

Evolutionary Relationship between K⁺ Channels and Symporters

Stewart R. Durell,* Yili Hao,* Tatsunosuke Nakamura,[#] Evert P. Bakker,[§] and H. Robert Guy*

*Laboratory of Experimental and Computational Biology, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-5677 USA; [#]Laboratory of Membrane Biochemistry, Faculty of Pharmaceutical Sciences, Chiba University, Inage-ku, Chiba 263, Japan; and [§]Abteilung Mikrobiologie, Universität Osnabrück, D-49069 Osnabrück, Germany

ABSTRACT The hypothesis is presented that at least four families of putative K⁺ symporter proteins, Trk and KtrAB from prokaryotes, Trk1,2 from fungi, and HKT1 from wheat, evolved from bacterial K⁺ channel proteins. Details of this hypothesis are organized around the recently determined crystal structure of a bacterial K⁺ channel: i.e., KcsA from *Streptomyces lividans*. Each of the four identical subunits of this channel has two fully transmembrane helices (designated M1 and M2), plus an intervening hairpin segment that determines the ion selectivity (designated P). The symporter sequences appear to contain four sequential M1-P-M2 motifs (MPM), which are likely to have arisen from gene duplication and fusion of the single MPM motif of a bacterial K⁺ channel subunit. The homology of MPM motifs is supported by a statistical comparison of the numerical profiles derived from multiple sequence alignments formed for each protein family. Furthermore, these quantitative results indicate that the KtrAB family of symporters has remained closest to the single-MPM ancestor protein. Strong sequence evidence is also found for homology between the cytoplasmic C-terminus of numerous bacterial K⁺ channels and the cytoplasm-resident TrkA and KtrA subunits of the Trk and KtrAB symporters, which in turn are homologous to known dinucleotide-binding domains of other proteins. The case for homology between bacterial K⁺ channels and the four families of K⁺ symporters is further supported by the accompanying manuscript, in which the patterns of residue conservation are demonstrated to be similar to each other and consistent with the known 3D structure of the KcsA K⁺ channel.

INTRODUCTION

Regulation of ion gradients across the plasma membrane is a requirement of all living cells. Much of this is accomplished by membrane channel proteins that allow ions to diffuse passively down their electrochemical gradients, and by membrane transport proteins that use energy to transport ions actively against their electrochemical gradients. There have been numerous suggestions that in some active ion transporters, ions may diffuse most of the way across the membrane through a “pore” (Jardetzky, 1966; Lauger, 1979; Su et al., 1996). This hypothesis has gained support from findings that several proteins that are homologous to transporters act as channels: the Cystic Fibrosis Conductance Regulator (CFTR) is a Cl[−] ion channel, even though its primary sequence is homologous to the ABC superfamily of transporters (Anderson et al., 1991); the glutamate transporters (Larsson et al., 1996) and norepinephrine transporters (Galli et al., 1998) apparently act as channels under some conditions; the Kef family of bacterial K⁺ channels (Booth et al., 1996) is homologous to the NapA Na⁺/H⁺ family of antiporters (Reizer et al., 1992); and the Kir inward rectifying K⁺ channel associates with a Sur protein that is homologous to the ABC transporters (Ashcroft and Gribble, 1998). Symporters are transporters in which the transport of one ion or molecule against its electrochemical

gradient is “powered” by the movement of another ion or molecule down its electrochemical gradient in the same direction through the membrane. Plausible mechanisms for symport, in which even the actively transported ion diffuses most of the way through the transmembrane protein, are discussed in the accompanying manuscript (Durell and Guy, 1999).

This report provides indirect evidence, from analysis of the sequences, for homology and common structural features between the superfamily of K⁺ channel proteins and four K⁺ symporter protein families. The four symporter families are 1) the K⁺-translocating TrkH subunit from the Trk systems of both bacteria and archaea (Schlösser et al., 1991, 1995; Stumpe et al., 1996), 2) the KtrB subunit from a recently described KtrAB system in eubacteria (Nakamura et al., 1998b) (previously identified as NtpJ by Takase et al., 1994, and Clayton et al., 1997), 3) the Trk1,2 proteins from yeasts and *Neurospora* (Gaber et al., 1988; Ko and Gaber, 1991; Lichtenberg-Fraté et al., 1996; Haro et al., 1999) and 4) the HKT1 protein from wheat (Schachtman and Schroeder, 1994; Wang et al., 1998) and a homologue from *Arabidopsis* (Washington University Genome Sequencing Center, 1998 [The *A. thaliana* Genome Sequencing Project, <http://genome.wustl.edu/gsc/arab/arabidopsis.html>]; Bevan et al., 1999 [EU *Arabidopsis* sequence project, unpublished; accession no. CAB39784]). For the purpose of this analysis, the fungi and plant symporters are grouped into a single eukaryotic family called Trk-euk. The current supposition is that functional Trk-euk proteins are formed from a single type of subunit, although the structural similarities outlined below may force some reconsideration. HKT1 symport in wheat is dependent on Na⁺ (Rubio et al., 1995; Diatloff et al., 1998), and TKH_p symport in the fission yeast *Schizo-*

Received for publication 8 February 1999 and in final form 3 May 1999.

Address reprint requests to Dr. H. Robert Guy, Laboratory of Experimental and Computational Biology, National Cancer Institute, National Institutes of Health, Bldg. 12B, Rm. B116, 12 South Drive, MSC 5677, Bethesda, MD 20892-5677. Tel.: 301-496-2068; Fax: 301-402-4724; E-mail: guy@ncl.nih.gov.

© 1999 by the Biophysical Society

0006-3495/99/08/775/14 \$2.00

saccharomyces pombe (which is closely related to the budding yeast Trk1,2 system) is dependent on H^+ (Lichtenberg-Fraté et al., 1996) (although a possible role for Na^+ has not been excluded). In comparison, the functional forms of the bacterial Trk and KtrAB systems are clearly more structurally complex; both comprise multiple subunit types (Stumpe et al., 1996; Nakamura et al., 1998). Trk cotransports H^+ with K^+ (Stumpe et al., 1996), whereas KtrAB is Na^+ linked (Tholema et al., 1999).

Additional evidence of the homology between these symporter and channel proteins comes from the development of 3D atomic-scale models of the transmembrane regions, which is presented in the accompanying paper. Specifically, it is found that the pattern of amino acid residue conservation within each symporter family is consistent with the structural fold and ion-selective mechanism employed by the superfamily of K^+ channel proteins.

The ability to compare the symporter and channel proteins is now greatly enhanced by the recently determined crystal structure of the transmembrane component of the KcsA K^+ channel from *Streptomyces lividans* (Doyle et al., 1998), which certifies the basic structural and functional roles of the different channel segments. Perhaps most importantly, this has confirmed the role of the P segment in forming the outer portion of the pore and the ion selectivity filter, which was previously predicted by indirect theoretical and experimental methods (see accompanying paper for details). Specifically, the four P segments (one from each of the four channel subunits) are arranged with fourfold symmetry around the axis of the pore, with each in the same hairpin conformation and dipping into the outer portion of the transmembrane region from the extracellular side. The first arm of the hairpin (P1) is an α -helix that slants toward the center of the channel, and the second arm (P2) is an extended α -structure (the backbone alternates between right- and left-handed α -helix conformations; Guy and Durrell, 1995) that rises out of the channel along the axis. Collectively, the four P2 segments form the narrowest portion of the pore, which consequently acts as the selectivity

filter. The K^+ binding sites are formed by the backbone carbonyl oxygen atoms of conserved "signature sequence" residues of the four P2 segments. The full P-segment hairpin (P1 + P2) is located between two hydrophobic transmembrane helices (M1 and M2) that together form the MPM (or 2TM) motif. This contrasts with the 6TM motif in many other types of K^+ channels (e.g., the voltage-gated *Shaker* channel protein), in which the MPM structure is preceded by four additional hydrophobic transmembrane segments (Uozumi et al., 1998; Shih and Goldin, 1997).

The first hint of homology between symporter and channel proteins came from the sequence analysis work of Jan and Jan (1994), who postulated that TrkH has two P-like segments similar to those of K^+ channels. This led Stumpe et al. (1996) to propose a transmembrane topology for TrkH that contained a MPM motif at both the N- and C-terminal ends of the transmembrane region of the sequence. While searching the databases for possible bacterial K^+ channels, the group of Guy found that some specific MPM channel sequences were actually more similar to portions of some K^+ symporters than to other K^+ channel proteins (see Fig. 1). Surprisingly, the matching portion in these symporter sequences was not at the P regions identified by the Jan and Jan group, but rather at an intermediate location. As described below, further sequence analyses led the Guy and Nakamura-Bakker groups independently to the notion that these symporters actually comprise four sequential MPM motifs (designated MPM_A, MPM_B, MPM_C, and MPM_D).

