Protein Science (1995), *4*:534–537. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Structural features of the uniporter/symporter/antiporter superfamily

VISALA CHEPURI GOSWITZ AND ROBERT J. BROOKER

Department of Genetics and Cell Biology and the Institute for Advanced Studies in Biological Process Technology, University of Minnesota, St. Paul, Minnesota 55108

(RECEIVED October 4, 1994; ACCEPTED December 16, 1994)

Abstract: The uniporter/symporter/antiporter superfamily is an evolutionarily related group of solute transporters. For the entire superfamily, we have used a new predictive program to identify the transmembrane domains. These transmembrane domains were then analyzed with regard to their overall hydrophobicity and amphipathicity. In addition, the lengths of the hydrophilic loops connecting the transmembrane domains were calculated. These data, together with structural information in the literature, were collectively used to produce a general model for the three-dimensional arrangement of the transmembrane domains.

Keywords: antiporter; lactose permease; symporter; uniporter

It has recently been noted that a large group of solute transporters are evolutionarily related to each other (Henderson, 1990; Griffith et al., 1992; Marger & Saier, 1993). From a bioenergetic point of view, this superfamily is particularly interesting and unusual. It contains members that function as uniporters, symporters, or antiporters. In addition, the solute specificity of the USA superfamily is also diverse. Substrates that are transported include antibiotics (i.e., H^+/tetracycline antiporter), sugars (e.g., mammalian glucose carrier and the H^+/lactose permease of *Escherichia coli*), Krebs cycle intermediates (i.e., H^+/citrate symporter), and phosphate/phosphate esters (i.e., hexosephosphate antiporter). It is worthwhile to note, however, that most Na^+/solute cotransporters do not appear to be evolutionarily related even though the USA superfamily does include a few Na^+/solute symporters (Reizer et al., 1994).

At the secondary structural level, hydropathy considerations have suggested that many members of the USA superfamily contain 12 transmembrane domains that are proposed to traverse the plasma membrane in an α-helical conformation. The overall structural pattern is consistent with a primordial gene containing 6 transmembrane domains that duplicated and fused to produce a protein containing 12 transmembrane domains (Maiden et al., 1987). Such an evolutionary event might be quite favorable if the primordial protein was a functional dimer. A gene duplication would thereby allow each half of the dimer to evolve independently of each other. This hypothesis is consistent with experiments conducted on the lactose permease. This protein has been shown to function as a monomer (Wright et al., 1983). Furthermore, when the two halves of the permease are expressed from different promoters, they are inserted into the membrane and assemble into a functional protein (Bibi & Kaback, 1990).

In our analysis aimed at proposing a tertiary arrangement for the transmembrane domains, we have considered 65 members of the USA superfamily. From evolutionary relationships, this group had been previously divided into five families and referred to as the major facilitator superfamily (MFS; see Marger & Saier, 1993). For our analysis, we have further divided family 1 into two subfamilies based upon sequence similarities. In addition, the large family 2 has been divided into bacterial, fungal/plant, and mammalian subgroups.

Discussion

Transmembrane domains of the USA superfamily differ with regard to hydrophobicity

For all of the members of the USA superfamily, we applied a new program to identify the segments within each protein that are likely to be transmembrane (i.e., MEMSAT; Jones et al., 1994). After we had identified the transmembrane domains, our first goal was to identify potential channel-lining domains. To accomplish this, we calculated the transmembrane hydrophobic indices according to Kyte and Doolittle (1982). In Table 1, the segments that are the most likely to be channel-lining domains (i.e., the least hydrophobic) are highlighted in bold type. Several interesting observations can be made. First, TM-3, TM-6, TM-9, and TM-12 tend to be fairly hydrophobic within all of the eight groups as well as the superfamily itself. Therefore, these domains are not likely to be channel-lining domains. In contrast, domains 2, 4, 5, 10, and 11 showed fairly uniform

Reprint requests to: Robert J. Brooker, Bioprocess Technical Institute, 240 Gortner Laboratories, 1479 Gortner Avenue, St. Paul, Minnesota 55108; e-mail: robert-b@molbio.cbs.umn.edu.

