Evolutionary mix-and-match with MFS transporters II

M. Gregor Madej and H. Ronald Kaback

Department of Physiology, Department of Microbiology, Immunology and Molecular Genetics, and Molecular Biology Institute, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

Contributed by H. Ronald Kaback, October 22, 2013 (sent for review October 14, 2013)

One fundamentally important problem for understanding the mechanism of coupling between substrate and H+ translocation with secondary active transport proteins is the identification and physical localization of residues involved in substrate and H+ binding. This information is exceptionally difficult to obtain with the Major Facilitator Superfamily (MFS) because of the broad sequence diversity of the members. The MFS is the largest and most diverse group of transporters, many of which are clinically important, and includes members from all kingdoms of life. A wide range of substrates is transported, in many instances against a concentration gradient so as to detect functionally homologous positions with similar transmembrane helices represents a common structural feature. An inverted triple-helix structural symmetry motif within the N- and C-terminal six-helix bundles suggests that the proteins may have arisen by intragenic multiplication. In the work presented here, the triple-helix motifs are aligned in combinatorial fashion so as to detect functionally homologous positions with known atomic structures of MFS members. Substrate and H+-binding sites in symporters that transport substrates, ranging from simple ions like phosphate to more complex peptides or disaccharides, are found to be in similar locations. It also appears likely that there is a homologous ordered kinetic mechanism for the H+-coupled MFS symporters.

The lactose permease from Escherichia coli (LacY), a galactoside/H+ symporter, arguably the most intensively studied secondary transporter known at present, is the paradigm of the MFS. LacY is comprised of 417 amino acid residues organized into two pseudosymmetrical six-helix bundles with the N and C termini on the cytoplasmic face of the membrane (Fig. 1). To determine which residues play an obligatory role in the mechanism and to create a library of mutants with a single-Cys residue at each position of the molecule for structure/function studies, each residue was replaced individually with Cys in a systematic fashion. Multiple transmembrane helices were disulfide bridged, thus allowing assembly of the transport protein into a deep central hydrophilic cavity surrounded by 12 mostly irregular transmembrane helices. This information is exceptionally difficult discerning underlying mechanistic principles is due to low sequence conservation. However, a common structural feature of MFS members, suggesting that they may have arisen by intragenic multiplication, is a repeat of four three-helix bundles organized in two pseudosymmetrical domains. An alignment of these triple-helix motifs in combinatorial fashion allows detection of functionally homologous positions. Thus, substrate and H+-binding sites in distantly related symporters are located at the same relative positions. The structural organization also suggests that an ordered kinetic mechanism similar to that determined for lactose permease may be operative in other MFS symporters.

Author contributions: M.G.M. and H.R.K. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

1To whom correspondence should be addressed. E-mail: r.kaback@mednet.ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1319754110/-/DCSupplemental.
Results and Discussion

Although it is generally believed that all MFS symporters operate by an alternating access mechanism, extensive biochemical and spectroscopic evidence is available for LacY only (reviewed in refs. 6 and 16). Furthermore, as a result of intensive study over three decades, an overall mechanism for coupling in LacY has been worked out.

An Overall Mechanism for Coupling in LacY.

i) Lactose/H+ symport in the uphill or downhill energetic modes is precisely the same reaction. The difference is in the rate-limiting step. For downhill symport, deprotonation is rate-limiting; for uphill transport, deprotonation is no longer limiting, and either dissociation of sugar or a conformational change that leads to deprotonation becomes limiting.ii) Sugar binding and dissociation—not ΔμH+—are the driving force for alternating access (ΔμH+ has no effect whatsoever on equilibrium exchange or counterflow).

iii) LacY must be protonated to bind sugar (the pKₐ for sugar binding is ~10.5).iv) Galactoside binds by an induced-fit mechanism, which powers a transition to an occluded state.
v) Sugar dissociates first.
The alignments are oriented with the LacY cytoplasmic side to the top. The plasmic loop). The numbers on the loops indicate the loops within symmetry motifs (white, cytoplasmic loop; gray, periplasmic loop). The orange; helices 10
motifs. Helix-triplets from LacY are aligned with FucP (helices 1
by the respective loop. See SI Appendix, Fig. S1 for a schematic representation.

Fig. 2. Schematic alignment of the helix-triplets in consecutive order in the sequence. (A) Helix-triplets (represented by colored boxes) from FucP, Pipt, XyIE, and PepT are aligned with LacY (helices 1–3, blue; helices 4–6, green; helices 7–9, orange; helices 10–12, yellow). The flags indicate the loops within symmetry motifs. Helix-triplets from LacY are aligned with FucP (B), Pipt (C), and PepT (D) (helices 1–3, blue; helices 4–6, green; helices 7–9, orange; helices 10–12, yellow). The alignments are oriented with the LacY cytoplasmic side to the top. The flags indicate the loops within symmetry motifs (white, cytoplasmic loop; gray, periplasmic loop). The numbers on the flags indicate the two helices that are connected by the respective loop. See SI Appendix, Fig. S1 for a schematic representation.

