doi:10.1016/j.jmb.2011.02.008 J. Mol. Biol. (2011) 407, 698–715

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

The Alternating-Access Mechanism of MFS Transporters Arises from Inverted-Topology Repeats

Sebastian Radestock and Lucy R. Forrest*

Computational Structural Biology Group, Max Planck Institute of Biophysics, Max-von-Laue-Str. 3, 60438 Frankfurt am Main, Germany

Received 22 October 2010; received in revised form 16 January 2011; accepted 2 February 2011
Available online 18 February 2011

Edited by J. Bowie

Keywords:
symmetry; outward-facing conformation; structural modeling; lactose permease; homology model

Lactose permease (LacY) is the prototype of the major facilitator superfamily (MFS) of secondary transporters. Available structures of LacY reveal a state in which the substrate is exposed to the cytoplasm but is occluded from the periplasm. However, the alternating-access transport mechanism requires the existence of a periplasm-facing state. We recently showed that inverted-topology structural repeats provide the foundation for the mechanisms of two transporter families with folds distinct from the MFS. Here, we generated a structural model of LacY by swapping the conformations of inverted-topology repeats identified in its two domains. The model exhibits all required properties of an outward-facing conformation, i.e., closure of the binding site to the cytoplasm and exposure to the periplasm. Furthermore, the model agrees with double electron–electron resonance distance changes, accessibility to cysteine-modifying reagents, cysteine cross-linking data, and a recent structure of a distantly related transporter. Analysis of the intradomain differences between the two states suggests a role for conserved sequence motifs in occluding the central pathway through kinking of the pore-lining helices. In addition, predicted re-pairing of critical salt-bridging residues in the binding sites agrees remarkably well with previous proposals, allowing a description of the proton/sugar transport mechanism. More fundamentally, our model demonstrates that inverted-topology repeats provide the foundation for the alternating-access mechanisms of MFS transporters.

Introduction

The major facilitator superfamily (MFS) of transporters constitutes the largest known family of secondary transporter proteins.[1,2] They are responsible for the ion-coupled transport of a wide range of substrates, from monoamines to sugars to peptides. Lactose permease (LacY), a proton/sugar cotransporter, is the prototype of these transporters, which typically contain 12 transmembrane (TM) helices, although some members have 14 TM helices.

Seminal X-ray crystallography studies solving the structures of LacY[3] and another MFS transporter, GlpT,[4] revealed that the 12 TM helices are arranged into two domains of 6 TM helices each[1,2] (Fig. 1a), which are related by a 2-fold pseudo-symmetry with an axis that runs normal to the membrane and between the two halves. The evolutionary relationship suggested by the structural similarity of the two domains is consistent with the presence of evolutionarily conserved sequence motifs between TM helices 2 and 3 in the N-terminal domain and the

*Corresponding author. E-mail address: lucy.forrest@biophys.mpg.de.

Abbreviations used: LacY, lactose permease; MFS, major facilitator superfamily; TM, transmembrane; EAAT, excitatory amino acid transporter; DEER, double electron–electron resonance; NPGal, 4-nitrophenyl-α-D-galactopyranoside; NEM, *N*-[^{14}C]ethylmaleimide; TDG, β-D-galactopyranosyl-1-thio-β-D-galactopyranoside; MD, molecular dynamics; DOPE, discrete optimized potential energy; SASA, solvent-accessible surface area.

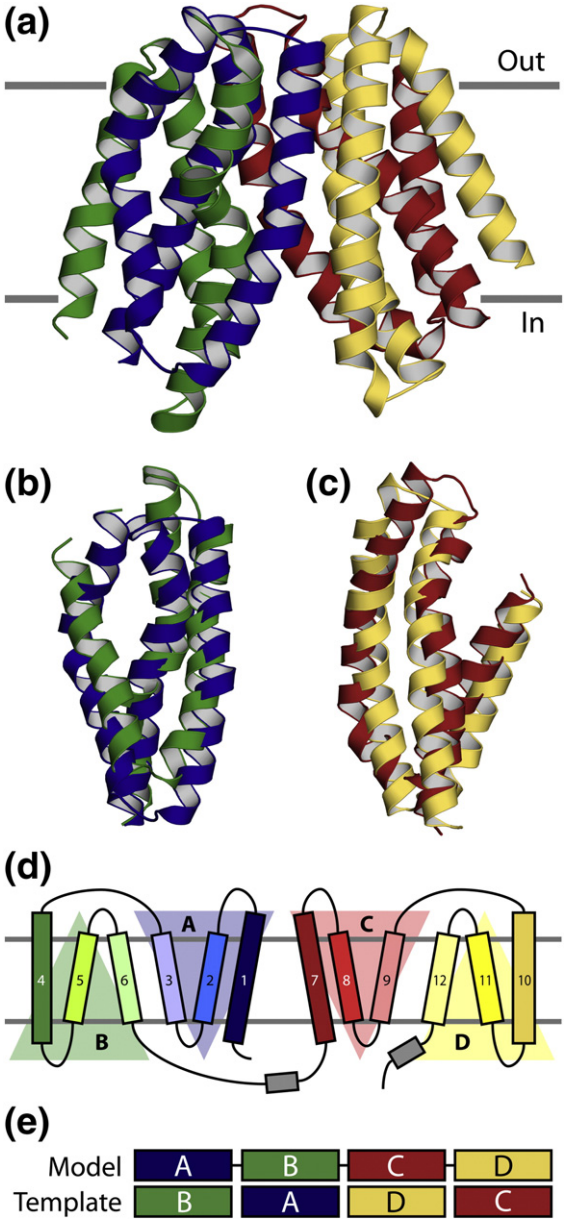

Fig. 1. Internal structural repeats in the LacY crystal structure. (a) The crystal structure viewed along the plane of the membrane with the periplasmic side to the top. The TM helices are shown in blue (repeat unit A), green (B), red (C), and yellow (D). Superposition of the two 3-TM helix repeat units from (b) the N-terminal domain and (c) the C-terminal domain, calculated using the structural alignment program SKA.[5,6] These and subsequent molecular figures were generated using PyMOL.[7] (d) A schematic of the TM topology of LacY emphasizing the structural relationships between the four segments A–D. (e) Schematic of the sequence alignment used to generate the repeat-swapped model, colored as in (a). For example, segment A was modeled using the structure of segment B as a template.

symmetry-related TM helices 8 and 9 in the C-terminal domain.[1,2]

Analysis of the LacY crystal structures revealed that the two 6-TM helix domains line a cavity that is open to the cytoplasm,[3] with a ligand binding site at the approximate center of the membrane.[3] The side chains important for sugar recognition are predominantly in the N-terminal domain.[8] Residues required for proton translocation are also approximately halfway across the membrane, although these residues are predominantly located in the C-terminal domain.[9]

Substrate-induced changes in side-chain reactivity probed through various cysteine-modifying reagents,[10–14] changes in the environment of site-directed fluorescent probes,[9,15,16] and spin labels,[17] as well as site-directed thiol cross-linking[18–20] and fluorescence resonance energy transfer measurements,[21,22] all suggest that LacY, like most secondary transporters,[23,24] undergoes a large conformational change during the functional transport cycle in order to expose the substrate to alternate sides of the membrane, according to the so-called alternating-access mechanism. Very recently, a structure of another *Escherichia coli* MFS transporter, L-fucose permease, was reported in an outward-facing state.[25] However, this protein is only distantly related (~10% sequence identity) to LacY; thus, a molecular understanding of the mechanism of transport by MFS transporters and LacY is still hampered by the lack of structures of the same protein in alternate states during the transport cycle.

Recently, we have shown that inverted-topology repeats in secondary transporters encode the structural information required for the generation of the two alternate states.[26,27] Specifically, these inverted-topology repeats were found, in any given state, to differ in the internal arrangement of the helices in a manner that was crucial for the formation of that state. By homology modeling of the two repeat units onto one another and thus by swapping of their conformations, we were able to generate models of previously unknown states of these transporters, demonstrating that the conversion between the two states occurs by interchanging the conformations of their inverted-topology repeats. The two transporters studied previously were representatives from the neurotransmitter sodium symporter[26] and excitatory amino acid transporter (EAAT) families,[27] whose folds differ significantly from the MFS transporter fold.

Here, we show that the MFS transporters also use a conformational change mechanism involving inverted-topology repeat units. Aside from the symmetry relationship between the two halves of the structure mentioned above, sequence analysis of MFS transporters has suggested that each of the two domains may have originated from a gene

duplication event.[28] Furthermore, analysis of the LacY crystal structure reveals the presence of inverted-topology repeat units within each of the two domains (Fig. 1b–d). That is, the first three TM helices (TMs 1–3, which we refer to as repeat unit A) are related in structure to the second three TM helices (TMs 4–6; repeat unit B) via a 2-fold symmetry axis that runs through the center of the 6-TM domain and parallel with the membrane plane (Fig. 1b). A similar relationship is found for the two pairs of 3-TM helices (repeat units C and D containing TMs 7–9 and 10–12, respectively) in the C-terminal domain (Fig. 1c).

We present a model of LacY generated by swapping the conformations of the repeat units in each half of the structure and show that it represents an outward-facing conformation of LacY. We also generated a second model of this state by homology modeling with the structure of FucP as a template. We compare these models with the extensive available experimental evidence and find both models in agreement with a large proportion of the data. In addition, the "repeat-swapped" model suggests the roles of conserved sequence motifs and moreover suggests an arrangement of key salt-bridge interactions in this outward-facing state of LacY, consistent with expectations from experiments. Finally, we discuss the implications of our results in terms of the role of inverted-topology repeats in MFS transporters.

Results

Examination of internal structural repeats in LacY

To characterize the internal symmetry of the MFS transporter fold, we analyzed various subdomains within the structure of LacY (Fig. 1a). The structural similarity between the two 6-TM helix domains is high, with a C^{α} root-mean-square deviation (RMSD) of 2.7 Å for the aligned residues, as determined by the structural alignment program SKA.[5,6] The two domains exhibit the same TM topology (Fig. 1d).

We also identified a structural relationship between the first three helices and the last three helices within each domain (Fig. 1b and c). Specifically, segment A (TMs 1–3) can be superimposed on segment B (TMs 4–6) with a C^{α} RMSD of 3.8 Å for the aligned residues (Fig. 1b), while segment C (TMs 7–9) can be superimposed on segment D (TMs 10–12) with an RMSD of 3.3 Å (Fig. 1c). This is consistent with earlier suggestions that the 6-TM helix domains of MFS transporters originated from a 3-TM precursor by gene duplication.[28] The two 3-TM segments within each domain have different TM topologies; therefore, one of them has to be rotated by 180° around an axis parallel to the membrane plane in order to achieve the fit.

