J. Membrane Biol. 159, 1–20 (1997)

The Journal of
Membrane
Biology

Topical Review

Membrane Topology Motifs in the SGLT Cotransporter Family

E. Turk, E.M. Wright
Department of Physiology, CHS Box 951751, UCLA School of Medicine, Los Angeles, CA 90095-1751, USA

Received: 18 May 1996

Abstract. Homologues of the Na^+/glucose cotransporter, the SGLT family, include sequences of mammalian, eubacterial, yeast, insect and nematode origin. The cotransported substrates are sugars, inositol, proline, pantothenate, iodide, urea and undetermined solutes. It is reasonable to expect that the SGLT family members share a similar or identical topology of membrane spanning elements, by virtue of their common ancestry and similar coupling of solute transport to downhill sodium flux. Here we examine their membrane topologies as deduced from diverse analyses of their primary sequences, and from their sequence correlations with the experimentally determined topology of the human Na^+/glucose cotransporter SGLT1. Our analyses indicate that all family members share a common core of 13 transmembrane helices, but that some, like SGLT1 itself, have one additional span appended to the C-terminus, and still others, two. One bacterial member incorporates an additional span at the N-terminus. Sequence comparisons indicative of common ancestry of the SGLT and the [Na^+ + Cl^-] transporter families are introduced, and evaluated in light of their topologies. New evidence concerning the previously asserted common ancestry of SGLT1 and an N-acetylglucosamine permease of the bacterial phosphotransferase system is considered. Finally, we analyze observations which lead us to conjecture that the experimental strategy most commonly employed to reveal the topology of bacterial transporters (i.e., the fusion of reporter enzymes such as phoA alkaline phosphatase, beta-lactamase or beta-galactosidase, to progressively C-truncated fragments of the transporter) has often instead so perturbed local topology as to have entirely missed pairs of adjacent membrane spans.

Key words: Phylogenetic analysis — Membrane insertion — Helical hairpin — Hydrophobicity — PredictProtein — Memsat

Introduction

The intestinal Na^+/glucose cotransporter SGLT1 was expression-cloned in 1987, and found to be the first member of a new gene family [22]. Over the past 10 years more than 30 members from bacteria, yeast, invertebrates and vertebrates have come to light by cDNA and genomic sequencing. Information on the evolutionary relationships of these proteins and the orientation of their secondary structure within the membrane, or membrane topology, can provide essential starting points in understanding the structure/function relationships of these important transporters. In this review we consider the SGLT1 homologues, the SGLT family, that have been identified to date in many species. We focus on their phylogeny and membrane topology. Unlike many transporter families comprised of members with 12 transmembrane helices, such as the major facilitator superfamily [38] that includes the mammalian facilitated sugar transporters and the bacterial H^+/sugar cotransporters (symporters), the SGLT members share a common core of 13 transmembrane helices. In addition to their core structure, the SGLT family members exhibit considerable functional similarity despite the fact that the transported solutes range from sugars to inorganic anions. Sequence and topological analyses provide evidence of a common ancestry for SGLT and a neurotransmitter

Correspondence to: E. Turk

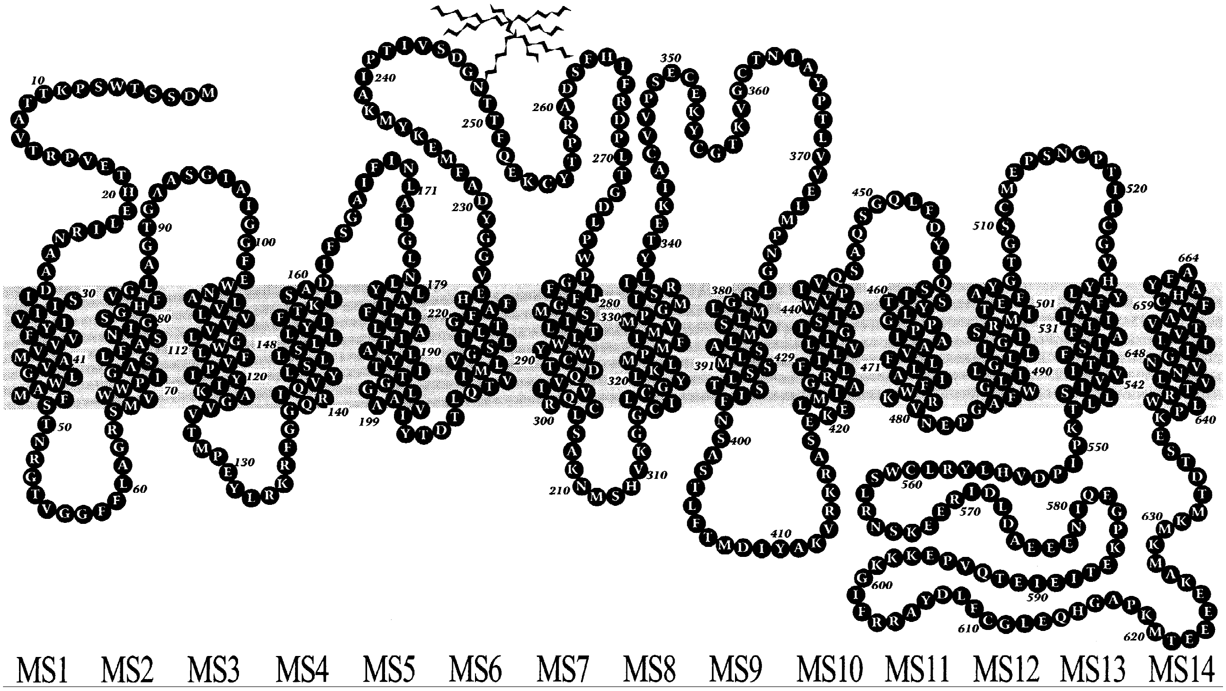


Fig. 1. Model of the membrane topology of the human Na^+/glucose cotransporter SGLT1 as determined experimentally and refined by prediction of the approximate transmembrane helix ends by analysis of interfacial hydrophobicity and reverse-turn propensity [64]. The Asn-linked carbohydrate tree is extracellular.

transporter family which includes the [Na^+ + Cl^-]-coupled GABA transporter, and for SGLT and a bacterial phosphotransferase system permease family. Finally we comment on the current methodologies used to examine the membrane structure of transport proteins. A full understanding of the topology of cotransporters is an essential prerequisite for planning definitive experiments to identify crucial residues in the transport pathway, for determining the helical packing of the transmembrane spans, and for elucidating the structural basis of coupling the transport of ion and substrate.

Topology of SGLT1

Experimental and computational analyses indicate the 14-transmembrane-span model of the human cotransporter SGLT1 shown in Fig. 1. All transmembrane spans are presumably alpha-helical [15, 26], the N-terminus is extracellular, the large highly charged C-terminal domain is cytoplasmic, and the hydrophobic C-terminus forms the 14th transmembrane span. This structural picture of SGLT1 was revealed primarily by N-glycosylation scanning mutagenesis, but essential information on the N-terminal region was supplied by functional expression in *Xenopus* oocytes, and experiments in vitro on the membrane insertion of the translation products of truncated mRNAs [64]. Membrane span prediction by two sophisticated computer algorithms also supported the 14-span model, and provided independent checks of the experimental results. Other evidence supporting this topology includes the localization of the large hydrophilic loop between spans 13 and 14 to the cytoplasm by immunogold electron microscopy ([60], also E.M. Cornford, E.M. Wright, *unpublished*). The C-terminus itself of the bacterial homologue putP, a proline cotransporter, was found by antibody recognition to be cytoplasmic [30], and this is consistent with SGLT1 topology since putP lacks the C-terminal fourteenth membrane span (*see Membrane Topology Motifs*).

We have recently constructed an SGLT1 mutant, bearing a Cys residue near the N-terminus, which expresses full functionality in oocytes. This Cys was found to be uniquely labelled with impermeant rhodamine maleimide in comparison with oocytes expressing wild type SGLT1, which also supports the extracellular localization of the N-terminus (B.H. Hirayama, D. Loo, E. Turk, E.M. Wright, *unpublished*). Rhodamine labelling of another cysteine mutant, Arg457Cys, is consistent with the extracellular disposition of the loop between spans 10 and 11 (D. Loo, B.H. Hirayama, E.M. Wright, *unpublished*).

Elucidation of SGLT1 membrane topology was not without difficulty. Only two N-glycosylation scanning mutants of the first set were glycosylated, because most consensuses translocated to the ER lumen were insuffi-

Table 1. SGLT cotransporter family members

Gene, or abbreviation	Organism	Transported solute	Size	Accession # and database
Eukaryotes:				
SGLT1	*Homo sapiens*	Glucose, galactose	664	P13866 sp
SGLT1	*Rattus norvegicus*	Glucose, galactose	665	P53790 sp
SGLT1	*Mus musculus*	Glucose, galactose	665	[a]
SGLT1	*Sus scrofa*	Glucose, galactose	662	P26429 sp[a]
SGLT1	*Oryctolagus cuniculus*	Glucose, galactose	662	P11170 sp
SGLT1	*Ovis aries*	Glucose, galactose	664	P53791 sp
SGLT2	*Sus scrofa*	Glucose	660	P31636 sp
SNST1	*Oryctolagus cuniculus*	Nucleoside/glucose(?)	672	P26430 sp
SNST1	*Homo sapiens*	Nucleoside/glucose(?)	672	P31639 sp
SNST1	*Rattus norvegicus*	Nucleoside/glucose(?)	670	P53792 sp
ST1	*Oryctolagus cuniculus*	?	674	D16226; 473969
RKD	*Oryctolagus cuniculus*	?	597	U08813; 520469
SMIT1	*Canis familiaris*	Myoinositol	718	P31637 sp
SMIT1	*Homo sapiens*	Myoinositol	718	P53794 sp
SMIT1	*Bos taurus*	Myoinositol	718	P53793 sp
NIS	*Rattus norvegicus*	Iodide	618	U60282; 1399954
NIS	*Homo sapiens*	Iodide	643	U66088; 1628579
HypDme	*Drosophila melanogaster*	?	623	U72716; 1763254
HypCel	*Caenorhabditis elegans*	?	602	Z73898; 1340019
HypCel2	*Caenorhabditis elegans*	?	830	Z81049; 1627681
DUR3	*Saccharomyces cerevisiae*	Urea	735	P33413 sp
Prokaryotes:				
SGLTV	*Vibrio parahaemolyticus*	Galactose, glucose	530	D78137; 1794165
HypE62	*Escherichia coli*	?	571	P31448 sp
HypSyn	*Synechocystis* sp.	?	512	D90913; 1653432
putP	*Escherichia coli*	Proline	502	X05653 gp
putP	*Salmonella typhimurium*	Proline	502	P10502 sp
putP	*Pseudomonas fluorescens*	Proline	494	D32069 gp
putP	*Haemophilus influenzae*	Proline	504	P45174 sp
putP	*Staphylococcus aureus*	Proline	497	U06451 gp
putP	*Bacillus subtilis*	Proline	>488	D50453; 1805394[b]
putP	*Rickettsia typhi*	Proline	489	L01134; 152490
panF	*Escherichia coli*	Pantothenate	482	P16256 sp
panF	*Haemophilus influenzae*	Pantothenate	477	P44963 sp
HypBac	*Bacillus subtilis*	?	513	P39599 sp[c]
HypXan	*Xanthobacter autotrophicus*	?	516	X86084; 763398
HypE59	*Escherichia coli*	?	549	P32705 sp
HypAeu	*Alcaligenes eutrophus*	?	>332	P31640 sp

Locus numbers are provided for 12 Genbank sequences which do not appear in either the Genpept (gp) or SwissProt (sp) databases. Two sequences appeared only recently in Genbank, and were not analyzed: putP of *Bacillus subtilis,* and HypCel2 of *Caenorhabditis elegans.* Homologues known from genomic sequencing have been given an abbreviation beginning with the prefix 'Hyp-', for hypothetical.

[a] D. Rhoads, private communiation.

[b] Translation of 26331-27794.

[c] Corrected for an apparent sequencing error. Frameshifting by adding a C after position 33269 (the middle of a GC-rich palindrome) in accession X73124 corrects the N-terminal 23 amino acid residues to a sequence of high % identity with the N-terminus of the closest homolog, HypXan.

ciently distant from membrane interface [40, 64]. The second set of mutants with a 42-residue insertion successfully revealed 12 C-terminal spans, but both glycosylation mutations flanking the N-terminal signal anchor were glycosylated and nonfunctional; topology had been perturbed by one of the mutations, thereby leaving indeterminate the N-terminus orientation and number of spans there. Orientation of the N-terminus has been found to be very sensitive to the distribution of positively charged residues near the signal anchor [12, 18, 47, 69]; this sensitivity has precluded the experimental determination of N-terminal topology of another transporter [62].

