Structural Model for 12-Helix Transporters Belonging to the Major Facilitator Superfamily

Teruhisa Hirai,1 Jürgen A. W. Heymann,1 Peter C. Maloney,2 and Sriram Subramaniam1*

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892,1 and Department of Physiology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 212052

Received 8 October 2002/Accepted 25 November 2002

The major facilitator superfamily includes a large collection of evolutionarily related proteins that have been implicated in the transport of a variety of solutes and metabolites across the membranes of organisms ranging from bacteria to humans. We have recently reported the three-dimensional structure, at 6.5 Å resolution, of the oxalate transporter, OxlT, a representative member of this superfamily. In the oxalate-bound state, 12 helices surround a central cavity to form a remarkably symmetrical structure that displays a well-defined pseudo twofold axis perpendicular to the plane of the membrane as well as two less pronounced, mutually perpendicular pseudo twofold axes in the plane of the membrane. Here, we combined this structural information with sequence information from other members of this protein family to arrive at models for the arrangement of helices in this superfamily of transport proteins. Our analysis narrows down the number of helix arrangements from about a billion starting possibilities to a single probable model for the relative spatial arrangement for the 12 helices, consistent both with our structural findings and with the majority of previous biochemical studies on members of this superfamily.

The movement of substrates across biological membranes is mediated by a large variety of membrane proteins that function as transporters. The major facilitator superfamily (MFS) (7, 19) is one of the largest classes of evolutionarily related transporters found in nature, representing a wide range of membrane proteins whose functions range from accumulating nutrients in bacteria to the cycling of neurotransmitters across human synaptic membranes. Sequence and biochemical analyses of several MFS proteins have suggested that most of these secondary active transporters are likely to contain 12 membrane-spanning segments (18). We have confirmed this prediction for the case of the oxalate transporter, OxlT, for which we recently reported a three-dimensional structure at a resolution of 6.5 Å (9).

At the present (6.5 Å) resolution, it is not possible to directly assign the 12 helices in the sequence of OxlT to the 12 transmembrane densities observed in the three-dimensional map. However, the unexpectedly high symmetry in the organization of the 12 helices leads to a significant reduction in the number of ways in which the transmembrane helices are likely to be arranged. By combining this structural information with selected biochemical and sequence information on other MFS proteins, we show that the number of helix assignments can be further reduced, leading to the deduction of a single, highly probable model for the spatial connectivity of the 12 helices. We propose that this arrangement is likely to reflect helix packing in the large variety of prokaryotic and eukaryotic transporters in the MFS.

Figure 1 shows a representation of the density map derived from our electron crystallographic studies together with an idealized representation of the 12 helical transmembrane segments fitted into the density map. The 12 helices fall naturally into three classes. One set of four helices (colored green), roughly perpendicular to the plane of the bilayer membrane, is found on the periphery of the molecule. These helices are furthest from the central cavity and do not appear to contribute any residues to the central substrate transport pathway. A second set of four helices (colored yellow) is tilted in a direction along the circumference of the molecule, with average tilts in the range of 10° to 20° from the direction of the membrane normal. The direction of the tilt is such that one face of these helices contributes residues to the central channel throughout its length. A third set of four helices (colored purple) is also part of the central cavity but tilted so that they cut across from the center of the cavity to the outside. The helices shown in purple display overall tilts of 20° to 30° and, in contrast to the helices shown in yellow, are tilted in such a way that they exclusively contact either the cytoplasmic or periplasmic half of the transmembrane channel.

Thus, the helices at the center of the protein on the cytoplasmic side are at the outer edges on the periplasmic side, and conversely, those at the center on the periplasmic side are at the outer edges on the cytoplasmic side. An important feature of our structural model is that residues from 8 of the 12 helices line the central cavity. We suppose that different transporters may use different combinations of residues from one or more of these eight helices to achieve the required substrate specificities.

The observation of three classes of helices in the structure is in remarkable agreement with the exhaustive biochemical studies of the TnJ0-encoded tetracycline transporter TetA(B), also a member of the MFS. As summarized in the work of Tamura et al. (21), the N-ethylmaleimide (NEM) reactivity of
FIG. 1. Top (A) and front (B) views of the three-dimensional density map of OxlT at 6.5 Å, contoured at 1.3σ, superimposed with models for the 12 helices. The helices are grouped into three sets: those colored green are nearly perpendicular to the plane of the membrane, those colored yellow contain bends in the transmembrane region, and those colored magenta contain both a bend and curve in the transmembrane region (from Hirai et al. [9]).
The close correspondence between the structure of OxlT and the NEM accessibility studies of related MFS proteins suggests that there is a simple connection between the two, leading to the assignment (in Fig. 1) of the helices shown in green to 3, 6, 9, and 12; yellow to 2, 5, 8, and 11; and the helices shown in purple to 1, 4, 7, and 10.