This arrangement of primary structure suggests the process of gene duplication, similar to the evolutionary schemes deduced for related Na^+ , Ca^{+2} , and some K^+ channel proteins. For example, the TWIK (or $2 \times 2TM$) type of K^+ channels have two MPM motifs within each of two identical subunits (Lesage et al., 1996), the yeast TOK (or DUK1) channel subunit has a 6TM motif followed by an MPM motif (Ketchum et al., 1995; Reid et al., 1996), and both Na^+ and Ca^{+2} channels have four consecutive 6TM motifs within their primary pore-forming subunits (Noda et al., 1984). Finally, the hypothesis of homology between the

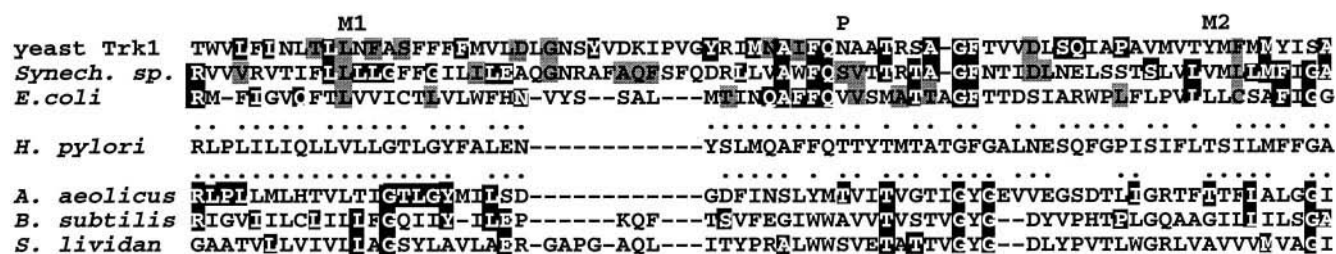


FIGURE 1 Alignment of the MPM_C motif from one member of each family of K^+ symporters (top three sequences) with the single MPM motif of four putative bacterial K^+ channel subunits. The residues with black backgrounds are identical to analogous residues from the putative K^+ channel of *Helicobacter pylori*, and residues with gray backgrounds are identical to residues in at least one of the other putative K^+ channel sequences. The dots indicate residues in the *H. pylori* sequence that are identical or similar to residues in the symporter (top) or other channel (bottom) sequences. Although these segments contain the ion-selective "signature sequence" of K^+ channels, the sequence from the putative *H. pylori* K^+ channel is actually more similar to the sequences of the K^+ symporters. However, the *H. pylori* sequence is classified as a K^+ channel because the portions of its sequence not shown here are similar to those of other putative bacterial K^+ channels, which have a single MPM motif followed by a long polar C-terminus containing a dinucleotide binding domain. The *Aquifex aeolicus* sequence is the most closely related K^+ channel that has been reported to date. The 2TM bacterial K^+ channel sequences from *Streptomyces lividans* and *Bacillus subtilis* contain the P segments most similar to those of the eukaryotic voltage-gated K^+ channels.

channel and symporter proteins is also supported by sequence similarity between the cytoplasmic domain of many of the bacterial K⁺ channels and the 120-residue NAD-binding domains in cytoplasmic subunits of the Trk and KtrAB symporter complexes, e.g., TrkA and KtrA (Schlösser et al., 1993; Nakamura et al., 1998).

METHODS

Sequence acquisition and alignments

The four families of homologous bacterial K⁺ channel and symporter sequences were obtained by a combination of motif and keyword searches of the NCBI's Genbank and microbial databases (see NCBI BLAST: Unfinished Microbial Genomes (<http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html>) and NCBI PSI-BLAST (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-psi_blast)). The motif searches were carried out using the ion-selective, P-segment of K⁺ channels as the seed for gapped BLAST and PSI-BLAST procedures (Altschul et al., 1997). The resultant multiple sequence alignments of MPM motifs were then manually adjusted to emphasize the common features among all proteins. This involved matching features within each protein family locally and between the four families globally. Because of variability within the loop regions, our primary concern was alignment of the three main segments (i.e., M1, P, and M2) with as few gaps as possible. In sum, the data consisted of 13 multiple sequence alignments of MPM motifs—one from the K⁺ channels and four from each of the three symporter families—each containing three subalignments corresponding to the M1, P, and M2 segments.

Quantification of homology

To quantify the similarity among motifs, each multiple sequence alignment was converted into a numerical profile matrix. This was carried out according to the methods described by Henikoff and Henikoff (1996) for creating a log-odds position-specific scoring matrix (PSSM). These procedures estimate the residue frequencies at each position for the entire population of related proteins in nature from the limited and nonrandomly sampled set of known sequences used in the alignments. Briefly, the steps were 1) weighting the observed counts of each residue by the calculated redundancy of the parent sequence in the multiple sequence alignment (Henikoff and Henikoff, 1994), 2) adding "imaginary" pseudo-counts to the sequence-weighted counts according to the residue diversity at each location and empirically determined residue substitution probabilities (BLOSUM; Henikoff and Henikoff, 1992), 3) normalizing these composite counts by the expected frequency of occurrence of the specific residue (estimated from the amino acid composition of the Swiss-Prot sequence database; Bairoch and Apweiler, 1998), and 4) taking the logarithm of these normalized counts to obtain the PSSM score for each of the 20 residues at each location. Throughout this analysis, effort was directed toward determining the sensitivity to the multiplication factor used for the total number of pseudo-counts, which determines the relative proportion to the weighted sequence counts of the alignments. Because the effect on the final results was minimal, the recommended value of 5 was used (Henikoff and Henikoff, 1996).

Quantification of the similarity of each pair of PSSMs of the same segment type was performed according to the methods of Pietrokovski (1996). This entailed calculating the Pearson's correlation coefficient for each pair of aligned profile columns and then adding the coefficients to obtain the total raw score. For the purpose of comparison, the raw scores were converted into Z scores, which is the number of standard deviations it is away from the mean of a distribution of best raw scores obtained by chance for unrelated protein families. Because there is a dependence of the raw score on the length of the segment, it is necessary to have a series of best chance score distributions corresponding to each possible segment length. Such distributions were calculated by full enumeration of every

possible pair of the 3670 PSSMs of multiply aligned sequences in the Blocks 10.1 database (Henikoff and Henikoff, 1991), which resulted in over 6.7 million chance scores. The database was previously modified by the removal of compositionally redundant blocks, and the sequence columns of one of each pair of PSSMs was randomly shuffled to eliminate bias in the results (Pietrokovski, 1996). Multiple collections of chance distributions were calculated for each set of trial PSSM creation parameters used to examine the sensitivity of the results (described above). Finally, assuming the distributions to be normal, the probability that a particular score would occur by chance was determined from the definite integral of the Gaussian probability distribution (Bevington, 1969). For example, the probability of obtaining Z scores of 2, 3, 4, 5, and 6 for unrelated protein families would be approximately 5×10^{-1} , 3×10^{-1} , 6×10^{-3} , 6×10^{-5} , and 4×10^{-9} %, respectively.

To provide a reference in the context of membrane proteins, comparisons of the bacterial 2TM channels and symporters were also made with the transmembrane segment blocks of 19 bacteriorhodopsin homologues and other ion channel proteins. Whereas the bacteriorhodopsin sequences were taken to be evolutionarily unrelated, the channel proteins, which included the TWIK family from *C. elegans*, IRK family from eukaryotes, and Na⁺ channel family were expected to have various degrees of homology. In the calculations, full enumeration was used to find the contiguous segment of at least four residues with the highest Z score. The only exception was for comparison of the bacterial 2TM channel and symporter families themselves, for which the relative overall alignments of the blocks were kept the same as shown in Fig. 2, A and B. Within this restriction, enumeration was again used to find the highest Z-scoring subsegment of at least four residues.

RESULTS AND DISCUSSION

Sequence alignment

Fig. 2, A and B, displays the global alignment of all of the MPM motifs used to study the evolutionary relationships within and between the four families of K⁺ channels and symporters. Although only consensus sequences are used in the figure for clarity, all analyses were conducted on the full set of multiply aligned sequences (see the Appendix for the list). The 13 consensus sequences correspond to the single MPM motif in the channels and the four MPM motifs from each of the three symporter families. In all parts of the figure, the spectrum from red to blue/black represents the range of residues from conserved to variable. Fig. 2 A presents the global pattern of conservation among the 13 consensus sequences, and Fig. 2 B presents the local pattern of conservation among the sequences used to generate each of the consensus sequences. As seen in Fig. 2 A, the alignments of the P and M2 segments were keyed to the highly conserved residues shown in red and orange. In contrast, alignment of the M1 segments was more difficult because of the lack of global conservation. Interestingly, the variable and highly hydrophobic nature of this segment in both the channel and symporter sequences is consistent with its position in the KcsA crystal structure, i.e., the four M1 helices are on the periphery of the protein, they are largely lipid exposed, and they do not directly form the pore structure. Consequently, these segments were initially aligned by simply matching the hydrophobic regions with as few insertions or deletions as possible. Then, as seen in Fig. 2 B, finer adjustments were made to align the conserved residues within each family.