Insertion of two Asn residues near the initiator methionine generated a consensus which was fully glycosylated, but this mutant retained full cotransport activity, which thus conclusively established that the N-terminus of wild type SGLT1 is extracellular. An extracellular

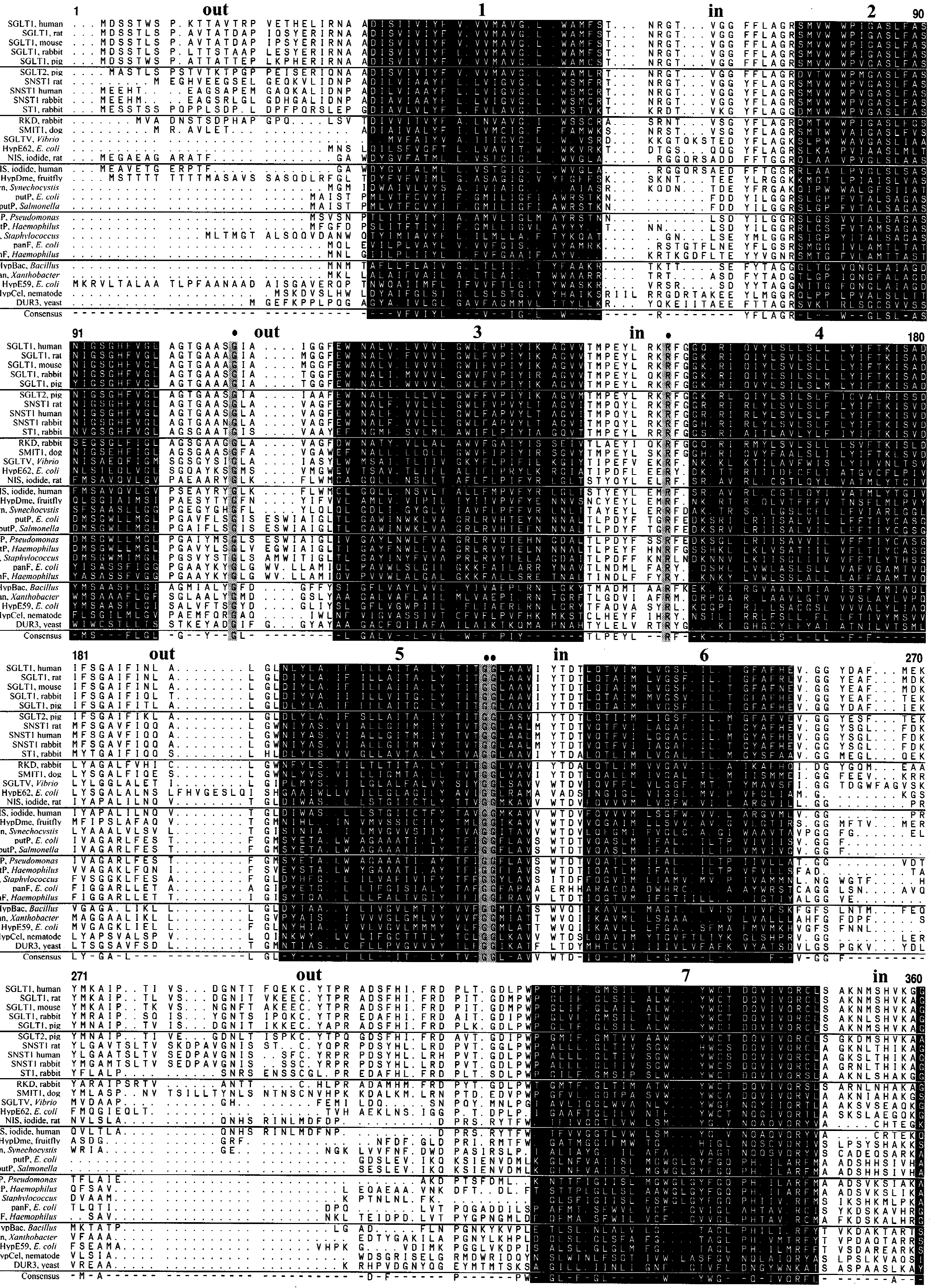


Fig. 2. Progressive sequence alignment of 30 SGLT homologues. Only the region encompassing the initial 13 transmembrane domains has been aligned; sequence divergence much beyond the thirteenth span is too extreme for meaningful alignment. The TREE program of Feng & Doolittle [17] was used for alignment. The consensus sequence was estimated, after proportionately down-weighting the contributions of homologues of high percent identity, using the GCG program Pretty [1]. Membrane spanning sequences are indicated by black columns, as based on the estimated helix

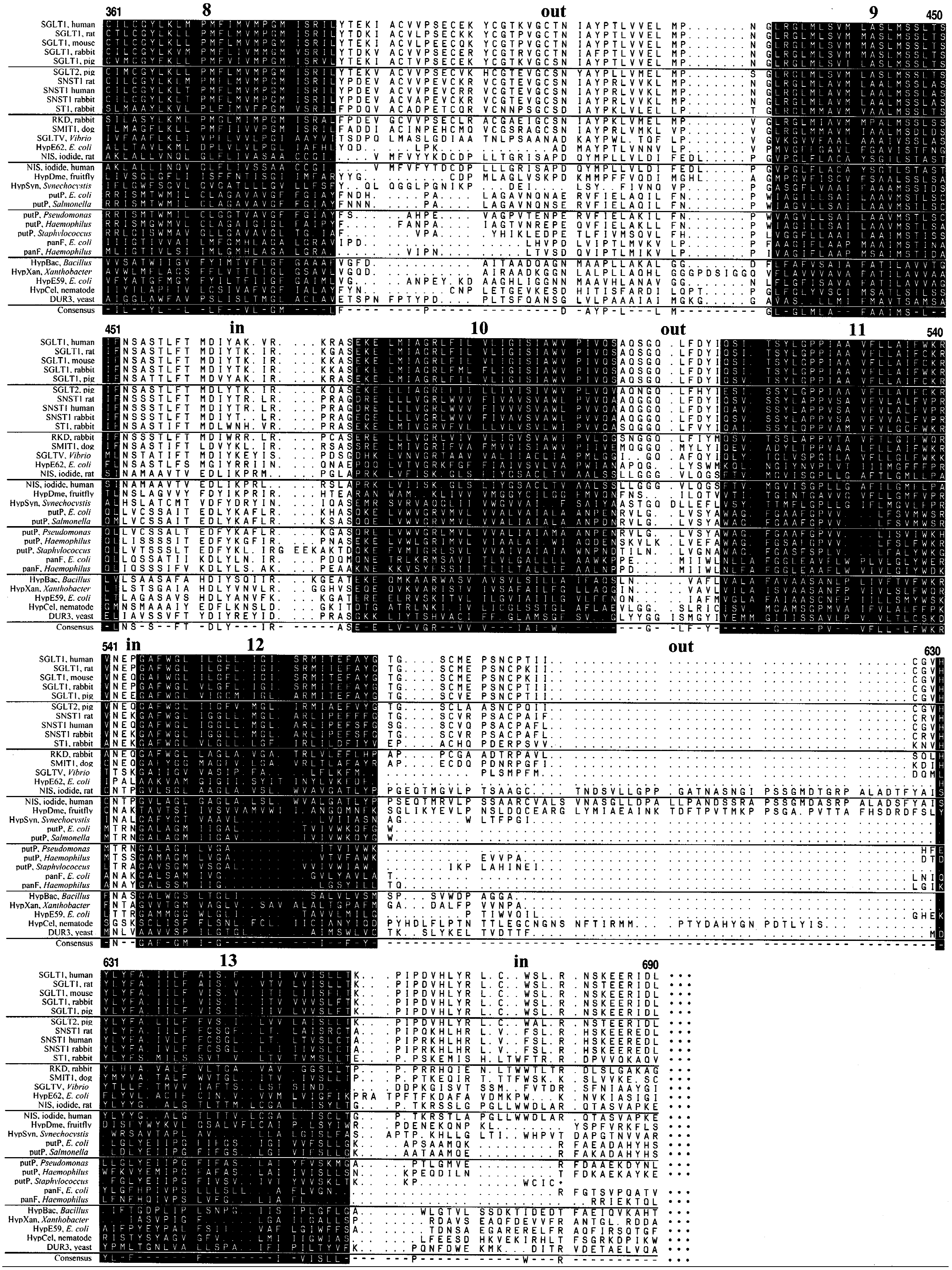


Fig. 2. Continued.

orientation of the N-terminus, though less common, is found in other membrane proteins [12], notably the seven-span G-protein coupled receptors [69], which usually lack a cleavable signal peptide. Our earlier conflicting results, however, necessitated in vitro translation experiments with truncated mRNAs to establish that the first two putative spans were actually retained within the membrane, rather than translocated completely into the ER lumen [64].

The C-terminus of SGLT1 apparently constitutes an exiting transmembrane span ending abruptly at the membrane outer interface. The C-terminus of the unrelated bacterial H^+/rhamnose transporter was experimentally found to form a similar abruptly terminating, exiting membrane span [62]. However, an 11-residue epitope fused to the C-terminus of rabbit SGLT1 was immunologically detected in the cytoplasm [65]. This fusion protein showed a 23-fold reduction in apparent K_m for sugar transport. The epitope's positive charges and high polarity may have impeded membrane insertion of the fourteenth hydrophobic domain [2, 18]. The observation (Membrane Topology Motifs) that all those SGLT homologues which incorporate exactly fourteen membrane spans lack any extracellular C-terminal sequence may reflect either thermodynamic constraints on unassisted insertion and translocation of a (short) polar C-terminus, or adverse consequences of an extracellular disposition of the carboxy terminus (such as carboxypeptidase susceptibility). (At least in bacteria, a hydrophilic sequence C-terminal to a potential exiting span, and not inserting spontaneously as part of a helical hairpin, must be of sufficient length to be recognized and translocated by the *Sec* translocation machinery [2, 32].) In this laboratory, a truncation mutant of human SGLT1 lacking the 16 C-terminal hydrophobic amino acids exhibited a 25-fold reduction in alpha-methyl-D-glucoside transport (E. Turk, E.M. Wright, *unpublished observations*). These two experiments—C-terminal epitope fusion and truncation—indicate that SGLT1 can still transport sugar whether the presumed fourteenth transmembrane domain is relocated to the cytoplasm or is lacking entirely.

Sequence Comparison and Phylogeny

Table 1 lists 21 eukaryotic and 16 prokaryotic sequences of the SGLT family (sequences not listed are incomplete, or are *Salmonella* and *Escherichia* putP variants). Homologues were taken to be those sequences exhibiting pairwise alignment scores 9 SD above that of their aligned randomizations [13, 55]. Sequence lengths vary from 477 to 830 residues, and eukaryotic forms are generally 100–200 residues longer than prokaryotic. Gene nomenclature in Table 1 for the controversial entities SNST1 and SGLT2 is based simply on historical precedence, and on apparent uncertainty of the predominant solute transported by SNST1. SNST1 in rabbits was first described as a nucleoside transporter [43], followed by demonstration of similarly modest transport of glucose by the human isoform [29]. SGLT2 was originally described as an amino acid cotransporter SAAT1 [31], but was later shown to transport glucose with vastly greater efficiency [37]. (Similarly, the [Na^+ + Cl^-]-coupled betaine transporter preferentially transports GABA [75], suggesting that nomenclature based on the solute chosen for transport assays may often be flawed.)

Figure 2 displays a multiple sequence alignment of 30 representatives of the SGLT family; some known sequences, including sheep SGLT1, human and cow SMIT1, and variants of bacterial putP, were not included due to software limitations. The sequences were aligned by the *progressive* sequence alignment method (''once a gap, always a gap'') of Feng & Doolittle in order to preserve phylogenetic relationships which can be obscured by other alignment algorithms [17]. The aligned sequences ranged from the N-terminus to just pass the 13th membrane span, since sequences more C-terminal than this are too divergent for alignment. Membrane span regions are shaded, based upon the transmembrane helix end estimates of SGLT1 [64]. Only four residues are perfectly conserved, and these correspond to human SGLT1 residues Gly95 in the second external domain, Arg135 in the second internal domain, and Gly195Gly196 in the inward half of the transmembrane span 5. Evolutionary conservation of glycines is not uncommon, and is usually taken to indicate an essential structural role.

A plot of the similarity of residue substitutions in the alignment of Fig. 2 is displayed in Fig. 3. The four peaks of greatest sequence conservation (>0.55) all reside within the cytoplasm or within the cytoplasmic half of a transmembrane domain. The tallest of these peaks corresponds to the conserved residues Gly195Gly196 in span 5. The other three peaks correspond to the cytoplasmic 'flag' consensus sequence (roughly YFLAGRSL, [64]) just proximal to span 2, and to the cytoplasmic halves of spans 9 and 11. Gly95 actually falls in the poorly conserved second external domain, which is particularly rich in residues bearing small side chains [64]. In general, cytoplasmic domains are seen in Fig. 3 to be shorter, and better conserved, than external domains, except for the poorly conserved cytoplasmic domain distal to span 13. The valleys of lowest conservation usually correspond to gaps introduced in the alignment. Two spans with the highest average level of conservation are 2 and 9, and it is interesting that span 2 is also the least 'conventional' in that it has low hydrophobicity, is not associated with a well-defined minimum in a reverse turn propensity plot [64], has the lowset span propensity predicted by the neural network PredictProtein (*see below*), and exhibits, overlapping

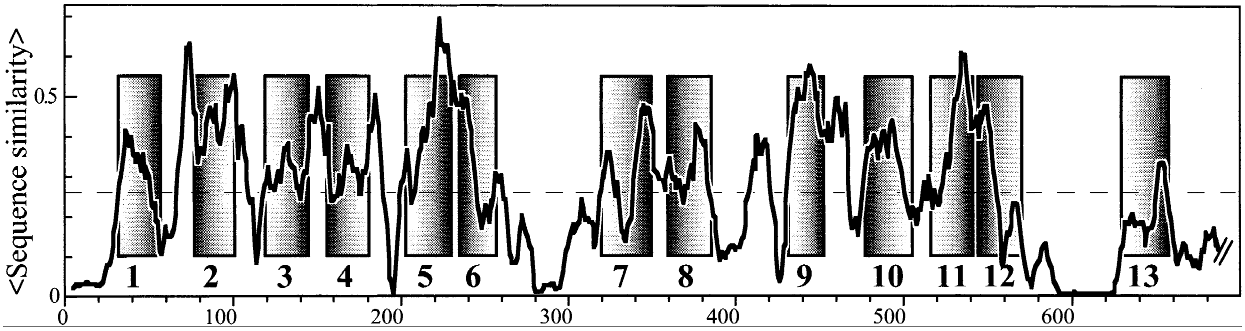


Fig. 3. Amino acid residue similarity of progressively aligned sequences of 30 SGLT homologues. The progressive sequence alignment of Fig. 2 was analyzed for amino acid conservation using an eight-residue window and the GCG program Plotsimilarity [1]. Rectangles indicate the transmembrane domains of human SGLT1, and rectangle shading indicates the orientation of the spans (dark = cytoplasmic, light = extracellular). Complete identity in a window would equal 1 on the similarity axis.

with the poorly conserved second external domain, an enrichment in small amino acid residues. Span 13 exhibits the lowest average level of conservation of the spans, just as its flanking hydrophilic regions are most poorly conserved.