Detection of Functionally Homologous Positions. MFS transporters consist of four symmetrically disposed triple-helix units that can be aligned individually and examined for conservation of functionally significant residues (Table 1; Fig. 2). The combinatorial alignment of the symmetry motifs allows detection of functionally homologous positions in different MFS transporters. The procedure relies on mapping known functional markers in the helix triplets followed by inferential mapping of unknown functional positions.

Arrangement and orientation of helix-triplets in PepT. The oligopeptidase/ H+ symporter (PepT) (19, 30, 31, 35) provides another variation. In PepT, the orientation of all four motifs is inverted with respect to LacY; all cytoplasmic loops superpose on periplasmic loops and periplasmic loops on their cytoplasmic counterparts in these two symporters (12). Unlike FucP, in XyIE and Pipt, only motifs A and B are inverted with respect to motifs C and D in LacY, thereby placing the C terminus of the first six-helix bundle (motifs A and B) and the N-terminal six-helix bundle (motifs C and D) on opposite sides of the protein in the alignment of the structure motifs (Fig. 2). This is not the case in the native XyIE or Pipt; rather, in these proteins, the N-terminal six-helix bundle predominate. The previous study (12) comparing LacY with FucP revealed that due to the different order of the symmetry motifs, functionally equivalent residues can be located at different positions in the protein. Superposition of the C-terminal six-helix bundle from XyIE or Pipt on the N-terminal bundle of LacY results in spatial alignment of substrate-binding residues in these three symporters (Fig. 3 A and B). However, an alignment of XyIE and Pipt with FucP can be achieved only if symmetry motif D is superposed on motif C in FucP and, respectively, symmetry motif C is superposed on motif D in FucP (Fig. 2; Table 1). The N-terminal six-helix bundles from XyIE, Pipt, and FucP superpose on each other without changing the order of symmetry motifs A and B. However, superposition on LacY succeeds only when symmetry motif C from LacY is superposed on symmetry motifs B from XyIE, Pipt, and FucP, and symmetry motif D from LacY with the symmetry motifs A from the other three symporters (Fig. 2; Table 1).

Altered orientation of helix-triplets. In FucP, the orientation of all four motifs is inverted with respect to LacY; all cytoplasmic loops superpose on periplasmic loops and periplasmic loops on their cytoplasmic counterparts in these two symporters (12). Unlike FucP, in XyIE and Pipt, only motifs A and B are inverted with respect to motifs C and D in LacY, thereby placing the C terminus of the first six-helix bundle (motifs A and B) and the N terminus of the following six-helix bundle (motifs C and D) on opposite sides of the protein in the alignment of the structure motifs (Fig. 2). This is not the case in the native XyIE or Pipt; rather, in these proteins, the N-terminal six-helix bundle is rotated 180° with respect to the membrane plane relative to LacY. This is compatible with the hypothesis that the symporters evolved by intragenic duplication and fusion of the helix-triplets (12, 36, 38). With single substrate and H+ binding-sites (5), the primordial MFS symporters may have exhibited substantial flexibility with regard to the orientation of the helix bundles to each other. This is not surprising considering that substrate-binding sites are generally located at the approximate middle of these symporters. Therefore, inversion of triple-helix motifs may be well tolerated. In this context, although the antipporter EmrE is not an MFS protein, it is a dimer with the binding site placed at the dimer interface in the middle of the complex. Moreover, it has been demonstrated (39) that the EmrE dimer is functional in both parallel and antiparallel configurations.

Arrangement and orientation of helix-triplets in PepT. The oligopeptidase/ H+ symporter (PepT) (19, 30, 31, 35) provides another variation to the order of the triple-helix units because the symmetry motifs

v) Upon sugar dissociation, there is a conformational change that causes Arg302 (helix IX) to approximate Glu325, leading to deprotonation of LacY.

As an initial step to investigate whether a similar overall mechanism might apply to other MFS symporters, we have searched for functional homologies between individual residues in additional structurally resolved symporters.

Fig. 2. Arrangement and orientation of helix-triplets in PepT. The oligopeptidase/ H+ symporter (PepT) (19, 30, 31, 35) provides another variation to the order of the triple-helix units because the symmetry motifs

12, 36, 38). With single substrate and H+ binding-sites (5), the primordial MFS symporters may have exhibited substantial flexibility with regard to the orientation of the helix bundles to each other. This is not surprising considering that substrate-binding sites are generally located at the approximate middle of these symporters. Therefore, inversion of triple-helix motifs may be well tolerated. In this context, although the antipporter EmrE is not an MFS protein, it is a dimer with the binding site placed at the dimer interface in the middle of the complex. Moreover, it has been demonstrated (39) that the EmrE dimer is functional in both parallel and antiparallel configurations.