Overall then, the structure of LacY can be described as containing two inverted-topology repeats (A and B, and C and D; Fig. 1d). However, by considering A and D together as a single noncontiguous unit, and by considering B plus C as a separate contiguous unit (with opposite topology to A plus D), we revealed differences in the orientations of the segments to one another. Specifically, fitting A onto B, and comparing the relative orientations that D and C adopt with respect to A and B, reveal a difference of an ~51° rotation about an axis running through the plane of the membrane

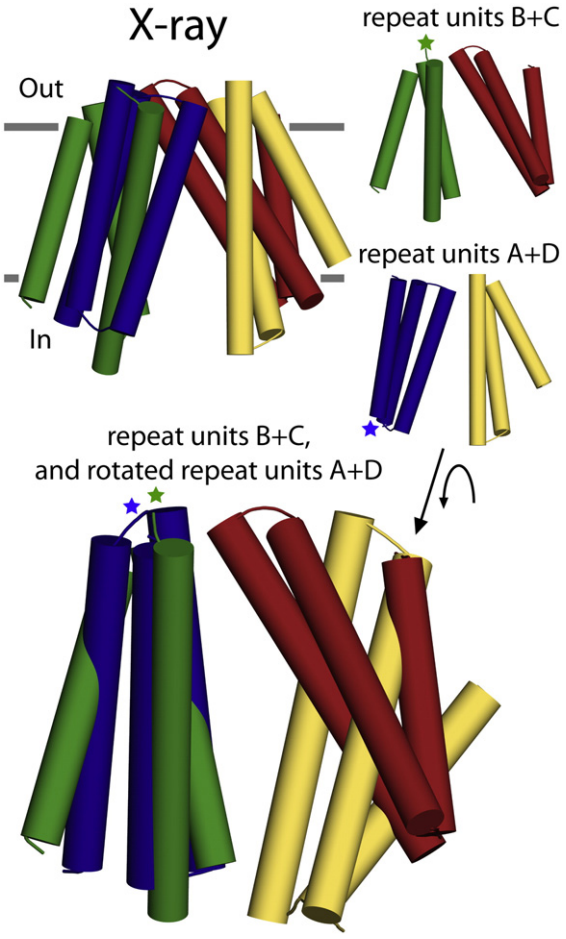

Fig. 2. Structural asymmetry of repeats in LacY. TM helices 1–12 of the crystal structure were separated into two segments, each containing two repeat units from the N-terminal and C-terminal domains. Superimposition of these two segments was generated by aligning the repeat units of the N-terminal domain, i.e., fitting repeat unit A (blue) onto repeat unit B (green), using the structural alignment program SKA.[5,6] This required an ~180° rotation of repeat unit A with respect to the membrane plane. The N-terminal end of each repeat is indicated by a star, in green for B plus C and in blue for A plus D. Note the different orientation that the repeat units from the C-terminal domain (red and yellow) adopt with respect to the repeat units from the N-terminal domain.

(Fig. 2). That is, within the repeat consisting of B plus C, the two repeats units are closest together at the N-terminal side of B, whereas in the repeat consisting of A plus D the two units approach each other on the opposite side of the repeat (Fig. 2). Such differences in conformation between inverted-topology repeats were shown to be critical for the conformational changes in neurotransmitter sodium symporter and EAAT.[26,27] That is, by allowing each repeat to exchange between two distinct conformations, the protein can create the two alternate states required for the transport mechanism.

A model built by swapping conformations of inverted-topology repeats is outward-facing

To demonstrate that the inverted-topology repeats encode the conformational change required for transport by MFS proteins, we tested whether the distinct conformations of their inverted-topology repeats could be used to model an alternate conformation of LacY using the approach developed previously, namely swapping of the conformations of the repeats. This was carried out for LacY here by modeling the conformation of segment A using the structure of segment B as a template and *vice versa*, while also modeling the conformation of segment C using the structure of segment D as a template and *vice versa* (Fig. 1e; Fig. S1).

The model of LacY generated by swapping the conformations of the repeats (referred to as the "repeat-swapped" model) represents a large conformational change compared to the crystal structure and is consistent with an outward-facing state of the protein (Fig. 3c and d). That is, the model is closed at the cytoplasmic side and open at the periplasmic side. This conformational change arises primarily from the difference in the orientation of the helices in the N-terminal domain with respect to the helices in the C-terminal domain (Fig. 2). Thus, swapping the conformations of the inverted-topology repeats changes the relative orientation of the two halves. Notably, the loop connecting the two domains (i.e., between TM helices 6 and 7), although not modeled here (see Materials and Methods), is sufficiently long and flexible that it is compatible with such a conformational change (Fig. S1).

The repeat-swapped model of LacY is extremely consistent with the recently reported crystal structure of FucP in an outward-facing conformation; the RMSD over the 12 TM helices is 3.2 Å (Fig. S2), compared to an RMSD of 4.9 Å between the inward-

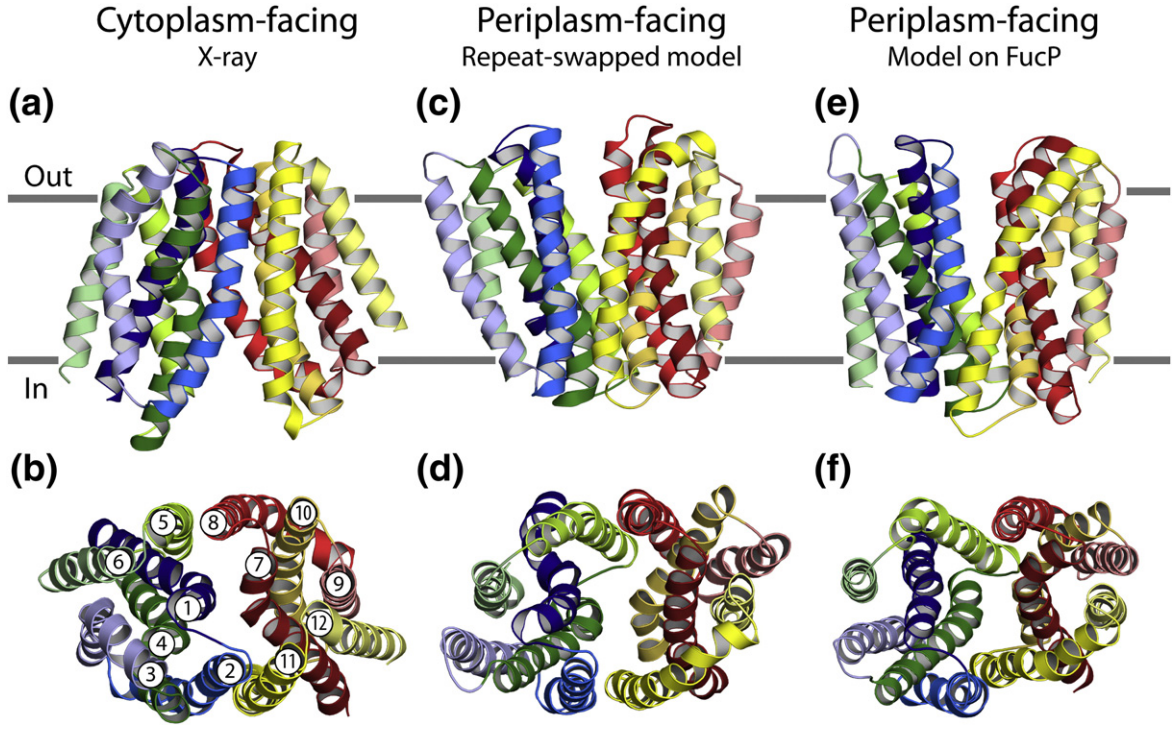

Fig. 3. Models of the outward-facing conformation of LacY generated by swapping the conformations of the repeats or by using the crystal structure of FucP in an outward-facing conformation as a template. The inward-facing crystal structure (a and b) is compared to the outward-facing repeat-swapped model (c and d) and to the outward-facing model based on FucP (e and f). The TDG molecule has been omitted for clarity. The transporter is viewed along the plane of the membrane (a, c, and e) or from the periplasm (b, d, and f).

facing crystal structure of LacY and the outward-facing crystal structure of FucP.

A homology model of LacY based on FucP is outward-facing

An outward-facing model of LacY can also be constructed using the outward-facing structure of FucP as a template (Fig. S3; see Materials and Methods). The corresponding model of LacY represents a large conformational change compared to the crystal structure of LacY and is also consistent with an outward-facing state of the protein (Fig. 3e and f). Distances between residues in the pore-lining helices of this FucP-based model are similar to those in the repeat-swapped model (Table S1), suggesting that they represent a similar state, whereas values for neither model match those for the inward-facing structure of LacY (Table S1), consistent with the LacY structure representing a different state from the models.

The Outward-facing models agree with distance changes from double electron–electron resonance measurements

Double electron–electron resonance (DEER) experiments, combined with site-directed spin labeling, have previously been used to measure interdomain distance changes induced by sugar binding to LacY.[17] Briefly, double-cysteine substitutions were made at the ends of TM helices on either the cytoplasmic side or the periplasmic side of LacY and were labeled with nitroxide. Distances between spin-labeled residue pairs were then measured in the presence and in the absence of the galactosidic sugar 4-nitrophenyl-α-D-galactopyranoside (NPGal), whose binding increased the population of transporters open to the periplasmic side, i.e., the state corresponding to our models.[17]

We calculated the predicted change in distance between those residue pairs by subtracting the distance in the inward-facing crystal structure of LacY from that in the outward-facing models. As shown in Table 1 and Fig. S4, the calculated distance changes between residues R73 and Q340, S136 and Q340, and N137 and Q340 on the cytoplasmic side of LacY, and between residues V105 and T310, I164 and T310, and I164 and S375 on the periplasmic side, are in qualitative agreement with the distance changes between the corresponding nitroxide-labeled cysteine-mutated residues measured using DEER.

Specifically, the experimentally determined distances between nitroxide labels on the cytoplasmic side of the transporter change by between −11 and −18 Å in the presence of substrate (Table 1), indicating that the two halves come closer together on this side and consequently close off the central cavity. For the repeat-swapped model, the corresponding calculated distance changes are in the range of −13 to −15 Å (Table 1; Fig. S4c), which is well within the likely variation in the DEER distances (see Table 1) compounded by the uncertainty of the model (see Materials and Methods), especially for residues in loop regions. For the model based on FucP, the corresponding calculated distance changes are in the range of −17 to −20 Å (Table 1; Fig. S4e).

On the periplasmic side of LacY, the experimentally determined distance changes are between 4 and 10 Å, indicating that the two domains become farther apart on this side, allowing the cavity to open up (Table 1). For the repeat-swapped model, the corresponding calculated distance changes are similar, i.e., in the range 8–11 Å (Table 1; Fig. S4d). For the model based on FucP, the corresponding calculated distance changes are in the range 11–13 Å (Table 1; Fig. S4f).

Table 1. Distance changes between residues during LacY conformational change

Residues	DEER[a] (Å)	Predicted (Å)	
		Repeat-swapped model[b]	Model based on FucP[b]
73 and 340	−14	−15±1	−17±1
136 and 340	−11	−15±1	−20±1
137 and 340	−18	−13±1	−17±1
105 and 310	+5	+11±2	+13±1
164 and 310	+10	+9±1	+13±1
164 and 375	+4	+8±1	+11±1

[a] Change in interspin distances determined by DEER electron paramagnetic resonance measurements of spin-labeled double-cysteine mutants of LacY for the conformational change expected upon addition of NPGal. Positive numbers imply that the distance is smaller in the inward-facing state than in the outward-facing state. The width at half-height of the distribution fitted to the distances of each state (i.e., with or without substrate), which can be expected to reflect the range of allowed conformations, is typically 4–6 Å.