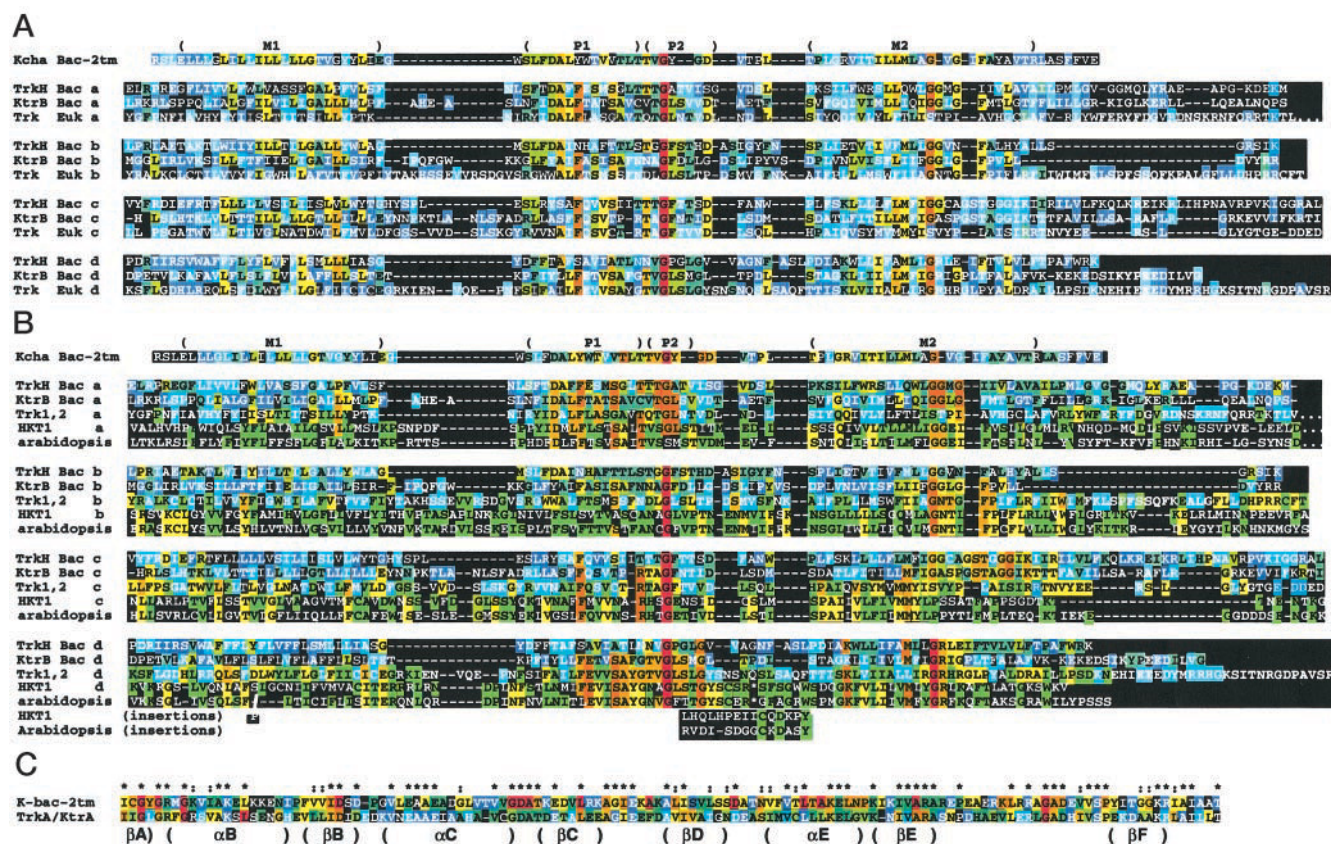


FIGURE 2 Consensus sequence alignments of putative 2TM bacterial K^+ channels with the three K^+ symporter families. See the Appendix for the list of sources used to generate the consensus sequences. (A) The residues are colored coded according to the number of identical residues in the 13 consensus sequences, i.e., red = 13; reddish orange = 10–12; yellowish orange = 8–9; yellow = 6–7; yellowish green = 5; green = 4; cyan = 3; blue = 2; black = 1. (B) The residues are colored according to the degree of conservation within each family or MPM motif, and between families of symporters. Colors were determined by counting the number of residue types that occur at each position in the alignment of the sequences for a given family. To reduce the influence of a sequence error, residues occurring only once in the channel, TrkH and KtrB sequences, were scored as 0.5; otherwise, residues were assigned the full score of 1.0. The color code for the channels is red = 1–1.5; reddish orange = 2–2.5; yellowish orange = 3–3.5; yellow = 4–4.5; yellowish green = 5–5.5; green = 6–6.5; cyan = 7–7.5; blue = 8–8.5; black > 8.5. The symporters were colored differently to identify residues that are conserved among or between families. The color code for the more highly conserved residues is red = 1–1.5, with identical consensus residues in all three families; reddish orange = 1–1.5, with identical consensus residues in two families; and yellowish orange < 2.5, with identical consensus residues in two families. The remaining residues that are not well conserved between families were colored as follows for KtrB and TrkH: yellow = 1–1.5; yellowish green = 2–2.5; green = 3–3.5; cyan = 4–4.5; blue = 5–5.5; black = > 5.5. The Trk1,2 family was scored differently because of the relatively low number of sequences, which are also distantly related. Each residue type was given a score of 1, even if it occurred only once. The color scheme for residues that were not conserved between families is yellow = 1; green = 2; cyan = 3; blue = 4; black = > 4. Furthermore, a Trk1,2 residue was colored yellowish orange if its score was 1 and identical residues occurred in the two plant sequences. Single sequences are shown for the two plant sequences. For these sequences, red and orange indicate residues that are identical to residues that are conserved between families for the other sequences, yellow indicates that the residue is identical in both plants and all Trk1,2 sequences, green indicates that the residue is identical in two of the three Trk-euk sequences, and black indicates that the residue is unique among the three sequences. Some insertions in MPM_D of the plant sequences are indicated below the consensus sequences. (C) Consensus sequences for dinucleotide binding regions of putative 2TM bacterial K^+ channels, and from the KtrA and TrkA subunits of the KtrAB and Trk symporters. Residues are colored according to the extent of identity among the sequences used to generate the consensus sequences. * and . indicate identical or similar residues, respectively. The assignment of α -helices and β -strands is according to the method of Schlösser et al. (1993).

Subsequently, special emphasis was given to the latter portion of the M1 segment, because this region in the KcsA structure packs closely to the crucial P segments in the crystal. For the 2TM bacterial K^+ channels, the two most highly conserved residues in the C-terminal half of M1 are a glycine located nine residues before the end and a glutamate located right at the end (see Fig. 2 B). Thus most M1 segments of the symporters were aligned so that a small residue, usually glycine, coincides with the channel glycine. In MPM_C and MPM_D of KtrB and Trk-euk, a glutamate or

aspartate at the end of M1 aligned with the highly conserved channel glutamate. In the few cases where these criteria were insufficient, the more highly conserved and/or more hydrophilic symporter residues in M1 were aligned with the more highly conserved channel residues that are oriented toward the protein and away from the lipid in the KcsA structure.

As indicated by the single red column in Fig. 2, A and B, the most highly conserved residue among the channel and symporter sequences is a glycine in the P region (the only

exception being a serine substitution in the MPM_A of the *Arabidopsis* protein). This provides an important evolutionary link between the protein families, in that this residue is known to play a major role in determining the ion selectivity for many classes of K⁺ channel proteins. For example, mutagenesis studies have found that this is the only residue in the Shaker K⁺ channel P segment that cannot be mutated to Cys in even one of the four subunits without loss of function (Lü and Miller, 1995). This can be explained by the findings in the KcsA structure that the backbone conformation of this glycine is energetically unfavorable for other types of residues and that the four backbone carbonyl oxygen atoms of the glycine residue—one from each of the four subunits—form an ion-binding site at the narrowest portion of the pore. Indeed, the functional significance of this glycine is further emphasized by the fact that this is the only residue that is identical among the set of 27 putative 2TM bacterial K⁺ channels (Fig. 2 B).

Next, the reddish orange columns in Fig. 2 A denote single residues, of the symporter P1 and M2 segments, that are identical among the consensus sequences in all but two locations. In P1 the phenylalanine aligns with the tyrosine (similarly aromatic) of the K⁺ channel consensus sequence. In the KcsA structure, that residue is a tryptophan, which combines with an adjacent P1 tryptophan and a P2 tyrosine in each of the four subunits to form an aromatic cuff around the selectivity filter (Doyle et al., 1998). For the M2 segment, the highly conserved residue is a glycine that is also well conserved among the channels. Its structural importance is indicated by the fact that in the KcsA structure it packs next to the innermost part of the P segment. This site may be important for channel gating, as well as selectivity, because the inner portion of M2 in the KcsA structure moves closer to the pore as the channel closes (Perozo et al., 1998, 1999).

Other residues that are conserved moderately well are indicated by letters in black type. These include 1) the threonine-rich region of P1 preceding the fully conserved glycine, which is strongly conserved among the K⁺ channels and, to a lesser degree, among the symporters; 2) a DAL sequence conserved in many of the symporters, lying just before the highly conserved aromatic P1 residue; 3) the ILLML consensus sequence preceding the conserved M2-glycine, which is strongly conserved among the channels and partly conserved among the symporters; and finally, 4) numerous leucines that appear to be well conserved in both the M1 and M2 segments. However, such leucine matches cannot be securely interpreted as indicative of homology, because leucine is the most frequently occurring residue in the hydrophobic regions of transmembrane helices (Hofmann and Stoffel, 1993). Note in Fig. 2 B, for example, that many of the M1 leucines are not well conserved within each family.

Conservation pattern

Related to the conservation of specific residue types, homologous relationships are also indicated by the similar

patterns of residue conservation for each of the protein families. This is demonstrated in Fig. 2 B, in which the consensus sequences are color coded according to the degree of conservation within each family (i.e., among the sequences used to develop the consensus sequences), or, in the case of the red and orange colors used for the symporters, by the similarity of the consensus sequences among the three families of symporters (see legend for code).

Red to orange denotes residues that are conserved in two or more families.

Yellow to green indicates residues that are well conserved within the family, but not conserved between the families.

Blue to black represents residues that are poorly conserved within the family.

As seen, the same general pattern of sequence conservation is repeated within each MPM core of the bacterial K⁺ channels and the three symporter families. More specifically, the poorly conserved M1 segments are followed by highly conserved P segments, which are followed by the center-region-conserved M2 segments. Furthermore, all linkers between segments are poorly conserved, with numerous insertions and deletions. Close inspection reveals that most of the globally conserved residues identified in Fig. 2 A are located within the regions of local conservation.

A separate comparison is also made in Fig. 2 B for the two plant symporters (wheat and *Arabidopsis*), which are colored according to the conservation between them and in relation to the fungal sequences (see legend for code).

Nucleotide binding domains and subunits

Another line of evidence supporting homology among these proteins involves the separate subunits of the KtrAB and Trk symporters that contain dinucleotide-binding domains: i.e., KtrA and TrkA. These are peripheral membrane proteins (Bossemeyer et al., 1989; Nakamura et al., 1998), which are probably located at the cytoplasmic side of the membrane. The KtrA subunit is homologous to the dinucleotide-binding site sequences of many other proteins and combines with the transmembrane KtrB protein to form a functional symporter. The TrkA subunit, however, is more complicated. Except for three archaeal TrkA species that also contain only one dinucleotide-binding domain, all other TrkAs have two dinucleotide-binding sites contained in each of two similar subdomains (Stumpe et al., 1996; Nakamura et al., 1998a; Kawarabayashi et al., 1998). In addition, TrkA interacts with multiple protein subunits, in addition to TrkH, to form the functional symporter (Dosch et al., 1991; Parra-Lopez et al., 1994; Stumpe et al., 1996; Nakamura et al., 1998a). It must be noted, however, that only the *E. coli* TrkA protein has actually been demonstrated to bind NAD⁺ and/or NADH in vitro (Schlösser et al., 1993). In addition, it is not yet known whether dinucleotides influence the transport activities of the proteins in vivo.