Arrangement and orientation of helix-triplets in PepT. The oligopeptidase/ H+ symporter (PepT) (19, 30, 31, 35) provides another variation to the order of the triple-helix units because the symmetry motifs
Helix-triplets A–D from PiPT, XyIE, PepT, and, for comparison, FucP are aligned to LacY. The triplets are color-coded as in Fig. 2. The values represent the similarity score and the functional similarity score according to SI Appendix, Eqs. S1 and S2 (Q/QAct). For comparison, the respective scores are also stated for the superposition of the N-terminal six helices (Nter) and the C-terminal six helices (Cter) with LacY for the respective model.

4. In all crystallographic structures with a bound substrate, the position corresponding to Glu269 in LacY couples the two six-helix bundles.

In addition to polar substrate ligands, aromatic side chains also play an important role in substrate binding. For example, Trp151 in LacY (8, 40) and Phe308, its counterpart in FucP (41), are essential for sugar binding (22), and they align with Trp416 in XyIE, Lys459 in PiPT, and Phe170 in PepTc. Irreversible Arg144 in LacY (42, 43) is in close proximity and on the cytoplasmic side of Trp151, and, in FucP, Arg312 (21) is functionally important and lies in a similar position relative to Phe308 (12).

Fig. 3. Overall architecture of the substrate binding sites compared with LacY. (A) C-terminal six-helix bundle of XyIE (colored in light olive) is superposed on the N-terminal six-helix bundle of LacY (side chains shown colored in light blue). The ligand of XyIE, xylose, is shown as a gold ball-and-stick model. Oxygen and nitrogen atoms are colored red and blue respectively. The superposition of XyIE and PiPT is provided in SI Appendix, Fig. S2.

Table 1. Combinations of symmetry motifs in MFS

<table>
<thead>
<tr>
<th></th>
<th>LacY: 1.00/1.00</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>PiPT</td>
<td>Nter: 0.22/0.42</td>
<td>C:</td>
<td>D:</td>
<td>B:</td>
</tr>
<tr>
<td></td>
<td>Cter: 0.23/0.36</td>
<td>0.16/0.52</td>
<td>0.21/0.54</td>
<td>0.16/0.41</td>
</tr>
<tr>
<td>XyIE</td>
<td>Nter: 0.2/0.33</td>
<td>C:</td>
<td>D:</td>
<td>B:</td>
</tr>
<tr>
<td></td>
<td>Cter: 0.2/0.33</td>
<td>0.2/0.65</td>
<td>0.2/0.58</td>
<td>0.1/0.50</td>
</tr>
<tr>
<td>PepT</td>
<td>Nter: 0.2/0.27</td>
<td>D:</td>
<td>C:</td>
<td>B:</td>
</tr>
<tr>
<td></td>
<td>Cter: 0.14/0.36</td>
<td>0.1/0.68</td>
<td>0.2/0.57</td>
<td>0.2/0.53</td>
</tr>
<tr>
<td>FucP</td>
<td>Nter: 0.2/0.48</td>
<td>D:</td>
<td>C:</td>
<td>B:</td>
</tr>
<tr>
<td></td>
<td>Cter: 0.2/0.50</td>
<td>0.2/0.72</td>
<td>0.1/0.52</td>
<td>0.1/0.60</td>
</tr>
</tbody>
</table>
gradient at a slow rate without H\(^+\) translocation (52). Although His322 is clearly involved in affinity for galactosides, it has been suggested that it is also involved in H\(^+\) translocation (54). Moreover, the notion has been put forward that structural water coordinated within a triad between His322, Tyr236, and Glu269 (two additional irreplaceable residues) may act as a cofactor for binding and galactoside/H\(^+\) symport by forming a hydronium ion intermediate during turnover (54, 55). In this regard, an analogous situation is apparent in PepT\(_{Gk}\) where Glu35, which corresponds to His322 in LacY, is not a direct substrate ligand. However, Glu35 is involved in substrate binding by positioning the guanidinium group of Arg36 with substrate (Figs. 4C and 6A) (19).

**Functional Correlations.** Homology of primary amino acid sequences reflects evolution and therefore provides a guide to structure, mechanism, and function. Proteins that are related by common descent are expected to exhibit homologous structures and functions proportional to the degree of their sequence similarity. This principle provides the motivation to define protein phylogenetic relationships and correlate specific residues with function. **A kinetic mechanism for LacY.** Varied experimental efforts have been undertaken with LacY to develop a kinetic scheme for lactose/H\(^+\) symport (5, 6, 46), and Fig. 5 depicts a simplified scheme. The sequence is initiated from the inward-facing conformer (HEi) in which the carboxyl group of Glu325 is protonated due