Table 1. *Hydrophobic index values for transmembrane domains*[a]

Group[b]	Transmembrane domain 1	2	3	4	5	6	7	8	9	10	11	12	Mean	*hsd*[c]
Family 1-A	**−2.3**	−1.8	**−2.0**	−1.6	−1.4	**−2.2**	**−2.1**	−1.6	**−2.1**	−1.9	−1.5	**−2.3**	−1.9	0.38/0.51
Family 1-B	**−2.0**	−1.3	**−2.1**	−1.8	−1.7	**−2.0**	**−2.0**	−1.9	**−2.1**	−1.9	−1.4	**−2.3**	−1.9	0.40/0.45
Family 2-A	−1.9	−1.7	−1.7	−1.7	−1.9	−1.9	−1.8	**−2.2**	−1.9	−1.8	−1.3	−1.7	−1.8	0.51/0.69
Family 2-B	−1.9	−1.8	**−2.3**	−1.7	−1.4	**−2.0**	−0.7	−1.9	**−2.0**	**−2.0**	−1.4	**−2.2**	−1.8	0.48/0.54
Family 2-C	−1.7	−1.7	−1.9	−1.6	−1.9	**−2.0**	−1.2	−1.8	**−2.2**	−1.8	−1.3	**−2.5**	−1.8	0.15/0.19
Family 3	−1.2	−1.6	**−2.3**	−1.5	**−2.0**	−1.7	−1.6	**−2.0**	**−2.2**	−1.9	−1.6	−1.9	−1.8	0.26/0.34
Family 4	−1.4	−1.3	**−2.1**	−1.6	−1.5	**−2.4**	−1.8	−1.2	**−2.1**	**−2.0**	−1.1	**−2.3**	−1.8	0.63/0.73
Family 5	−1.7	−1.9	**−2.2**	−1.9	−1.8	**−2.1**	−1.7	−1.7	−1.8	−1.9	−1.5	**−2.3**	−1.9	0.45/0.66
Superfamily	−1.7	−1.6	**−2.1**	−1.7	−1.7	**−2.1**	−1.7	−1.8	**−2.0**	−1.9	−1.4	**−2.2**	−1.8	0.40/0.46

[a] Hydrophobic indices were calculated as described in the Electronic Appendix. Values of −2.0 and below are presented in bold type.

[b] The 65 family members are described in Table S1 in the Electronic Appendix. Family 1-A and 1-B are bacterial antibiotic antiporters; family 2-A are bacterial hexose symporters; family 2-B are fungal and plant sugar symporters; family 2-C are mammalian glucose facilitators; family 3 are Krebs cycle (i.e., citrate) symporters; family 4 are sugar/phosphate antiporters; and family 5 are disaccharide (i.e., lactose) and trisaccharide symporters. Families 3, 4, and 5 are all bacterial transporters.

[c] To compare the values within a group, a multiple comparison Tukey test was conducted. The *hsd* value indicates the amount by which two values (within a row) must differ to be judged significantly different. The upper and lower values are at the 5% and 1% significance levels, respectively.

values, which were significantly less hydrophobic compared with segments 3, 6, 9, and 12. Segment 11, in particular, was commonly observed to be the least hydrophobic domain among many of the groups. Therefore, domains 2, 4, 5, 10, and 11 are good candidates for channel-lining domains.

Among the remaining segments (e.g., 1, 7, and 8), a greater variation in hydrophobicity values was noted. For example, TM-1 and TM-7 are rather hydrophobic among antibiotic transporters (i.e., family 1-A and 1-B) but not in the other four families. However, when averaged over the entire superfamily, domains 1, 7, and 8 were observed to be substantially less hydrophobic than segments 3, 6, 9, and 12. For this reason, TM-1, TM-7, and TM-8 should also be considered as putative channel-lining domains. One way to explain the variability seen among potential channel-lining domains would be that different groups within the superfamily utilize different transmembrane domains as a hydrophilic surface for solute binding. In the case of the lactose permease, most sugar specificity mutants have been located on transmembrane domains in the second half of the protein (Brooker & Wilson, 1985; Markgraf et al., 1985; Collins et al., 1989; Franco et al., 1989). However, other families may use the first half, or combinations of the first and second half, to form a solute recognition site.

