permease of *Escherichia coli*: site-directed mutagenesis studies on the mechanism of β-galactoside/H^+ symport. Res. Microbiol. **141:**290–308.
161. **Roise, D.** 1993. The amphipathic helix in mitochondrial targeting sequences, p. 257–283. *In* R. M. Epand (ed.), The amphipathic helix. CRC Press, Inc., Boca Raton, Fla.
162. **Romano, A. H., and M. H. Saier, Jr.** 1992. Evolution of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. I. Physiologic and organismic considerations, p. 143–170. *In* R. P. Mortlock (ed.), The evolution of metabolic function. CRC Press, Inc., Boca Raton, Fla.
163. **Rosen, B. P.** 1990. The plasmid-encoded arsenical resistance pump: an anion-translocating ATPase. Res. Microbiol. **141:**336–341.
164. **Rosen, B. P., S. Dey, D. Dou, G. Ji, P. Kaur, M. Y. Ksenzenko, S. Silver, and J. Wu.** 1992. Evolution of an ion-translocating ATPase. Ann. N.Y. Acad. Sci. **671:**257–272.
165. **Ruhfel, R. E., D. A. Manias, and G. M. Dunny.** 1993. Cloning and characterization of a region of the *Enterococcus faecalis* conjugative plasmid, pCF10, encoding a sex pheromone-binding function. J. Bacteriol. **175:**5253–5259.
166. **Saier, M. H., Jr.** 1985. Mechanisms and regulation of carbohydrate transport in bacteria. Academic Press, Inc., New York.
167. **Saier, M. H., Jr.** 1987. Enzymes in metabolic pathways. A comparative study of mechanism, structure, evolution, and control. Harper & Row Publishers, Inc., New York.
168. **Saier, M. H., Jr. (ed.).** 1990. 6th Forum Microbiol. Coupling of energy to transmembrane solute translocation in bacteria. Res. Microbiol. **141:**281–396.
169. **Saier, M. H., Jr.** 1990. Evolution of permease diversity and energy coupling mechanisms: an introduction. 6th Forum Microbiol. Coupling of energy to transmembrane solute translocation in bacteria. Res. Microbiol. **141:**281–286.
170. **Saier, M. H., Jr.** 1991. Mechanisms of carbohydrate transport in bacteria and comparisons with those in eukaryotes, p. 833–890. *In* P. L. Yeagle (ed.), The structure of biological membranes. CRC Press, Inc., Boca Raton, Fla.
171. **Saier, M. H., Jr.** Convergence and divergence in the evolution of transport proteins. Bioessays, in press.
172. **Saier, M. H., Jr., M. J. Fagan, C. Hoischen, and J. Reizer.** 1993. Transport mechanisms in gram-positive bacteria, p. 133–156. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria, biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
173. **Saier, M. H., Jr., F. C. Grenier, C. A. Lee, and E. B. Waygood.** 1985. Evidence for the evolutionary relatedness of the proteins of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. J. Cell. Biochem. **27:**43–56.

174. **Saier, M. H., Jr., and P. McCaldon.** 1988. Statistical and functional analyses of viral and cellular proteins with N-terminal amphipathic α-helices with large hydrophobic moments. Importance to macromolecular recognition and organellar targeting. J. Bacteriol. **170:**2296–2300.
175. **Saier, M. H., Jr., A. Reizer, G. M. Pao, L.-F. Wu, J. Reizer, and A. H. Romano.** 1992. Evolution of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. II. Molecular considerations, p. 171–204. *In* R. P. Mortlock (ed.), The evolution of metabolic function. CRC Press, Inc., Boca Raton, Fla.
176. **Saier, M. H., Jr., and J. Reizer.** 1990. Domain shuffling during evolution of the proteins of the bacterial phosphotransferase system. Res. Microbiol. **141:**1033–1038.
177. **Saier, M. H., Jr., and J. Reizer.** 1991. Families and superfamilies of transport proteins common to prokaryotes and eukaryotes. Curr. Opin. Struct. Biol. **1:**362–368.
178. **Saier, M. H., Jr., and J. Reizer.** 1992. Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. J. Bacteriol. **174:**1433–1438.