Database searches indicated that the closest sequences to the KtrA and TrkA subunits are C-terminal portions of some 2TM bacterial K⁺ channels and of some members of the Kef family of K⁺ channels (Munro et al., 1991; Stumpe et al., 1996). The close homology between these sequences is evident in the alignment of representative samples shown in Fig. 2 C, in which there is 44% identity over a stretch of ~120 residues. This is the longest segment for which the sequences can be aligned unambiguously. It contains only one complete dinucleotide-binding domain, which appears to be a chimera between the N-terminal segment of the NAD⁺-binding domain of malate/lactate dehydrogenase-like proteins and the C-terminal segment of the NAD⁺-binding domain from the glyceraldehyde-3-phosphate dehydrogenase-like proteins (Schlösser et al., 1993; Stumpe et al., 1996). Such findings suggest that the small KtrA and TrkA subunits may have derived from the cleavage of a covalently attached C-terminal region of an ancestral 2TM K⁺ channel.

Although most eukaryotic and some bacterial K⁺ channels (including the KcsA protein) lack an intrinsic dinucleotide-binding domain, various other K⁺ channels are found to have more distantly homologous sequences at the C-termini. These include some putative bacterial channels of the 6TM type (e.g., Kch from *E. coli*; Parra-Lopez et al., 1994), the high-conductance *Slo*-type channel from animal cells (Parra-Lopez et al., 1994; Stumpe et al., 1996), and the newly identified channel-like sequence from *Aquifex aeolicus* (Deckert et al., 1998). Moreover, proper β -subunits from a variety of plant and animal K⁺ channels have redox function; and some, such as the *Shaker* K⁺ channel β -subunit, align nicely with the eight-stranded β -barrel structure of NAD(P)H-dependent oxidoreductases (McCormack and McCormack, 1994; Jan and Jan, 1997).

Statistical analysis

The statistical analysis was intended to determine the following: the degree of homology among 1) the three symporter families, 2) the bacterial 2TM channels and the symporters, and 3) the four MPM motifs of each symporter family. As described in the Methods, the results are given as Z scores, which are the number of standard deviations the raw score is from the mean of best chance alignments for segments of the same length. The greater the Z score, the more similar the sequence profiles are, and the less likely the alignment is to occur by chance. As also described, the alignment of the bacterial 2TM channel and symporter motif blocks was the same as represented by the consensus sequences in Fig. 2, and the reported score is the highest of all possible subsegments of at least four contiguous residues. Furthermore, the linkers in each motif have been excluded because of their extreme variability, leaving the three primary segments (i.e., M1, P, and M2) to be treated individually.

For interpretation of these results it is important to consider that membrane proteins share some basic properties

independent of their evolutionary relationships. For example, our experience with this methodology suggests that comparison of any two transmembrane segments in which nonpolar residues predominate will result in a positive similarity score. Thus, to determine a baseline control of this effect, each segment block of the bacterial 2TM channels and symporters was compared to each of the seven transmembrane segments of an alignment of 19 bacteriorhodopsin homologs (Horn et al., 1998). This latter family was judged an ideal membrane protein control for the following reasons: 1) they are bacterial proteins, 2) the structure of one member of the family is known (i.e., bacteriorhodopsin), 3) they lack P segments and are unrelated to K⁺ channels, 4) they have multiple transmembrane segments, 5) there are numerous homologs, and 6) the transmembrane segments of the homologs can be aligned with little ambiguity.

The comparisons of the three symporter families, i.e., KtrB, TrkH, and Trk-euk, are shown in Table 1. Only motifs at the same positions in the sequences were compared, rather than considering all possible cross-terms of motifs from different gene duplications. A general measure of the similarity for each of the three family comparisons is obtained by simply taking the average of the 12 scores of each group. This results in the similar average Z scores of 7.6 and 7.9 for the KtrB versus TrkH and Trk-euk comparisons, respectively, and the relatively low value of 5.0 for the

TABLE 1 Statistical analysis of the similarity of the symporter families

Symporters	Segment		
	M1	P	M2
KtrB vs. TrkH			
Motif A	6.8	7.9	7.1
Motif B	<u>4.3</u>	7.3	8.4
Motif C	5.6	6.4	12.1
Motif D	7.9	7.7	9.8
KtrB vs. Trk-euk			
Motif A	6.1	10.3	6.2
Motif B	5.9	10.0	8.0
Motif C	6.8	10.9	6.5
Motif D	5.1	10.7	8.0
TrkH vs. Trk-euk			
Motif A	<u>4.9</u>	<u>7.5</u>	<u>0.8</u>
Motif B	5.0	<u>3.5</u>	<u>6.1</u>
Motif C	<u>3.5</u>	<u>6.1</u>	<u>3.5</u>
Motif D	<u>4.6</u>	<u>7.6</u>	<u>6.5</u>
Control: Mean of all symporter segments vs. all 7 TMs of bacteriorhodopsin homologs.			
	3.8	2.2	3.2

In Tables 1–3, the values are given as Z scores, which are the number of standard deviations the raw score is from the mean of a distribution of raw scores for segments of the same length calculated from a database of unrelated protein families (see Methods). The more positive the Z score, the more likely it is that the two compared families are homologous. Control values for transmembrane helices of unrelated proteins are given at the bottom. In Table 1, the highest score among the three families for each segment is indicated in bold, and the lowest is underlined and italicized.

TrkH versus Trk-euk comparison. Considering that the average for all of the comparisons with bacteriorhodopsin is 3.1, these results indicate statistically significant sequence similarities among almost all of the corresponding segments of the symporters. Furthermore, among the symporters, the fact that the Z scores are consistently lowest for the TrkH versus Trk-euk comparison supports the hypothesis that the KtrB family is more like the presumed common ancestor.

At greater detail, it is interesting that in some instances the degrees of similarity for the three families depend upon which motif segments are being compared. For example, when the KtrB family is compared to the Trk-euk family, the four P segments are conserved substantially better than are the other two segments. (This pattern is similar to that found when different families of K⁺ channels are compared as shown in Table 2.) In contrast, when the KtrB family is compared to the TrkH family, most of the M2 segments are conserved to a greater extent than are the P segments. When related to the three-dimensional structure of KcsA, this indicates that the structures of the Trk-euk proteins are more

similar to the KtrB proteins in the outer half of the transmembrane region (where the pore is formed by the P segments), and the structures of the TrkH proteins are more similar to the KtrB proteins at the inner half of the transmembrane region (where the pore is formed by the M2 segments). Overall, the M1 segments are found to have the least degree of conservation; the average of the four scores is 6.2 and 6.0 for the KtrB versus TrkH and Trk-euk comparisons, respectively, and 4.5 for the TrkH versus Trk-euk comparison. Again, this is consistent with the lesser structural role the M1 segment plays in forming the pore in the KcsA crystal structure.

Table 2 shows the results when the M1, P, and M2 segments of the bacterial 2TM K⁺ channels are compared with the proposed analogous segments of the symporters. These results support our hypotheses that the four putative MPM motifs of the symporters are related to the MPM motif of the bacterial K⁺ channels and that the KtrB family is closest to the presumed ancestor. Specifically, three of the four MPM motifs of the KtrB symporters are found closest to the single MPM motif of the channels, scoring substantially higher (at least 2.0 points) than the control comparisons with the bacteriorhodopsin segments. The exception is the MPM_B motif, which instead scores highest for the TrkH family. As is expected for the shift from prokaryotic to eukaryotic species, the Trk-euk family is clearly the most distant from the bacterial K⁺ channels, with only one segment scoring more than two points higher than the control. It is also seen that in general the evolutionary distance between the channels and symporters is larger than among the three symporter families themselves (Table 1). Using a simple measure, the 12-score averages in Table 2 for the similarities between the channels and symporters are 6.0, 5.4, and 4.5 for the KtrB, TrkH, and Trk-euk families, respectively. The only exception is the score for the TrkH versus Trk-euk symporter families (i.e., 5.0), which indicates a greater distance than that between the channels and the KtrB and TrkH families.

To provide further insight into the calculated evolutionary distances, the bacterial 2TM K⁺ channel sequences were compared to three other ion channel families. These were 1) the relatively similar TWIK or 2 × 2TM family of K⁺ channels from *C. elegans* (which has two consecutive MPM motifs per subunit), 2) the more distantly related IRK K⁺ channel family from eukaryotes, and 3) the homologous S5-P-S6 regions of the Na⁺ channel family (which have P segments selective for Na⁺ instead of K⁺). As expected, the scores of the P segments of the 2 × 2TM K⁺ channels were significantly closer to those of the bacterial 2TM channels than were those of the symporters; however, the scores for the M1 and M2 segments were about the same as for those of the KtrB family. Surprisingly, for the IRK family the M1 and M2 scores were about two points lower than the averages for the KtrB symporters, and the score for the ion-selective P segments was only slightly higher (i.e., 0.6 and 0.3 greater than the averages for the KtrB and TrkH families). Moreover, the scores for the analogous regions of the

TABLE 2 Statistical analysis of the similarity of 2TM bacterial K⁺ channel segments to the three symporter families, eukaryote 2×2TM and 2TM inwardly rectifying K⁺ channels, Na⁺ channels, and bacteriorhodopsin homologs

