Repeat-swap homology modeling of secondary active transporters: updated protocol and prediction of elevator-type mechanisms

Ariela Vergara-Jaque[†‡], Cristina Fenollar-Ferrer[‡], Desirée Kaufmann and Lucy R. Forrest *

Computational Structural Biology Section, Porter Neuroscience Research Center, National Institute of Neurological Disorders and Stroke – National Institutes of Health, Bethesda, MD, USA

Secondary active transporters are critical for neurotransmitter clearance and recycling during synaptic transmission and uptake of nutrients. These proteins mediate the movement of solutes against their concentration gradients, by using the energy released in the movement of ions down pre-existing concentration gradients. To achieve this, transporters conform to the so-called alternating-access hypothesis, whereby the protein adopts at least two conformations in which the substrate binding sites are exposed to one or other side of the membrane, but not both simultaneously. Structures of a bacterial homolog of neuronal glutamate transporters, Glt_{Ph}, in several different conformational states have revealed that the protein structure is asymmetric in the outward- and inward-open states, and that the conformational change connecting them involves a elevator-like movement of a substrate binding domain across the membrane. The structural asymmetry is created by inverted-topology repeats, i.e., structural repeats with similar overall folds whose transmembrane topologies are related to each other by two-fold pseudo-symmetry around an axis parallel to the membrane plane. Inverted repeats have been found in around three-quarters of secondary transporter folds. Moreover, the (a)symmetry of these systems has been successfully used as a bioinformatic tool, called "repeat-swap modeling" to predict structural models of a transporter in one conformation using the known structure of the transporter in the complementary conformation as a template. Here, we describe an updated repeat-swap homology modeling protocol, and calibrate the accuracy of the method using Glt_{Ph}, for which both inward- and outward-facing conformations are known. We then apply this repeat-swap homology modeling procedure to a concentrative nucleoside transporter, VcCNT, which has a three-dimensional arrangement related to that of Glt_{Ph}. The repeat-swapped model of VcCNT predicts that nucleoside transport also occurs via an elevator-like mechanism.

Keywords: neurotransmitter, membrane protein, secondary transport, alternating access, asymmetry exchange, glutamate, concentrative nucleoside transporter

Introduction

Neurotransmitters are required for signal transduction between neurons; however, to prevent continuous signaling and neuronal toxicity, they are rapidly removed from the synaptic cleft by membrane transporters. Transporter proteins are also required for TM movement of essential nutrients such as citrate, into the brain. The involvement of transporters in diseases such as depression and autism, render them important therapeutic targets for neurological and psychiatric disorders. Secondary transporter proteins mediate the movement of neurotransmitters and nutrients against their concentration gradients by using the energy released in the movement of ions down pre-existing concentration gradients. To achieve this, transporters conform to the so-called alternating-access hypothesis, whereby the protein adopts at least two conformations in which the substrate binding sites are exposed to one or other side of the membrane, but not both simultaneously.

A significant bottleneck in the design of therapeutic treatments, however, is the limited amount of structural information available for these transporters. The overexpression and crystallization of membrane proteins is notoriously challenging (Bolla et al., 2012). Transporter proteins in particular, tend to crystallize in only one of the two major states, making it difficult to resolve their structures in the other conformations, and limiting our understanding of transporter mechanism. For example, the structure of a sodium-coupled dicarboxylate transporter from *Vibrio cholerae*, VcINDY, which is related to the neuronal citrate transporter NaCT, has only been determined in one conformation to date (Mancusso et al., 2012). In this regard, Glt_{Ph}, a homolog of the neuronal EAATs of the SLC1 family, from *Pyrococcus horikoshii*, has become an important model system. The relatively large number of X-ray structures of Glt_{Ph} in different states (Yernool et al., 2004; Reyes et al., 2009; Verdon and Boudker, 2012; Verdon et al., 2014) makes this transporter an ideal system for the study of transport mechanism in neuronal transporters, from a structural point of view.

Analysis of the structure of Glt_{Ph} revealed an inverted-repeat architecture (Yernool et al., 2004; Crisman et al., 2009; Boudker and Verdon, 2010), a feature observed in around three-quarters of secondary active transporters of known fold (see Forrest et al., 2011; Forrest, 2015 for reviews). Specifically, large segments of the polypeptide chain are related by structure, but are oriented with the opposite TM topology. This confers the structure with an inherent two-fold pseudo-symmetry, which turns out to be related to the ability of the transporter to adopt two major conformations during the transport cycle, namely outward- and inward-facing states. Specifically, the two repeats adopt a distinctive asymmetry in their conformations, which results in the overall formation of, e.g., an outward-facing state (Fleishman et al., 2006; Chen et al., 2007; Forrest et al., 2008; Crisman et al., 2009; Radestock and Forrest, 2011; Liao et al., 2012). By exchanging their conformations, the two repeats can create a new asymmetric state that is open to the opposite side of the membrane, e.g., inward-facing (Fleishman et al., 2006; Chen et al., 2007; Forrest et al., 2008; Crisman et al., 2009; Radestock and Forrest, 2011;

Liao et al., 2012; Schushan et al., 2012). We refer to this process of alternating access by inverted-topology transporters as "asymmetry exchange."