---

**Fig. 4.** Functional alignment of XylE (A), PiPT (B), and PepT (C) to LacY helix-triplets (stereo-view). The Cα atom trace is shown as wire and colored according to the helix-triplet (Fig. 2 and Table 1) except for PiPT in B. Significantly equivalent residue pairs are shown as sticks in the same color. Labels of LacY residues are indicated in blue.
to the low dielectric of the local environment, as evidenced from the X-ray structures. Because the apparent pKₐ of Glu325 (helix X) is ~10.5 (54), deprotonation is likely due to the transient proximity of Arg302 (helix IX) (56) (step 1). Subsequently, the cytoplasmic cavity closes (step 2), resulting in the formation of the apo-intermediate (En). This conformer can relax back to the inward-facing or open to the periplasmic side where it is reprotonated, which may involve Tyr236 (helix VII), His322 (helix X), and Glu269 (helix VIII) (55) (step 3). The protein then assumes a conformation symmetrical to the inward-facing conformer to reorient the helices for transfer of the H⁺ to negatively charged Glu325 and binding of sugar from the peri-plasmic side (HEo, steps 4 and 5). All of the specificity of LacY for substrate is directed toward the galactopyranosyl end of the substrate, and binding begins with a nonspecific hydrophobic interaction between the bottom of the galactopyranosyl ring with the hydrophobic surface of the indole ring of Trp151 (40, 57). Once the galactoside is oriented properly, formation of the binding site occurs. Arg144 (helix V) and Glu126 (helix IV), as well as Glu269 (helix VIII), interact with specific OH groups on the galactopyranosyl ring. Thus, galactopyranoside binding probably involves induced-fit (58). As the protonated galactoside-bound LacY ternary complex forms, the protein closes around the sugar to form an occluded state (HENs, step 6). This conformer can either relax back to the outward-facing conformation or the cytoplasmic cavity opens with release of sugar to the cytoplasm (step 7). With release of sugar, the initial inward-facing conformation is restored (step 8). By reorientation of helices, Tyr236 (helix VII) is displaced from a position between Arg302 and protonated Glu269, thereby causing deprotonation of Glu325 (Ei+H⁺) with reorientation of the cycle (step 1) (41). This ordered scheme was proposed originally in 1979 (59) and has been continuously refined since.

**The ordered kinetic mechanism.** There is an abundance of evidence for the ordered mechanism as depicted in Fig. 5 (reviewed in refs. 5 and 46). As stated above, the pKₐ for galactoside binding is ~10.5. It is apparent that LacY is protonated at physiologic pH before substrate binding, and recent experiments supporting this conclusion more directly have been presented (60). Critically, neutral replacements for Glu325 yield mutants that are totally unable to carry out any reaction that involves net H⁺ transport but catalyze equilibrium exchange and counterflow as well as or better than the WT. As indicated in Fig. 5, a satisfying explanation for the behavior of Glu325 neutral replacement mutants that provides very strong evidence for the ordered mechanism shown is that the mutants can oscillate within the shaded portion of the scheme, but cannot deprotonate, and therefore cannot form the apo-intermediate and complete the cycle. In other words, the order of release in the WT must be dissociation of the galactoside first, followed by deprotonation. Because FucP, Xyle, and PepT all have a carboxyl group corresponding to Glu325 in LacY that displays a similar phenotype when neutralized by mutagenesis, it is apparent that these symporters and perhaps many other MFS symporters likely operate by a similar ordered kinetic mechanism.

**A common mechanistic pattern.** As discussed, the initial step in the cycle for LacY involves deprotonation of Glu325, which is thought to cause closure of the cytoplasmic cavity (61). As a result, the central cavity becomes closed on both sides. Although this conformation has not been observed crystallographically with LacY, such a conformation was observed with PepTα (30), and a homology model for the LacY apo-intermediate is also available (41). The apo-intermediate can relax to the outside-open conformation, and the symporter reprotonates from the outside. An open-outward crystallographic model of FucP (21) and theoretical models for LacY (38) have been presented. Binding of substrate to the protonated open-outward conformer causes the protein to close around the substrate to form the occluded state. This conformation has been described in Xyle (20) and PiPT (26). Interestingly, in both structures, the substrate-liganding side chains are located in the N- and C-terminal.

---

**Fig. 5.** Transport cycle of LacY. Overview of the postulated steps in the transport model. Inward-facing (blue) and outward-facing (green) conformations are separated by the apo-intermediate conformational cluster (gray) or by the occluded-intermediate conformational cluster (orange). Substrate (S) and H⁺ are indicated. Steps are numbered consecutively: Substrate translocating transitions are indicated by blue arrows (steps 5–8) and transitions recycling the outward-open cavity are indicated by red arrows (steps 1–4). All steps are reversible (indicated by double-headed arrows). The blue-shaded area demarcates the equilibrium-exchange reaction. Examples of experimental coordinates (transporter and PDBID) associated with respective conformations are indicated.