	Bacterial 2TM K ⁺ channel segments		
	M1	P	M2
Symporters			
KtrB A	6.8	5.9	5.8
KtrB B	5.5	<u>3.5</u>	5.0
KtrB C	8.1	5.4	6.3
KtrB D	7.4	5.1	7.7
TrkH A	6.5	5.3	<u>2.3</u>
TrkH B	6.2	5.4	5.8
TrkH C	5.6	5.3	4.9
TrkH D	7.2	5.3	5.0
Trk-euk A	<u>4.9</u>	4.5	<u>4.5</u>
Trk-euk B	5.2	<u>3.6</u>	5.5
Trk-euk C	6.4	4.5	<u>2.4</u>
Trk-euk D	<u>3.6</u>	<u>3.7</u>	4.8
Controls with other K⁺ channels			
2×2TM	6.5	10.6	6.3
MPM _A			
2×2TM	7.2	11.7	5.1
MPM _B			
Euk 2TM IRK	<u>4.5</u>	5.6	<u>4.7</u>
Controls with Na⁺ channel S5, P, and S6 segments			
Repeat I	<u>3.5</u>	<u>3.1</u>	<u>4.3</u>
Repeat II	5.1	<u>2.5</u>	<u>4.7</u>
Repeat III	<u>4.4</u>	<u>2.5</u>	<u>3.1</u>
Repeat IV	<u>4.5</u>	<u>2.0</u>	<u>4.3</u>
Controls with bacteriorhodopsin homologs			
Mean of 7 TMs	4.1	2.8	3.8

Bold indicates that the Z value is at least 2.0 greater than the control with bacteriorhodopsin homologs; underlined italics indicate that the Z score is less than 1.0 greater than the control.

four motifs of the Na⁺ channel family were on average no greater than those for the unrelated bacteriorhodopsin family. Despite the difference in P-segment ion selectivity, this is somewhat surprising, because the voltage-gated Na⁺ channels are thought to have evolved from voltage-gated Ca⁺² channels, which in turn are thought to have evolved from voltage-gated K⁺ channels (Strong et al. 1993). Thus the finding that the KtrB and TrkH families score substantially higher than do the distantly related IRK and Na⁺ channel families supports the hypothesis that the symporter and bacterial 2TM channel families are homologous.

Table 3 displays the calculated similarities of the four MPM motifs within each of the three symporter families individually. The 18-score averages from Table 3 are 6.8, 5.2, and 4.4 for the KtrB, TrkH, and Trk-euk families, respectively. Thus comparison with Table 2 indicates that the four symporter MPM motifs are almost as similar to the bacterial 2TM K⁺ channel MPM motifs as they are to each other. For example, the average score from Table 3 is 0.8 greater than that from Table 2 for the KtrB family, but is 0.2 and 0.1 smaller for the TrkH and Trk-euk families. As can

be seen by the pattern of bold numbers, all of the KtrB segments score substantially higher to each other than to the bacteriorhodopsin controls. This strongly supports the premise that the four MPM motifs are indeed homologous and are likely due to gene duplications. Although Table 3 indicates less similarity for the M1 and M2 segments of the other two symporter families, the strong case for mutual homology with KtrB seen in Table 1 supports the extension of this conclusion to the TrkH and Trk-euk proteins. In addition, the fact that the majority of the scores in Table 3 are highest for the KtrB family (15 of 18) is consistent with the hypothesis that this family is the closest to the common ancestor, because it indicates the least divergence of the four gene repeats. Likewise, the finding that 12 of the 18 scores are lowest for the eukaryotic Trk-euk family is consistent with it being the most divergent from the prokaryotic progenitor. Unfortunately, the pattern of conservation is not clear enough to predict the order of the motif duplications. That is, the pattern of high and low scores is not uniform among the three segments of the MPM motifs, nor is it uniform for the three families. For example, MPM_A and MPM_D are most similar in the KtrB family, but are the least similar for the TrkH and Trk-euk families.

TABLE 3 Statistical analysis of the similarity between the four motifs of the TrkH, KtrAB, and Trk-euk symporters individually

Symporters	Segment		
	M1	P	M2
KtrB			
Motifs			
A vs. B	6.4	5.8	8.0
A vs. C	7.3	5.7	5.9
A vs. D	7.4	7.4	9.9
B vs. C	6.9	6.9	6.1
B vs. D	6.8	6.4	6.0
C vs. D	8.7	4.2	7.4
TrkH			
Motifs			
A vs. B	5.6	5.7	<u>3.3</u>
A vs. C	4.3	6.3	4.5
A vs. D	4.8	5.6	<u>3.6</u>
B vs. C	<u>5.0</u>	5.5	5.9
B vs. D	6.4	5.4	6.3
C vs. D	5.4	4.9	5.7
Trk-euk			
Motifs			
A vs. B	6.9	4.8	4.9
A vs. C	<u>4.1</u>	6.4	<u>3.1</u>
A vs. D	<u>3.7</u>	<u>3.2</u>	<u>4.0</u>
B vs. C	6.0	5.1	<u>2.8</u>
B vs. D	<u>4.0</u>	4.8	<u>3.6</u>
C vs. D	<u>4.6</u>	5.3	<u>2.3</u>
Control: Mean of all symporter segments with all 7 TMs of bacteriorhodopsin homologs			
	3.8	2.2	3.2

Bold indicates that the Z value is at least 2.0 greater than the control with bacteriorhodopsin homologs; underlined italics indicate that the Z score is less than 1.0 greater than the control.

Evolutionary tree

Based on this analysis of the sequences, an evolutionary relationship between the different channel and symporter families is deduced as shown in Fig. 3. Specifically, a single prototype MPM transmembrane motif (*left*) underwent a fourfold gene duplication and gene fusion to form a K⁺ symporter protein ancestor (*center*). Furthermore, the cytoplasmic dinucleotide domain of the K⁺ channel ancestor may have split off to form a separate dinucleotide-binding subunit that associates with the symporters. Most members of the KtrB family (*right*) of eubacteria have remained similar to this ancestral protein. However, KtrB's from two *Mycoplasma* species contain additional extracellular domains between the M1 and P1 segments of the first three MPM motifs, and KtrB (NtpJ) from *Trepanoma pallidum* contains two additional transmembrane domains preceding the intracellular N-terminus (not shown). The TrkH family (*top*) in bacteria and archaeobacteria, which also has two additional transmembrane helices at the N-terminal (unique and different from those in *T. pallidum* KtrB), has diverged more than have most members of the KtrB family. The TrkA subunit probably underwent an internal gene duplication to produce two dinucleotide-binding domains. The Trk1,2 family in fungi (*bottom*) has diverged even more. Its members have an extra long cytoplasmic loop between MPM_A and MPM_B, and a smaller, linker-like insert between MPM_C and MPM_D. The two plant sequences (*bottom right*) are only slightly closer to the Trk1,2 sequences than to KtrB and should probably be considered a separate family. At present, the eukaryotic symporters are still not known to have a dinucleotide-binding subunit.

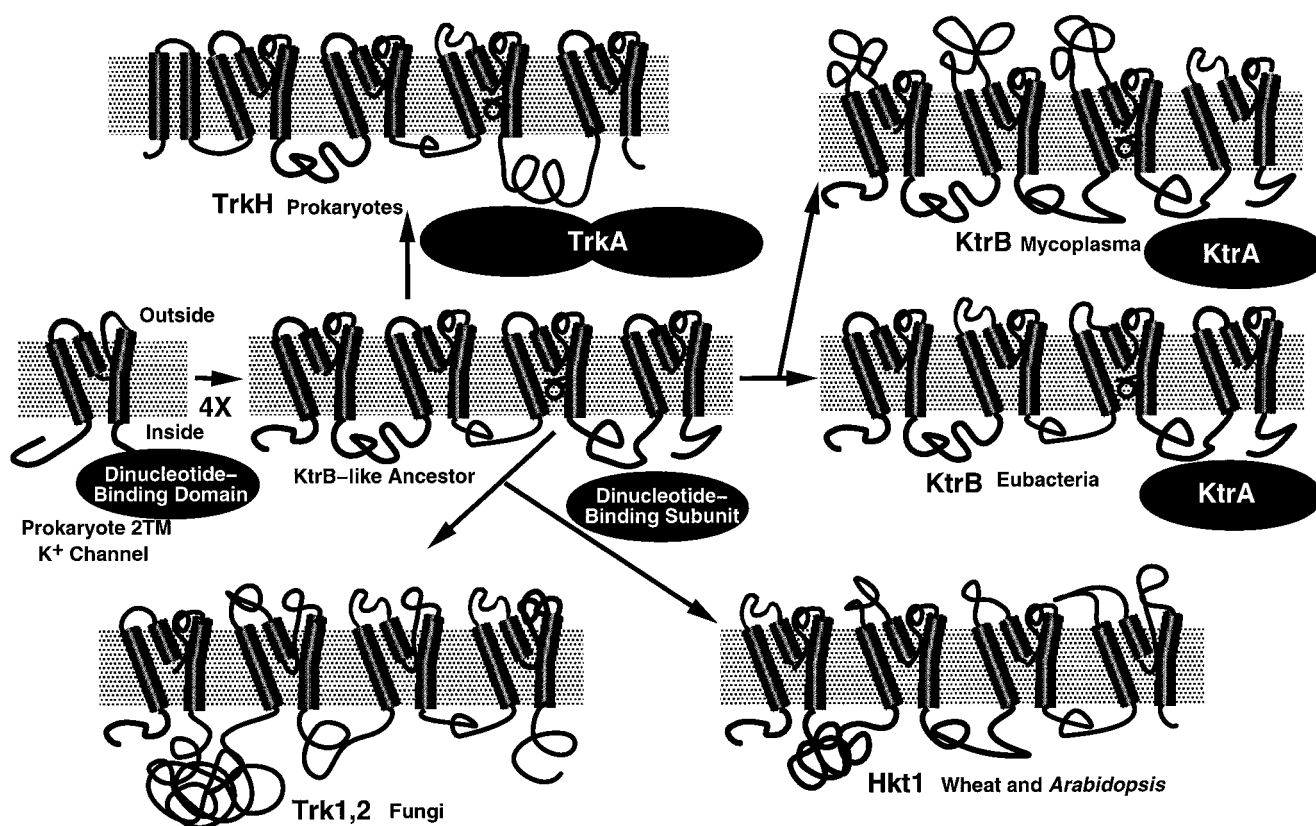


FIGURE 3 Proposed schematic for the evolutionary development of the K⁺-symporter families from the 2TM-type of prokaryote K⁺ channels. The four MPM motif-containing KtrB-like ancestor (*middle*) derives from a single MPM motif-containing K⁺ channel subunit (*left*) by gene duplication and fusion, and a dinucleotide-binding subunit derives from the cytoplasmic, C-terminal domain of the K⁺ channel. Further evolution leads to the current KtrAB symporter family (*right*), the prokaryote TrkH and TrkA subunits (with two dinucleotide-binding domains) (*top*), and the eukaryote Trk1,2 (fungi) and HKT1 (wheat and *Arabidopsis*) families (*bottom*). *Mycoplasma* KtrB's develop with longer segments linking the M1 and P1 segments in the first three MPM motifs. See the text for details.