Although this assignment of helices is based on the TetA(B) studies (21), it is also consistent with NEM accessibility studies of related MFS proteins (2). The most hydrophobic and least conserved segments in MFS proteins are likely to be the transmembrane regions, and a third set (helices 1, 4, 7, and 10) that show significant reactivity to NEM throughout the length of the transmembrane region, a second set (helices 2, 5, 8, and 11) that show significant reactivity to NEM along one face of the entire length of the transmembrane region, and a first set (helices 3, 6, 9, and 12) that show limited reactivity to NEM along one face of the transmembrane region, and a third set (helices 1, 4, 7, and 10) that show significant reactivity to NEM along one face of only one half of the membrane-spanning region. In the last set, the sidedness of NEM reactivity is such that helices 1 and 7 show reactivity on the cytoplasmic side, while helices 4 and 10 show display reactivity on the periplasmic side. Closely comparable results have also come from this type of analysis for the bacterial lactose transporter LacY (1), also an MFS protein.

The number of possible helix arrangements allowed by the above constraints can be estimated as follows. There are four possible locations each for helices 1, 2, and 3, and 12, the helices shown in yellow to 2, 5, 8, and 11, and the helices shown in purple to 1, 4, 7, and 10. Although this assignment of helices is based on the TetA(B) studies (21), it is also consistent with NEM accessibility studies of LacY (1) and with predictions of the likely locations of helices 3, 6, 9, and 12 based on the finding that they are the most hydrophobic and least conserved segments in MFS proteins (2).

The use of additional constraints leads to a further reduction in the possible arrangements for the helices. As a starting point for helix assignment, we considered the locations of helices 1, 2, and 3 (4 × 4 × 4 = 64); in turn, these are coupled with two possible locations for each of helices 4, 5, and 6 (2 × 2 × 2 = 8). There are only two possible locations for the latter set, because the presence of the twofold axis automatically determines the locations of helices 8, 9, and 10 once the positions of helices 1, 2, and 3 are specified. Therefore, we arrive at total of 512 (64 × 8) possible helix arrangements. Since the bundle of 12 helices can be positioned into the present density map in two possible orientations (up or down, i.e., with N and C termini facing upwards or downwards), we obtained a final estimate of 2 × 512 possible unique helix assignments, based solely on identification of the three classes of helices. This is a significant reduction from the theoretical maximum of 2 × 12! (≈ 9.58 × 10⁸) possible ways to arrange the 12 helices into the map. The use of additional constraints leads to a further reduction in the possible arrangements for the helices. As a starting point for helix assignment, we considered the locations of helices 1, 2, and 3 (4 × 4 × 4 = 64); in turn, these are coupled with two possible locations for each of helices 4, 5, and 6 (2 × 2 × 2 = 8). There are only two possible locations for the latter set, because the presence of the twofold axis automatically determines the locations of helices 8, 9, and 10 once the positions of helices 1, 2, and 3 are specified. Therefore, we arrive at total of 512 (64 × 8) possible helix arrangements.

The use of additional constraints leads to a further reduction in the possible arrangements for the helices. As a starting point for helix assignment, we considered the locations of helices 1, 2, and 3 (4 × 4 × 4 = 64); in turn, these are coupled with two possible locations for each of helices 4, 5, and 6 (2 × 2 × 2 = 8). There are only two possible locations for the latter set, because the presence of the twofold axis automatically determines the locations of helices 8, 9, and 10 once the positions of helices 1, 2, and 3 are specified. Therefore, we arrive at total of 512 (64 × 8) possible helix arrangements. Since the bundle of 12 helices can be positioned into the present density map in two possible orientations (up or down, i.e., with N and C termini facing upwards or downwards), we obtained a final estimate of 2 × 512 possible unique helix assignments, based solely on identification of the three classes of helices. This is a significant reduction from the theoretical maximum of 2 × 12! (≈ 9.58 × 10⁸) possible ways to arrange the 12 helices into the map.

The use of additional constraints leads to a further reduction in the possible arrangements for the helices. As a starting point for helix assignment, we considered the locations of helices 1, 2, and 3 (4 × 4 × 4 = 64); in turn, these are coupled with two possible locations for each of helices 4, 5, and 6 (2 × 2 × 2 = 8). There are only two possible locations for the latter set, because the presence of the twofold axis automatically determines the locations of helices 8, 9, and 10 once the positions of helices 1, 2, and 3 are specified. Therefore, we arrive at total of 512 (64 × 8) possible helix arrangements. Since the bundle of 12 helices can be positioned into the present density map in two possible orientations (up or down, i.e., with N and C termini facing upwards or downwards), we obtained a final estimate of 2 × 512 possible unique helix assignments, based solely on identification of the three classes of helices. This is a significant reduction from the theoretical maximum of 2 × 12! (≈ 9.58 × 10⁸) possible ways to arrange the 12 helices into the map.