An unrooted phylogenetic tree was constructed from the alignment in Fig. 2, and the evolutionary relationships of the SGLT family members are depicted in Fig. 4. The mammalian homologues (SGLT1, SGLT2, SNST1, ST1, RKD and SMIT1) cluster together, except for the NIS iodide transporter, which is very divergent. On the NIS branch are transporters of unknown function from fruitfly, nematode and a cyanobacterium. Branching from a point between the SGLT1 and NIS branches are HypE62 and the SGLTV Na^+/galactose transporter of the pathogenic marine bacterium *Vibrio parahaemolyticus*. HypE62 may be the *E. coli* isoform of SGLTV because the entire genome (4.6 Mb) of *E. coli* has been sequenced, and just four SGLT members are detected (HypE62, HypE59, putP and panF). SGLTV and HypE62 may then be members of the SGLT1 branch.

The bacterial putP (proline) and panF (pantothenate) transporters form two other main branches. The fourth *E. coli* homologue, HypE59, maps to another distinct branch, of unknown solute specificity. The most divergent member of all is the yeast DUR3 urea transporter, which shares only 22% identity with its closest homologue, HypBac.

The entire genome of *Saccharomyces cerevisiae* has been sequenced (16 Mb), but DUR3 is the only detectable SGLT homologue. Similarly, HypSyn is the only homologue detected in the completed genome sequence (3.6 Mb) of the photosynthetic cyanobacterium *Synechocystis* sp., strain PCC6803. The completed genomic sequence (1.8 Mb) of *Haemophilus influenzae* encodes putP and panF only; however, a 101 residue protein fragment (HI1315) contains a perfect 'flag' consensus [64] and is 49% identical to the N-terminus of HypE62, ending with span 3 (*not shown*). Frameshifting the 3′ DNA sequence, to compensate for any sequencing errors, does not reveal more coding sequence similar to HypE62, and it is unknown whether this small homologous sequence represents a complete functional protein, a pseudogene fragment, or an indication of a subcloning transposition error. The genomes of the archaebacterium *Methanococcus jannaschii* (1.8 Mb) and the Gram-positive eubacterium *Mycoplasma genitalium* (0.6 Mb) have been completely sequenced, and those of the eubacteria *Neisseria gonorrhoeae* (2.1 Mb) and *Streptococcus pyogenes* (1.8 Mb, Gram-positive) are >93% complete, but no SGLT members were detected.[1]

Membrane Topology Motifs

In the course of experimentally mapping the topology of human SGLT1, we found two computational prediction methods to be of significant utility in evaluating the experimental results. The more useful of these is the trained neural network PredictProtein [53] which accurately predicts transmembrane helices from single sequences or, preferably, multiple sequence alignments (97% accuracy was claimed, but we have found it to be apparently lower with transporters of aqueous solutes). The algorithm Memsat [27] calculates scores for a variety of topologies of the protein under scrutiny, and capitalizes on (i) the pronounced statistical preferences [39, 49] of various amino acids to reside extracellularly or intracellularly, mid-helix, inner helix or outer helix, (i.e., in one of five states), and (ii) the powerful topological constraint that proximate spans must be antiparallel, which thereby globally propagates the effect of introducing each new test span to all helix and loop scores. PredictProtein and Memsat predictions can complement

[1] The URL http://www.mcs.anl.gov/home/gaasterl/genomes.html provides an excellent list of links to genome projects.

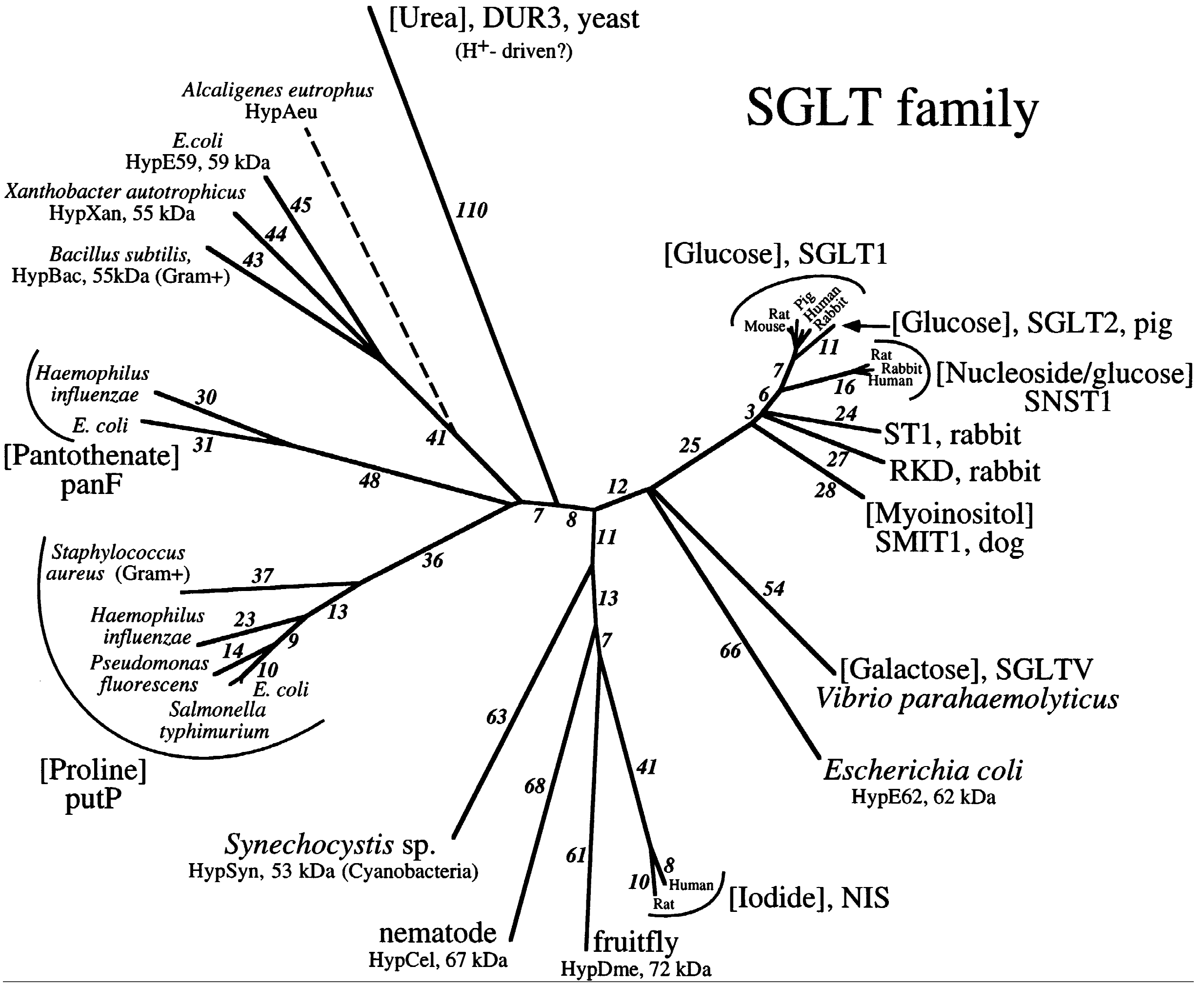


Fig. 4. Unrooted phylogenetic tree of the 30 homologues progressively aligned in Fig. 2. Text in square brackets indicates the known or presumed cotransported substrate. Molecular weights are given for those transporters known only from genomic sequencing and for which the cotransported substrate is unknown. The Felsenstein program Drawtree of the PHYLIP package [16] was used to graphically represent the phylogenetic distances (arbitrary units) calculated from the sequence alignment in Fig. 2, as determined by the Feng & Doolittle program TREE [17]. The partial sequence of HypAeu encodes only its first six transmembrane spans; the dashed line indicates its tree position and distance after a progressive alignment of the N-terminal region up to the sixth transmembrane span with all the other homologues.

each other, for while the former was trained extensively with no assumptions of hydrophobicity or arbitrary 5-state divisions and can weigh evolutionary (multiple sequence alignment) data, it does not incorporate the global topological constraint of the latter. While PredictProtein loses some accuracy when applied to a sequence in isolation rather than as part of a multiple sequence alignment, it was constrained for the special purposes of this paper to consider only the protein to be analyzed (except for Fig. 10), since in the SGLT family the number of transmembrane spans at the C-terminus is variable (as will be shown below).

We have found the interfacial hydrophobicity (IFH) scales of Jacobs and White [26], determined empirically using actual phospholipid bilayers rather than an octanol/water system, to yield hydrophobicity plots more interpretable than those based on the Eisenberg [14] or Kyte/Doolittle [33] scales. IFH plots additionally reveal the potential for increased hydrophobicity which results from satisfaction of potential H-bonds of polar side chains by the formation of residue-residue H-bonds (*vs.* residue-water H-bonds) upon partition of helices into the hydrocarbon or interfacial layers of the membrane. Using IFH plots in combination with reverse-turn propensity plots, White and Jacobs devised a means to predict transmembrane helix ends, as evaluated against the known X-ray diffraction structure of a bacterial photosynthetic reaction center [71].

As a test and demonstration of the foregoing predictive methodologies, we have applied them in Fig. 5 to the

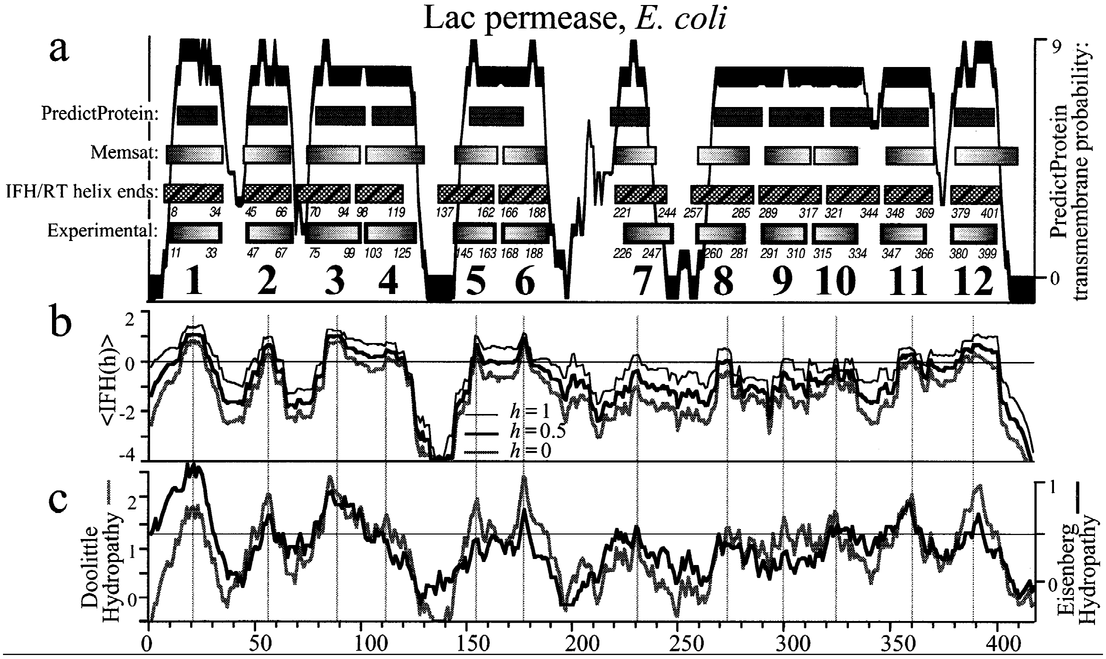


Fig. 5. Comparison with the experimentally determined topology of the lactose transporter lac permease of *E. coli* demonstrates that the combined interpretation of the results of the neural network algorithm PredictProtein [53] and the statistically based algorithm Memsat [27] can accurately predict the membrane topology of a difficult multispanning membrane protein. (*a*) The probability (arbitrary units) of transmembrane helix content as determined by the trained neural network PredictProtein was predicted for *E. coli* lac permease alone ('aligned' to itself). The reliability (certainty) of the prediction at each residue is indicated by the thickness of the plot, thickness increasing with reliability. (The thick curve represents the darkened area between two curves — that of the probability *P* and of the function *P-R*/10, where *R* is reliability.) The topmost set of horizontal bars represents PredictProtein's final prediction based upon the probability plot. The second set of horizontal bars represents the membrane spans and their orientations predicted by the algorithm Memsat, where shading indicates the orientation of the spans (dark = cytoplasmic, light = extracellular). Based upon comparison of the results from Memsat and PredictProtein, twelve membrane helices were deduced and their approximate helix ends were estimated from interfacial hydrophobicity (IFH) and reverse turn (RT) propensity plots, by the method of White & Jacobs [71]. The lowermost set of horizontal bars represents the topology as experimentally determined with alkaline phosphatase fusions [8]. (*b*) Interfacial hydrophobicity (IFH) analysis of lac permease. The IFH averages of a scanning 19AA window are shown for three assumptions of side chain-side chain satisfaction of hydrogen bonds ranging from none ($h = 0$) to complete ($h = 1$). Vertical dashed lines have been drawn through the residue corresponding to the highest value of IFH(0.5) for each of 12 deduced membrane spans and extended to the hydrophobicity plots in *c*. (*c*) Hydrophobicity plots of lac permease using a scanning 19AA window and the hydrophobicity scales of Eisenberg [14] or Kyte & Doolittle [33].