[b] Distances from the models are reported as mean±standard deviation over the 100 minimized top-scoring models after subtracting the distances in the crystal structure. Distances in the crystal structure and models were calculated between the C^{α} atoms of the corresponding residues.

Predicted solvent accessibility changes agree with changes in *N*-[^{14}C]ethylmaleimide reactivity

Both the outward-facing repeat-swapped model and the model based on FucP predict significant changes in solvent accessibility. A striking example are the periplasmic halves of the pore-lining helices 2 and 11, and 5 and 8, all of which have one face that is accessible to the extracellular solution in both models but is buried in the crystal structure (Fig. 4). In contrast, the cytoplasmic halves of these helices are accessible to intracellular solution in the crystal

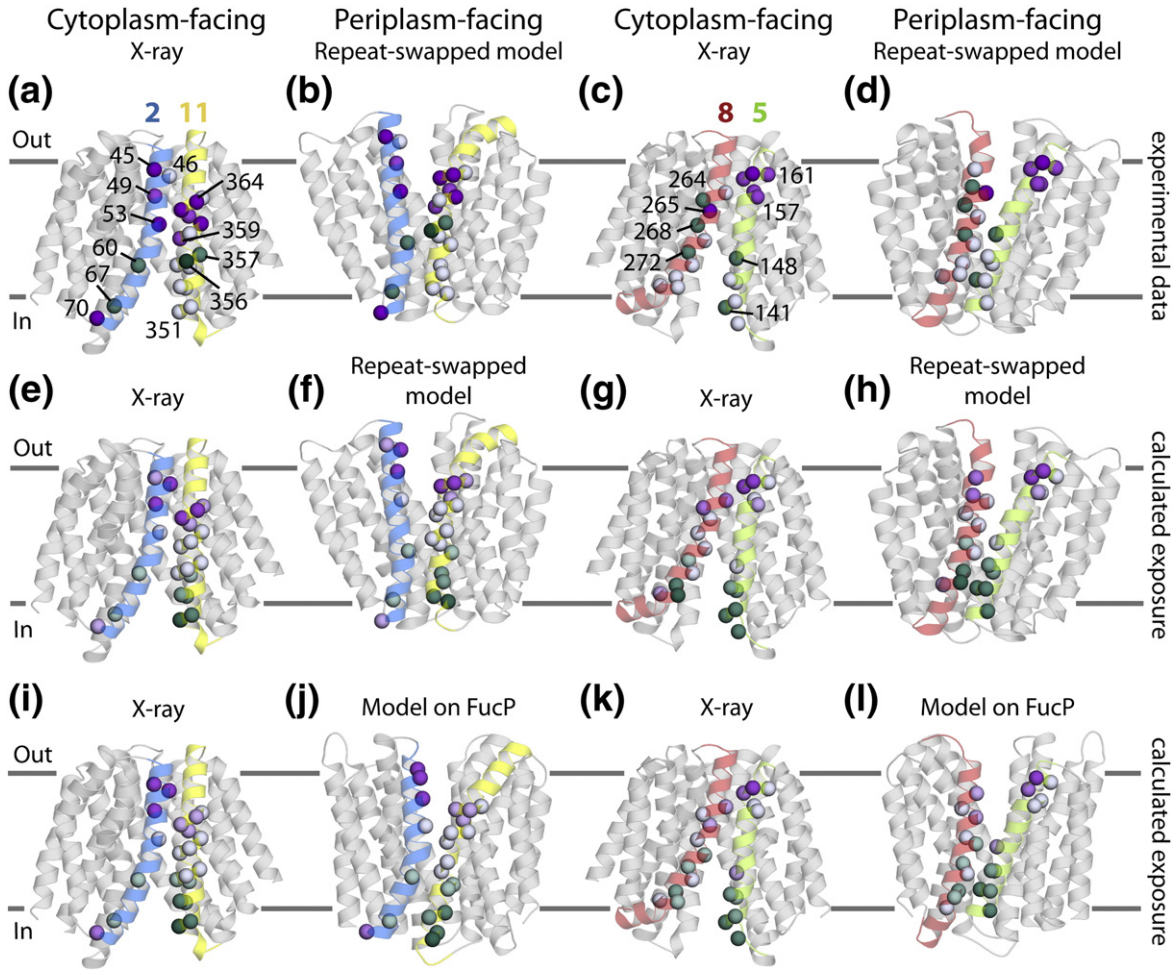

Fig. 4. Comparison of predicted changes in solvent accessibility from the repeat-swapped model and the model on FucP with experimentally determined changes in the reactivity of single-cysteine mutants with NEM. (a–d) Changes in the NEM reactivity of single-cysteine mutants in (a and b) TM helices 2 (blue) and 11 (yellow), and (c and d) in TM helices 5 (red) and 8 (green) are mapped onto the inward-facing crystal structure (a and c) and the outward-facing repeat-swapped model (b and d). (e–h) Predicted changes in the solvent accessibility of the same residues are mapped onto the inward-facing crystal structure (e and g) and the outward-facing repeat-swapped model (f and h). (i–l) Predicted changes in the solvent accessibility of the same residues are mapped onto the inward-facing crystal structure (e and g) and the outward-facing model based on FucP (f and h). Coloring of accessibility changes ranges from purple (increases in the presence of TDG) to green (decreases in the presence of TDG), through white (minimal changes). Amino acid positions discussed in the text are labeled.

structure, whereas they are buried in both models (Fig. 4). Qualitatively similar observations have been made by experiments, where the reactivity of single-cysteine mutants in TM helices 2, 5, 8, and 11 with the cysteine-modifying reagent *N*-[^{14}C]ethylmaleimide (NEM) has been measured in the presence and in the absence of the galactosidic sugar β-D-galactopyranosyl-1-thio-β-D-galactopyranoside (TDG).[10] As for NPGal, TDG binding increases the population of LacY transporters that are open at the periplasmic side, so that accessibility measurements in the presence of TDG are expected to match those in our outward-facing models better than those in the crystal structure.

We next consider the accessibility of these helices in more detail, beginning with TM 2 (see Fig. 4a and b). In the presence of TDG, cysteine residues at positions 45 and 53 in the periplasmic half of TM 2 become reactive; labeling at position 49 increases; and labeling at positions 60 and 67 in the cytoplasmic half of TM 2 decreases.[29] Interestingly, a cysteine residue at position 70, despite being at the very cytoplasmic end of TM 2, becomes more labeled in the presence of substrate.[30] A cysteine residue at position 46 is reactive, but its reactivity is unaffected by the presence of substrate.[29] Calculated changes in surface exposure from the repeat-swapped model are consistent with almost all of

these data: T45, F49, and L70 become exposed upon the conformational change; S53 becomes slightly exposed upon the conformational change; while Q60 and S67 become more buried (Fig. 4e and f). The only deviation from the experimental data is G46, which would be expected to be exposed in both states based on its reactivity, but is in fact only accessible to solution in the model; this may be because the introduced cysteine is less easily buried in the interface with TM 11 than the native glycine in the crystal structure. Calculated changes in surface exposure from the model on FucP are qualitatively the same as those for the repeat-swapped model, except for S53, which becomes slightly more buried upon the conformational change (Fig. 4i and j).

In the periplasmic half of TM 11, cysteine residues at positions 359–364 become more labeled in the presence of TDG, while positions 356 and 357 become less reactive[30] (Fig. 4a and b). In the cytoplasmic half of this helix, cysteine residues at positions 347, 348, 350, 351, and 354, as well as native C355, are reactive, but their reactivity is unaffected by substrate. The increased exposure in the periplasmic half of TM 11 is captured by the repeat-swapped model (Fig. 4e and f). For example, M362, I363, and F364 become exposed, and Q359 experiences a slight increase in surface accessibility. However, F356, F357, and A361 do not exhibit any change in surface exposure (Fig. 4e and f) even though NEM reactivity at these positions was significantly altered upon binding of TDG (Fig. 4a and b).[30] Moreover, A347, T348, Y350, L351, and F354 show a decrease in surface exposure (Fig. 4e and f). These latter deviations from the experimental data may be explained not by errors in the model but rather by the assumption that the inward-facing crystal structure is equal to the ligand-free state, which is presumably rather a dynamic ensemble, possibly containing a significant population of occluded states. Calculated changes in surface exposure from the model on FucP are qualitatively the same as for the repeat-swapped model, except for C355 and Q359, which become slightly buried upon the conformational change (Fig. 4i and j) even though NEM reactivity is unaffected or increased by substrate, respectively.

For two residues in the cytoplasmic half of TM 5 (Fig. 4c and d), binding of TDG completely abolishes labeling (of native C148) or causes weak protection of an introduced cysteine (at position 141).[31] Cysteine residues at positions 139, 142, 144, and 145 are also reactive, but their reactivity is not significantly altered by substrate. By contrast, there is a significant increase in the labeling of cysteine residues at positions 157–161 in the periplasmic half of TM 5.[31] In agreement with these data, calculated surface exposure changes based on the repeat-swapped model predict that G141 becomes buried upon the conformational change; I157, G159, and I160 become significantly more exposed; and V158 and M161 become very slightly more accessible (Fig. 4g and h). However, the model also predicts that C148 would be slightly more exposed upon the conformational change, in contrast to the experimentally observed decreases (Fig. 4e and f); the measured decrease may reflect direct steric inhibition by the substrate, which binds to the adjacent residue R144 (see Fig. S4d) but is not represented in our outward-facing model. Finally, we also find a discrepancy in the accessibilities of E139, R142, R144, and M145, which, in contrast to the experimental evidence (Fig. 4e and f), are predicted to become less exposed in the outward-facing conformation (Fig. 4g and h). Calculated changes in surface exposure from the model based on FucP are qualitatively the same as for the repeat-swapped model (Fig. 4k and l), with the exception that I157 is less exposed than in the repeat-swapped model, and that C148 is significantly exposed even though NEM reactivity is decreased by substrate.

In TM 8, TDG binding decreases the reactivity of cysteine residues at positions 264, 268, and 272, while the reactivity at position 265 is markedly increased (Fig. 4c and d).[32] Cysteine residues at positions 262, 269, 273, 276, 277, and 279 are reactive, but their reactivity is unaffected by substrate binding. In agreement with these data, N272 becomes buried upon the conformational change (Fig. 4g and h). However, calculated surface exposure changes for V264, T265, and G268 do not accommodate the experimental data (Fig. 4g and h). Calculated changes in surface exposure from the model based on FucP are qualitatively the same as for the repeat-swapped model (Fig. 4k and l).