179. **Saier, M. H., Jr., B. Schnierow, Y. Yamada, and G. A. Daniels.** 1989. Structures, evolution and membrane insertion of the integral membrane permease proteins of the bacterial phosphotransferase system, p. 771–780. *In* A. Kotyk, J. Skoda, V. Paces, and V. Kostka (ed.), Highlights of modern biochemistry. VSP International Science Publishers, Zeist, The Netherlands.
180. **Saier, M. H., Jr., R. Tam, A. Reizer, and J. Reizer.** Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. Mol. Microbiol., in press.
181. **Saier, M. H., Jr., P. K. Werner, and M. Müller.** 1989. Insertion of proteins into bacterial membranes: mechanism, characteristics and comparisons with the eucaryotic process. Microbiol. Rev. **53:**333–366.
182. **Saier, M. H., Jr., L.-F. Wu, G. Sweet, A. Reizer, and J. Reizer.** 1990. Evolution of permease diversity and energy coupling mechanisms with special reference to the bacterial phosphotransferase system. Biochim. Biophys. Acta **1018:**248–251.
183. **Saier, M. H., Jr., M. Yamada, J. Lengeler, B. Erni, K. Suda, P. Argos, K. Schnetz, B. Rak, C. A. Lee, G. C. Stewart, K. G. Peri, and E. B. Waygood.** 1988. Sugar permeases of the bacterial phosphoenolpyruvate-dependent phosphotransferase system: sequence comparisons. FASEB J. **2:**199–208.
184. **Saier, M. H., Jr., M. Yamada, K. Suda, B. Erni, B. Rak, J. Lengeler, G. C. Stewart, E. B. Waygood, and G. Rapoport.** 1988. Bacterial proteins with N-terminal leader sequences resembling mitochondrial targeting sequences of eukaryotes. Biochimie **70:** 1743–1748.
185. **Saimi, Y., B. Martinac, M. C. Gustin, M. R. Culbertson, J. Adler, and C. Kung.** 1988. Ion channels in *Paramecium*, yeast and *Escherichia coli*. Trends Biochem. Sci. **13:**304–309.
186. **Salzberg, S., and S. Cost.** 1992. Predicting protein secondary structure with a nearest-neighbor algorithm. J. Mol. Biol. **227:** 371–374.
187. **Saraste, M.** 1990. Structural features of cytochrome oxidase. Q. Rev. Biophys. **23:**331–366.
188. **Schnierow, B., M. Yamada, and M. H. Saier, Jr.** 1989. Partial nucleotide sequence of the *pts* operon in *Salmonella typhimurium*: comparative analyses in five bacteria genera. Mol. Microbiol. **3:**113–118.
189. **Schroeder, J. I., J. M. Ward, and W. Gassmann.** Perspectives on the physiology and structure of inward-rectifying K^+ channels in higher plants: biophysical implications for K^+ uptake. Submitted for publication.
190. **Schulein, R., I. Gentschev, H. J. Mollenkopf, and W. Goebel.** 1992. A topological model for the haemolysin translocator protein HlyD. Mol. Gen. Genet. **234:**155–163.
191. **Schulz, G. E.** 1988. A critical evaluation of methods for prediction of protein secondary structures. Annu. Rev. Biophys. Biophys. Chem. **17:**1–21.
192. **Schumacher, M. A., J. R. Macdonald, J. Björkman, S. L. Mowbray, and R. G. Brennan.** 1993. Structural analysis of the purine repressor, an *Escherichia coli* DNA-binding protein. J. Biol. Chem. **268:**12282–12288.
193. **Seol, W., and A. J. Shatkin.** 1993. Membrane topology model of *Escherichia coli* α-ketoglutarate permease by PhoA fusion analysis. J. Bacteriol. **175:**565–567.
194. **Serrano, R.** 1988. Structure and function of proton translocating ATPase in plasma membranes of plants and fungi. Biochim. Biophys. Acta **947:**1–28.
195. **Shai, Y., D. Bach, and A. Yanovsky.** 1990. Channel formation properties of synthetic pardaxin and analogues. J. Biol. Chem. **265:**20202–20209.
196. **Silver, S., G. Ji, S. Bröer, S. Dey, D. Dou, and B. P. Rosen.** 1993. Orphan enzyme or patriarch of a new tribe: the arsenic resistance ATPase of bacterial plasmids. Mol. Microbiol. **8:**637–642.
197. **Sipos, L., and G. von Heijne.** 1993. Predicting the topology of eukaryotic membrane proteins. Eur. J. Biochem. **213:**1333–1340.