best studied cotransport protein, the lac permease of *E. coli*. Lac permease has been variously predicted on the basis of hydrophobicity to encompass from 8- to 14-spans [70], but is now known from extensive analysis to comprise 12 spans [28]. In Fig. 5, the upper plot of span probability as determined by PredictProtein also incorporates information on the reliability (certainty) of the prediction at each residue, represented as a directly-related thickening of the curve. PredictProtein's final estimate of the occurrence of 11 spans is indicated in the first row of bars. Memsat's highest scoring topology, 12 spans, is shown in the second row. The bottom row shows the experimentally determined span locations. Clearly, spans 5 and 6 were misinterpreted by PredictProtein as a single span. Memsat correctly predicts two spans there, and brief study of the PredictProtein probability *curve* indicates the broad peak of probability in this region to be consistent with two spans. It is noteworthy that PredictProtein has calculated high probability, with high reliability, of four membrane spans in the region now known to include four spans, numbers 7 to 10. As apparent in the Eisenberg and Kyte/Doolittle hydrophobicity plots in the lower panel of Fig. 5, this region is of only modest hydrophobicity, thereby making span predictions based on hydrophobicity alone very uncertain. For comparison, the middle panel presents an IFH hydrophobicity plot. The Memsat and PredictProtein algorithms clearly predict the correct 12-span topology of lac permease remarkably well.

The extracellular orientation of the N-terminus of SGLT1 implies a similar N-terminal orientation in the other SGLT family members. Lodish' group has empirically devised a simple calculation, based on the charge distribution flanking the N-terminal span, which is a good predictor of the orientation of eukaryotic N-termini [21]. A positive difference (⩾0; *see* Table 2) in the net charge in two flanking windows indicates an extracellular N-terminus. Table 2 shows results for 23 SGLT

Table 2. Predictions of N-terminus orientation (H. Lodish)

SGLT homologue	$\langle N_{15}\rangle$	$\langle C_{15}\rangle$	$\Delta(C_{15}-N_{15})$	Predicted N-terminus
SGLT1, human	−0.5	2	2.5	out
SGLT2, pig	−2	2	4	out
SNST1, rabbit	−2.5	3	5.5	out
ST1, rabbit	−3	3	6	out
RKD, rabbit	−1.5	3	4.5	out
SMIT1, human	0	2	2	out
SGLTV, *Vibrio parahaemolyticus*	0	1	1	out
HypE62, *Escherichia coli*	0	1	1	out
NIS, rat	−2	1	3	out
HypDme, fruitfly	0	3	3	out
HypCel, nematode	0.5	4.5	4	out
HypSyn, *Synechocystis* sp.	−1	1	2	out
putP, *Escherichia coli*	0	1	1	out
putP, *Haemophilus influenzae*	−1	1	2	out
putP, *Pseudomonas fluorescens*	0	1	1	out
putP, *Staphylococcus aureus*	−1	1	2	out
panF, *Escherichia coli*	−1	2	3	out
panF, *Haemophilus influenzae*	0	1	1	out
HypE59, post cleavage, *E. coli*[a]	−1	3	4	out
HypBac, *Bacillus subtilis*	0	2	2	out
HypXan, *Xanthobacter autotrophicus*	1	1	0	out
DUR3, yeast	1	0	−1	(in)
HypE59, signal peptide, *E. coli*[a]	2	−1	−3	in
HypAeu, *Alcaligenes eutrophus*[b]	3	−1	−4	in

The net charge was calculated in each of two 15-residue windows flanking the apparent center of the hydrophobic signal anchor; each window begins with the charged residue closest to the helix center, and extends outward. Histidine is assigned a value of +0.5. The initiator methionine is assigned a value of +1 in eukaryotes, and 0 in prokaryotes. The difference in window net charge is predictive of the membrane orientation of the N-terminus. One incorrect prediction of N-terminal disposition is indicated in parentheses.

[a] HypE59 encodes an apparent cleavable signal peptide.

[b] HypAeu encodes an additional N-terminal transmembrane span (see text).

members representative of the semimajor branches of the phylogenetic tree in Fig. 4; we have applied the calculation to prokaryotic sequences also, with a corrective assignment of zero charge to the initiator methionine, since its amine group is formylated during synthesis and insertion [19, 66]. Twenty one are predicted to have an extracellular N-terminus, but the N-terminus of DUR3, was incorrectly (as assessed by multiple sequence alignment) predicted as intracellular. The three positive charges in the DUR3 C-terminal window cluster within 5 residues of the signal anchor, and reducing the window size to 10 [21] yields a calculated difference of +1, consistent with an extracellular N-terminus. As discussed below, HypE59 encodes an N-terminal signal peptide, and HypAeu an additional N-terminal membrane span. Their charge difference scores strongly indicate intracellular N-termini (Table 2), concurring with the results of PredictProtein, Memsat and hydrophobicity plots. It is evident that, with the exception of HypAeu, all members of the family have an extracellular N-terminus.

Figure 6 presents the predictive data bearing on the topologies of six SGLT family members, selected as representative of the four topological motifs found. The graphs have been aligned on span 1, as indicated by the left grey ribbon. A second grey ribbon runs through span 12, for visual reference. Each of the six panels shows an upper IFH plot, a lower PredictProtein plot, and a horizontal row of bars indicating the Memsat prediction. SGLT1 is predicted to comprise fourteen spans. The presence of the fourteenth, with which the protein abruptly terminates, is strongly indicated by PredictProtein (with high reliability) and Memsat. Span 2 is weakly indicated by PredictProtein, and only experimental determination of the orientation of the N-terminus resolved its presence [64]. (Note in Fig. 6 that PredictProtein strongly indicates span 2 in the NIS iodide and panF pantothenate transporters.) Fourteen spans are found in all the mammalian SGLT homologues, except the NIS iodide transporter. Topological models of 14-span representatives of the semi-major mammalian branches of the phylogenetic tree are schematically depicted in the middle column of Fig. 7; these models are drawn with spans normalized to 20 residues, and with aqueous domains drawn to this scale. One bacterial homologue, the SGLTV galactose transporter, bears 14 spans also (Fig. 7).

HypE62, a possible *E. coli* isoform of SGLTV, evidently encodes 15 spans, the additional one appearing C-terminally. The 15th span is strongly indicated by PredictProtein and Memsat, and its presence is also clearly apparent in the IFH plot (Fig. 6). The yeast DUR3 urea transporter is the only other SGLT member encoding 15 spans (*data not shown*). Both HypE62 and DUR3 are schematically modelled in the right column of Fig. 7.

A thirteen-membrane-span motif is found to be common to the remaining SGLT members, as demonstrated in Fig. 6 for three representative members, panF, putP and NIS. (No homologue with the classical transporter 12-span motif was indicated algorithmically or by charge calculations of the N-terminal span.) PredictProtein fails to detect helix 2 in putP and helix 6 in panF (Fig. 6) when applied to each sequence in isolation, but indicates both helices when applied to a multiple sequence alignment of the panF and putP family members (*not shown*). Representative 13-span members are schematically modelled

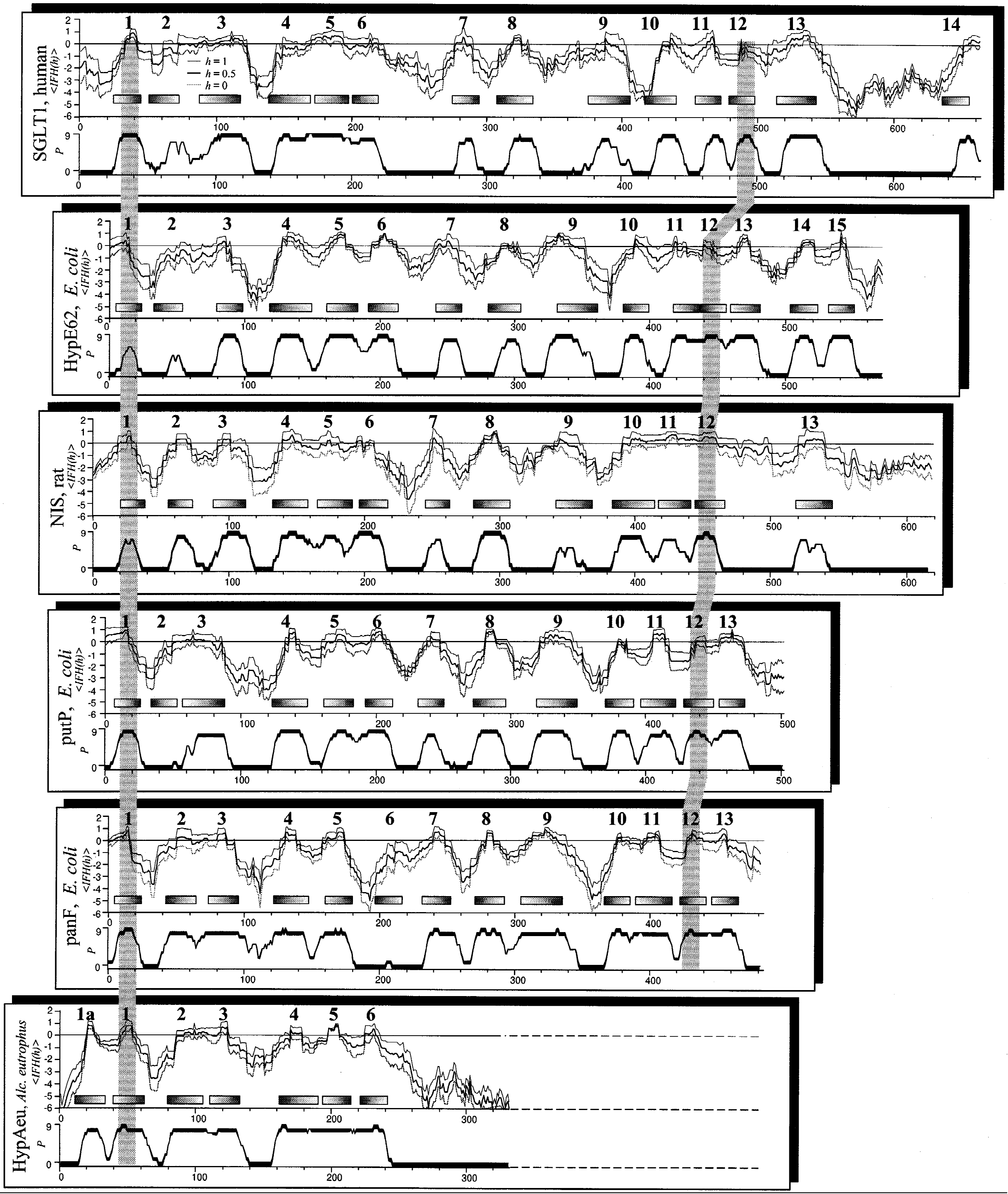


Fig. 6. Transmembrane helix predictions for six representative members of the SGLT cotransporter family indicate a common core of 13 transmembrane spans with additional flanking transmembrane helices in some homologues. Interfacial hydrophobicity plots [26] represent IFH(h) averages of a scanning 19AA window for three assumptions of side chain-side chain satisfaction of hydrogen-bonds ranging from none (h = 0) to complete (h = 1). The probability (P, arbitrary units) of transmembrane helix content was determined by the trained neural network Predict-Protein [53] for each cotransporter alone ('aligned' to itself). The reliability (certainty) of the prediction at each residue is indicated by the thickness of the plot, thickness increasing with reliability. Each set of horizontal bars represents the membrane spans and their orientations as predicted by the algorithm Memsat (minloop = 1, minhelix = 18, maxhelix = 31, minscore = −1950), where shading indicates the orientation of the spans (dark = cytoplasmic, light = extracellular). Two vertical grey bands are drawn through transmembrane spans 1 and 12 to help visual orientation. While the bacterial panF and putP transporters, and the rat NIS iodide transporter incorporate the common core of 13 transmembrane spans, the SGLT1 transporter incorporates a fourteenth span at its C-terminus. A fifteenth span is incorporated at the C-terminus of the 13-span core of bacterial HypE62. The partial sequence of bacterial HypAeu shows incorporation of an additional span at its N-terminus.