The use of additional constraints leads to a further reduction in the possible arrangements for the helices. As a starting point for helix assignment, we considered the locations of helices 1, 2, and 3 (4 × 4 × 4 = 64); in turn, these are coupled with two possible locations for each of helices 4, 5, and 6 (2 × 2 × 2 = 8). There are only two possible locations for the latter set, because the presence of the twofold axis automatically determines the locations of helices 8, 9, and 10 once the positions of helices 1, 2, and 3 are specified. Therefore, we arrive at total of 512 (64 × 8) possible helix arrangements. Since the bundle of 12 helices can be positioned into the present density map in two possible orientations (up or down, i.e., with N and C termini facing upwards or downwards), we obtained a final estimate of 2 × 512 possible unique helix assignments, based solely on identification of the three classes of helices. This is a significant reduction from the theoretical maximum of 2 × 12! (≈ 9.58 × 10⁸) possible ways to arrange the 12 helices into the map.

The use of additional constraints leads to a further reduction in the possible arrangements for the helices. As a starting point for helix assignment, we considered the locations of helices 1, 2, and 3 (4 × 4 × 4 = 64); in turn, these are coupled with two possible locations for each of helices 4, 5, and 6 (2 × 2 × 2 = 8). There are only two possible locations for the latter set, because the presence of the twofold axis automatically determines the locations of helices 8, 9, and 10 once the positions of helices 1, 2, and 3 are specified. Therefore, we arrive at total of 512 (64 × 8) possible helix arrangements. Since the bundle of 12 helices can be positioned into the present density map in two possible orientations (up or down, i.e., with N and C termini facing upwards or downwards), we obtained a final estimate of 2 × 512 possible unique helix assignments, based solely on identification of the three classes of helices. This is a significant reduction from the theoretical maximum of 2 × 12! (≈ 9.58 × 10⁸) possible ways to arrange the 12 helices into the map.
To accommodate this restriction, we propose that the cytoplasmic ends of helices 2 and 3 are close to one another. This value of eight residues is typical for the length of loops connecting neighboring helical segments found in the structures of several membrane proteins (http://www.mpib-frankfurt.mpg.de/michel/public/memprotstruct/, membrane proteins of known structure).

Interestingly, the three other loop regions that we would predict to be approximately symmetry related, i.e., the 5-6, 8-9, and 11-12 loops, are also among the shortest loops in MFS proteins, averaging 8 to 11 residues in length (2). Since helices 2 and 3 must correspond to the helices colored yellow and green, respectively, in Fig. 1, it follows from the above discussion that there are only four possible different arrangements of helices 1, 2, 3, 4, 7, 8, 9, and 10 (Fig. 3).

For each of the four arrangements noted in Fig. 3C, there are only four possible ways in which helices 5 and 6 can be positioned, since helix 5 must correspond to one of the two unassigned helices in yellow and helix 6 must correspond to one of the two unassigned helices in green. The positions of helices 5 and 6 uniquely specify the locations of helices 11 and 12, leading to a total of 16 possible relative spatial arrangements, i.e., four sets of four arrangements, of the 12 helices. From each set, only one arrangement has the helix triplets 1, 2, and 3, 4, 5, and 6, 7, 8, and 9, and 10, 11, and 12, in structurally equivalent environments, i.e., there are only four arrangements in which the 1-2 and 2-3 loop regions are in symmetry-related positions to the 3-4 and 4-5 loops, the 6-7 and 7-8 loops, and the 10-11 and 11-12 loops, respectively. We have argued previously (9) that this is likely to be a feature of MFS proteins, given the high level of symmetry in the architecture.

These final four arrangements (Fig. 4) have interesting differences. Thus, only the models in Fig. 4A and Fig. 4D have a short loop connecting helices 3 and 4 and a long loop connecting helices 6 and 7. We favor these two models over those in Fig. 4B and Fig. 4C because of the overwhelming evidence from sequence analyses that the average length of the 3-4 loop is about eight residues, versus six for OxIT (8), 14 for LacY (1), seven for TetA(B) (21), and 10 for GLUT1 (10) residues, and that the 6-7 loop is almost never shorter than about 30 residues, versus 31 for OxIT (8), 34 for LacY (14), 35 for TetA(B) (21), and 65 for GLUT1 (10).