**Fig. 6.** Conformational changes in the substrate binding site of PepT (see SI Appendix, Fig. S3 for stero-view version of this figure). (A) Inward-open, substrate-bound state of PepTα (PDB ID code: 4IKZ). (B) Inward-open, substrate-free state of PepTα (PDB ID code: 4IKV). The position of the substrate detected in the substrate-bound state is indicated as white profile. (C) Inward-occluded conformation of PepTα (PDB ID code: 2XUT). The colors of the bars at the bottom indicate the mechanistic affiliation of the respective states with regard to the mechanistic model shown in Fig. 5.
halves of the symporters. The occluded conformation can either reopen to the outside or open to the inside with release of substrate. In LacY, this global conformational change is postulated to be initiated by a localized scissors-like movement between helices V and I that is induced by galactoside binding (24). High-resolution X-ray structures of the inward-facing conformation of PepT\(_{\text{Gk}}\) (19), with bound inhibitor (pdbID: 4IKZ), show helix V with Asn166 displaced from its corresponding position in XyIE (Asn415 on helix XI). It is likely that Asn166 ligates substrate in the fully occluded state and is no longer in contact with alafosfalin in the inward-facing conformation of PepT\(_{\text{Gk}}\) (Fig. 6f). Asn166 is presumably important for substrate binding in PepT\(_{\text{Gk}}\) (19), like Arg144 in LacY (43). Therefore, movement of helix V may initiate release of substrate on the cytoplasmic side. As discussed above with respect to LacY, helices V and I cross in the approximate middle of the membrane where Cys154 (helix V) and Gly24 (helix I) are in close proximity (8, 44). The C154G mutant binds ligand with high affinity but catalyzes almost no transport (44, 62). The X-ray structures of PepT (18, 19, 30) demonstrate that helix V crosses helix I in the approximate middle of the membrane in such a manner that Ala171 (helix V in PepT\(_{\text{Gk}}\)) lies close to Gly42 (helix I in PepT\(_{\text{Gk}}\)), thereby emphasizing the homology to this position in LacY. In the inward-facing conformation of PepT\(_{\text{Gk}}\) without bound substrate (pdbID: 4IKV) (Fig. 6f), further distortion of the substrate-binding site is observed. An induced-fit mechanism has already been suggested for LacY (9). Galactoside is thought to induce the formation of its binding site, primarily through side-chain movements without a global conformation change. Such a scenario is consistent with crystallographic structures observed with PepT\(_{\text{Gk}}\) (19).

It is obvious that the formation of the open-outward cavity to initiate another round of transport requires a conformation in which substrate is released without rebinding. The structure of the open-inward conformation of PepT\(_{\text{Gk}}\) (pdbID: 4APS) (31) provides structural insight in this regard. Here, the substrate-capping Tyr residue is rotated back toward the center of the cavity, blocking access of substrate. Further stabilization of this conformation is represented in the crystal structure of PepT\(_{\text{Gk}}\) (30) where the Tyr residue is stabilized in a conformation blocking access to substrate by H-bonding to an Arg32 (Fig. 6c).

Taken together, the striking similarities within the MFS class suggest that a common mechanistic pattern may be used for catalysis of symport by additional transporters in this superfamily.

Materials and Methods

Structure and Sequence Alignments. The structural symmetry motifs according to Radestock and Forrest (38) were generated from the crystallographic coordinates of LacY (A, Thr7-Asn102; B, Leu104-Phe187; C, Lys220-Ser309; D, Ala311-Leu400) (10), FuXP (A, Arg22-Met115; B, Asn116-Thr229; C, Arg258-Ala355; D, Gly347-Phe431) (21), XyIE (A, Tyr5-Ile112; B, Trp152-Phe237; C, Gly276-Thr365; D, Gly369-Glu465) (22), PPT (A, Pro30-His125; B, Thr126-Arg229; C, Thr306-leu410; D, Gly411-Arg518) (26), and PepT (A, His211-Leu211; C, Ala286-Ser392; D, His398-Met492) (19), and the superimposition of the helix-triplets was carried out as described in Madelj et al. (12). In brief, computer programs COOT v0.7 (63) and UCSF-Chimera (64) were used for the structure-guided sequence alignment. No positional restraints were applied to functionally significant residues in the sequence or structure alignments. The superposition was visually inspected for the conservation of LacY functional markers. As functional markers, 22 polar residues in LacY were used where mutations to Cys cause greater than 50% (SI Appendix, Eq. 52; sAA\(_{25-50}\)) or 75% (SI Appendix, Eq. 52; sAA\(_{25-50}\)) inhibition of the transport rate with Cys-less LacY (7, 12).