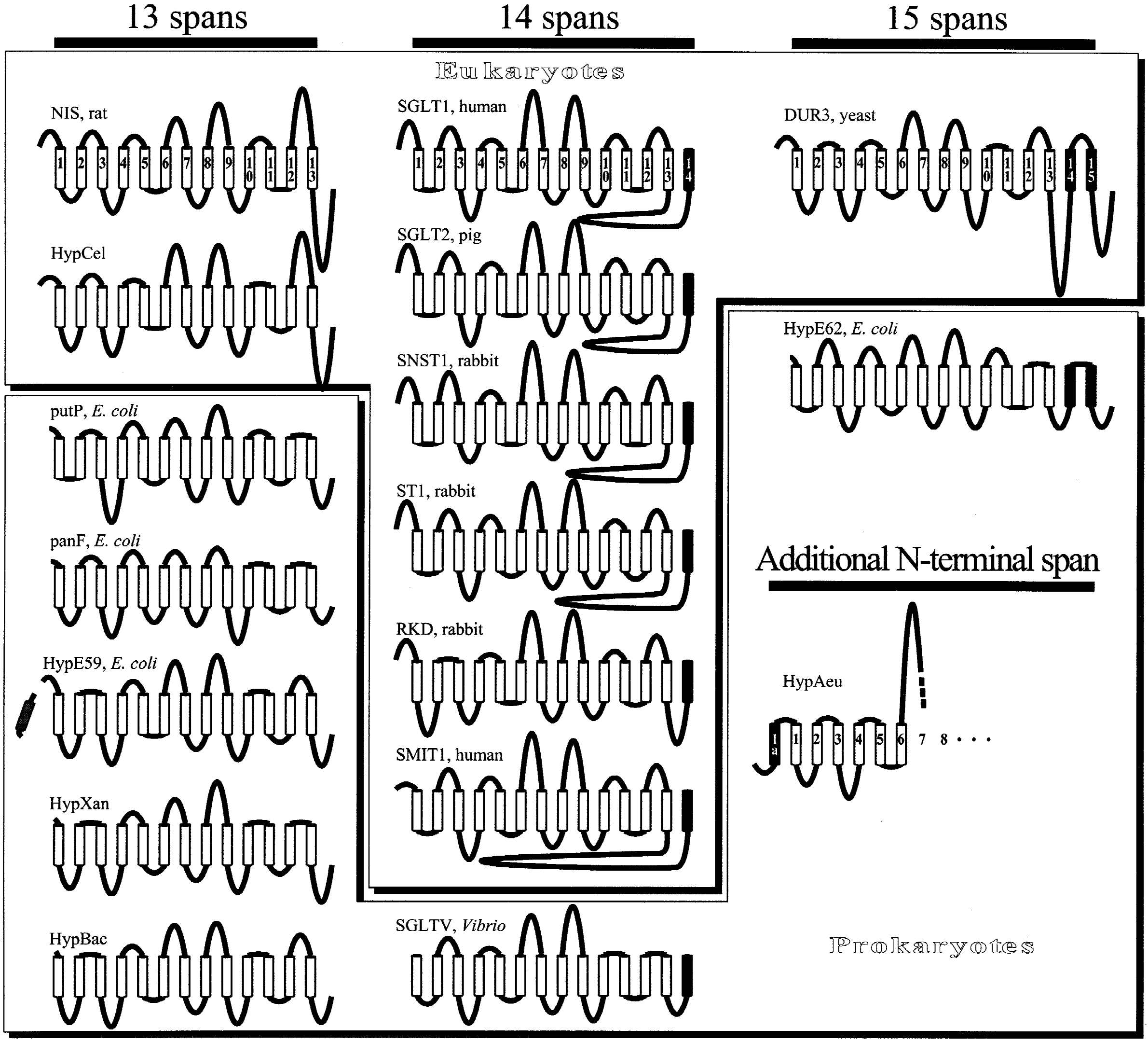


Fig. 7. Schematized membrane topologies of representative SGLT homologues. For each homologue displayed, the membrane topologies predicted by Memsat and PredictProtein were examined in the light of the experimentally established topology of SGLT1. Multiple sequence alignments identified or confirmed correlating transmembrane spans between homologues. For simplifying the schematization of topology, each membrane span was normalized to a 20 amino acid residue length centered on the middle of the span as predicted by Memsat. The residues in loops between spans are drawn in loop lengths proportional to this scale (i.e., as if these residues' lengths were also of an alpha-helical value). Additional spans which supplement the core 13-span motif are shown darkened. The HypE59 homologue evidently encodes a cleavable signal peptide (grey). In addition to many examples of 13- and 14-spanners, two homologues, HypE62 of *E. coli* and the urea transporter of yeast clearly encoded 15 spans. The bacterial HypAeu homologue sequence is known only for the N-terminal region, which incorporates an additional transmembrane span at the N-terminus.

in the left column of Fig. 7. Lack of a 14th span places the C-terminus in the cytoplasm, as supported by antibody experiments on the C-terminus of the putP proline transporter [30]. Recent experimental mapping of putP confirms that it comprises 13 spans, not 12 as originally proposed from hydropathy plots (H. Jung and M. Quick, *personal communication*). The external loop between spans 12 and 13 of the eukaryotic members of the NIS branch (NIS, HypDme and HypCel) is significantly larger, by about 40 residues, than that of the other 13-span homologues (*see* Fig. 7).

One of the 13-spanners, HypE59 of *E. coli,* evidently encodes a signal peptide (Fig. 7, and compare the N-terminal sequence of HypE59 in Fig. 2 with that of its isoforms, HypXan and HypBac). This hydrophobic addition is interpreted (*data not shown*) to be a signal peptide, not a span, because (i) the hydrophobic region is relatively short, and appears in an IFH plot as an isolated, sharp peak, flanked abruptly by deep hydrophilic valleys, unlike any of the true spans, (ii) it is enriched in alanine [68], (iii) PredictProtein finds essentially no indication of a span there and (iv) a putative signal cleavage consensus

[11] occurs where one would be anticipated (the 21-residue putative cleaved signal peptide is MKRVLTA-LAATLPFAANAADA).

The sequence of HypAeu (partial) evidently does encode an additional membrane span at the N-terminus (Fig. 7). A domain N-terminal to the span aligning with span 1 of the other homologues is very hydrophobic, PredictProtein strongly indicates a spanner there (Fig. 6), no signal cleavage consensus is apparent, and the flanking charge distribution is solidly consistent with an intracellular N-terminus (Table 2). HypAeu and the signal-peptide bearing homologue HypE59 map to the same phylogenetic branch (Fig. 4).

SGLT Functional Similarities

Detailed functional studies conducted on SGLT family members, as expressed in heterologous systems, indicate clearly that they share remarkable similarities in function. Three isoforms of the high affinity Na^+/glucose cotransporter SGLT1 (rabbit, human and rat), the low affinity Na^+/glucose cotransporter SGLT2, the Na^+/myoinositol cotransporter SMIT1, and the Na^+/iodide cotransporter NIS have all been expressed in *Xenopus laevis* oocytes and studied extensively in this laboratory using radioactive tracer and electrophysiological techniques [23, 24, 44–46]. In the absence of substrate, all show Na^+ leaks, exhibit fast voltage-dependent current transients, and, of those studied so far, behave as low conductance water channels. In the presence of substrate, they are all Na^+/substrate cotransporters with ordered (Na^+ on first), voltage-dependent kinetics, and cotransport substantial amounts of water.

Less detailed studies have been conducted on the functional properties of bacterial SGLT family members. However, the kinetics and specificity of SGLTV are very similar to the mammalian SGLT1 [56, 57], and there are common features between the *E. coli* putP cotransporters and the mammalian members of the family [20, 51, 74]. We expect that more detailed studies of the prokaryotic and yeast homologues will reveal further functional similarities among the diverse SGLT members.

SGLT Homology with the [Na^+ + Cl^-]-coupled Transporter Family

Homology means common ancestry, and thus there are no degrees of homology; proteins are either homologous or not [13]. The probability of homology of two sequences can be estimated statistically by scoring (for similarity and identity) their optimal alignment, then scoring many alignments of one with random sequence scrambles of the other, and finally calculating the mean and standard deviation (SD) of the scrambled alignment scores. If the genuine and the mean scramble scores differ by $\geqslant 9$ SD ($P \leqslant 10^{-19}$) the proteins are considered definitely homologous [13, 55]. Differences of 5 to 9 SD imply homology, and in such cases the question of homology must be decided on the basis of whether the proteins share physiological aspects [13].

Figure 8 presents a case for homology of the SGLT Na^+ cotransporter family and the [Na^+ + Cl^-] transporter family. The aligned sequence fragments, and a schematic of their topological contexts, of the mouse [Na^+ + Cl^-]-coupled GABA transporter GAT-2 and the pig Na^+/glucose cotransporter SGLT1 appear in Fig. 8*a*. This sequence alignment scores 7.2 SD ($P < 10^{-12.5}$) greater than chance, and the two sequences concur well in their respective topological placements, both locally (microtopology) and globally (in their occurrence with reference to the complete protein, or macrotopology). Similar results are seen for a longer sequence alignment between GAT-2 and the dog Na^+/myoinositol cotransporter SMIT1 (Fig. 8*b*). This alignment scores 6.3 SD ($P < 10^{-10}$) greater than chance, and again the micro- and macrotopological concurrences are close, involving a region which includes spans 9 of 11 of SMIT1. There is also considerable functional similarity between these two families: for example, SGLT1, SMIT1 and GAT-1 share common kinetics of Na^+/substrate cotransport, Na^+ leak or Na^+ uniport in the absence of substrate, presteady-state current transients, Na^+/substrate/water cotransport, and behavior as water channels in the absence of substrate [35, 73]. Based on the high SD values of the alignments, their micro- and macrotopological correspondence, and the shared familial physiological functions, we conclude that the SGLT and the [Na^+ + Cl^-] transporter families arose from a common ancestor.

Recently, two members of the [Na^+ + Cl^-] transporter family, the GABA transporter GAT-1 [5] and the glycine transporter GLYT1 [42] were both found experimentally to incorporate 12 spans. The first hydrophobic domain, considered previously to form the N-terminal span, was found in both transporters not to span the membrane, but to remain on the cytoplasmic side. Comparison of the schematic models of these two families (Fig. 8) suggests an evolutionary relationship between this domain and the N-terminal span of the SGLT family. However, this [Na^+ + Cl^-] transporter domain contains an essential Arg, and other charged and polar residues, unlike the highly nonpolar character of the SGLT N-terminal span, indicating that the spans' functions are likely very different.

SGLT homology with the 12-span [Na^+ + Cl^-] transporter family, considered with the prevalence of the 12-span motif among many seemingly unrelated transporters, suggests that the SGLT family arose from a 12-span ancestor. The unusual nature of span 2 (discussed previously) possibly conflicts with the requirements of an insertion signal sequence. Addition of span 1 may have relieved signalling constraints on span 2, freeing its evo-

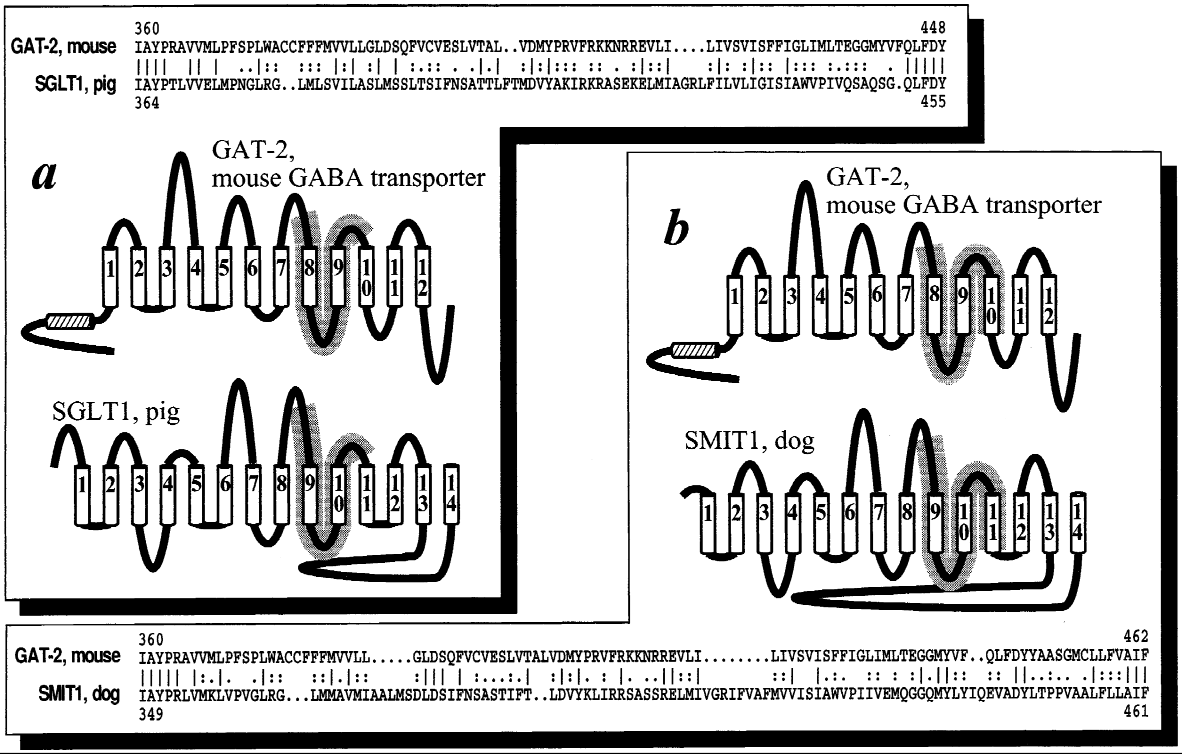
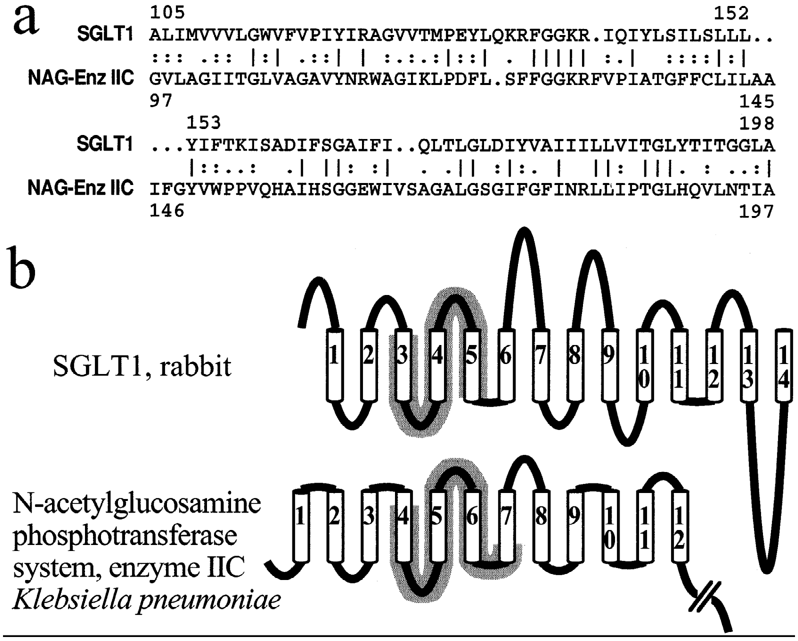