The models match cross-linking data for the cytoplasmic half of LacY

Long-range distances within LacY have been measured experimentally by cysteine cross-linking and Mn(II) binding studies on double-cysteine and double-histidine mutants, respectively[33] (see Materials and Methods). To enable a comparison between experimentally determined distances and calculated distances in the crystal structure and in the model, we divided the data into two classes (i.e., pairs of residues in the cytoplasmic half *versus* the periplasmic half of the transporter) (see Materials and Methods). As described previously,[33] comparison of the experimental data with the distances calculated for the crystal structure revealed a better agreement for residues on the periplasmic side (Fig. 5b) than for residues on the cytoplasmic side of the transporter (Fig. 5a). Interestingly, the comparison between the experimental data and the calculated distances in the repeat-swapped model leads to the converse observation, i.e., the agreement between experiment and prediction is

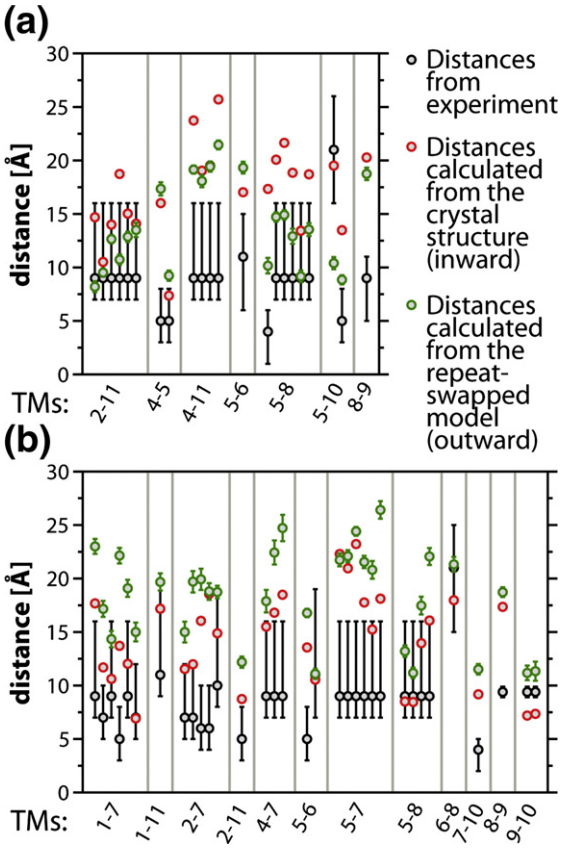

Fig. 5. Comparison of distances extrapolated from cysteine cross-linking and Mn(II) binding studies with distances calculated in the crystal structure and in the repeat-swapped model. Distances between pairs of residues are sorted according to the TM helices to which they belong (e.g., the section marked "2–11" includes six pairs of residues; one residue in each pair is in TM 2, and the other is in TM 11). Distances are shown for residues in either the cytoplasmic side (a) or the periplasmic side (b) of the transporter (see Materials and Methods for details). Experimentally determined distances (O) are those implied by cysteine cross-linking and Mn(II) binding studies, with error bars reflecting the distance range expected to be consistent with a cross-link (see Materials and Methods for details). Calculated distances from structural models are shown for the same pairs of residues, measured either in the crystal structure (O) or in the repeat-swapped model (O), where error bars reflect the standard deviation over 100 minimized selected models.

better for distances on the cytoplasmic side (Fig. 5a) than for distances on the periplasmic side of the transporter (Fig. 5b). The comparison between the experimental data and the calculated distances in the model based on FucP leads qualitatively and quantitatively to the same results (data not shown).

More specifically, for 82.5% of the residue pairs on the periplasmic side, the experimental data are closer to the values calculated for the crystal structure than to those calculated for the repeat-swapped model (Fig. 5b). Indeed, for 17.5% of the periplasmic pairs, the magnitude of the difference from the experimental data for the crystal structure is half that for the repeat-swapped model. Similar numbers are obtained from the model based on FucP (80% and 27.5%, respectively). Thus, the crystal structure provides a significantly better representation of the measured distances than the models do for residue pairs on the periplasmic side of the membrane. By contrast, of the cross-linked pairs on the cytoplasmic side, 81.8% match the distances in the repeat-swapped model better than the distances in the crystal structure (Fig. 5a). Moreover, for 40.9% of the pairs on the cytoplasmic side, the difference in the calculated distances for the repeat-swapped model is at least half that for the crystal structure. Similar numbers are obtained from the model based on FucP (81.8% and 45.5%, respectively). Thus, on the cytoplasmic side, the models provide a significantly better representation of the experimental data than the crystal structure does.

The outward-facing repeat-swapped model suggests a role for conserved sequence motifs in the closure of the substrate pathway

Analysis of the LacY crystal structure reveals that closure of the transporter at the periplasmic side is achieved by contacts between the periplasmic halves of the pore-lining TM helices 1 and 7. This helix–helix contact is facilitated by kinks in these helices located at P31 (~48°) and A244 (~36°), respectively (Fig. 6a and b; Fig. S1 and Table S2), which enable the helices to run more parallel with each other for one or two turns, even though the remainder of the helices are almost at right angles to one another (Fig. 3a and b, dark blue and dark red). An additional proline residue at position 28 in TM 1 may facilitate the kink formation at P31. This PXXP motif has also been found in another family of MFS transporters (i.e., multidrug resistance transporters).[34] Although the residue at the position of the kink in TM 7 (A244) is not a proline in *E. coli* LacY, a proline is found at this position in homologous LacY sequences (Fig. S5a and b). This observation suggests that position 244 is the location of a vestigial proline that was later replaced during the evolution of *E. coli* LacY while still retaining a local environment that can stabilize a kink in the helix.[35]

In the repeat-swapped model of the outward-facing conformation of LacY, closure at the cytoplasmic side is, by symmetry, predicted to be facilitated by kinks in TM helices 4 and 10 located at E126 (~52°) and C333 (~30°), respectively (Figs. 3c and d and 6c and d; Fig. S1 and Table S2). In support of these predicted kinks, a proline residue is found in TM 4 precisely at the position that corresponds to P28 in TM 1 (i.e., at P122). Moreover,

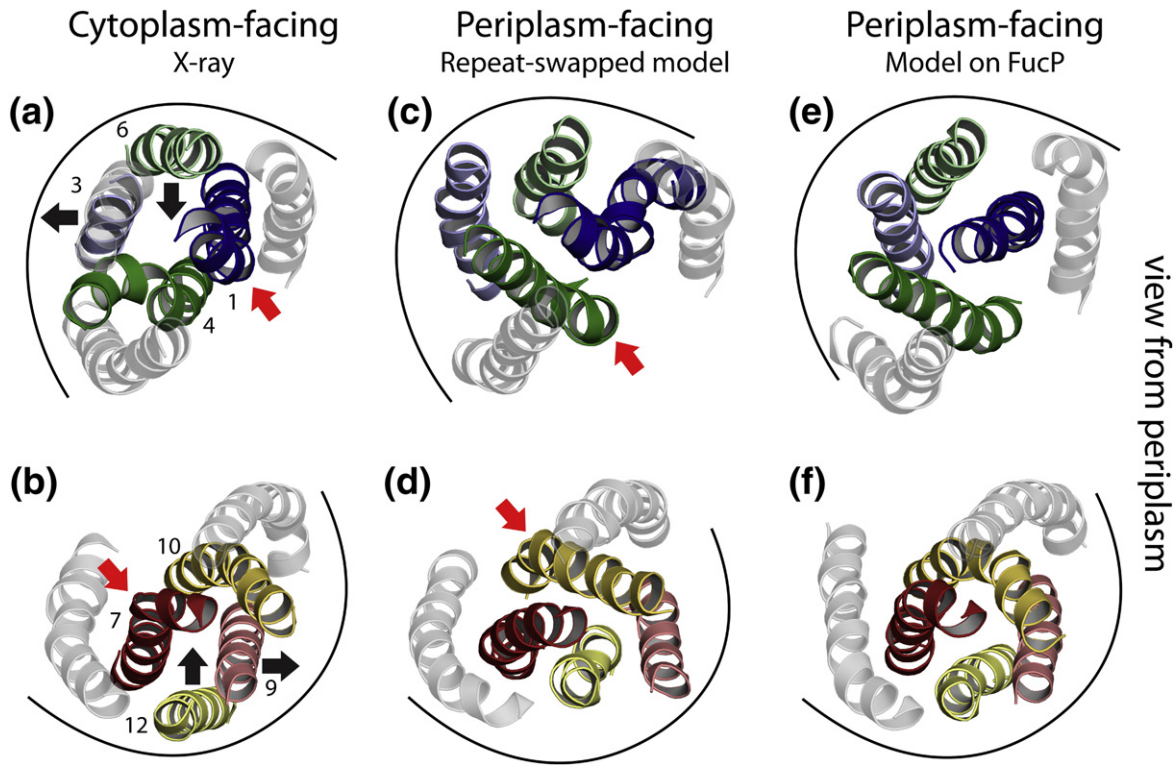

Fig. 6. Intradomain conformational changes. The N-terminal domain (a, c, e) and the C-terminal domain (b, d, f) from the crystal structure (a and b), from the repeat-swapped model (c and d), and from the model based on FucP (e and f) viewed from the periplasm. Comparison between the crystal structure and the repeat-swapped model (a–d) indicates that upon the conformational change between the inward-facing state and the outward-facing state, closure of the transporter at the cytoplasmic side and the periplasmic side is facilitated by kink formation in the pore-lining TM helices 1 and 7, and 4 and 10 (red arrows). In addition, a change in the location of the peripheral helices (TMs 3 and 6 from the N-terminal domain and TMs 9 and 12 from the C-terminal domain) can be observed (black arrows). Such changes are not observed in the model based on FucP. Helices are colored according to Fig. 1d. A black line indicates the approximate lipid interface.

molecular dynamics (MD) simulations have shown that this region of TM 4 is highly mobile in the inward-facing conformation,[36] indicating that it is a good candidate for such a conformational change. In TM 10, the kink is predicted to be at a position that does not contain a proline or an ancestral proline in homologous transporters; however, a glycine at position 332 and a proline at position 327 might facilitate the predicted structural distortions (Fig. S1). In summary, the outward-facing repeat-swapped model of LacY predicts that the role of the MFS sequence motifs is to form a well-defined occlusion of the central pathway by enabling bending of the pore-lining helices at each end of the pathway.

In the model based on FucP, the kinks in TM helices 4 and 10 are much less pronounced (Fig. 6e and f). In particular, the conformation of the C-terminal domain of the outward-facing model on FucP is more similar to the C-terminal domain of the inward-facing crystal structure than to that of the outward-facing repeat-swapped model (Fig. 6b, d, and f; Table S2). Nevertheless, although smaller than in LacY (Table S2), the kinks in TMs 4 and 10 of the FucP-based model (and also in the structure of FucP itself) also help to close off the cytoplasmic access pathway. Thus, these conserved motifs appear play a role, albeit to varying degrees, in closing off the substrate pathway during the conformational change of MFS transporters.