An alternative way to distinguish transmembrane segments is based upon amphipathicity (Eisenberg et al., 1984). The results of this analysis are included in the Electronic Appendix. For the entire superfamily, helices 3, 6, 9, and 12 were observed to be the least amphipathic consistent with the notion that they are not channel-lining domains. (The amphipathic index values were 0.35, 0.34, 0.31, and 0.32, respectively.) Helices 2, 4, and 11 were the most amphipathic (with index values of 0.43, 0.43, and 0.44) and segments 1, 5, 7, 8, and 10 were moderately amphipathic (with index values of 0.39, 0.38, 0.37, 0.40, and 0.36) within the entire superfamily. Taken together, these results are also consistent with a model in which segments 1, 2, 4, 5, 7, 8, 10, and 11 are channel-lining domains, whereas 3, 6, 9, and 12 perform a scaffolding function.

Hydrophilic loop length

The length of the hydrophilic loops connecting transmembrane domains in the USA superfamily may provide clues concerning protein tertiary structure. Short loops would be adequate to connect adjacent transmembrane domains that are close together in the tertiary structure although longer hydrophilic loops could also be involved. However, relatively long loops would be necessary when two transmembrane domains are adjacent in the primary sequence but far apart in the tertiary structure. With these ideas in mind, we calculated the loop lengths between the 12 transmembrane segments in the USA superfamily (see Table S3 in the Electronic Appendix). For the entire superfamily, loops 2/3, 3/4, 5/6, 8/9, 9/10, and 11/12 are relatively short; the average numbers of amino acids in these loops are 8.8, 6.0, 8.1, 8.5, 7.3, and 10.7, respectively. These results are consistent with a model in which transmembrane domains 2 and 3, 3 and 4, 5 and 6, 8 and 9, 9 and 10, and 11 and 12 are adjacent to each other in the tertiary structure. Loops 1/2, 4/5, 7/8, and 10/11 are of moderate length containing 20.8, 16.8, 14.9, and 13.9 amino acids, respectively. These results are compatible with a model in which transmembrane segments 1 and 2, 4 and 5, 7 and 8, and 10 and 11 are not side by side in the tertiary model. Nevertheless, their relative distances apart might not be expected to be as far as the distance between TM-6 and TM-7, which are always connected by a very long loop averaging 52 amino acids in length.

A comparison between the structural features of the photosynthetic reaction center and the USA superfamily

Although we do not have crystallographic data concerning any members of the USA superfamily, it may be the case that nonrelated proteins with α-helical transmembrane domains may share common structural features. In this regard, it is instructive to examine the photosynthetic reaction center (PSRC) from

Rhodopseudomonas viridis even though PSRC is not involved in solute transport. The structures of the L- and M-subunits of PSRC, which form the core of the membrane-spanning domains, are particularly relevant. The L- and M-subunits are homologous and each subunit contains five transmembrane domains folded in an α-helical conformation (Deisenhofer & Michel, 1989). In the crystal structure, these two subunits dimerize to form a complex of 10 transmembrane segments. It is interesting to consider that certain aspects of the 5 + 5 transmembrane arrangement in PSRC may be analogous to the 6 + 6 arrangement in members of the USA superfamily. For example, the structure of the LM dimer has rotational symmetry. The two subunits can be superpositioned by rotation of 180° around an axis running perpendicular to the membrane surface (central local symmetry axis). Therefore, a 180° rotation superpositions transmembrane helix A of the L-subunit over transmembrane helix A of the M-subunit.

A second aspect worth examining in PSRC is the relationship between hydrophilic loop length and the arrangement of the transmembrane helices. The helices in both subunits are arranged side by side in a concave layer in the order A, B, C, E, and D. Three out of four connecting hydrophilic loops range in length from 27 to 36 amino acids. However, a short loop (4 or 5 amino acids) connects transmembrane domains B and C. These two α-helices are arranged side by side in a roughly antiparallel manner. In contrast, transmembrane domains C and D (which are adjacent in the primary sequence) are not adjacent in the tertiary structure. These two helices are connected by a 31-amino acid loop.

A general model for the USA superfamily

Due to their evolutionary relationship, it is expected that members of the USA superfamily will exhibit significant similarities in their overall structure. Therefore, we have used the information presented in this paper to propose a general model for the tertiary arrangement of the 12 transmembrane domains. This model is shown in Figure 1 and is based, in part, on the common structural features within the superfamily. For example, the data concerning hydrophobicity and amphipathicity suggest a model in which helices 3, 6, 9, and 12 do not line the channel, whereas the other eight are potential "channel-lining" candidates. Although studies that identify channel-lining segments are lacking, recent biochemical experiments on *uhpT* (*E. coli*) suggest that transmembrane 7 is in the translocation pathway (Yan & Maloney, 1993). In addition, our model is consistent with the lengths of the hydrophilic connecting loops. Short loop lengths observed between 2/3, 3/4, 5/6, 8/9, 9/10, and 11/12 are connecting transmembrane domains, which would lie side by side and antiparallel to each other. Moderate-sized loops connect domains that are not adjacent in the tertiary structure but, nevertheless, are relatively close to each other. In contrast, TM-6 and TM-7, which are connected by a long hydrophilic loop, are far apart in the tertiary structure.