Discrimination of Defective Alignments. For discrimination of defective alignments, a similarity score, Q, was defined. A reduction matrix (SI Appendix, Table S1) of similar side chains was defined based on clustering of reduction groups in similar folds, omitting interfac ing for different levels of reduction (65). This reduction matrix was modified with reweighting of packing for values of helix-helix contacts in the membrane, their natural occurrence in α-helical membrane proteins (66), and hydrophobicity (67) (SI Appendix, Fig. S4 and Table S1). The main difference is that His was grouped with positively charged residues Arg and Lys instead of with Asn in the first reduction level. Additionally, groups introducing helix irregularities and small side chains (Gly, Pro) are grouped with Cys at the last reduction level based on similar helix packing and hydrophobicity to Ser and Thr. Due to high abundance in helical membrane proteins, the hydrophobic side chains I, F, V, and L were grouped. The following values for the scoring of sequence conservation were used: conserved residue = 20, first reduction level = 10, second reduction level = 7, and third reduction level = 4. The sum of the scoring values (red Value) was divided by the total number of analyzed positions multiplied by the highest score (v = 20) for a conserved residue (SI Appendix, Eq. S1). For 100% conservation, Q should equal 1. The quantification of functionally overlapping positions was performed by weighting the scoring of functionally critical positions (AA\(_{25-50}\), activity inhibited by at least 75% upon mutation to Cys) and functionally relevant positions (AA\(_{50-100}\), activity inhibited by at least 50% upon mutation to Cys) using the reduction matrix (SI Appendix, Table S1) and an arbitrary weighting factor of 5 for functionally critical positions (SI Appendix, Eq. S2). For 100% conserved transporters with 100% conserved functional positions, this value, Q\(_{\text{AA}}\), should equal 1. The best scoring structure alignments were inspected visually for conservation of the geometry in the functional site, i.e., if functional groups face each other in the alignment.

Acknowledgments. We are deeply indebted to Arthur Karlin for editorial contributions. This work was supported by National Institutes of Health Grants DK55113, DK68493, and GM073210, as well as National Science Foundation Grant MCB–1129551 (to H.R.K.).

19. We are deeply indebted to Arthur Karlin for editorial contributions. This work was supported by National Institutes of Health Grants DK55113, DK68493, and GM073210, as well as National Science Foundation Grant MCB–1129551 (to H.R.K.).
Supplementary Information
Mix-and-Match Evolution II

M. Gregor Madej\textsuperscript{1} and H. Ronald Kaback\textsuperscript{1,2,3*}

\textsuperscript{1}Department of Physiology, \textsuperscript{2}Department of Microbiology, Immunology & Molecular Genetics, \textsuperscript{3}Molecular Biology Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095

Running title: Structure-Function Correlations in the MFS

Keywords: membrane transport | symport | MFS | sequence alignment | bioenergetics

*Corresponding author: H. Ronald Kaback, rkaback@mednet.ucla.edu
Figure S1: Clustering of amino acid properties (three-dimensional representation). The coordinates are computed from the relative occurrence [%] in alpha-helical membrane proteins for x-coordinate (1), the packing value [val.] for y-coordinate (1) and \( \Delta \) hydrophobicity [kcal/mol] for z-coordinate (2 and http://blanco.biomol.uci.edu/hydrophobicity_scales.html). The amino acid identity is indicated by single letter code and the positions are colored according to \( \Delta \) hydrophobicity (from violet to red). Hydrophobic side chains I, F, V and L display similar helical packing and hydrophobicity properties paired with high abundance in in alpha-helical membrane proteins.
Figure S2: Structure superposition of PiPT and XylE. (A) Schematic representation of crystallographic structures of PiPT (PDBid 4J05; gray) and XylE (PDBid 4GBY; rainbow). A satisfying structure alignment is generated (rmsd achieved: 2.6 Å, number of residues reference (4GBY): 475, number of residues moving (4J05): 422, number of aligned residues: 346, sequence identity: 19.9%) where the putative H⁺-binding-sites (B, blue square) and the substrate binding-sites (C, red square) superpose. In the detailed representations residue labels of XylE are shown in green color and residue labels of PiPT are shown in magenta color. Xylose is shown a transparent shape for better visually.
Figure S3: Conformational changes in the substrate binding site of PepT. (A) Inward-open, substrate-bound state of PepT_{Gk} (pdbID: 4IKZ). (B) Inward-open, substrate-free state of PepT_{Gk} (pdbID: 4IKV). The position of the substrate detected in the substrate-bound state is indicated as white profile. (C) Inward-occluded conformation of PepT_{So} (pdbID: 2XUT). The colors of the bars on the right side indicate the mechanistic affiliation of the respective states with regard to the mechanistic model shown in Fig. 5.
Figure S4: Orientation of helix-triplets. Helix-triplets from FucP, PiPT, XylE and PepT are aligned with LacY (helices 1–3, blue; helices 4–6, green; helices 7–9, orange; helices 10–12, yellow) according to Fig. 2, the order of the helices (roman numerals) is shown according to the alignment to LacY. The N- and the C-termini of the helix-triplets are indicated by red and blue rectangles respectively. A red dot marks the N-terminus of the protein and a blue dot marks the C-terminus. The central middle loop is shown as a green line. Although the orientation of homologous helix-triplets may be flipped relatively to LacY, The N-, the C-termini and the middle loop are always on the cytoplasmic side.
Table S1: Reduction matrix for sequence simplification. Grouping of residues based on reasonable simplification of protein sequence with no interlacing for different levels of reduction (modified from ref. 2). Letter grouped without spaces represent a reduction clade and are treated as similar in scoring of sequence alignments.