Predicted changes in the location of the peripheral helices

Comparison of the conformations of the MFS transporter domains in the inward-facing crystal structure and in the outward-facing repeat-swapped model suggests changes in the relative location of the peripheral TM helices, i.e., TMs 3 and 6 in the N-terminal domain, and TMs 9 and 12 in the C-terminal domain (Fig. 6a–d). Specifically, the repeat-swapped model predicts a shear movement of the peripheral helices upon the conformational change, i.e., in going from the inward-facing state to

the outward-facing state, TM 3 would move away from TM 1, while TM 6 would come closer to TM 4 (Fig. 6a and c). Analogously, TM 9 in the C-terminal domain would move away from TM 7, and TM 12 would come closer to TM 10 (Fig. 6b and d). The outward-facing model based on FucP does not predict these changes in the location of the peripheral helices (Fig. 6e and f).

Predicted changes in salt-bridge interactions are consistent with the proposed mechanism

Residues that play a critical role in the mechanism of proton/substrate symport have been identified by extensive site-directed and cysteine-scanning mutageneses, alongside functional studies.[37,38] Structural interpretation of these data has been facilitated by the structures of LacY in both the TDG-bound and unbound states,[3,8,39] as well as by MD simulations based on those structures.[36,40–42] Although we recognize that the expected accuracy of our outward-facing models is limited (see Materials and Methods), we were interested to note that changes in the interactions between binding site residues predicted by the repeat-swapped model, which arise in part from the changes in the relative positions of the peripheral helices described above, are in remarkably good agreement with the transport mechanism suggested in those earlier studies.

One of the most important residues in the substrate binding site is E269 from TM 8.[43,44] In the inward-facing crystal structure, this residue interacts with the substrate TDG, as well as with R144 from TM 5 (Fig. S6d), and thereby provides a direct link between the sugar and both the N-terminal domain and the C-terminal domain.[3,45] However, biochemical evidence suggests that, after protonation, E269 no longer interacts with R144, but instead shares the proton with H322 from TM 10.[46–49] Such an interaction between E269 and H322 is absent in the sugar-bound and apo structures of the inward-facing state (Fig. 7a and

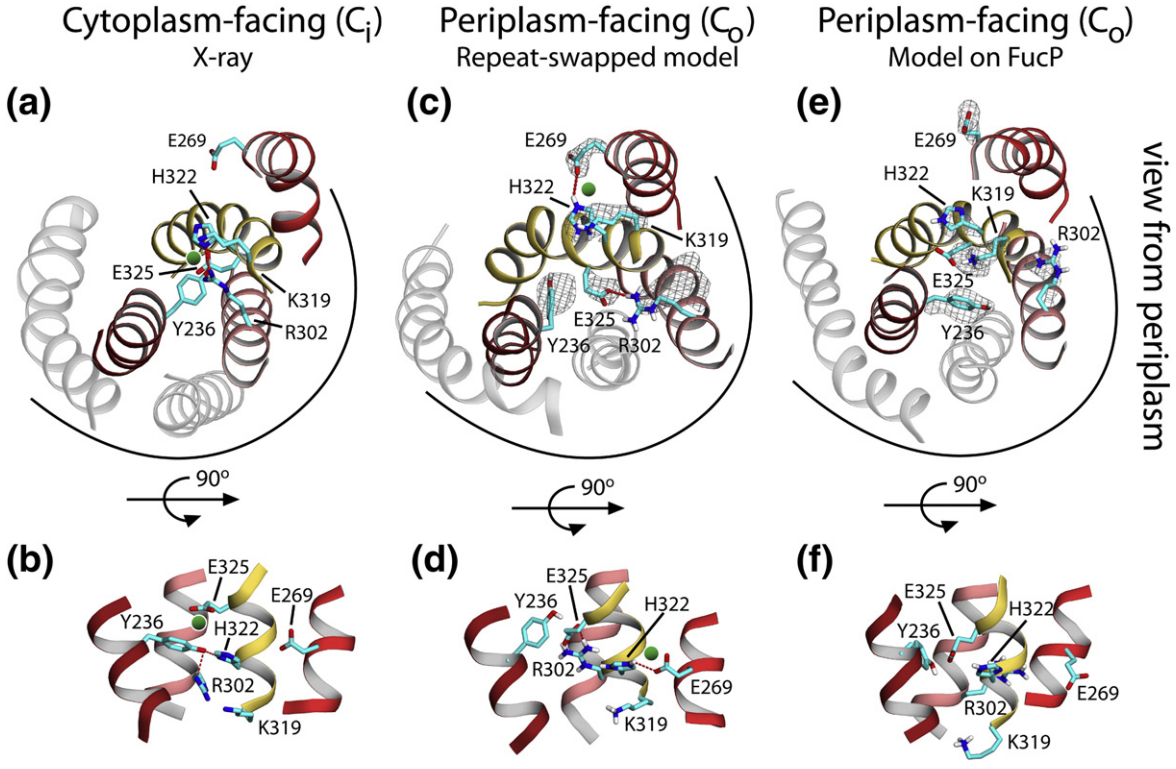

Fig. 7. Predicted changes in salt-bridge interactions. Key salt bridges in the crystal structure (a and b), in the repeat-swapped model (c and d), and in the model based on FucP (e and f) are shown for two different views, i.e., from the periplasm (upper panel) or along the plane of the membrane (lower panel). E325 is assumed to be protonated (green circle) in the state represented by the crystal structure (a and b). In the outward-facing state, a proton is assumed to be shared between E269 and H322 (see the text for details); such an interaction is only found in the repeat-swapped model (c and d) but not in the model based on FucP (e and f). In addition, an interaction between E325 and R302 can be found in the repeat-swapped model (c and d) but is not observed in the crystal structure (a and b) or the model on FucP (e and f). Helices are colored according to Fig. 1d. Putative hydrogen bonds are shown as dashed red lines. A black line indicates the approximate lipid interface. The statistical distribution of side-chain conformers within the ensemble of 100 minimized selected models is represented using reweighted atomic density maps,[50] as calculated by Xplor-NIH.[51,52]

b; Fig. S6d–i), but is present in the outward-facing state predicted by our repeat-swapped model (Fig. 7c and d; Fig. S6b and c), consistent with the proposed mechanism.[46,49] Interestingly, E269 and H322 do not interact directly in the model based on FucP (Fig. 7e and f; Fig. S7).

Concomitant with the formation of the E269-H322 salt bridge, two residues within the C-terminal domain (namely R302 from TM 9 and E325 from TM 10) are also predicted to form a direct interaction in the outward-facing repeat-swapped model (Fig. 7c and d; Fig. S6b and c), unlike in the inward-facing structures, where the two residues are separated by Y236 from TM 7 (Fig. S6f and i). The orientation of this tyrosine appears to be governed, in part, by an interaction with H322. Breaking this Y236-H322 interaction therefore appears to be associated with formation of both E269-H322 and E325-R302 salt bridges (Fig. 7c and d; Fig. S6b and c). These rearrangements appear to result from the changes in the position of the peripheral TM helices 9 and 12 described above. That is, an outward translation of TM 9 away from TM 7, and an inward translation of TM 12 toward TM 10 (Fig. 6b), act to separate R302 from Y236, thus enabling R302 to instead directly interact with E325 (Fig. 7d). The model based on FucP does not reproduce this interaction between E325 and R302 (Fig. 7e and f; Fig. S7b and c).

Discussion

The alternate state of LacY can be modeled using inverted-topology repeats

All available crystal structures of LacY represent the inward-facing state, even though theoretical considerations, as well as significant biochemical and biophysical evidence, support the presence of an outward-facing conformation to expose the sugar binding site to the periplasm, similar to that recently reported for a distantly related MFS transporter, FucP. The global conformational change implied by these data is a relative rotation of rigid N-terminal and C-terminal domains, consistent with the previously proposed rocker-switch mechanism.[45,53,54] Indeed, a simple model built by a 60° rotation of the N-terminal and C-terminal domains relative to one another is in agreement with much of the experimental data.[17] Here, we demonstrate that the outward-facing conformation of LacY is, in fact, a symmetry-related version of the inward-facing structure, an observation made recently for two other major secondary transporter folds.[26,27] Specifically, a model of LacY built by swapping the conformations of the repeats from the inward-facing structures corresponds to an outward-facing state of the protein, similar to a homology model built using FucP as a template. This repeat-swapped model predicts that the overall conformational change involves an ~51° rotation of the two domains relative to each other (similar to earlier proposals) (Fig. 2) but, unlike the FucP-based model, also indicates that conformational changes of the pore-lining helices and translations of peripheral TM helices may occur within each of the two domains (Fig. 6a–d).

Experimental evidence in support of the outward-facing models of LacY

During the conformational change predicted by both the repeat-swapped model and the model based on FucP, residues in loops on the periplasmic side of LacY become farther apart, while on the cytoplasmic side, the loops approach each other. The direction and magnitude of those changes are consistent with distance changes determined from DEER measurements upon binding of NPGal (Table 1; Fig. S4).[17] Quantitative differences originate from the fact that the experimentally derived distances are measured between spin labels, whereas calculated distances are between C^{α} atoms. In addition, the labeled residues are all located in or close to loop regions, whose positions are more difficult to predict accurately than the TM segments.[55] Moreover, while transporter flexibility is considered, to some extent, for the outward-facing conformations by averaging distances over the 100 minimized selected models, we did not consider flexibility in the crystal structure. That LacY is likely to be flexible in each state is reflected in the broadness of the distance distributions measured by DEER[17] (see legend to Table 1).

The outward-facing models also predict significant changes in solvent accessibility, i.e., the cytoplasmic halves of the pore-lining TM helices 2, 5, 8, and 11 are accessible in the inward-facing state, while they are buried in the outward-facing models (Fig. 4e–h). The reverse is true for the periplasmic halves of these helices (Fig. 4e–h). These predictions are consistent with the NEM reactivity of cysteine residues introduced into these regions, i.e., the reactivity of positions in the cytoplasmic halves of TM helices 2, 5, 8, and 11 often decreases upon binding of TDG, whereas the reactivity in the periplasmic halves typically increases (Fig. 4a–d).

The predicted conformational changes are also supported by data from cysteine cross-linking and Mn(II) binding studies.[33] The distances implied by the cross-linking tend to be underestimates,[56] especially for a molecule as dynamic as LacY,[57–59] since cross-linking is an irreversible process that will trap any conformation in which a suitable distance is achieved, even if those conformations occur rarely. This being said, it is interesting to note that the measured distances between residue pairs on the cytoplasmic side of the transporter agree rather well

with the outward-facing models (Fig. 5a), whereas distances on the periplasmic side of the transporter agree more closely with the inward-facing crystal structure (Fig. 5b). Since the conformation of LacY was not arrested in any way before the cross-linking experiments, the experimental results probably capture fluctuations between the inward- and outward-facing states of the transporter.

Although the repeat-swapped model recapitulates the major features of the outward-facing FucP crystal structure (Fig. S2) and, consequently, also a homology model of LacY built on FucP (Table S1; Figs. 3 and 4), there are nevertheless differences between the two models. The most striking differences are found in the predicted conformational changes within the N-terminal and C-terminal domains (Fig. 6c–f): the repeat-swapped model predicts the formation of pronounced kinks in TM helices 4 and 10, as well as changes in the location of the peripheral helices 3, 6, 9, and 12. In contrast, such conformational changes are not predicted by the model on FucP.