198. **Smith, B. L., and P. Agre.** 1991. Erythrocyte M28,000 transmembrane protein exists as a multisubunit oligomer similar to a channel protein. J. Biol. Chem. **266:**6407–6415.
199. **Smith, R. L., J. L. Banks, M. D. Snavely, and M. E. Maguire.** 1993. Sequence and topology of the CorA magnesium transport systems of *Salmonella typhimurium* and *Escherichia coli*. Identification of a new class of transport protein. J. Biol. Chem. **268:**14071–14080.
200. **Spurlino, J. C., G.-Y. Lu, and F. A. Quiocho.** 1991. The 2.3-Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J. Biol. Chem. **266:**5202–5219.
201. **Stock, J. B., A. J. Ninfa, and A. M. Stock.** 1989. Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol. Rev. **53:**450–490.
202. **Stühmer, W.** 1991. Structure-function studies of voltage-gated ion channels. Annu. Rev. Biophys. Biophys. Chem. **20:**65–78.
203. **Sugawara, E., and H. Nikaido.** 1992. Pore-forming activity of OmpA protein of *Escherichia coli*. J. Biol. Chem. **267:**2507–2511.
204. **Sugiyama, J. E., S. Mahmoodian, and G. R. Jacobson.** 1991. Membrane topology analysis of *Escherichia coli* mannitol permease by using a nested-deletion method to create *mtlA-phoA* fusions. Proc. Natl. Acad. Sci. USA **88:**9603–9607.
205. **Sutrina, S. L., P. Reddy, M. H. Saier, Jr., and J. Reizer.** 1990. The glucose permease of *Bacillus subtilis* is a single polypeptide chain that functions to energize the sucrose permease. J. Biol. Chem. **265:**18581–18589.
206. **Sutrina, S. L., K. Schnetz, B. Rak, and M. H. Saier, Jr.** 1990. Mechanism of sugar transport and phosphorylation via permeases of the bacterial phosphotransferase system: catalytic residues in the β-glucoside-specific permease as defined by site-specific mutagenesis. Res. Microbiol. **141:**368–374.
207. **Takase, K., I. Yamato, and Y. Kakinuma.** 1993. Cloning and sequencing of the genes coding for the A and B subunits of vacuolar-type Na^+-ATPase from *Enterococcus hirae*. Coexistence of vacuolar- and F_0F_1-type ATPases in one bacterial cell. J. Biol. Chem. **268:**11610–11616.
208. **Tam, R., and M. H. Saier, Jr.** 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol. Rev. **57:**320–346.
209. **Tam, R., and M. H. Saier, Jr.** 1993. A bacterial periplasmic receptor homologue with catalytic activity: cyclohexadienyl dehydratase of *Pseudomonas aeruginosa* is homologous to receptors specific for polar amino acids. Res. Microbiol. **144:**165–169.
210. **Tamm, L. K., and S. R. Tatulian.** 1993. Orientation of functional and non-functional PTS permease signal sequences in lipid bilayers. A polarized attenuated total reflection infrared study. Biochemistry **32:**7720–7726.
211. **Tanimoto, K., F. Y. An, and D. B. Clewell.** 1993. Characterization of the *traC* determinant of the *Enterococcus faecalis* hemolysin-bacteriocin plasmid pAD1: binding of sex pheromone. J. Bacteriol. **175:**5260–5264.
212. **Tisa, L. S., and J. Adler.** 1992. Calcium ions are involved in *Escherichia coli* chemotaxis. Proc. Natl. Acad. Sci. USA **89:** 11804–11808.
213. **Tisa, L. S., B. M. Olivera, and J. Adler.** 1993. Inhibition of *Escherichia coli* chemotaxis by ω-conotoxin, a calcium ion channel blocker. J. Bacteriol. **175:**1235–1238.

213a. **Titgemeyer, F., M. Eddy, and M. H. Saier, Jr.** Unpublished results.
214. **Tomich, J.** 1993. Amphipathic helixes in channel-forming structures, p. 221–254. *In* R. M. Epand (ed.), The amphipathic helix. CRC Press, Inc., Boca Raton, Fla.
215. **Trandinh, C. C., G. M. Pao, and M. H. Saier, Jr.** 1992. Structural and evolutionary relationships among the immunophilins: two ubiquitous families of peptidyl-prolyl *cis-trans* isomerases. FASEB J. **6**:3410–3420.