CONCLUSIONS

Paleontologists often search for evidence of links between distantly related groups of organisms. For example, the discovery of a subgroup family of dinosaurs that have feathers can establish the evolutionary link with modern-day birds (Ji et al., 1998). Although there is no fossil record for molecular evolution, a similar method can be used to establish links of distantly related proteins: i.e., by determining subgroups that have intermediary sequences, structures, and/or functions. In this and the accompanying paper, it is argued that the bacterial KtrAB and 2TM K⁺ channel protein families serve such a function, in that they link the K⁺ channels with the distantly related K⁺ symporter proteins.

Although we believe the sequence comparison and model building methods presented in the accompanying and present papers can be generalized constructively to other protein systems, care must be taken to avoid certain pitfalls. For example, studies and intuition concur on the benefit of using profiles of families over individual sequences to identify the homology of distantly related proteins (Tatusov et al., 1994; Henikoff and Henikoff, 1996). Unfortunately, however, this is not an automatic procedure. Beyond selec-

tion of the specific profiling and comparison scoring methods, judgment is required in selecting the range of related sequences that make up each family group. Although it is obvious that a profile of nearly identical sequences does not contain much added information, it can also be detrimental to form a profile of too diverse a grouping (as might occur in a larger superfamily). For example, comparison of the KtrB symporter and bacterial 2TM K⁺ channel profiles convincingly indicates an evolutionary relationship between these two protein families. However, the results are considerably more tentative for the Trk-euk symporter family, in which the scores of the M1 and M2 segments are not very similar to those of the K⁺ channels, or even to themselves in the different MPM motif repeats. Likewise, comparisons of the symporters to distantly related families of K⁺ channels, such as the IRK family, indicate little similarity (data not reported). Thus a profile that combined all of the symport families and/or that combined all of the K⁺ channel families would result in a weaker similarity score than that of KtrB versus bacterial 2TM K⁺ channels. This could lead to the erroneous conclusion that the symporter and channel proteins are not homologous. Rather, the case for the Trk-euk symporters being related to the channels comes indi-

rectly through the strong score similarity that its P segment profiles have with the KtrB family (Table 1). The observation that the M1-P-M2 segments of bacterial 2TM K^+ channels score no better with Na^+ channel S5-P-S6 segments than they do with transmembrane segments of bacteriorhodopsin homologs suggests that this procedure is unable to detect distant homology for protein families in which the primary functional property (in this case ion selectivity of P segment) has changed.

It is important to note that there are other shared sequence properties between the bacterial 2TM K^+ channel and symporter families indicative of an evolutionary relationship that are not quantified by the calculations presented here. For example, although the statistical analysis strongly suggests that the four MPM motifs of the KtrB symporters are homologous to each other as well as to that of the bacterial 2TM K^+ channels, it does not take into account that the three constituent segments (i.e., M1, P, and M2) are always in the same order. Furthermore, no score is provided for the probability of finding the same number of MPM motifs in the symporter sequences as there are single-motif subunits in the channels. Similarly, no quantification is made for finding the similar patterns of residue conservation and polarity among the MPM motifs of the channels and symporters: e.g., the P segments are the most well conserved, whereas the M1 segments are the least well conserved. And finally, the statistical analysis also does not take into account that several highly conserved residues known to be functionally important in the channel proteins (most notably the glycines of the P2 segment responsible for ion selection) are also highly conserved in each of the four MPM motifs of the symporters. In the accompanying paper it is shown how these properties justify building 3D atomic-scale models for the three symporter families in which the four MTM motifs each have the same general fold as the single KcsA K^+ channel subunit seen in the crystal structure.

An essential note of caution is that the data used for analysis in this paper are mostly from recently determined

nucleic acid sequences. In only a few cases have experiments already been conducted to establish that the encoded proteins are actually expressed and that they have channel or symporter functions as predicted. This is particularly true for the putative 2TM bacterial K^+ channels that contain C-terminal dinucleotide-binding domains. At present, there are no published data demonstrating that these specific genes encode functional channels rather than other types of transport proteins.

An important question is whether the transmembrane topology proposed here carries over to other families of transporters. Unfortunately, the simple method of constructing a hydropathy plot to predict the transmembrane topologies of these proteins is not very reliable and is not designed to identify P segments. To date, the transporter proteins that have been studied most extensively do not appear to have P segments. For example, the lactose permease protein has been experimentally determined to have 12 fully transmembrane segments (Lee and Manoil, 1996). Likewise, cryoelectron microscopy studies have indicated that the H^+ (Auer et al., 1998) and Ca^{+2} (Zhang et al., 1998) P-type pumps each have 10 fully transmembrane segments. In addition, the Kef proteins appear to form a different class of bacterial K^+ channels that lack the classic K^+ channel P-segment "signature sequence," but which do appear to have a dinucleotide-binding C-terminus (Booth et al., 1996). In fact, their transmembrane sequences appear to be more similar to those of the NapA Na^+/H^+ antiporters than to the channels (Reizer et al., 1992).

We thank Clifford Slayman for many helpful comments and assistance. Some preliminary sequences sequence data were obtained from the Institute for Genomic Research website at <http://www.tigr.org> and the NCBI website at <http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html>.

The work in Osnabrück was supported by the Deutsche Forschungsgemeinschaft (SFB171) and the Fonds der Chemischen Industrie.

APPENDIX

Sources of sequences for Fig. 2, A and B

	Accession no.	DBSOURCE:	Locus, Accession
Prokaryote 2TM K^+ channels			
Organism			
<i>Archaeoglobus fulgidus</i>	2648884	GENBANK:	AE000988, AE000988
<i>Archaeoglobus fulgidus</i>	2649899	GENBANK:	AE001055, AE001055
<i>Aquifex aeolicus</i>	2983007	GENBANK:	AE000683, AE000683
<i>Bacillus caldotenax</i>	39429	EMBL:	BCLCTB, X05066
<i>Bacillus stearothermophilus</i>	39974	EMBL:	BSLCTB, X05067
<i>Bacillus subtilis</i>	1934805	EMBL:	BSZ93936, Z93936
<i>Bacillus subtilis</i>	2522014	DDBJ:	AB007638, AB007638
<i>Helicobacter pylori</i>	2313603	GENBANK:	HPAE000564, AE000564
<i>Streptomyces coelicolor</i>	2808791	EMBL:	SC7H1, AL021411
<i>Streptomyces coelicolor</i>	3413418	EMBL:	SC10H5, AL031232
<i>Streptomyces lividans</i>	2127577	PIR:	S60172
<i>Methanobacterium thermoautotrophicum</i>	2621577	GENBANK:	AE000834, AE000834
<i>Methanobacterium thermoautotrophicum</i>	2622639	GENBANK:	AE000912, AE000912