A distinguishing feature of the models in Fig. 4A and 4D is that in Fig. 4A, helices 2 and 11 are next to each other, while

FIG. 3. Illustration of intermediate steps in the reasoning to arrive at helix arrangement in MFS proteins. Once we had made the assumption (see text) that helices 1 and 7 are at the center on the periplasmic side, there were four possible ways in which these helices could be placed into the density map, of which one arrangement is shown in A as viewed from the cytoplasmic side. The $\times$ and $\blacklozenge$ symbols indicate whether the ends of the helices correspond to the N- or C-terminal side, respectively. The four assignments arise from the presence of the three twofold pseudo rotational symmetry axes (two in the plane and the other two perpendicular to the membrane). For each arrangement of helices 1 and 7, there are two ways in which 4 and 10 can be positioned, as shown in B. Based on the argument that the cytoplasmic ends of helices 2 and 3 must be in close proximity and the conclusion (9) that helix 2 must be one of the helices shown in yellow while helix 3 must be one of the helices shown in green, we arrived at two further possible locations of helices 2 and 3 for each of the arrangements of helices 1, 4, 7, and 10 shown in B. Since the locations of helices 8 and 9 are uniquely specified by the locations of helices 2 and 3, we therefore arrived at only four possible relative arrangements of these eight helices (1, 2, 3, 4, 7, 8, 9, and 10). As discussed above, there is a fourfold degeneracy in C because of the four ways in which this arrangement can be placed on the map.
in Fig. 4D, they are on either side of the central cavity. Because of the strong biochemical (15, 28) and genetic (11, 3) evidence in favor of helices 2 and 11 being close to each other, we therefore favor the model in Fig. 4A as the most probable arrangement of helices in MFS proteins. The helix connectivity of this final, probable model is illustrated in Fig. 5.

Our proposed architecture for MFS proteins is in very good agreement with the majority of the helix proximities deduced from either the presence of ion pairs or cross-linking studies. The proposed associations between helices in LacY based on the existence of ion pairs include Asp237 (helix7)-Lys358 (helix11) (16), Asp240 (helix7)-Lys319 (helix10) (17), Glu269 (helix8)-His322 (helix 10) (13), Glu269 (helix8)-Arg302 (helix 9) (5), Arg302 (helix 9)-His322 (helix 10), Arg302 (helix 9)-Glu325 (helix 10) (6), and Glu126 (helix4)-Arg144 (helix 5) (32). Each of these associations can be accommodated in principle in the either of the models shown in Fig. 4A and Fig. 4D. Most of the cross-linking data between helices 1 and 7 (29, 27), 2 and 7 (29, 30), 1 and 11 (24), 4 and 7 and 4 and 11 (25), 5 and 7, 5 and 8, and 5 and 10 (26), and 6 and 8 (4) can also be accommodated with these two models, while the evidence for cross-linking between helices 3 and 7 (23) or 1 and 12 (24) is not consistent with either model. As noted already, the primary distinguishing feature between the models in Fig. 4A and 4D is the different relative arrangements of helices 2 and 11.

The high symmetry in the molecular architecture of OxlT in our proposed model is not only consistent with the idea that MFS proteins are the result of a gene duplication (7, 20) but suggests that MFS proteins may have developed from an ancestral precursor that contained three transmembrane segments. The architecture of the 12 transmembrane helices is such that each set of six helices is located in a different half of the molecule, and consequently the two halves form a nearly flat interface at their boundary that constitutes the transport channel pathway (Fig. 5). Within each half, the two sets of three-helix units appear to be in structurally equivalent positions. For each three-helix unit, one highly tilted helix (colored magenta) interdigitates into the space occupied by the neighboring three-helix unit. This arrangement of helices represents a novel fold for a membrane protein, and its provocative sym-
FIG. 5. View from the cytoplasmic side (A) and front (B) of the probable arrangement of helices in proteins belonging to the MFS, corresponding to the model in Fig. 4A. The helices are shown as idealized straight cylinders, and putative regions connecting neighboring helices are shown as unstructured loops. The boundary between the two halves of the protein is indicated by a transparent plane. The approach used to deduce the helix assignment is discussed in detail in the text. While this assignment scheme represents a unique relative arrangement of the 12 helices, it is fourfold degenerate in the sense that there are four ways to match this arrangement to the density map in Fig. 1 due to the presence of three pseudo-rotational twofold symmetry axes in the molecule.
metry suggests a plausible explanation for bidirectional substrate transport by proteins in this family.

We thank D. Bliss for assistance with preparation of figures. This work was supported by a grant to S.S. from the intramural program at the National Institutes of Health.

REFERENCES