In fact, based on a comparison of the LacY and FucP structures, it has been suggested that rigid-body motions are sufficient to describe the conformational change by MFS transporters.[25] We consider it unlikely that the N- and C-terminal domains move entirely as rigid bodies, for several reasons: first, the structures of low-resolution occluded conformations of other MFS transporters show the two domains to be curved around the substrate binding pocket, rather than adopting parallel orientations;[53,60] second, LacY is extremely dynamic in nature, as shown by rapid rates of hydrogen/deuterium exchange;[57–59] third, pore-lining helices may need to kink or bend in order to help occlude the pathway (Figs. 3a–d and 6a–d; Fig. S6a, b, d, and e); and, finally, subtle conformation changes within the domains predicted by our repeat-swapped model (Fig. 6a–d) appear to allow for side-chain movements (Fig. 7a–d) of the type required for the proton/sugar transport mechanism, none of which has been explained by a model based on FucP or a model created by a simple rotation of the domains with respect to one another.[3] Specifically, lateral movements of the peripheral helices may help separate Y236 and R302, allowing formation of direct interactions between E235 and R302, and between H322 and E269 (Fig. 7). Nevertheless, we cannot rule out that such subtle conformational changes are specific to a given transporter, nor that the resolution of our model is insufficiently high to give an accurate prediction.

A mechanism of alternating access in MFS transporters

The mechanism of proton/sugar symport by LacY can be described using a simple scheme in which the transporter is initially in the outward-facing state, becomes protonated, recognizes and binds a sugar molecule from the periplasm, then changes its conformation to the inward-facing state, releases the sugar molecule to the cytoplasm, becomes deprotonated, and finally converts back into the outward-facing state.[38] LacY is believed to be unstable in the outward-facing state and thus is likely to be protonated immediately.[37] It has been proposed that this proton is shared between E269 from TM 8 and H322 from TM 10, based on biochemical evidence that the two residues interact.[46,49] In agreement with this proposal, in our repeat-swapped model of this state, unlike in the inward-facing crystal structures, an interaction between these residues is observed (Fig. 7c and d). The repeat-swapped model also predicts that another important salt bridge, namely between R144 from TM 5 and E126 from TM 4, is present in the outward-facing state (Fig. S6a), consistent with biochemical data.[61–63] In this protonated outward-facing conformation, LacY should now be ready to recognize and bind a sugar molecule from the periplasm.

It has been suggested that the sugar molecule is recognized and initially bound by W151 from TM 5, as well as by R144 and E126.[64–66] Indeed, in the TDG-bound crystal structure, R144 and E126 both interact with the sugar (Fig. S6d cf. Fig. S6g), indicating that substrate binding interrupts the interaction that they shared in the outward-facing state. This holo structure also shows that E269 is recruited into the sugar binding site (Fig. S6d), whereas this residue interacts with H322 in the outward-facing repeat-swapped model (Fig. S6g). Thus, upon substrate binding, the interaction between E269 and H322 is broken, and interactions between E269, R144, and the sugar molecule are formed. These events probably also enable H322 to approach E325, so that they can now share the proton (Fig. 7a and b), implying that the role of H322 is to shuttle the proton between the two glutamate residues E269 and E325 in response to the presence of substrate.

There is also evidence that, in the outward-facing state, E325 interacts directly with R302 from TM 9,[67] an interaction that is predicted by our outward-facing repeat-swapped model (Fig. 7c and d). According to the inward-facing crystal structures, subsequent protonation of E325 (by interaction with H322) likely renders its interaction with R302 less favorable, while concomitantly allowing formation of the interactions between Y236, H322, and R302 (Fig. 7a and b). Presumably, the net effect of these new interactions is to favor the transition to the inward-facing state.

After the conformational change from the outward-facing state to the inward-facing state, the sugar molecule can diffuse into the cytoplasm. Comparison of the sugar-bound and sugar-free crystal structures indicates that substrate unbinding

breaks the interaction between R144 and E269 (i.e., across the two domains), and R144 instead forms its interaction with E126 in the same domain (Fig. S6d and g). These events presumably enable the transporter to convert back into the outward-facing state. Indeed, MD simulations have indicated that manipulation of E269 either by protonation[36] or by loss of interactions with the substrate[40] leads to closure of the transporter at the cytoplasmic side.

Throughout these unbinding and reconversion steps, it is likely that E325 is deprotonated, presumably due to the reestablishment of the interaction between E325 and R302 (Fig. 7c and d). Indeed, there is evidence that interactions with R302 drive E325 to release a proton, namely that mutations at either residue cause the same specific effects on sugar binding, exchange, and counterflow.[67] We predict that this salt-bridge formation is facilitated by Y236 moving away from H322 and R302 (Fig. 7c and d).

Clearly, predicted changes in salt-bridge interactions from the repeat-swapped model are consistent with the previously proposed mechanism, which was deduced from extensive biochemical and biophysical studies.[37,38] However, the model allows for an advancement of our understanding and rationalization of these data. That is, the basis for the observed changes in salt-bridge interactions is apparently not encoded by the major conformational change, i.e., the rocking movement of the two halves relative to each other, but rather by the predicted smaller conformational changes within the individual domains. Specifically, changes in the salt-bridge interactions upon the conformational change appear to be associated with a shear motion of the peripheral TM helices 3, 6, 9, and 12. As a striking example, the existence of an interaction between E325 and R302 has been suggested from biochemical data, and the model predicts that this interaction can be formed by the shear motion of the TM helices 9 and 12. In this context, the repeat-swapped model might be more valuable than the homology model based on FucP.

The role of symmetry in alternating access

We have shown here that exploiting the symmetry relationship within the structure of LacY allows the major conformational change required for alternating access to be modeled. This demonstrates that the inverted-topology repeats in the MFS transporter structures contain within them the essential ingredients for the formation of the two alternate states required for transport. In combination with studies on the LeuT fold[26] and EAATs,[27] this work therefore provides strong evidence that inverted-topology repeats are a key factor in the alternating-access mechanism of the large majority of secondary transporters.

Materials and Methods

Constructing the preliminary sequence alignment

A pairwise sequence alignment between the two halves of the LacY sequence was first constructed by extracting the pairs of residues found when superposing the structural repeats onto one another. To do so, we superposed repeat unit A (residues 7–102) on repeat unit B (residues 103–187), and we superposed repeat unit C (residues 220–309) on repeat unit D (residues 310–399) using the structure alignment program SKA,[5,6] resulting in two pairs of alignments (Fig. 1c and d). These fragments of sequence alignment were then joined together to obtain a pairwise alignment between the model sequence (LacY) and the corresponding template, which is the LacY sequence whose repeats are rearranged so that they have the order BADC.

This preliminary alignment was adjusted manually to remove gaps in the TM helices. Specifically, the C-terminal part of TM 3, the N-terminal part of TM 5, the N-terminal part of TM 8, and the N-terminal part of TM 10 were shifted by 1 residue, whereas the N-terminal part of TM 6 and the C-terminal part of TM 7 were shifted by −1 residue. Even without further adjustments to the preliminary alignment, the corresponding model is entirely consistent with a periplasm-facing state of LacY, i.e., access to the core of the protein is closed on the cytoplasmic side and open on the periplasmic side (Fig. 3b).

Optimization of the alignment

Refinements were made to the preliminary alignment: (i) to match the secondary structure observed in the crystal structure with the helical regions in the template as far as possible; (ii) to maintain the pseudo-symmetry between the two MFS transporter domains; and (iii) to maintain the relative positions of the MFS transporter motifs and the side chains important for sugar recognition and for formation of the proton binding site at the same positions as observed in the crystal structure. The Rate4-Site algorithm[68] implemented in the ConSurf Web server[69] was used to quantify the evolutionary conservation of each amino acid in the alignment; mapping these values onto the resultant model helped guide the placement of variable residues so that they face the lipid environment, as well as the placement of more conserved residues so that they face the interior. These changes increased the ConQuass scores[70] of the resultant models from 0.00 ± 0.01 (mean ± standard deviation over 1000 Modeller models) for the initial alignment to 0.12 ± 0.01 for the final alignment, reflecting a more realistic model from physicochemical and biological perspectives while not affecting the major conformational change. Specifically, adjustments to the alignment were as follows:

1. A shift of the sequence in TM 3 by 3 residues allowed the secondary structure of the N-terminal half of TM 3 observed in the crystal structure to be matched with the helical regions in the template. Similarly, a shift of the sequence in TM 6 by −3 residues allowed the secondary structure of the N-terminal half of TM 6 observed in the crystal

structure to be matched with the helical regions in the template.
2. A shift of the sequence in TM 11 by −3 residues allowed the secondary structure of TM 11 observed in the crystal structure to be matched with the helical regions in the template.
3. A shift of the sequence in TMs 7 and 10 by −2 and 2 residues, respectively, allowed the pseudo-symmetry between the two MFS transporter domains in TM 1, 4, 7, and 10 regions to be maintained.
4. A shift of the sequence in TMs 2 and 8 by −4 and 3 residues, respectively, placed the MFS transporter motifs between TM helices 2 and 3, and between TM helices 8 and 9, in the same relative positions as observed in the crystal structure. This also maintained the pseudo-symmetry between the two MFS transporter domains in these regions.
5. A shift of the sequence in TM 5 by 4 residues allowed the relative positions of the side chains of R144 and W151, which are important for sugar recognition, to be similar to those observed in the crystal structure.

The final alignment (Fig. S1) has a sequence identity of 10% for both the N-terminal domain and the C-terminal domain (excluding the loop between TM helices 6 and 7). Typically, homology models built with this sequence identity range can be expected to have an error of ~3 Å in the positions of the C^{α} atoms.[55]

Building the repeat-swapped model

The repeat-swapped model of LacY was constructed using Modeller (version 9.7).[71] As a template, the corresponding segment of the LacY crystal structure in an inward-facing conformation (Protein Data Bank ID 1PV7) was used.[3] This structure is in complex with TDG and has a resolution of 3.6 Å. Models were built using the optimized sequence alignment (Fig. S1). The loop between TM helices 6 and 7 was not modeled because its length is out of the range accessible to current loop prediction methodologies.

To further conserve the internal structure of the model, we added restraints for segments in which the template differed slightly from the X-ray structure. Specifically, we constrained the N-terminal ends of TM helices 2, 8, and 11, and the C-terminal ends of TM helices 1, 2, and 3, to be helical. In addition, the distance between the side-chain oxygen or nitrogen atoms of the charged pairs E325 and R302, D240 and K319, and D237 and K358 was constrained to be 2.8 ± 0.1 Å. This step prevented some of their side chains from pointing to the hydrophobic lipid core, which is currently not disfavored by the energy functions used in Modeller as the lipid bilayer is not represented explicitly.