	Accession no.	DBSOURCE:	Locus, Accession
<i>Methanococcus jannaschii</i>	2493595	SW-PROT:	Y13B_METJA, Q57604
<i>Methanococcus jannaschii</i>	2493596	SWIPROT:	YD57_METJA, Q58752
<i>Mycobacterium tuberculosis</i>	2827610	EMBL:	MTV014, AL021646
<i>Pyrococcus horikoshii</i>	3131937	DDBJ:	AB009522, AB009522
<i>Synechocystis sp.</i>	1652235	DDBJ:	D90904, D90904
<i>Synechocystis sp.</i>	1652933	DDBJ:	D90909, D90909
Unfinished genomes		Source contig	
<i>Chlorobium tepidum</i>		gnl TIGR C.tepidum_292	
<i>Deinococcus radiodurans</i>		gnl TIGR gdr_171	
<i>Deinococcus radiodurans</i>		gnl TIGR gdr_52	
<i>Pseudomonas aeruginosa</i>		gnl PAGP Contig502	
<i>Pseudomonas aeruginosa</i>		gnl PAGP Contig536	
<i>Pyrococcus furiosus</i>		gnl UCHGR MM64-00907 00907	
<i>Thermotoga maritima</i>		gnl TIGR BTMDS92R	
<i>Vibrio cholerae</i>		gnl TIGR GVCCZ38F	
Prokaryote TrkH			
Organism			
<i>Archaeoglobus fulgidus</i>	2649764	GENBANK:	AE001046, AE001046
<i>Escherichia coli</i>	136239	SW-PROT:	TRKG_ECOLI, P23849
<i>Escherichia coli</i>	1174773	SW-PROT:	TRKH_ECOLI, P21166
<i>Haemophilus influenzae</i>	3212204	GENBANK:	U32755, U32755
<i>Methanobacterium thermoautotrophicum</i>	2622378	GENBANK:	AE000893, AE000893
<i>Methanococcus jannaschii</i>	2129329	PIR:	D64485 gi 2129329
<i>Pyrococcus horikoshii</i>	3132095	DDBJ:	AB009526, AB009526
<i>Thermoanaerobacter ethanolicus</i>	2581795	GENBANK:	AF001974, AF001974
<i>Vibrio alginolyticus</i>	3288671	DDBJ:	D86411, D86411
Unfinished genome		Source contig	
<i>Actinobacillus actinomycetemcomitans</i>		gnl OUACGT A.actin_Contig574	
<i>Neisseria gonorrhoeae</i>		gnl OUACGT Contig202	
<i>Neisseria meningitidis</i>		gnl Sanger Contig395	
<i>Porphyromonas gingivalis W83</i>		gnl TIGR P.gingivalis_107	
<i>Pseudomonas aeruginosa</i>		gnl PAGP Contig284	
<i>Pyrococcus furiosus</i>		gnl UCHGR MM1-MM1 02861	
<i>Streptococcus pneumoniae</i>		gnl TIGR stp_4251	
<i>Yersinia pestis</i>		gnl Sanger Contig1193.0	
Bacterial KtrB			
Organism			
<i>Aquifex aeolicus</i>	2983882	GENBANK:	AE000743, AE000743
<i>Bacillus subtilis</i>	2633721	EMBL:	BSUB0008, Z99111
<i>Bacillus subtilis</i>	2635594	EMBL:	BSUB0016, Z99119
<i>Borrelia burgdorferi</i>	2688660	GENBANK:	AE001172, AE001172
<i>Enterococcus hirae</i>	1075716	PIR:	G53610
<i>Enterococcus hirae</i>	1171790	SW-PROT:	NTPJ_ENTHR, P43440
<i>Mycoplasma genitalium</i>	1361729	PIR:	F64235
<i>Mycoplasma pneumoniae</i>	2496386	SW-PROT:	Y322_MYCPN, P75323
<i>Synechocystis sp.</i>	1653100	DDBJ:	D90911, D90911
<i>Treponema pallidum</i>	3322404	GENBANK:	AE001199, AE001199
<i>Vibrio alginolyticus</i>	3288679	DDBJ:	D89592, D89592
Unfinished genome		Source contig	
<i>Campylobacter jejuni</i>		gnl Sanger campylo_Cj.seq	
<i>Deinococcus radiodurans</i>		gnl TIGR gdr_54	
<i>Deinococcus radiodurans</i>		gnl TIGR gdr_54	
<i>Enterococcus faecalis</i>		gnl TIGR gef_6231	
<i>Enterococcus faecalis</i>		gnl TIGR gef_6205	
<i>Streptococcus pneumoniae</i>		gnl TIGR stp_4172	
<i>Streptococcus pyogenes</i>		gnl OUACGT Contig142	
Eukaryotic Trk-euk			
Fungi (Trk1, 2)			
Organism			
<i>Neurospora crassa</i>	3724137	EMBL:	NCR9758, AJ009758
<i>Saccharomyces bayanus</i>	136230	SW-PROT:	TRK1_SACUV, P28569
<i>Saccharomyces cerevisiae</i>	136231	SW-PROT:	TRK1_YEAST, P12685

	Accession no.	DBSOURCE:	Locus, Accession
<i>Saccharomyces cerevisiae</i>	136232	SW-PROT:	TRK2_YEAST, P28584
<i>Schizosaccharomyces pombe</i>	1717749	SW-PROT:	TRK_SCHPO, P47946
<i>Schizosaccharomyces pombe</i>	1076908	PIR:	S50225
<i>Schizosaccharomyces pombe</i>	1351299	SW-PROT:	TRKH_SCHPO, Q10065
Plant (HKT1)			
Organism			
<i>Triticum aestivum</i>	567062	GENBANK:	TAU16709, U16709
<i>Arabidopsis thaliana</i> *	CAB39784	EMBL:	ATF24G24, AL049488.1
<i>Arabidopsis thaliana</i> *	3695407	GENBANK:	T9A4, AF096373

*The deduced protein sequences differed in the data bases for this protein because of different assignments of the exon boundaries. We used the exon boundaries that produced the sequence most similar to that of HKT1 from *Triticum aestivum*. This corresponds to the CAB39784 sequence for all but the last exon, for which the sequence from 3695407 was used. Using these boundaries, the deduced protein sequences are identical from both sources.

Sources of sequences of dinucleotide-binding domains for Fig. 2 C

	Accession no.	DBSOURCE:	Locus, Accession
K ⁺ channels			
Organism			
<i>Archaeoglobus fulgidus</i>	2648884	GENBANK:	AE000988, AE000988
<i>Methanococcus jannaschii</i>	2493595	SW-PROT:	Y13B_METJA, Q57604
<i>Methanococcus jannaschii</i>	2493596	SW-PROT:	YD57_METJA, Q58752
<i>Mycobacterium tuberculosis</i>	2827610	EMBL:	MTV014,
<i>Synechocystis</i> sp.	1652235	DDBJ:	D90904, D90904
<i>Synechocystis</i> sp.	1652933	DDBJ:	D90909, D90909
Unfinished genome		Source contig	
<i>Deinococcus radiodurans</i>		gnl TIGR gdr_171	
KtrA/TrkA			
Organism			
<i>Archaeoglobus fulgidus</i>	2649897	GENBANK:	AE001055, AE001055
<i>Aquifex aeolicus</i>	2983887	GENBANK:	AE000743, AE000743
<i>Bacillus subtilis</i>	2635593	EMBL:	BSUB0016, Z99119
<i>Bacillus subtilis</i>	732401	SW-PROT:	YKQB_BACSU, P39760
<i>Borrelia burgdorferi</i>	2688659	GENBANK:	AE001172, AE001172
<i>Methanobacterium thermoautotrophicum</i>	2622379	GENBANK:	AE000893, AE000893
<i>Methanococcus jannaschii</i>	2129328	PIR:	H64437
<i>Methanococcus jannaschii</i>	1591749	GENBANK:	U67553, U67553
<i>Mycobacterium tuberculosis</i>	2181983	EMBL:	MTCY05A6, Z96072
<i>Mycobacterium tuberculosis</i>	2181982	EMBL:	MTCY05A6, Z96072
<i>Mycoplasma genitalium</i>	1351544	SW-PROT:	Y323_MYCGE, P47565
<i>Mycoplasma pneumoniae</i>	2496387	SW-PROT:	Y323_MYCPN, P75322
<i>Streptomyces coelicolor</i>	2815361	EMBL:	SC2E9, AL021530
<i>Synechocystis</i> sp.	1001490	DDBJ:	SYCSLRB, D64000
<i>Synechocystis</i> sp.	1001359	DDBJ:	SYCSLLH,
<i>Thermoanaerobacter ethanolicus</i>	2581796	GENBANK:	AF001974, AF001974
<i>Vibrio alginolyticus</i>	3288678	DDBJ:	D89592, D89592

REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Anderson, M. P., R. J. Gregory, S. Thompson, D. W. Souza, S. Paul, R. C. Mulligan, A. E. Smith, and M. J. Welsh. 1991. Demonstration that CFTR is a chloride channel by alteration of its anion selectivity. *Science*. 253:202–205.
- Ashcroft, F. M., and F. M. Gribble. 1998. Correlating structure and function in ATP-sensitive K⁺ channels. *Trends Neurosci.* 21(7):288–294.
- Auer, M., G. A. Scarborough, and W. Kuhlbrandt. 1998. Three-dimensional map of the plasma membrane H⁺-ATPase in the open conformation. *Nature*. 392:840–843.
- Bairoch, A., and R. Apweiler. 1998. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Res.* 26:38–42.
- Bevington, P. R. 1969. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York.
- Booth, I. R., M. Jones, D. McLaggan, Y. Nikolaev, L. Ness, C. Wood, S. Miller, S. Töttemeyer, and G. Ferguson. 1996. Bacterial ion channels. In *Transport Processes in Eukaryotic and Prokaryotic Organisms*. W. N. Konings, H. R. Kaback, and J. S. Lolkema, editors. Elsevier, New York. 693–729.
- Bossemeyer, D., A. Borchard, D. C. Dosch, G. C. Helmer, W. Epstein, I. R. Booth, and E. P. Bakker. 1989. K⁺-transport protein TrkA of *Escherichia coli* is a peripheral membrane protein that requires other trk gene products for attachment to the cytoplasmic membrane. *J. Biol. Chem.* 264:16403–16410.
- Clayton, R. A., O. White, K. A. Ketchum, and J. C. Venter. 1997. The first genome from the third domain of life. *Nature*. 387:459–462.
- Deckert, G., P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen, and R. V. Swanson.