360
448
GAT-2, mouse
IAYPRAVVMLPFSPLWACCFFFMVVLLGLDSQFVCVESLVTAL..VDMYPRVFRKKNRREVLI....LIVSVISFFIGLIMLTEGGMYVFQLFDY
SGLT1, pig
IAYPTLVVELMPNGLRG..LMLSVILASLMSSLTSIFNSATTLFTMDVYAKIRKRASEKELMIAGRLFILVLIGISIAWVPIVQSAQSG.QLFDY
364
455
a
GAT-2,
mouse GABA transporter
SGLT1, pig
b
GAT-2,
mouse GABA transporter
SMIT1, dog
360
462
GAT-2, mouse
IAYPRAVVMLPFSPLWACCFFFMVVLL.....GLDSQFVCVESLVTALVDMYPRVFRKKNRREVLI........LIVSVISFFIGLIMLTEGGMYVF..QLFDYYAASGMCLLFVAIF
SMIT1, dog
IAYPRLVMKLVPVGLRG...LMMAVMIAALMSDLDSIFNSASTIFT..LDVYKLIRRSASSRELMIVGRIFVAFMVVISIAWVPIIVEMQGGQMYLYIQEVADYLTPPVAALFLLAIF
349
461

Fig. 8. Topological comparisons strongly corroborate a putative homology, first detected by sequence similarity, between the 12-span [Na^+ + Cl^-] transporter family and 14-span SGLT transporter family. (*a*) The aligned sequence fragments of the mouse [Na^+ + Cl^-]-coupled GABA transporter GAT-2 (P31649, SwissProt) and pig SGLT1, as shown at the upper left, were analyzed by the Pearson and Lipman program RDF2 [48] for statistical significance of apparent homology. Sequence comparison scores for 1000 random local shuffles in 10-residue windows (a conservative measure to compensate for the bias in spans for hydrophobic residues) were calculated. Their mean optimized score was found to be 7.2 standard deviations (SD) less than that of the unshuffled alignment displayed. Schematized membrane topologies of SGLT1 and GAT-2 were constructed as described in Fig. 7. The compared sequence fragments are indicated in the topological models by the greyed regions in the upper left. (*b*) Similarly, the aligned sequence fragments, shown in the lower right, of the [Na^+ + Cl^+]-coupled GABA transporter GAT-2 and canine SMIT1 (which shares only ~50% identity with SGLT1) were also analyzed by RDF2; the mean shuffled scores and that of the native alignment displayed differed by 6.3 SD. The aligned sequence fragments are schematically indicated in the context of membrane topology (lower right models) by the greyed regions. A ''hydrophobic'' domain thought previously to form the first membrane span of GAT-1 and GLYT1, and recently demonstrated to be nonspanning [5, 42], is indicated by the hatched cylinder in the GAT-2 models.

lution to more efficiently perform its (unknown) function. The predominance of the 13-span motif among the major branches of the SGLT family tree (Fig. 4) is consistent with this notion.

SGLT Homology with the Phosphotransferase Permease Family

It was previously reported [52] that a bacterial N-acetylglucosamine permease (the enzyme IIC domain of the N-acetylglucosamine phosphotransferase (PTS) system in *E. coli*) and rabbit SGLT1 exhibited regions of similarity scoring at 7.1 SD, thereby indicating probable homology. Fig. 9*a* presents an alignment of rabbit SGLT1 with the more recently available N-acetylglucosamine permease of *Klebsiella*. This alignment scores at 8.3 SD ($P < 10^{-16}$), virtually establishing homology. Overall, the microtopological correspondence (Fig. 9*b*) is good, but diverges at the C-ends of the alignment: the bacterial sequence includes a second cytoplasmic domain and extends into the next helix, while the SGLT sequence does not. The macrotopologies do not obviously correspond, except in that both aligned sequences lie nearer the N-terminus. The previous evidence (Figs. 6 and 7) that membrane spans can be appended at either end of a protein may help to explain this discrepancy. The alternative is that these regions scoring at 8.3 SD have evolved convergently. A test of the apparent homology of the SGLT and PTS IIC families would be provided by finding a third family of transporters exhibiting two domains of probable homology, each domain aligning with a different family. The topological occurrences of these domains' sequences from the three



a
b
SGLT1, rabbit
1
2
3
4
5
6
7
8
9
10
11
12
13
14
N-acetylglucosamine phosphotransferase system, enzyme IIC *Klebsiella pneumoniae*
1
2
3
4
5
6
7
8
9
10
11
12

```
            105                                            152
SGLT1       ALIMVVVLGWVFVPIYIRAGVVTMPEYLQKRFGGKR.IQIYLSILSLLL..
            ::: .:: |:| .::| | :.:.:|::| . ||||| :.|  ::::|:|
NAG-Enz IIC GVLAGIITGLVAGAVYNRWAGIKLPDFL.SFFGGKRFVPIATGFFCLILAA
            97                                              145
               153                                          198
SGLT1       ...YIFTKISADIFSGAIFI..QLTLGLDIYVAIIILLVITGLYTITGGLA
               |::..:  .| ||: :|  . .|| :|:. |  ||: |||. : ..:|
NAG-Enz IIC IFGYVWPPVQHAIHSGGEWIVSAGALGSGIFGFINRLLIPTGLHQVLNTIA
            146                                             197
```

Fig. 9. Apparent homology, as indicated by sequence alignment score, of the SGLT transporter family and the enzyme IIC domain of transporters of the bacterial phosphotransferase system is only crudely supported by comparison of local membrane topologies. (*a*) The aligned segments of rabbit SGLT1 and the enzyme IIC region of the N-acetylglucosamine transporter of *Klebsiella pneumoniae* (P45604, SwissProt) score 8.3 SD higher (program RDF2) than the mean optimized sequence comparison scores for 1000 random local shuffles in 10-residue windows. (*b*) The aligned sequence fragments are schematically indicated by the greyed regions in the context of models of membrane topology. Schematized membrane topologies of SGLT1 and the N-acetylglucosamine transporter were constructed as described in Fig. 7.

pair-alignments would need to be macrotopologically consistent with one another if the three families indeed shared one common ancestry.

Experimental Probing or Perturbation of Topology?

Topological mappings of bacterial integral membrane proteins have commonly relied on expression of a reporter enzyme fused to progressive C-truncations of the protein. Reporters have included alkaline phosphatase (phoA), beta-lactamase (bla) and beta-galactosidase (lacZ) (*see* [6] for a review). Protease susceptibility and antibody recognition, with intact or permeabilized membranes, have been extensively used for mapping eukaryotic multispanning membrane proteins [10, 63, 72]. N-glycosylation scanning mutagenesis has been useful for eukaryotic proteins [5, 41, 42, 64]. More recently, membrane permeant and impermeant thiol reagents have been used with cysteine scanning mutagenesis [36].

Though powerful, these methods have their individual drawbacks, and those which rely on changing the sequence of the mapped protein, whether by insertion, deletion, truncation, fusion or missense mutation, all implicitly hinge on the presumption that local topology remains unperturbed by the changes imposed. Here we consider just those methods which rely upon changes to the encoding polynucleotide sequence. The smaller the change introduced, the less likely topology will be perturbed.

Cysteine scanning mutagenesis seems the least perturbing mutagenic method, owing to cysteine's small and relatively nonpolar nature. The unique chemical reactivity of the sulfhydryl group makes this method particularly attractive, although steric hindrance may be problematic [58, 61]. The protein wild type to be analyzed should be free of cysteine residues or be mutated to eliminate them, which may destroy function and compromise confidence in subsequent mapping results. The P-glycoprotein, for example, was successfully mapped by cysteine scanning mutagenesis after eliminating naturally occurring cysteines [36]. Although human SGLT1 contains 15 cysteines, eight residing extracellularly, impermeant rhodamine maleimide did not label intact oocytes expressing SGLT1 more than noninjected oocytes, consistent with anticipated extracellular cystine disulfides. The mutant Arg457Cys, presenting an extracellular sulfhydryl in the loop between spans 10 and 11, was labelled proportionate with its level of expression (D. Loo, B.H. Hirayama, E.M. Wright, *unpublished*), indicating that extracellular cysteine scanning mutagenesis may sometimes be cautiously applied without eliminating wild-type cysteines.

N-glycosylation scanning mutagenesis can be nonperturbing, often requiring mutation of only a single residue to generate an N-glycosylation consensus (NXS/T). However the consensus must reside sufficiently far (>12–14 residues) from the membrane interface [40], and even then the N-glycosyltransferase may fail to glycosylate it, due presumably to steric hindrance by secondary structure [5]. Insertion of additional residues to distance the consensus from the membrane, increase steric recognition, or both, is often required [64], but increases the possibility of topology perturbation.

Mutations near the N-terminal span introduce the special problem that the distribution of positive charges may be altered, thereby reversing span orientation during the first insertion event [12, 18, 47, 69]. This phenomenon has prevented the elucidation of the orientation of the N-terminal span (as well as leaving indeterminate whether one or two spans occur there) in at least one transporter [62]. As described earlier, the N-terminus of SGLT1 was similarly perturbed by an insertion near the signal anchor [64].

The expression of a reporter enzyme fused to progressive C-truncations of the protein (truncation/fusion methodology) is potentially the most perturbing method of mapping topology, since extensive polytopic regions are replaced by a large hydrophilic reporter. Indeed, fusing the reporter at a point N-terminal to topogenic (in the wild type) positively charged residues in intracellular loops causes inversion of the local topology [6, 67]. And yet, notably, the 12-span topology of lac permease was successfully determined by this means [8, 28]. Several members of the bacterial phosphotransferase system permease family, including the *E. coli* mannitol [59], mannose [25] and glucose permeases [7], have also been mapped by this strategy. The mannose and mannitol permeases were determined to incorporate six spans, and the glucose permease eight, in contrast to the general trend of many apparently unrelated transporter families to exhibit a 12-span motif. We examine truncation/fusion mapping here in some depth because of its widespread application to many transporters, and because we observe that it apparently fails to detect pairs of helices in many of them, yielding a false model of membrane topology.

Experiments in *E. coli* demonstrate that domains within the same protein can insert into the membrane by different mechanisms ([34]; *see* [67] *for review*). Typically, a pair of spans inserts thermodynamically *en bloc* in a ‘helical hairpin’ configuration (independent of the *Sec*-translocation machinery) when the intervening hydrophilic loop to be translocated is short (<25 residues) and bears few positively charged residues [2, 3, 9, 15]. When the translocated loop is long (>60 residues), the loop and its ‘outgoing’ span are fully *Sec*-dependent for insertion and translocation, with the C-terminal ‘incoming’ span acting in a stop-transfer function [2, 9, 32]. Sequences of intermediate length are partially *Sec*-dependent and are translocated inefficiently [2]; tellingly, they are also under-represented in nature [68].

For bacterial polytopic membrane proteins then, outgoing spans may be classed by their degree of *Sec*-dependence during insertion in the wild type. Therefore, with regard to topological mapping experiments using truncation/fusions, the translocation of a reporter enzyme fused at a point following a ‘‘Sec_{wt}-independent’’ outgoing transmembrane span (one which in the wild type is inserted *Sec*-independently, paired with its counterpart incoming span) can be viewed as a fortuitous accident which occurred only because the hydrophobicity, sequence context or other characteristic [54] of the outgoing span was sufficiently enabling for recognition by, and activation of, the *Sec* translocation machinery. For example, like lac permease, the melibiose permease of *E. coli* was found by truncation/fusions to comprise 12 spans, with all translocated loops no longer than 23 residues [50], which indicates that every span most likely inserts *Sec*-independently in the wild type. Analyses by Memsat and PredictProtein concur with the 12-span model. The entire melibiose permease was in fact recently shown by several experimental approaches to insert fully *Sec*-independently [4]. Therefore, the truncation/fusion mapping experiments had succeeded despite the completely different insertion mechanisms used in the experimental and wild type situations. Can such a fortuitously correct result be expected from the truncation/fusion mapping of the topology of all or most other transporters?