<table>
<thead>
<tr>
<th>level</th>
<th>score</th>
<th>reduction alphabet</th>
<th>symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20</td>
<td>L I V F Y W M G P AT S C N Q E D H R K</td>
<td>*</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>L I V F Y W M G P A T S C N Q E D H R K</td>
<td>!</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>L I V F Y W M G P A T S C N Q E D H R K</td>
<td>:</td>
</tr>
</tbody>
</table>

Equation S1: Similarity score.

\[ Q = \frac{\sum \text{redValue}}{(AA_{all} - AA_{LIVF}) \cdot 20} \]

\( Q \): similarity score, \( \text{redValue} \): score assigned from the reduction matrix (SI Appendix, Table S1), \( AA_{all} \): number of all residues in the alignment, \( AA_{LIVF} \): number of not compared side-chains. For 100% conservation, \( Q \) should equal 1.

Equation S2: Functional similarity score.

\[ Q_{act} = \left( \frac{sAA_{0-25\%} \cdot 5}{AA_{0-25\%}} + \frac{sAA_{25-50\%}}{AA_{25-50\%}} \right) + 6 \]

\( Q_{act} \): functional similarity score, \( sAA_{0-25\%} \): number of aligned critical positions in the alignment, \( AA_{0-25\%} \): number of all critical positions in the alignment, \( sAA_{25-50\%} \): number of aligned functionally relevant positions in the alignment, \( AA_{25-50\%} \): number of all functionally relevant positions in the alignment. Critical positions are positions where mutations to Cys cause greater than 50% and functionally relevant positions are positions where mutations to Cys cause 75% inhibition of the transport rate with Cys-less LacY (3, 4). For 100% conserved transporters with 100% conserved functional positions this value, \( Q_{act} \), should equal 1.
Sequence Alignments

LacY / PiPT

LacY-A.pdb._chain_B/1-97  NTNFWMFGLFFPYFIMAYFPFF....PIWHLDINHIS......... 36
PiPT-c.pdb._chain_A/306-410  TWHNFRNLGMLGWLVDIAFYGINLNVSVVLAQFGAKTGTDYDKLF 50

LacY-A.pdb._chain_B/1-97  .KSDTGIIFAFAILSFLLFQPLFGLLSDKLGLRKYLLWIIITGMLVMAFP 85
PiPT-c.pdb._chain_A/306-410  QLATGNNIVTLGFLPGYFTTLF..LIDIVG.....RKQLQFGIMSSLG 93

LacY-A.pdb._chain_B/1-97  FIFIFGPLLQYN 97
PiPT-c.pdb._chain_A/306-410  FLAILAGEIDHI 105

LacY-B.pdb._chain_B/1-83  LGVSIVGGI.YLGFCFNAGAPAVEAFI...EKVRNRSNFEFGRARM..FGCV 46
PiPT-D.pdb._chain_A/411-499  GKPPLLACFTMFQFFFFNGAFNTTITFIVAAELFPTIRAS.AHGISAAGKC 50

LacY-B.pdb._chain_B/1-83  GAILSSLVFQNLKAGITSAVLWIFPSSTCLGFISTFTLI 89
PiPT-D.pdb._chain_A/411-499  NQGWGFSVGSLTVIGLTVIFTSGHK...RKLQGTHVDKAWRIIGLSLIPAGFT 100

LacY-C.pdb._chain_B/1-184  LFRQPKL.WFLSLYVGVSCT.YDVFDQQFANFTSFATGEQGTRVFGYVT 50
PiPT-B.pdb._chain_A/126-229  WDGNRVLTWITICCRVLGIGGPPHPSMTVSDR.ANIRRCLLCFIFA 50

LacY-C.pdb._chain_B/1-184  TMGE.LNASIMFFAFLINRIGGNK.ALLLAG'TMSVRIIGSSFATSALEV 100
PiPT-B.pdb._chain_A/126-229  NQGWGFSVGSLTVITIISGKH...RLKSGHTDVKAWRIIGLSLIPAGFT 100
LacY / PepT