1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature*. 392(6674):353–358.
- Diatloff, E., R. Kumar, and D. P. Schachtman. 1998. Site directed mutagenesis reduces the Na⁺ affinity of HKT1, an Na⁺ energized high affinity K⁺ transporter. *FEBS Lett.* 432:31–36.
- Dosch, D. C., G. L. Helmer, S. H. Sutton, F. F. Salvacion, and W. Epstein. 1991. Genetic analysis of potassium transport loci in *Escherichia coli*: evidence for three constitutive systems mediating uptake potassium. *J. Bacteriol.* 173:687–696.
- Doyle, D. A., J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon. 1998. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science*. 280:69–77.
- Durell, S. R., and H. R. Guy. 1999. Structural models of the KtrB, TrkH and Trk1,2 symporters based on the structure of the KcsA K⁺ channel. *Biophys. J.* 77:785–803.
- Gaber, R. F., C. A. Styles, and G. R. Fink. 1988. TRK1 encodes a plasma membrane protein required for high-affinity potassium transport in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8:2848–2859.
- Galli, A., R. D. Blakely, and L. J. DeFelice. 1998. Patch-clamp and amperometric recordings from norepinephrine transporters: channel activity and voltage-dependent uptake. *Proc. Natl. Acad. Sci. USA*. 95: 13260–13265.
- Guy, H. R., and S. R. Durell. 1995. Structural model of Na⁺, Ca²⁺, and K⁺ channels. In *Ion Channels and Genetic Diseases*. D. Dawson, editor. The Rockefeller University Press, New York. 1–16.
- Haro, R., L. Sainz, F. Rubio, and A. Rodriguez-Navarro. 1999. Cloning of two genes encoding potassium transporters in *Neurospora crassa* and expression on the corresponding cDNAs in *Saccharomyces cerevisiae*. *Mol. Microbiol.* 31:511–520.
- Henikoff, J. G., and S. Henikoff. 1996. Using substitution probabilities to improve position-specific scoring matrices. *Comput. Appl. Biosci.* 12: 135–143.
- Henikoff, S., and J. G. Henikoff. 1991. Automated assembly of protein blocks for database searching. *Nucleic Acids Res.* 19:6565–6572.
- Henikoff, S., and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA*. 89:10915–10919.
- Henikoff, S., and J. G. Henikoff. 1994. Position-based sequence weights. *J. Mol. Biol.* 243:574–578.
- Hofmann, K., and W. Stoffel. 1993. TMbase—a database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler*. 347:166.
- Horn, F., J. Weare, M. W. Beukers, S. Hörsch, A. Bairoch, W. Chen, Ø. Edvardsen, F. Campagne, and G. Vriend. 1998. GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.* 26: 277–281.
- Jan, L. Y., and Y. N. Jan. 1994. Potassium channels and their evolving gates. *Nature*. 371:119–122.
- Jan, L. Y., and Y. N. Jan. 1997. Cloned potassium channels from eukaryotes and prokaryotes. *Annu. Rev. Neurosci.* 20:91–123.
- Jardetzky, O. 1966. Simple allosteric model for membrane pumps. *Nature*. 211:969–970.
- Ji, O., P. J. Currie, M. A. Norell, and S.-A. Ji. 1998. Two feathered dinosaurs from northeastern China. *Nature*. 393:753–761.
- Kawarabayashi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3 (supplement). *DNA Res.* 5:147–155.
- Ketchum, K. A., W. J. Joiner, A. J. Sellers, L. K. Kaczmarek, and S. A. Goldstein. 1995. A new family of outwardly rectifying potassium channel proteins with two pore domains in tandem. *Nature*. 376:690–695.
- Ko, C. H., and R. F. Gaber. 1991. TRK1 and TRK2 encode structurally related K⁺ transporters in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 11:4266–4273.
- Larsson, H. P., S. A. Picard, F. S. Werblin, and H. Lecar. 1996. Noise analysis of the glutamate-activated current in photoreceptors. *Biophys. J.* 70:733–742.
- Lauger, P. 1979. A channel mechanism for electrogenic ion pumps. *Biochim. Biophys. Acta*. 552:143–161.
- Lee, M., and C. Manoil. 1996. Molecular genetic analysis of membrane protein topology. In *Transport Processes in Eukaryotic and Prokaryotic Organisms*. W. N. Konings, H. R. Kaback, and J. S. Lolkema, editors. Elsevier, New York. 189–202.
- Lesage, F., R. Reyes, M. Fink, F. Duprat, E. Guillemare, and M. Lazdunski. 1996. Dimerization of TWIK-1 K⁺ channel subunits via a disulfide bridge. *EMBO J.* 15:6400–6407.
- Lichtenberg-Fraté, H., J. D. Reid, M. Heyer, and M. Hofer. 1996. The SpTRK gene encodes a potassium-specific transport protein TKHp in *Schizosaccharomyces pombe*. *J. Membr. Biol.* 152:169–181.
- Lü, Q., and C. Miller. 1995. Silver as a probe of pore-forming residues in a potassium channel. *Science*. 268:304–307.
- McCormack, T., and K. McCormack. 1994. Shaker K⁺ channel β subunits belong to an NAD(P)H-dependent oxidoreductase superfamily. *Cell*. 79:1133–1135.
- Munro, A. W., G. Y. Ritchie, A. J. Lamb, R. M. Douglas, and I. R. Booth. 1991. The cloning and DNA sequence of the gene for the glutathione-regulated potassium-efflux system KefC of *Escherichia coli*. *Mol. Microbiol.* 5:607–616.
- Nakamura, T., N. Yamamuro, S. Stumpe, T. Unemoto, and E. P. Bakker. 1998a. Cloning of the trkAH gene cluster and characterization of the Trk K(+) uptake system of *Vibrio alginolyticus*. *Microbiology*. 144: 2281–2289.
- Nakamura, T., R. Yuda, T. Unemoto, and E. P. Bakker. 1998b. KtrAB, a new type of bacterial K⁺-uptake system from *Vibrio alginolyticus*. *J. Bacteriol.* 180:3491–3494.
- Noda, M., S. Shimizu, T. Tanabe, T. Takai, T. Kayano, T. Ikeda, H. Takahashi, H. Nakayama, Y. Kanaoka, and N. Minamino, K. Kangawa, H. Matsuo, M. A. Raftery, T. Hirose, S. Inayama, H. Hayashida, T. Miyata, and S. Numa. 1984. Primary structure of *Electrophorus electricus* sodium channel deduced from cDNA sequence. *Nature*. 312: 121–127.
- Parra-Lopez, C., R. Lin, A. Aspedon, and E. A. Groisman. 1994. A Salmonella protein that is required for resistance to antimicrobial peptides and transport of potassium. *EMBO J.* 13:3964–3972.
- Perozo, E., D. M. Cortes, and L. G. Cuello. 1998. Three-dimensional architecture and gating mechanism of a K⁺ channel studied by EPR spectroscopy. *Nature Struct. Biol.* 5:459–469.
- Perozo, E., D. M. Cortes, and L. G. Cuello. 1999. Structural rearrangements underlying activation gating in the *Streptomyces* K⁺ channel. *Biophys. J.* 76:A149.
- Petrokovski, S. 1996. Searching databases of conserved sequence regions by aligning protein multiple-alignments. *Nucleic Acids Res.* 24: 3836–3845.
- Reid, J. D., W. Lukas, R. Shafaatian, A. Bertl, C. Scheurmann-Kettner, H. R. Guy, and R. A. North. 1996. The *S. cerevisiae* outwardly-rectifying potassium channel (DUK1) identifies a new family of channels with duplicated pore domains. *Receptors Channels*. 4:51–62.
- 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*. *FEMS Microbiol. Lett.* 73:161–163.
- Rubio, F., W. Gassmann, and J. I. Schroeder. 1995. Sodium-driven potassium uptake by the plant potassium transporter HKT1 and mutations conferring salt tolerance. *Science*. 270:1660–1663.
- Schachtman, D. P., and J. I. Schroeder. 1994. Structure and transport mechanism of a high-affinity potassium uptake transporter from higher plants. *Nature*. 370:655–658.
- Schlösser, A., A. Hamann, D. Bossemeyer, E. Schneider, and E. P. Bakker. 1993. NAD⁺ binding to the *Escherichia coli* K(+) uptake protein TrkA and sequence similarity between TrkA and domains of a family of dehydrogenases suggest a role for NAD⁺ in bacterial transport. *Mol. Microbiol.* 9:533–543.
- Schlösser, A., S. Klüttig, A. Hamann, and E. P. Bakker. 1991. Subcloning, nucleotide sequence, and expression of trkG, a gene that encodes an integral membrane protein involved in potassium uptake via the Trk system of *Escherichia coli*. *J. Bacteriol.* 173:3170–3176.
- Schlösser, A., M. Meldorf, S. Stumpe, E. P. Bakker, and W. Epstein. 1995. TrkH and its homolog, TrkG, determine the specificity and kinetics of

- cation transport by the Trk system of *Escherichia coli*. *J. Bacteriol.* 177:1908–1910.
- Shih, T. M., and A. L. Goldin. 1997. Topology of the *Shaker* potassium channel probed with hydrophilic epitope insertions. *J. Cell Biol.* 136: 1037–1045.
- Strong, M., K. G. Chandy, and G. A. Gutman. 1993. Molecular evolution of voltage-sensitive ion channel genes: on the origins of electrical excitability. *Mol. Biol. Evol.* 10:221–242.
- Stumpe, A., A. Schlösser, M. Schleyer, and E. P. Bakker. 1996. K⁺ circulation across the prokaryotic cell membrane: K⁺-uptake systems. In *Transport Processes In Eukaryotic and Prokaryotic Organisms*. W. N. Konings, H. R. Kaback, and J. S. Lolkema, editors. Elsevier, New York. 473–499.
- Su, A., S. Mager, S. L. Mayo, and H. A. Lester. 1996. A multi-substrate single-file model for ion-coupled transporters. *Biophys. J.* 70:762–777.
- Takase, K., S. Kakinuma, I. Yamato, K. Konishi, K. Igarashi, and Y. Kakinuma. 1994. Sequencing and characterization of the ntp gene cluster for vacuolar-type Na(+)-translocating ATPase of *Enterococcus hirae*. *J. Biol. Chem.* 269:11037–11044.
- Tatusov, R. L., S. F. Altschul, and E. V. Koonin. 1994. Detection of conserved segments in proteins: iterative scanning of sequence databases with alignment blocks. *Proc. Natl. Acad. Sci. USA.* 91:12091–12095.
- Tholema, N., E. P. Bakker, A. Suzuki, and T. Nakamura. 1999. Change to alanine of one out of four selectivity filter glycines in KtrB causes a two orders of magnitude decrease in the affinities for both K⁺ and Na⁺ of the Na⁺ dependent K⁺-uptake system KtrAB from *Vibrio alginolyticus*. *FEBS Lett.* 450:217–220.
- Uozumi, N., T. Nakamura, J. I. Schroeder, and S. Muto. 1998. Determination of transmembrane topology of an inward-rectifying potassium channel from *Arabidopsis thaliana* based on functional expression in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* 95:9773–9778.
- Wang, T. B., W. Gassmann, F. Rubio, J. I. Schroeder, and A. D. Glass. 1998. Rapid up-regulation of HKT1, a high-affinity potassium transporter gene, in roots of barley and wheat following withdrawal of potassium. *Plant Physiol.* 118:651–659.
- Zhang, P., C. Toyoshima, K. Yonekura, N. M. Green, and D. L. Stokes. 1998. Structure of the calcium pump from sarcoplasmic reticulum at 8-Å resolution. *Nature.* 392:835–839.