While comparing SGLT1 and the N-acetylglucosamine phosphotransferase system permease for Fig. 9, we noted a gross discrepancy between the permease’s putative topology as predicted by PredictProtein and Memsat (12 spans) and the experimental topology (8 spans) of its homologue, the glucose phosphotransferase system permease (44% identity). Figure 10 compares the glucose permease 8-span topology, as determined by truncation/fusions [7] with a 12-span topology indicated by PredictProtein and Memsat. The PredictProtein plot of span probability in the upper panel shows strong indications of several spans flanking the experimentally determined pair of spans labelled 7e and 8e. For the predicted spans labelled 8 and 12, there correspond very prominent, broad peaks in each of the three hydrophobicity plots (IFH, Eisenberg and Kyte/Doolittle scales). In fact, predicted spans 8 and 12 correspond to domains more hydrophobic, as measured by any of the three hydrophobicity scales, than most of the experimentally determined spans (Fig. 9). How is it that such hydrophobic domains are experimentally nonspanning? Spans 8 and 12 share two features which may bear upon these peculiar results: (i) they are ‘incoming’ spans, and (ii) the predicted ‘outgoing’ span immediately preceding either of them (spans 7 and 11, Fig. 9) is of modest (and C-terminally decreasing) hydrophobicity.

Probing the glucose permease by progressively truncating the C-terminus and fusing it to phosphatase or beta-galactosidase [7] results in expression of a moderately hydrophobic outgoing span (7 or 11) isolated from its insertion-assisting incoming span. Because of their moderate hydrophobicity or some aspect of their sequence context [54], spans 7 and 11 may be ineffective in being recognized by and activating the *Sec*-machinery

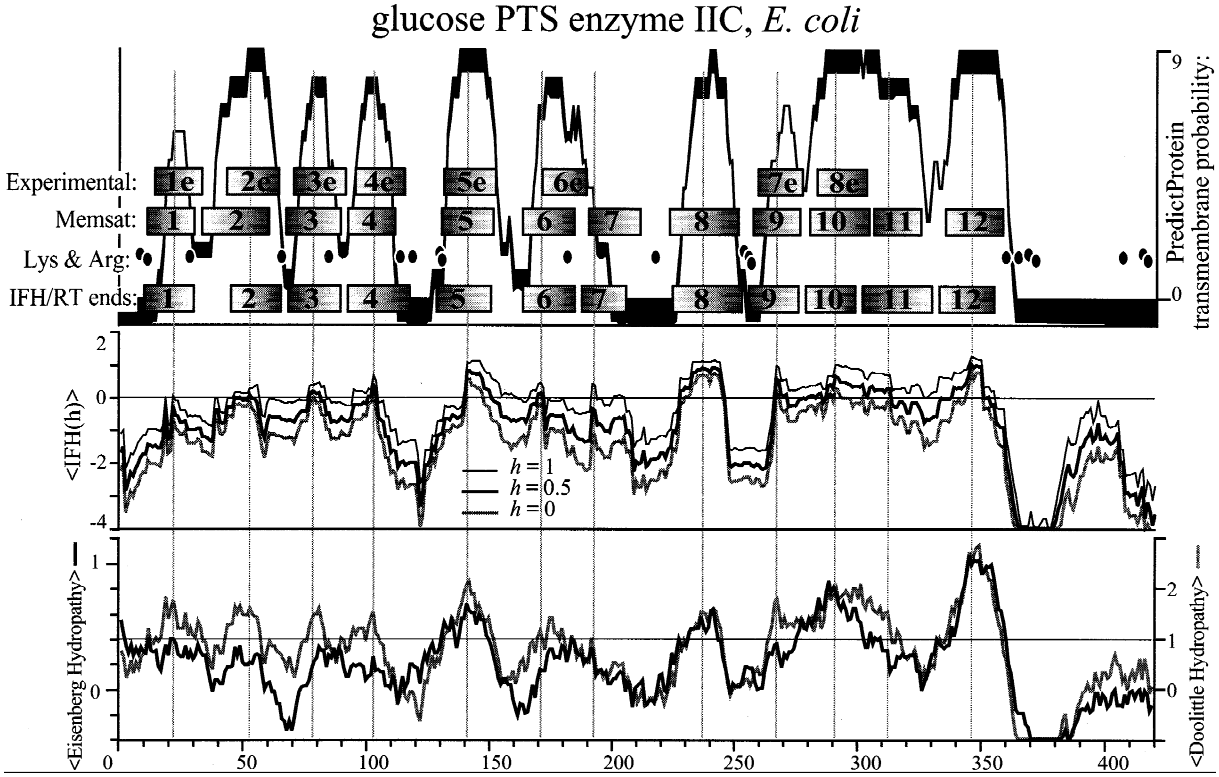


Fig. 10. The large disparity between (i) the experimentally determined topology of the *E. coli* glucose PTS permease and (ii) its putative topology as indicated by each of two computer algorithms and three hydrophobicity plots, encourages the conjecture that protein fusion experiments intended to reveal topology may sometimes instead so perturb it as to entirely miss pairs of adjacent membrane spans. The three panels show analyses of the enzyme IIC (transporter) domain of the *E. coli* glucose phosphotransferase system (P05053, SwissProt). The lowest panel plots hydrophobicity (19 AA window) as measured using the Eisenberg [14] or Doolittle [33] hydrophobicity scales. The middle panel plots interfacial hydrophobicity (19 AA window) determined for three assumptions of side chain-side chain satisfaction of hydrogen-bonds; none ($h = 0$, lowest curve) 50% ($h = 0.5$, middle curve) and complete ($h = 1$, upper curve). Vertical dashed lines have been drawn through the residue corresponding to the highest value of IFH(0.5) for each. The top panel plots probability (arbitrary units) of transmembrane helix content as determined by application of the trained neural network PredictProtein [53] to a multiple sequence alignment of the *E. coli* glucose transporter and 15 other close homologues in the phosphotransferase system family. The reliability (certainty) of the prediction at each residue is indicated by the thickness of the plot, thickness increasing with reliability. The four rows of schematically depicted data in the top panel represent, starting from the top: *Experimental* shows the membrane spans detected by phosphatase and beta-galactosidase fusions to C-terminally truncated *E. coli* glucose transporter IIC domain [7]; *Memsat* shows the highest scoring prediction by that algorithm, consistent with the experimentally determined orientations of the N- and C-termini and with the PredictProtein probabilities; *Lys & Arg* shows the occurrences of these two residues as ellipses; and *IFH/RT ends* shows the approximate transmembrane helix ends, as determined by interfacial hydrophobicity and reverse turn propensity [71] of a set of 12 transmembrane helices inferred by cross-checking (i) the PredictProtein probability plot (ii) the highest scoring Memsat predictions, (iii) the IFH(h) hydrophobicity plot, (iv) the distribution of Lys and Arg residues ('positive-inside' rule), and (v) experimental data [7] on the N-terminus.

for translocation of the reporter. (We note that outgoing helices 7 and 9 of lac permease (Fig. 5) are also of low hydrophobicity, yet were well mapped by reporter fusions.)

A consequence of a ''Sec_{wt}-independent'' outgoing span's incapacity to activate the *Sec*-machinery for reporter translocation is that there the truncation/fusion methodology will miss entirely a *pair* of helices constituting that helical hairpin insertion in the wild type. We conjecture that the homologous N-acetylglucosamine and glucose permeases each incorporate 12 spans, two more pairs of helices (four) in addition to the eight detected experimentally in the latter. The tight localization of three positively charged residues to the cytoplasmic loop between putative spans 8 and 9 (positive-inside rule [66–68]) further supports the existence of spans 7 and 8 between spans 6e and 7e (Fig. 10). Similarly, the cluster of 4 positive charges appearing immediately after span 12 is consistent with the presence of spans 11 and 12 (Fig. 10). Our analysis of other bacterial transport proteins that have been topologically mapped by the progressive truncation/fusion method indicates that other pairs of spans may have gone undetected, leading us to suggest that some proteins mapped by the translocated reporter method may well need reevaluation with a less perturbing methodology. For example, phoA truncation/

fusion experiments on the mannitol permease of *E. coli* (a phosphotransferase IIC domain) indicate six spans [59], yet application of Memsat and PredictProtein (*not shown*) indicates as many as 12. In fact, reexamination of the original phoA experimental results reveals modest phosphatase activity peaks at two regions in the protein sequence which corresponds to the midpoints of two predicted hairpin-span pairs (near residues 90 and 220 in Fig. 3 of reference [59]), thus indicating at least 4 additional spans.

The 42-residue insert used to map the topology of SGLT1 [64] presumably also required the eukaryotic translocation machinery-equivalent of the *Sec* mechanism for transmembrane movement into the ER lumen. However, the analyses by PredictProtein and Memsat indicate that the experimental topology is indeed correct, and therefore that, like lac permease, each outgoing span preceding the insert was of a nature appropriate for recognition by and activation of the translocation machinery.

Conclusions

The SGLT cotransporter family members are diverse in size (477–830 residues) and in sequence, retaining only four invariant residues. The most common topological motif of 13 transmembrane spans occurs in the NIS, putP, panF and HypE59 phylogenetic branches. A 14th C-terminal span appears in the SGLT1 branch members and *Vibrio* SGLTV, and a 15th C-terminal span appears in bacterial HypE62 and yeast DUR3. All members retain an extracellular N-terminus except HypAeu, which incorporates an additional N-terminal span. An N-terminal signal peptide appears in HypE59.

Graphical representation of the output of the neural network Predict Protein has been of greater value than hydrophobicity plots alone in assessing putative membrane topologies, particularly when checked against the topology-constrained predictions of the Memsat algorithm. The programs' accurate prediction of the topology of a difficult and well-characterized protein, the lac permease, illustrates this effectively. Span predictions are useful for the design of experiments aimed to determine helix topology and stacking within the membrane. Based on these computational procedures, we conjecture that protein truncation/fusion experiments intended to reveal topology may sometimes instead so perturb it as to entirely miss adjacent pairs of membrane spans. Some members of the phosphotransferase permease family, for example, mapped by truncation/fusions may need topology reevaluation by a method such as cysteine scanning mutagenesis which does not eliminate entire spans when the reporter is incorporated, and we encourage the pursuit of such experiments to verify or falsify our conjecture. We note that the topological analysis of transporter proteins, whether by hydrophobicity plots, computer algorithms or experimental mutations, is inherently more difficult than that of other membrane proteins by virtue of transporters' functional requirements for hydrophilic residues in their spanning domains in order to enable transmembrane passage of aqueous solutes.

Topology comparisons can be of use in deciding questions of homology. Sequence comparisons described here of SGLT and [Na^+ + Cl^-]-coupled GABA transporter family members indicate their homology. The similar function of these families, i.e., utilization of the transmembrane sodium gradient to effect solute uptake, together with the concurrence of the respective local and macroscopic topologies to which the compared sequence fragments map, virtually certify the common ancestry of these families. A sequence fragment comparison strongly indicates homology of the SGLT and PTS permease families, but discrepancies in the comparison of the local and macroscopic topologies to which the fragments map make a claim for common ancestry less certain.

Knowledge of span topology should assist the design of experiments to determine helix bundling. Short of obtaining protein crystals for X-ray crystallography, the relative proximities of membrane spans will be of considerable interest.

We thank R.F. Doolittle for kindly providing a preprint copy of his chapter in Methods in Enzymology [17]. This work was supported by the National Institutes of Health, grants DK-19567, DK-44582 and NS-25554.