216. **Trumpower, B. L.** 1990. Cytochrome bc_1 complexes of microorganisms. Microbiol. Rev. **54**:101–129.
217. **Tsien, R. W., P. Hess, E. W. McCleskey, and R. L. Rosenberg.** 1987. Calcium channels: mechanisms of selectivity, permeation, and block. Annu. Rev. Biophys. Biophys. Chem. **16**:265–290.
218. **Turner, R. J., Y. Hou, J. H. Weiner, and D. E. Taylor.** 1992. The arsenical ATPase efflux pump mediates tellurite resistance. J. Bacteriol. **174**:3092–3094.
219. **Uhl, G. R.** 1992. Neurotransmitter transporters (plus): a promising new gene family. Trends Neurosci. **15**:265–268.
220. **Unwin, N.** 1993. Nicotinic acetylcholine receptor at 9 Å resolution. J. Mol. Biol. **229**:1101–1124.
221. **van Hoek, A. N., and A. S. Verkman.** 1992. Functional reconstitution of the isolated erythrocyte water channel CHIP28. J. Biol. Chem. **267**:18267–18269.
222. **Van Rosmalen, M., and M. H. Saier, Jr.** 1993. Structural and evolutionary relationships among two families of bacterial periplasmic chaperone proteins which function cooperatively in fimbrial assembly. Res. Microbiol. **144**:507–527.
223. **Vartak, N. B., J. Reizer, A. Reizer, J. T. Gripp, E. A. Groisman, L.-F. Wu, J. M. Tomich, and M. H. Saier, Jr.** 1991. Sequence and evolution of the FruR protein of *Salmonella typhimurium*: a pleiotropic transcriptional regulatory protein possessing both activator and repressor functions which is homologous to the periplasmic ribose-binding protein. Res. Microbiol. **142**:951–963.
224. **Vázquez, M., O. Santana, and C. Quinto.** 1993. The NodI and NodJ proteins from *Rhizobium* and *Bradyrhizobium* strains are similar to capsular polysaccharide secretion proteins from Gram-negative bacteria. Mol. Microbiol. **8**:369–377.
225. **Verma, D. P. S.** 1992. Soybean nodulin-26: a channel protein conserved from bacteria to mammals, p. 113–117. *In* D. T. Cooke and D. T. Clarkson (ed.), Transport and receptor proteins of plant membranes. Plenum Press, New York.
226. **Voegele, R., G. Sweet, and W. Boos.** 1992. Isolation and characterization of the glycerol transport protein, GlpF, from *Escherichia coli*, abstr. P212, p. 57. BioEngineering, February 1992.
227. **von Heijne, G.** 1986. Mitochondrial targeting sequences may form amphiphilic helices. EMBO J. **5**:1335–1342.
228. **von Heijne, G.** 1986. Towards a comparative anatomy of N-terminal topogenic protein sequences. J. Mol. Biol. **189**:239–242.
229. **von Heijne, G.** 1988. Transcending the impenetrable: how proteins come to terms with membranes. Biochim. Biophys. Acta **947**:307–333.
230. **von Heijne, G.** 1991. Computer analysis of DNA and protein sequences. Eur. J. Biochem. **199**:253–256.
231. **von Heijne, G.** 1992. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J. Mol. Biol. **225**:487–494.
232. **Vulpe, C., B. Levinson, S. Whitney, S. Packman, and J. Gitschier.** 1993. Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase. Nat. Genet. **3**:7–13.
233. **Vyas, N. K., M. N. Vyas, and F. A. Quiocho.** 1987. A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis. Nature (London) **327**:635–638.
234. **Vyas, N. K., M. N. Vyas, and F. A. Quiocho.** 1991. Comparison of the periplasmic receptors for L-arabinose, D-glucose/D-galactose, and D-ribose. Structural and functional similarity. J. Biol. Chem. **266**:5226–5237.
235. **Walker, J. E.** 1992. The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains. Q. Rev. Biophys. **25**:253–324.
236. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.
237. **Wang, D. N., W. Kühlbrandt, V. E. Sarabia, and R. A. F. Reithmeier.** 1993. Two-dimensional structure of the membrane domain of human Band 3, the anion transport protein of the erythrocyte membrane. EMBO J. **12**:2233–2239.