The structure with the lowest Discrete Optimized Potential Energy (DOPE) score of 1000 optimization cycles was selected as the representative. This model has reasonable quality according to PROCHECK,[72] with only one and two residues occupying generously allowed or disallowed regions of the Ramachandran plot, respectively. In addition to the representative model, the 100 models with the lowest DOPE scores were selected for further refinement by minimization (see the text below) and subsequent analysis.

Building the homology model of LacY based on FucP

A pairwise sequence alignment between FucP and LacY was obtained from the AlignMe server†. Specifically, an alignment of hydrophobicity profiles was made,[73] with additional input from secondary structure predictions. This preliminary pairwise alignment was adjusted manually to remove gaps in the TM helix regions and to optimize the burial or exposure of conserved or variable residues, respectively, according to ConSurf.[69] The ConQuass scores[70] of the resultant models increased from −0.07 ± 0.01 (mean ± standard deviation over 200 models) to 0.13 ± 0.01 for the final alignment. Specifically, the C-terminal part of TM 1, the N-terminal part of TM 3, the N-terminal part of TM 4, and the N-terminal part of TM 8 were shifted by 1 residue; the N-terminal part of TM 12 was shifted by 2 residues; whereas the N-terminal part of TM 1, the C-terminal part of TM 3, the C-terminal part of TM 5, the C-terminal part of TM 6, and the C-terminal part of TM 10 were shifted by −1 residue. Additional reasons for the adjustments were as follows:

1. Adjustment of of the alignment of TM 1 by −1 residue allowed a proline residue, which is structurally important in the LacY crystal structure, to be aligned with a proline residue in the template FucP.
2. Adjustment of the alignment of TM 4 allowed a proline residue to be aligned to a residue that is involved in the formation of a kink in the template.
3. Adjustment of the TM 8 region placed the MFS transporter motif between TM helices 8 and 9 (i.e., in the same relative position as observed in the crystal structure of LacY). It also maintained the pseudo-symmetry between the two MFS transporter domains in these regions.
4. Adjustments in TMs 11 and 12 allowed residues observed to be helical in the LacY crystal structure to be matched with helical regions in the template.
5. Adjustment of TM 10 allowed the relative position of the side chain of H322 to be similar to that observed in the crystal structure.

The final alignment (Fig. S3) has a sequence identity of ~10%. The model of LacY was constructed from this alignment using Modeller (version 9.7).[71] The 3.14-Å-resolution structure of FucP in an outward-facing conformation (Protein Data Bank ID 3O7Q) was used as a template.[25]

As for the repeat-swapped model, the structure with the lowest Modeller (DOPE) score of 1000 optimization cycles was selected as the representative. This model has reasonable quality according to PROCHECK,[72] with three and two residues occupying generously allowed or disallowed regions of the Ramachandran plot, respectively. In addition to the representative model, the 100 models with the lowest DOPE scores were selected for further

† http://www.bioinfo.mpg.de/AlignMe

refinement by minimization (see the text below) and subsequent analysis.

Energy minimization of the models

All models were energy minimized prior to analysis using CHARMM (version 3.4).[74] Prior to minimization, the positions of all hydrogen atoms were predicted using REDUCE (version 2.21).[75] Where necessary, REDUCE was also used to flip the orientations of asparagine, glutamine, or histidine side chains in order to optimize their hydrogen-bonding interactions. The subsequent CHARMM minimization was performed in three stages, each with 250 steps of steepest-descent minimization, followed by 250 steps of conjugated-gradient minimization. In the first stage, all heavy atoms were kept fixed, and only the hydrogen atoms were allowed to move. In the second stage, the heavy atoms of the backbone were kept fixed, and only the hydrogen and side-chain atoms were allowed to move. In the final stage, no constraints were applied, and all atoms were allowed to move. The minimization was performed using the CHARMM22 force field.[76] To account for the fact that the central cavity is filled with water, we used a high dielectric constant ($\varepsilon = 80$) for the calculation of all electrostatic interactions.

The final minimized selected models are available from the Protein Model Databank‡[77] under accession codes PM0076824 (repeat-swapped model) and PM0077183 (homology model based on FucP).

Multiple-sequence alignment

Using the LacY sequence as the query, we extracted 250 similar sequences from the nonredundant protein database (as of December 2, 2009) using PSIBLAST,[78] as provided in the NCBI BLAST package (version 2.2.17).[79] For the search, five PSIBLAST iterations were performed. From the PSIBLAST hits, the most informative sequences were selected according to a sequence identity cutoff of 60% using CD-HIT.[80] This resulted in 49 sequences, which were aligned using MUSCLE.[81] A sequence logo was constructed from this multiple alignment using TEXshade.[82]

Comparison with experimental data

The properties of the inward-facing crystal structure and of the outward-facing models were compared with various experimental data. For comparison with measured distances from DEER experiments on site-directed spin-labeled cysteines, distances were calculated between the C^{α} atoms of the corresponding residues, either averaged over the 100 minimized selected repeat-swapped models or models on FucP or taken from the crystal structure itself. For comparison with cross-linking data, distances were calculated between pairs of C^{β} atoms and averaged over the 100 selected models. For glycine residues, a virtual C^{β} atom was placed by rotating the N atom along the C^{α}–C bond by −120°. The cysteine cross-linking experiments used five different cross-link types (direct disulfide, 1,3-propanediyl-bis-methanoethiosulfonate, *N*,*N*′-*o*-phenylenedimaleimide, *N*,*N*′-*p*-phenylenedimaleimide, and 1,6-bis-maleimidohexane), and the C^{β}–C^{β} distances between the cross-linked residues were assumed to be 3–5, 3–6, 2–6, 2–10, or 6–16 Å, respectively. For the Mn(II) binding experiments, the equivalent C^{β}–C^{β} distances were assumed to be 8.9–9.4 Å.[33]

To discriminate between distances on the cytoplasmic side of the transporter and distances on the periplasmic side of the transporter, we defined a plane parallel with the membrane that divides the transporter into two halves. This plane is equidistant from the limits of the hydrophobic core, which in turn were defined according to the Orientation of Proteins in Membranes database.[83] A residue was considered to be on the cytoplasmic or periplasmic side when the position of its C^{α} atom was >5 Å from the central plane in the direction of the cytoplasm or periplasm, respectively. A pair of residues was classified as being on the cytoplasmic or periplasmic side when at least one of the two residues was on the cytoplasmic or periplasmic side, respectively, and the other residue was not on the opposite side (although it could be in the 10-Å-wide central region).

Changes in accessibility to the solvent were predicted by calculating the solvent-accessible surface area (SASA) for each residue in the minimized representative model (either for the repeat-swapped model or for the model based on FucP) as a percentage of the SASA of the same amino acid type (X) in a GXG tripeptide. From this, the equivalent percentage of the SASA in the crystal structure was subtracted, so that the accessibility change is positive for residues that become exposed in the outward-facing model and is negative for residues that become buried.

The distribution of side-chain conformations within the ensemble of 100 minimized selected models was visualized using reweighted atomic density maps,[50] calculated with Xplor-NIH.[51,52] In general, the density is given by the probability of atomic occupancy in each volume element when the different conformations are superimposed. Contour surfaces have been drawn at 50% of the maximum occupancy using PyMOL.[7]

‡ http://mi.caspur.it/PMDB

Acknowledgements

We thank Gary Rudnick, H. Ron Kaback, and Vladimir Kasho for helpful discussions, and H. Ron Kaback, Vladimir Kasho, and Mark Girvin for generously providing a compiled list of distance constraints. This work was supported, in part, by the DFG Collaborative Research Center 807 "Transport and Communication across Biological Membranes" (L.R.F.).

Supplementary Data

Supplementary data to this article can be found online at doi:10.1016/j.jmb.2011.02.008

References

1. Pao, S. S., Paulsen, I. T. & Saier, M. H., Jr (1998). Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–34.
2. Saier, M. H., Jr, Beatty, J. T., Goffeau, A., Harley, K. T., Heijne, W. H., Huang, S. C. *et al.* (1999). The major facilitator superfamily. *J. Mol. Microbiol. Biotechnol.* **1**, 257–279.
3. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. & Iwata, S. (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*, **301**, 610–615.
4. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. (2003). Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science*, **301**, 616–620.
5. Petrey, D. & Honig, B. (2003). GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. *Methods Enzymol.* **374**, 492–509.
6. Petrey, D., Xiang, Z. X., Tang, C. L., Xie, L., Gimpelev, M., Mitros, T. *et al.* (2003). Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins*, **53**, 430–435.
7. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, Palo Alto, CA.
8. Mirza, O., Guan, L., Verner, G., Iwata, S. & Kaback, H. R. (2006). Structural evidence for induced fit and a mechanism for sugar/H^+ symport in LacY. *EMBO J.* **25**, 1177–1183.
9. Smirnova, I., Kasho, V., Sugihara, J., Choe, J. Y. & Kaback, H. R. (2009). Residues in the H^+ translocation site define the pK_a for sugar binding to LacY. *Biochemistry*, **48**, 8852–8860.
10. Kaback, H. R., Dunten, R., Frillingos, S., Venkatesan, P., Kwaw, I., Zhang, W. & Ermolova, N. (2007). Site-directed alkylation and the alternating access model for LacY. *Proc. Natl Acad. Sci. USA*, **104**, 491–494.
11. Nie, Y., Ermolova, N. & Kaback, H. R. (2007). Site-directed alkylation of LacY: effect of the proton electrochemical gradient. *J. Mol. Biol.* **374**, 356–364.
12. Nie, W., Sabetfard, F. E. & Kaback, H. R. (2008). The Cys154→Gly mutation in LacY causes constitutive opening of the hydrophilic periplasmic pathway. *J. Mol. Biol.* **379**, 695–703.
13. Nie, Y. L. & Kaback, H. R. (2010). Sugar binding induces the same global conformational change in purified LacY as in the native bacterial membrane. *Proc. Natl Acad. Sci. USA*, **107**, 9903–9908.
14. Liu, Z. Y., Madej, M. G. & Kaback, H. R. (2010). Helix dynamics in LacY: helices II and IV. *J. Mol. Biol.* **396**, 617–626.
15. Smirnova, I. N., Kasho, V. & Kaback, H. R. (2008). Protonation and sugar binding to LacY. *Proc. Natl Acad. Sci. USA*, **105**, 8896–8901.
16. Smirnova, I., Kasho, V., Sugihara, J. & Kaback, H. R. (2009). Probing of the rates of alternating access in LacY with Trp fluorescence. *Proc. Natl Acad. Sci. USA*, **106**, 21561–21566.
17. Smirnova, I., Kasho, V., Choe, J. Y., Altenbach, C., Hubbell, W. L. & Kaback, H. R. (2007). Sugar binding induces an outward facing conformation of LacY. *Proc. Natl. Acad. Sci. USA*, **104**, 16504–16509.
18. Zhou, Y., Guan, L., Freites, J. A. & Kaback, H. R. (2008). Opening and closing of the periplasmic gate in lactose permease. *Proc. Natl Acad. Sci. USA*, **105**, 3774–3778.
19. Zhou, Y. G., Nie, Y. L. & Kaback, H. R. (2009). Residues gating the periplasmic pathway of LacY. *J. Mol. Biol.* **394**, 219–225.
20. Nie, Y., Zhou, Y. G. & Kaback, H. R. (2009). Clogging the periplasmic pathway in LacY. *Biochemistry*, **48**, 738–743.
21. Smirnova, I. N., Kasho, V. N. & Kaback, H. R. (2006). Direct sugar binding to LacY measured by resonance energy transfer. *Biochemistry*, **45**, 15279–15287.
22. Majumdar, D. S., Smirnova, I., Kasho, V., Nir, E., Kong, X. X., Weiss, S. & Kaback, H. R. (2007). Single-molecule FRET reveals sugar-induced conformational dynamics in LacY. *Proc. Natl Acad. Sci. USA*, **104**, 12640–12645.
23. Forrest, L. R. & Rudnick, G. (2009). The Rocking Bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology*, **24**, 377–386.
24. Forrest, L. R., Kraemer, R. & Ziegler, C. (2011). The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta*, **1807**, 167–188.
25. Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H. *et al.* (2010). Structure of a fucose transporter in an outward-open conformation. *Nature*, **467**, 734–738.
26. Forrest, L. R., Zhang, Y. W., Jacobs, M. T., Gesmonde, J., Xie, L., Honig, B. H. & Rudnick, G. (2008). Mechanism for alternating access in neurotransmitter transporters. *Proc. Natl Acad. Sci. USA*, **105**, 10338–10343.
27. Crisman, T. J., Qu, S., Kanner, B. I. & Forrest, L. R. (2009). Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats. *Proc. Natl Acad. Sci. USA*, **106**, 20752–20757.
28. Hvorup, R. N. & Saier, M. H., Jr (2002). Sequence similarity between the channel-forming domains of voltage-gated ion channel proteins and the C-terminal domains of secondary carriers of the major facilitator superfamily. *Microbiology*, **148**, 3760–3762.
29. Venkatesan, P., Liu, Z., Hu, Y. & Kaback, H. R. (2000). Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: *N*-ethylmaleimide-sensitive face of helix II. *Biochemistry*, **39**, 10649–10655.
30. Ermolova, N., Madhvani, R. V. & Kaback, H. R. (2006). Site-directed alkylation of cysteine replacements in the lactose permease of *Escherichia coli*: helices I, III, VI, and XI. *Biochemistry*, **45**, 4182–4189.
31. Kwaw, I., Zen, K. C., Hu, Y. & Kaback, H. R. (2001). Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: helices IV and V that contain the major determinants for substrate binding. *Biochemistry*, **40**, 10491–10499.
32. Frillingos, S. & Kaback, H. R. (1997). The role of helix VIII in the lactose permease of *Escherichia coli*: II. Site-directed sulfhydryl modification. *Protein Sci.* **6**, 438–443.
33. Sorgen, P. L., Hu, Y., Guan, L., Kaback, H. R. & Girvin, M. E. (2002). An approach to membrane protein structure without crystals. *Proc. Natl Acad. Sci. USA*, **99**, 14037–14040.