Aside from general features that are common to the entire USA superfamily, our model is also based on other types of information. Like the LM dimer in PSRC, the tertiary structure depicted in Figure 1 exhibits rotational symmetry between the two halves of the protein (i.e., transmembrane segments 1–6 and 7–12). For example, rotation of 180° around the central local symmetry axis superpositions transmembrane helix 7 over helix 1. And finally, in order to specify the actual arrangement of particular segments, we have used information from mutant studies of the lactose permease of *E. coli*. In that protein, helical interactions have been proposed from studies concerned with putative ionic interactions between certain charged residues. The transmembrane domains that have been suggested to interact are: TM-7/TM-10; TM-7/TM-11; TM-8/TM-10; and TM-9/TM-10 (King et al., 1991; Sahin-Toth et al., 1992; Jung et al., 1993; Lee et al., 1993). In our general model, the information concerning the lactose permease generates a specific arrangement that fits nicely with loop length, the identification of "channel-lining" domains, and rotational symmetry for the two halves of the protein.

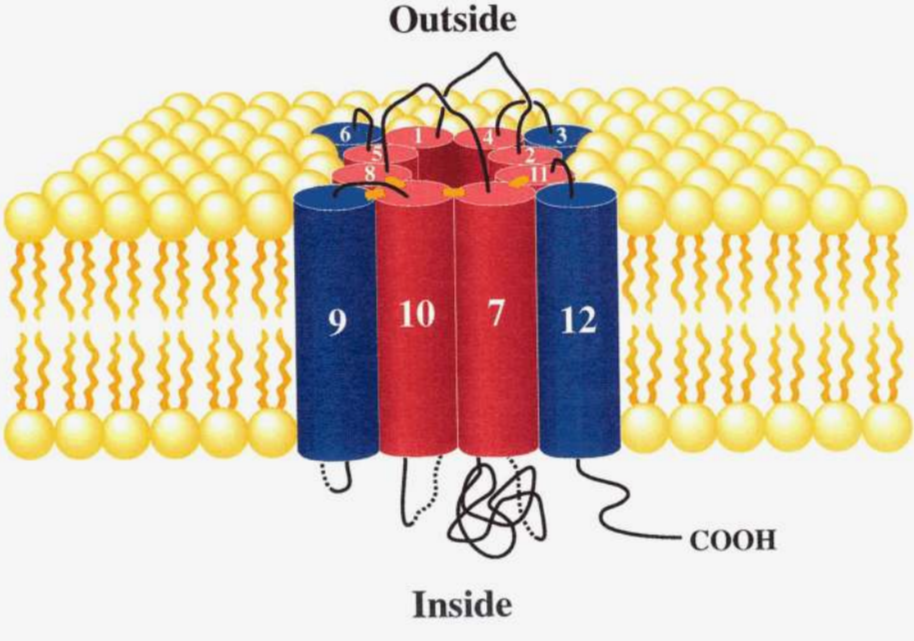

Fig. 1. General model for the arrangement of transmembrane domains within the USA superfamily. The model depicts 12 cylindrical transmembrane domains connected by linear hydrophilic loops. Yellow bars connecting TM-7/TM-10, TM-7/TM-11, TM-8/TM-10, and TM-9/TM-10 are indicative of putative ionic interactions within the lactose permease.

As it is drawn, the model in Figure 1 suggests that eight of the helices line the channel and four do not. However, it is worth considering that there may be different degrees to which a segment may line the channel. For example, certain helices may be pushed away from the channel so that fewer side chains on these helices would have access to the channel. Likewise, even though helices 3, 6, 9, and 12 are very hydrophobic, it is possible that gaps between channel-lining helices could provide limited access of side chains on these four helices to the channel lumen. Finally, the likelihood should be considered that helices may not align perfectly perpendicular to the plane of the lipid bilayer.