References

1. Anonymous 1994. Program Manual for the Wisconsin Package, Version 8, Genetics Computer Group, Madison, Wisconsin
2. Andersson, H., von Heijne, G. 1993. Sec dependent and sec independent assembly of *E. coli* inner membrane proteins: the topological rules depend on chain length. *EMBO J.* **12:**683–691
3. Andersson, H., von Heijne, G. 1994. Positively charged residues influence the degree of SecA dependence in protein translocation across the *E. coli* inner membrane. *FEBS Lett.* **347:**169–172
4. Bassilana, M., Gwizdek, C. 1996. In vivo membrane assembly of the *E. coli* polytopic protein, melibiose permease, occurs via a Sec-independent process which requires the protonmotive force. *EMBO J.* **15:**5202–5208
5. Bennett, E.R., Kanner, B.I. 1997. The membrane topology of GAT-1, a [Na^+ + Cl^-]-coupled gamma-aminobutyric acid transporter from rat brain. *J. Biol. Chem.* **272:**1203–1210
6. Boyd, D., Traxler, B., Jander, G., Prinz, W., Beckwith, J. 1993. Gene fusion approaches to membrane protein topology. *Soc. Gen. Physiol. Ser.* **48:**23–37
7. Buhr, A., Erni, B. 1993. Membrane topology of the glucose transporter of *Escherichia coli. J. Biol. Chem.* **268:**11599–11603
8. Calamia, J., Manoil, C. 1990. lac permease of *Escherichia coli:* topology and sequence elements promoting membrane insertion. *Proc. Natl. Acad. Sci. USA* **87:**4937–4941
9. Cao, G., Cheng, S., Whitley, P., von Heijne, G., Kuhn, A., Dalbey, R.E. 1994. Synergistic insertion of two hydrophobic regions drives

Sec-independent membrane protein assembly. *J. Biol. Chem.* **269:**26898–26903

10. Conti-Fine, B.M., Lei, S.J., McLane, K.E. 1996. Antibodies as tools to study the structure of membrane proteins: The case of the nicotinic acetylcholine receptor. *Annu. Rev. Biophys. Biomol. Struct.* **25:**197–229
11. Creighton, T.E. 1993. Proteins: Structures and Molecular Properties. W.H. Freeman, New York
12. Dalbey, R.E., Kuhn, A., von Heijne, G. 1995. Directioality in protein translocation across membranes: the N-tail phenomenon. *Trends Cell Biol.* **5:**380–383
13. Doolittle, R.F. 1986. Of Urfs and Orfs: A Primer on How to Analyze Derived Amino Acid Sequences. University Science Books, Mill Valley, CA
14. Eisenberg, D., Schwarz, E., Komaromy, M., Wall, R. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179:**125–142
15. Engelman, D.M., Steitz, T.A. 1981. The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. *Cell* **23:**411–422
16. Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) Version 3.5c, Distributed by the author. Dept. Genetics, U. Washington, Seattle
17. Feng, D.F., Doolittle, R.F. 1996. Progressive alignment of amino acid sequences and construction of phylogenetic trees from them. *Methods Enzymol.* **266:**368–382
18. Gafvelin, G., Sakaguchi, M., Andersson, H., von Heijne, G. 1997. Topological rules for membrane protein assembly in eukaryotic cells. *J. Biol. Chem.* **272:**6119–6127
19. Gafvelin, G., von Heijne, G. 1994. Topological ‘‘frustration’’ in multispanning. *E. coli* inner membrane proteins. *Cell* **77:**401–412
20. Hanada, K., Yoshida, T., Yamato, I., Anraku, Y. 1992. Sodium ion and proline binding sites in the Na^+/proline symport carrier of *Escherichia coli. Biochim. Biophys. Acta* **1105:**61–66
21. Hartmann, E., Rapoport, T.A., Lodish, H.F. 1989. Predicting the orientation of eukaryotic membrane-spanning proteins. *Proc. Natl. Acad. Sci. USA* **86:**5786–5790
22. Hediger, M.A., Coady, M.J., Ikeda, T.S., Wright, E.M. 1987. Expression cloning and cDNA sequencing of the Na^+/glucose cotransporter. *Nature* **330:**379–381
23. Hirayama, B.A., Loo, D.D.F., Wright, E.M. 1997. Cation effects on protein conformation and transport in the Na^+/glucose cotransporter. *J. Biol. Chem.* **272:**2110–2115
24. Hirayama, B.A., Lostao, M.P., Panayotova-Heiermann, M., Loo, D.D.F., Turk, E., Wright, E.M. 1996. Kinetic and specificity differences between rat, human, and rabbit Na^+-glucose cotransporters (SGLT-1). *Am. J. Physiol.* **270:**G919–G926
25. Huber, F., Erni, B. 1996. Membrane topology of the mannose transporter of *Escherichia coli* K12. *Eur. J. Biochem.* **239:**810–817
26. Jacobs, R.E., White, S.H. 1989. The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. *Biochemistry* **28:**3421–3437
27. Jones, D.T., Taylor, W.R., Thornton, J.M. 1994. A model recognition approach to the prediction of all-helical membarne protein structure and topology. *Biochemistry* **33:**3038–3049
28. Kaback, H.R., Frillingos, S., Jung, H., Jung, K., Prive, G.G., Ujwal, M.L., Weitzman, C., Wu, J., Zen, K. 1994. The lactose permease meets Frankenstein. *J. Exp. Biol.* **196:**183–195
29. Kanai, Y., Lee, W.S., You, G., Brown, D., Hediger, M.A. 1994. The human kidney low affinity Na^+/glucose cotransporter SGLT2. Delineation of the major renal reabsorptive mechanism for D-glucose. *J. Clin. Invest.* **93:**397–404
30. Komeiji, Y., Hanada, K., Yamato, I., Anraku, Y. 1989. Orientation of the carboxyl terminus of the Na^+/proline symport carrier in *Escherichia coli. FEBS Lett.* **256:**135–138
31. Kong, C.T., Yet, S.F., Lever, J.E. 1993. Cloning and expression of a mammalian Na^+/amino acid cotransporter with sequence similarity to Na^+/glucose cotransporters. *J. Biol. Chem.* **268:**1509–1512
32. Kuhn, A. 1988. Alterations in the extracellular domain of M13 procoat protein make its membrane insertion dependent on secA and secY. *Eur. J. Biochem.* **177:**267–271
33. Kyte, J., Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:**105–132
34. Lee, J.I., Kuhn, A., Dalbey, R.E. 1992. Distinct domains of an oligotopic membrane protein are Sec-dependent and Sec-independent for membrane insertion. *J. Biol. Chem.* **267:**938–943
35. Loo, D.D.F., Zeuthen, T., Chandy, G., Wright, E.M. 1996. Cotransport of water by the Na^+/glucose cotransporter. *Proc. Natl. Acad. Sci. USA* **93:**13367–13370
36. Loo, T.W., Clarke, D.M. 1995. Membrane topology of a cysteineless mutant of human P-glycoprotein, *J. Biol. Chem.* **270:**843–848
37. Mackenzie, B., Panayotova-Heiermann, M., Loo, D.D.F., Lever, J.E., Wright, E.M. 1994. SAAT1 is a low affinity Na^+/glucose cotransporter and not an amino acid transporter. A reinterpretation. *J. Biol. Chem.* **269:**22488–22491
38. Marger, M.D., Saier, M.H. 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport [see comments]. *Trends Biochem. Sci.* **18:**13–20
39. Nakashima, H., Nishikawa, K. 1992. The amino acid composition is different between the cytoplasmic and extracellular sides in membrane proteins. *FEBS Lett.* **303:**141–146
40. Nilsson, I.M., von Heijne, G. 1993. Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.* **268:**5798–5801
41. Olender, E.H., Simoni, R.D. 1992. The intracellular targeting and membrane topology of 3-hydroxy-3-methylglutaryl-CoA reductase. *J. Biol. Chem.* **267:**4223–4235
42. Olivares, L., Aragon, C., Gimenez, C., Zafra, F. 1997. Analysis of the transmembrane topology of the glycine transporter GLYT1. *J. Biol. Chem.* **272:**1211–1217
43. Pajor, A.M., Wright, E.M. 1992. Cloning and functional expression of a mammalian Na^+/nucleoside cotransporter. *J. Biol. Chem.* **267:**3557–3560
44. Panayotova-Heiermann, M., Loo, D.D.F., Wright, E.M. 1995. Kinetics of steady-state currents and charge movements associated with the rat Na^+/glucose cotransporter. *J. Biol. Chem.* **270:**27099–27105
45. Parent, L., Supplisson, S., Loo, D.D.F., Wright, E.M. 1992. Electrogenic properties of the cloned Na^+/glucose cotransporter: I. Voltage-clamp studies. *J. Membrane Biol.* **125:**49–62
46. Parent, L., Supplisson, S., Loo, D.D.F., Wright, E.M. 1992. Electrogenic properties of the cloned Na^+/glucose cotransporter: II. A transport model under nonrapid equilibrium conditions. *J. Membrane Biol.* **125:**63–79
47. Parks, G.D., Lamb, R.A. 1993. Role of NH_2-terminal positively charged residues in establishing membrane protein topology. *J. Biol. Chem.* **268:**19101–19109
48. Pearson, W.R., Miller, W. 1992. Dynamic programming algorithms for biological sequence comparison. *Methods Enzymol.* **210:**575–601
49. Persson, B., Argos, P. 1996. Topology prediction of membrane proteins. *Protein Science* **5:**363–371
50. Pourcher, T., Bibi, E., Kaback, H.R., Leblanc, G. 1996. Membrane topology of the melibiose permease of *Escherichia coli* studied by *melB-phoA* fusion analysis. *Biochemistry* **35:**4161–4168
51. Quick, M., Tebbe, S., Jung, H. 1996. Ser57 in the Na^+/proline permease of *Escherichia coli* is critical for high-affinity proline uptake. *Eur. J. Biochem.* **239:**732–736

52. Reizer, J., Reizer, A., Saier, M.H. 1990. The Na$^+$/pantothenate symporter (PanF) of *Escherichia coil* is homologous to the Na$^+$/proline symporter (PutP) of *E. coli* and the Na$^+$/glucose symporters of mammals. *Res. Microbiol.* **141:**1069–1072
53. Rost, B., Sander, C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19:**55–72
54. Saaf, A., Andersson, H., Gafvelin, G., von Heijne, G. 1995. SecA-dependence of the translocation of a large periplasmic loop in the *Escherichia coli* MalF inner membrane protein is a function of sequence context. *Mol. Membr. Biol.* **12:**209–215
55. Saier, M.H. 1994. Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution. *Microbiol. Rev.* **58:**71–93
56. Sarker, R.I., Ogawa, W., Shimamoto, T., Tsuchiya, T. 1997. Primary structure and properties of the Na$^+$/glucose symporter (Sg1S) of *Vibrio parahaemolyticus. J. Bacteriol.* **179:**1805–1808
57. Sarker, R.I., Ogawa, W., Tsuda, M., Tanaka, S., Tsuchiya, T. 1996. Properties of a Na$^+$/galactose (glucose) symport system in *Vibrio parahaemolyticus. Biochim. Biophys. Acta* **1279:**149–156
58. Spruijt, R.B., Wolfs, C.J., Verver, J.W., Hemminga, M.A. 1996. Accessibility and environment probing using cysteine residues introduced along the putative transmembrane domain of the major coat protein of bacteriophage M13. *Biochemistry* **35:**10383–10391
59. Sugiyama, J.E., Mahmoodian, S., Jacobson, G.R. 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
60. Takata, K., Kasahara, T., Kasahara, M., Ezaki, O., Hirano, H. 1991. Localization of Na$^+$-dependent active type and erythrocyte/HepG2-glucose transporters in rat kidney: immunofluorescence and immunogold study. *J. Histochem. Cytochem.* **39:**287–298
61. Takeda, K., Shigemura, A., Hamada, S., Gu, W., Fang, D., Sasa, K., Hachiya, K. 1992. Dependence of reaction rate of 5,5′-dithiobis-(2-nitrobenzoic acid) to free sulfhydryl groups of bovine serum albumin and ovalbumin on the protein conformations. *J. Protein Chem.* **11:**187–192
62. Tate, C.G., Henderson, P.J. 1993. Membrane topology of the L-rhamnose-H$^+$ transport protein (RhaT) from enterobacteria. *J. Biol. Chem.* **268:**26850–26857
63. Traxler, B., Boyd, D., Beckwith, J. 1993. The topological analysis of intergral cytoplasmic membrane proteins. *J. Membrane Biol.* **132:**1–11.
64. Turk, E., Kerner, C.J., Lostao, M.P., Wright, E.M. 1996. Membrane topology of the human Na$^+$/glucose cotransporter SGLT1. *J. Biol. Chem.* **271:**1925–1934
65. Turner, J.R., Lencer, W.I., Carlson, S., Madara, J.L. 1996. Carboxy-terminal vesicular stomatitis virus G protein-tagged intestinal Na$^+$-dependent glucose cotransporter (SGLT1): maintenance of surface expression and global transport function with selective perturbation of transport kinetics and polarized expression. *J. Biol. Chem.* **271:**7738–7744
66. von Heijne, G. 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341:**456–458
67. von Heijne, G. 1994. Membrane proteins: from sequence to structure. *Annu. Rev. Biophys. Biomol. Struct.* **23:**167–192
68. von Heijne, G., Gavel, Y. 1988. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **174:**671–678
69. Wallin, E., von Heijne, G. 1995. Properties of N-terminal tails in G-protein coupled receptors: a statistical study. *Protein Eng.* **8:**693–698
70. White, S.H. 1994. *Membrane Protein Structure* (White, S.H., ed) pp. 97–124, Oxford University Press, New York
71. White, S.H., Jacobs, R.E. 1990. Observations concerning topology and locations of helix ends of membrane proteins of known structure. *J. Membrane Biol.* **115:**145–158
72. Wilkinson, B.M., Critchley, A.J., Stirling, C.J. 1996. Determination of the transmembrane topology of yeast Sec61p, an essential component of the endoplasmic reticulum translocation complex. *J. Biol. Chem.* **271:**25590–25597
73. Wright, E.M., Loo, D.D.F., Turk, E., Hirayama, B.A. 1996. Sodium cotransporters. *Curr. Opin. Cell Biol.* **8:**468–473
74. Yamato, I., Kotani, M., Oka, Y., Anraku, Y. 1994. Site-specific alteration of arginine 376, the unique positively charged amino acid residue in the mid-membrane-spanning regions of the proline carrier of *Escherichia coli. J. Biol. Chem.* **269:**5720–5724
75. Yamauchi, A., Uchida, S., Kwon, H.M., Preston, A.S., Robey, R.B., Garcia-Perez, A., Burg, M.B., Handler, J.S. 1992. Cloning of a Na$^+$- and Cl$^-$-dependent betaine transporter that is regulated by hypertonicity. *J. Biol. Chem.* **267:**649–652