34. Vardy, E., Steiner-Mordoch, S. & Schuldiner, S. (2005). Characterization of bacterial drug antiporters homologous to mammalian neurotransmitter transporters. *J. Bacteriol.* **187**, 7518–7525.
35. Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P. & Bowie, J. U. (2004). The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc. Natl Acad. Sci. USA*, **101**, 959–963.
36. Yin, Y., Jensen, M. O., Tajkhorshid, E. & Schulten, K. (2006). Sugar binding and protein conformational changes in lactose permease. *Biophys. J.* **91**, 3972–3985.
37. Kaback, H. R., Sahin-Toth, M. & Weinglass, A. B. (2001). The kamikaze approach to membrane transport. *Nat. Rev. Mol. Cell Biol.* **2**, 610–620.
38. Guan, L. & Kaback, H. R. (2006). Lessons from lactose permease. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 67–91.
39. Guan, L., Mirza, O., Verner, G., Iwata, S. & Kaback, H. R. (2007). Structural determination of wild-type lactose permease. *Proc. Natl Acad. Sci. USA*, **104**, 15294–15298.
40. Holyoake, J. & Sansom, M. S. P. (2007). Conformational change in an MFS protein: MD simulations of LacY. *Structure*, **15**, 873–884.
41. Jensen, M. O., Yin, Y., Tajkhorshid, E. & Schulten, K. (2007). Sugar transport across lactose permease probed by steered molecular dynamics. *Biophys. J.* **93**, 92–102.
42. Klauda, J. B. & Brooks, B. R. (2007). Sugar binding in lactose permease: anomeric state of a disaccharide influences binding structure. *J. Mol. Biol.* **367**, 1523–1534.
43. Franco, P. J. & Brooker, R. J. (1994). Functional roles of Glu-269 and Glu-325 within the lactose permease of *Escherichia coli*. *J. Biol. Chem.* **269**, 7379–7386.
44. Ujwal, M. L., Sahin-Toth, M., Persson, B. & Kaback, H. R. (1994). Role of glutamate-269 in the lactose permease of *Escherichia coli*. *Mol. Membr. Biol.* **11**, 9–16.
45. Abramson, J., Iwata, S. & Kaback, H. R. (2004). Lactose permease as a paradigm for membrane transport proteins. *Mol. Membr. Biol.* **21**, 227–236.
46. Sahin-Tóth, M., Karlin, A. & Kaback, H. R. (2000). Unraveling the mechanism of the lactose permease of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **97**, 10729–10732.
47. Jung, K., Jung, H., Wu, J., Privé, G. G. & Kaback, H. R. (1993). Use of site-directed fluorescence labeling to study proximity relationships in the lactose permease of *Escherichia coli*. *Biochemistry*, **32**, 12273–12278.
48. Frillingos, S., Sun, J., Gonzalez, A. & Kaback, H. R. (1997). Cysteine-scanning mutagenesis of helix II and flanking hydrophilic domains in the lactose permease of *Escherichia coli*. *Biochemistry*, **36**, 269–273.
49. He, M. M. & Kaback, H. R. (1997). Interaction between residues Glu 269 (helix VIII) and His 322 (helix X) of the lactose permease of *Escherichia coli* is essential for substrate binding. *Biochemistry*, **36**, 13688–13692.
50. Schwieters, C. D. & Clore, G. M. (2002). Reweighted atomic densities to represent ensembles of NMR structures. *J. Biomol. NMR*, **23**, 221–225.
51. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73.
52. Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. (2006). Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47–62.
53. Hirai, T. & Subramaniam, S. (2004). Structure and transport mechanism of the bacterial oxalate transporter OxlT. *Biophys. J.* **87**, 3600–3607.
54. Law, C. J., Maloney, P. C. & Wang, D. N. (2008). Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* **62**, 289–305.
55. Forrest, L. R., Tang, C. L. & Honig, B. (2006). On the accuracy of homology modeling and alignment methods applied to membrane proteins. *Biophys. J.* **91**, 508–517.
56. Careaga, C. L. & Falke, J. J. (1992). Thermal motions of surface alpha-helices in the D-galactose chemosensory receptor—detection by disulfide trapping. *J. Mol. Biol.* **226**, 1219–1235.
57. Patzlaff, J. S., Moeller, J. A., Barry, B. A. & Brooker, R. J. (1998). Fourier transform infrared analysis of purified lactose permease: a monodisperse lactose permease preparation is stably folded, alpha-helical, and highly accessible to deuterium exchange. *Biochemistry*, **37**, 15363–15375.
58. le Coutre, J., Kaback, H. R., Patel, C. K. N., Heginbotham, L. & Miller, C. (1998). Fourier transform infrared spectroscopy reveals a rigid alpha-helical assembly for the tetrameric *Streptomyces lividans* K^+ channel. *Proc. Natl Acad. Sci. USA*, **95**, 6114–6117.
59. Sayeed, W. M. H. & Baenziger, J. E. (2009). Structural characterization of the osmosensor ProP. *Biochim. Biophys. Acta*, **1788**, 1108–1115.
60. Yin, Y., He, X., Szewczyk, P., Nguyen, T. & Chang, G. (2006). Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science*, **312**, 741–744.
61. Venkatesan, P. & Kaback, H. R. (1998). The substrate-binding site in the lactose permease of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **95**, 9802–9807.
62. Wolin, C. D. & Kaback, H. R. (2000). Thiol cross-linking of transmembrane domains IV and V in the lactose permease of *Escherichia coli*. *Biochemistry*, **39**, 6130–6135.
63. Zhao, M., Zen, K. C., Hubbell, W. L. & Kaback, H. R. (1999). Proximity between Glu126 and Arg144 in the lactose permease of *Escherichia coli*. *Biochemistry*, **38**, 7407–7412.
64. Guan, L., Hu, Y. L. & Kaback, H. R. (2003). Aromatic stacking in the sugar binding site of the lactose permease. *Biochemistry*, **42**, 1377–1382.
65. Vazquez-Ibar, J. L., Guan, L., Svrakic, M. & Kaback, H. R. (2003). Exploiting luminescence spectroscopy to elucidate the interaction between sugar and a tryptophan residue in the lactose permease of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **100**, 12706–12711.
66. Vazquez-Ibar, J. L., Guan, L., Weinglass, A. B., Verner, G., Gordillo, R. & Kaback, H. R. (2004). Sugar recognition by the lactose permease of *Escherichia coli*. *J. Biol. Chem.* **279**, 49214–49221.
67. Sahin-Tóth, M. & Kaback, H. R. (2001). Arg 302 facilitates deprotonation of Glu 325 in the transport mechanism of the lactose permease from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **98**, 6068–6073.
68. Pupko, T., Bell, R. E., Mayrose, I., Glaser, F. & Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the

identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, **18**, S71–S77.
69. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.
70. Kalman, M. & Ben-Tal, N. (2010). Quality assessment of protein model-structures using evolutionary conservation. *Bioinformatics*, **26**, 1299–1307.
71. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
72. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
73. Khafizov, K., Staritzbichler, R., Stamm, M. & Forrest, L. R. (2010). A study of the evolution of inverted-topology repeats from LeuT-fold transporters using AlignMe. *Biochemistry*, **49**, 10702–10713.
74. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
75. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **285**, 1735–1747.
76. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J. *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
77. Castrignanò, T., De Meo, P. D., Cozzetto, D., Talamo, I. G. & Tramontano, A. (2006). The PMDB Protein Model Database. *Nucleic Acids Res.* **34**, D306–D309.
78. Altschul, S., Madden, T., Schaffer, A., Zhang, J. H., Zhang, Z., Miller, W. & Lipman, D. (1998). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *FASEB J.* **12**, A1326.
79. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410.
80. Li, W. Z. & Godzik, A. (2006). CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
81. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.
82. Beitz, E. (2000). TEXshade: shading and labeling of multiple sequence alignments using LATEX2e. *Bioinformatics*, **16**, 135–139.
83. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. (2006). OPM: Orientations of Proteins in Membranes database. *Bioinformatics*, **22**, 623–625.