LacY-C.pdb._chain_B/1-91
Xyle-B-coot-0.pdb._chain_A/125-221

KLWFLSLVYGVSCTYDVFDQQFANFFTSFFAT. . . .GEQGTRVFGYVTMG 48
YVPEFVIYRIGGIVGGLASLSPMYIAELAPAHIGRKLVSNQFAIIFG 50

LacY-C.pdb._chain_B/1-91
Xyle-B-coot-0.pdb._chain_A/125-221

ELLNASIMFFAPLIINRI....GGKNALLAGTMSVRIIGSSFATS 91
QLLYVCVNYFIARSGDASWLNTDGWRYM.FASECIALLMLLYTVP 97

LacY-D.pdb._chain_B/1-101
Xyle-A-coot-0.pdb._chain_A/5-110

GSSSFATSALEVILKLTHM.FEVPFLLVG.C.FKYIT.SQFVRFSATIY 46
NSSYIFSVIATLGGFLGQTAVSITVESLSNTVFAQLQNSAANS 50

LacY-D.pdb._chain_B/1-101
Xyle-A-coot-0.pdb._chain_A/5-110

LVCFCFF.KQLAMIFMSVLAGNMYESIGFGAYLVLGLVALGFTLVFT 95
LLGFCVAALIGCIIGGALGGYSNRFGRRSLKIAAVLFFISGVGSAWP 100

LacY-D.pdb._chain_B/1-101
Xyle-A-coot-0.pdb._chain_A/5-110

LSGPFP 101
ELGFTS 106

LacY-A.pdb._chain_B/1-96
Pept_4IKZ_D.pdb._chain_A/1-101

TNFWMFGLFFYYFFIMGAYFPPPFIWLDINHIS.KSDTGIIFAAILF 49
GGLVHPIWLVLASYFIVYGRELCLPGVLSATTKLAPAFSQTMSLWFSD 50

LacY-A.pdb._chain_B/1-96
Pept_4IKZ_D.pdb._chain_A/1-101

SLFFQPLFGSSLK.DLGLRKLYLWITGLMVFAAFFIFIFGPLLQ...Y 95
NAFAQAINQAVRLVFYTPPTAFFTIGGAAALVGLILLALAPIGRMLK 100

LacY-A.pdb._chain_B/1-96
Pept_4IKZ_D.pdb._chain_A/1-101

N 96
G 101
<table>
<thead>
<tr>
<th>Protein</th>
<th>Chain</th>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>LacY-B.pdb</td>
<td>B</td>
<td>83</td>
<td>36</td>
<td>LVGSIVGGYLG......FCFNAGAPAVEAFKVSRRS......N.FE.......</td>
</tr>
<tr>
<td>PepT_4IKZ_B</td>
<td>A</td>
<td>1-100</td>
<td>82</td>
<td>FGRARMFGCVGALGASIVGIMFTINNNQFVFWLGSQCAVLLAVFE......</td>
</tr>
<tr>
<td>LacY-C.pdb</td>
<td>B</td>
<td>88</td>
<td>49</td>
<td>LWFLSLYVIGVSCTVDFDQQFANFFTSF.FATGEOQTRVGYVTMGEL</td>
</tr>
<tr>
<td>PepT_4IKZ_C</td>
<td>A</td>
<td>1-102</td>
<td>50</td>
<td>AGFSIFYMGINLGAFLAPLVVGTAGMYFHLGFLAAVGMFLGLVVFVA</td>
</tr>
<tr>
<td>LacY-C.pdb</td>
<td>B</td>
<td>88</td>
<td>50</td>
<td>LWNASMFF.APILINRIG.........G.KNA..LLLAGTIMSVRIIGS</td>
</tr>
<tr>
<td>PepT_4IKZ_C</td>
<td>A</td>
<td>1-102</td>
<td>100</td>
<td>LNPLFIIIAPVFAWVWKLQCRQPTIPQFALGLFAFLSFIVILVPGH</td>
</tr>
<tr>
<td>LacY-D.pdb</td>
<td>B</td>
<td>89</td>
<td>40</td>
<td>SALEVVILKTLHMFEVPFLLVCGFKYITSQ............FEVRFSATIY</td>
</tr>
<tr>
<td>PepT_4IKZ_A</td>
<td>A</td>
<td>1-99</td>
<td>50</td>
<td>HKPGLFOTAFFEFPSWYY.YGM.RAILVYYMYYEVSQGGLDHEHLLAIM</td>
</tr>
<tr>
<td>LacY-D.pdb</td>
<td>B</td>
<td>89</td>
<td>40</td>
<td>LVCFCFFKQLAMIFMSVLAGNMYIESIGFQGAYVLGLVALGFTLSVFT</td>
</tr>
<tr>
<td>PepT_4IKZ_A</td>
<td>A</td>
<td>1-99</td>
<td>99</td>
<td>SYIYALVYMGIIGGLWADVFTSGAVFYPYGGILIMAGHIALAIPGVA</td>
</tr>
</tbody>
</table>
REFERENCES