238. **Waser, M., D. Hess-Bienz, K. Davies, and M. Solioz.** 1992. Cloning and disruption of a putative Na^+/H^+-antiporter gene of *Enterococcus hirae*. J. Biol. Chem. **267**:5396–5400.
239. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5164.
240. **Weickert, M. J., and S. Adhya.** 1992. A family of bacterial regulators homologous to Gal and Lac repressors. J. Biol. Chem. **267**:15869–15874.
241. **Weiss, M. S., A. Kreusch, E. Schiltz, U. Nestel, W. Welte, J. Weckesser, and G. E. Schulz.** 1991. The structure of porin from *Rhodobacter capsulatus* at 1.8 Å resolution. FEBS Lett. **280**:379–382.
242. **Wilson, S., and R. Drew.** 1991. Cloning and DNA sequence of *amiC*, a new gene regulating expression of the *Pseudomonas aeruginosa* aliphatic amidase, and purification of the *amiC* product. J. Bacteriol. **173**:4914–4921.
243. **Wootton, J. C., and M. H. Drummond.** 1989. The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. Protein Eng. **2**:535–543.
244. **Wu, J., L. S. Tisa, and B. P. Rosen.** 1992. Membrane topology of the ArsB protein, the membrane subunit of an anion-translocating ATPase. J. Biol. Chem. **267**:12570–12576.
245. **Wu, L.-F., A. Reizer, J. Reizer, B. Cai, J. M. Tomich, and M. H. Saier, Jr.** 1991. Nucleotide sequence of the *Rhodobacter capsulatus fruK* gene, which encodes fructose-1-phosphate kinase: evidence for a kinase superfamily including both phosphofructokinases of *Escherichia coli*. J. Bacteriol. **173**:3117–3127.
246. **Wu, L.-F., and M. H. Saier, Jr.** 1990. On the evolutionary origins of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. Mol. Microbiol. **4**:1219–1222.
247. **Wu, L.-F., and M. H. Saier, Jr.** 1990. Nucleotide sequence of the *fruA* gene encoding the fructose permease of the *Rhodobacter capsulatus* phosphotransferase system, and analyses of the deduced protein sequence. J. Bacteriol. **172**:7167–7178.
248. **Wu, L.-F., and M. H. Saier, Jr.** 1991. Differences in codon usage among genes encoding proteins of different function in *Rhodobacter capsulatus*. Res. Microbiol. **142**:943–949.
249. **Wu, L.-F., J. M. Tomich, and M. H. Saier, Jr.** 1990. Structure and evolution of a multidomain, multiphosphoryl transfer protein: nucleotide sequence of the *fruB(HI)* gene in *Rhodobacter capsulatus* and comparisons with homologous genes from other organisms. J. Mol. Biol. **213**:687–703.
249a. **Xu, X., and M. H. Saier, Jr.** Unpublished results.
250. **Yamada, M., and M. H. Saier, Jr.** 1987. Glucitol-specific enzymes of the phosphotransferase system in *Escherichia coli*. Nucleotide sequence of the *gut* operon. J. Biol. Chem. **262**:5455–5463.
251. **Yamada, M., and M. H. Saier, Jr.** 1988. Positive and negative regulators for glucitol (*gut*) operon expression in *Escherichia coli*. J. Mol. Biol. **203**:569–583.
252. **Yamada, M., Y. Yamada, and M. H. Saier, Jr.** 1990. Nucleotide sequence and expression of the *gutQ* gene within the glucitol operon of *Escherichia coli*. DNA Sequence **1**:141–145.
253. **Yamada, Y., Y.-Y. Chang, G. A. Daniels, L.-F. Wu, J. M. Tomich, M. Yamada, and M. H. Saier, Jr.** 1991. Insertion of the mannitol permease into the membrane of Escherichia coli. Possible involvement of an N-terminal amphiphilic sequence. J. Biol. Chem. **266**:17863–17871.
254. **Yaron, A., and Naider, F.** 1993. Proline-dependent structural and biological properties of peptides and proteins. Crit. Rev. Biochem. Mol. Biol. **28**:31–81.
255. **Zani, M.-L., T. Pourcher, and G. Leblanc.** 1993. Mutagenesis of acidic residues in putative membrane-spanning segments of the melibiose permease of *Escherichia coli*. II. Effect on cationic selectivity and coupling properties. J. Biol. Chem. **268**:3216–3221.