Supplementary material in Electronic Appendix

The Electronic Appendix (SUPLEMNT directory, Goswitz.SUP subdirectory) contains an alignment of the amino acid sequences for all 65 members in the USA superfamily that we have analyzed. The predicted transmembrane domains are underlined. The Electronic Appendix also contains methods for the calculation of transmembrane domains, hydrophobicity, amphipathicity, and loop length. Table S1 lists the 65 members of the superfamily; Table S2 describes the calculated amphipathicity values; and Table S3 gives the values for hydrophilic loop lengths.

Acknowledgments

This work was supported from grant AI24204 from the National Institutes of Health.

References

Bibi E, Kaback HR. 1990. In vivo expression of the lacY gene in two segments leads to functional lac permease. *Proc Natl Acad Sci USA 87*:4325-4329.

Brooker RJ, Wilson TH. 1985. Isolation and nucleotide sequencing of lactose carrier mutants that transport maltose. *Proc Natl Acad Sci USA 82*:3959-3963.

Collins JC, Permuth SF, Brooker RJ. 1989. Isolation and characterization of lactose permease mutants with an enhanced recognition of maltose and diminished recognition of cellobiose. *J Biol Chem 264*:14698-14703.

Deisenhofer J, Michel H. 1989. The photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis*. *EMBO J 8*:2149-2170.

Eisenberg D, Weiss RM, Terwilliger JC. 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc Natl Acad Sci USA 81*:140-144.

Franco PJ, Eelkema JA, Brooker RJ. 1989. Isolation and characterization of thiodigalactoside-resistant mutants of the lactose permease which possess and enhanced recognition for maltose. *J Biol Chem 264*:15988-15992.

Griffith JK, Baker ME, Rouch DA, Page MGP, Skurray RA, Paulsen IT, Chater KF, Baldwin SA, Henderson PJF. 1992. Membrane transport proteins: Implications of sequence comparisons. *Curr Opin Cell Biol 4*: 684-695.

Henderson PJF. 1990. Proton-linked sugar transport systems in bacteria. *J Bioenerg Biomembr 22*:525-569.

Jones DT, Taylor WR, Thornton JM. 1994. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry 33*:3038-3049.

Jung K, Jung H, Wu J, Prive GG, Kaback HR. 1993. Use of site-directed fluorescence labeling to study proximity relationships in the lactose permease of *Escherichia coli*. *Biochemistry 32*:12273-12278.

King SC, Hansen CL, Wilson TH. 1991. The interaction between aspartic acid 237 and lysine 358 in the lactose carrier of *Escherichia coli*. *Biochim Biophys Acta 1062*:177-186.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol 157*:105-132.

Lee JI, Hwang PP, Wilson TH. 1993. Lys-319 interacts with both Glu-269 and Asp-240 in the lactose carrier of *Escherichia coli*. *J Biol Chem 268*:20007-20015.

Maiden MCJ, Davis EO, Baldwin SA, Moore DCM, Henderson PJF. 1987. Mammalian and bacterial sugar transport proteins are homologous. *Nature 325*:641-643.

Marger MD, Saier MH Jr. 1993. A major superfamily of transmembrane facilitators that catalyze uniport, symport and antiport. *Trends Biochem Sci 18*:13-20.

Markgraf M, Bocklage H, Muller-Hill B. 1985. A change of threonine 266 to isoleucine in the lac permease of *Escherichia coli* diminishes the transport of lactose and increases the transport of maltose. *Mol Gen Genet 198*:473-475.

Reizer J, Reizer A, Saier MH Jr. 1994. A functional superfamily of sodium/solute symporters. *Biochim Biophys Acta 1197*:133-166.

Sahin-Toth M, Dunten RL, Gonzalez A, Kaback HR. 1992. Functional interactions between putative intramembrane charged residues in the lactose permease of *Escherichia coli*. *Proc Natl Acad Sci USA 89*:10547-10551.

Wright JK, Weigel U, Lustig A, Bocklage H, Mieschendahl M, Muller-Hill B, Overath P. 1983. Does the lactose carrier of *Escherichia coli* function as a monomer? *FEBS Lett 162*:11-15.

Yan RT, Maloney PC. 1993. Identification of a residue in the translocation pathway of a membrane carrier. *Cell 75*:37-44.