The asymmetry-exchange concept has been successfully used as a bioinformatics tool to obtain models of a given conformation using the X-ray structure of the transporter in the complementary conformation as a template (Forrest et al., 2008; Crisman et al., 2009; Radestock and Forrest, 2011; Liao et al., 2012; Schushan et al., 2012; Yaffe et al., 2013; Fowler et al., 2015). In this so-called repeat-swap homology modeling technique, the first structural repeat is modeled using the second structural repeat as a template, while the second repeat is concurrently modeled using the first repeat as a template. Given the difficulties of transporter crystallization mentioned above, the ability to predict one state of a transporter once structural data is available for the complementary state provides a very valuable tool.

This repeat-swap technique was first applied to the bacterial amino-acid transporter LeuT from *Aquifex aeolicus*, which is a homolog of neuronal NSSs from the SLC6 family, such as SERT. Comparison of the outward-facing X-ray structure with the inward-facing model of LeuT accurately predicted the cytoplasmic pathway in NSS transporters (Forrest et al., 2008), and led to the proposed "rocking-bundle" mechanism (Forrest and Rudnick, 2009). This mechanism involves the movement of a so-called "transport" domain (which in the case of LeuT, consists of a four-helix bundle, leading to the name "rocking bundle"), relative to a "scaffold" domain, which remains essentially static with respect to the membrane. Although the rocking-bundle concept is somewhat simplistic, the major states of the transport cycle in LeuT-like transporters were subsequently found to be consistent with the overall changes predicted by the repeat-swap mechanism, at least with respective to the relative positions of domains (Weyand et al., 2008; Shimamura et al., 2010; Krishnamurthy and Gouaux, 2012). Similar rigid-body movements, also referred to as a "rocker-switch" mechanism, have been proposed for other transporter families such as the major facilitator superfamily (Abramson et al., 2003; Radestock and Forrest, 2011). In all these cases, structures of the end states confirm that the asymmetry exchange hypothesis predicts the changes in relative positions of domains (see e.g., Dang et al., 2010; Radestock and Forrest, 2011), and predicts the helices that line the access pathways.

One of the most dramatic conformational changes identified in a secondary transporter has been for Glt_{Ph}. The mechanism for Glt_{Ph} is unlike the aforementioned mechanisms, in which the substrate is expected to maintain its position, while domains in the protein called "gates" open and close on either side of the substrate binding site. Comparison of outward- and inward-facing structures of Glt_{Ph} instead reveals a mechanism in which a domain containing the substrate binding site (known as the "transport" domain) moves up and down within the protein, reminiscent of an elevator or a piston, in order to transport the substrate and ion binding sites across the hydrophobic barrier provided by the protein and membrane (Reyes et al., 2009). Specifically, the transport domain moves perpendicular to the membrane plane by ∼16 Å, and rotates by ∼40°. Remarkably, even this extensive conformational change was predicted by a

repeat-swap modeling procedure, in which a cytoplasm-facing state of the protein was modeled using the inverted repeats from a Glt$_{Ph}$ crystal structure in an outward-facing conformation as a template (Crisman et al., 2009; Forrest et al., 2011).

The procedure used by Crisman et al. (2009) was developed *ad hoc* since the goal of that study was to provide proof of principle that repeat-swapping could lead to an alternate-conformation model of a transporter with such a complex fold. In the present work, we describe a refined and standardized repeat-swap homology modeling protocol. We then focus on the application of this protocol to proteins that potentially undergo an elevator-like movement during their transport cycle, although the updated protocol is applicable to any protein with an inverted-repeat topology, independent of the type of conformational mechanism. The availability of X-ray structures of Glt$_{Ph}$ in both outward- and inward-open conformations makes this protein a suitable model system on which to validate, refine, and calibrate the protocol. We therefore use the protocol to predict the outward-facing conformation of Glt$_{Ph}$ based on an inward-facing conformation, i.e., the converse of the model built during the previous work (Crisman et al., 2009). The model constructed using the refined protocol differs by ∼4 Å (averaged over all Cα positions) from the X-ray structure in the same conformation, which is a remarkably small error given that the conformational changes are on the order of 15–20 Å. We demonstrate that a large proportion of that "error" is in fact a difference in the extent of the excursion, which depends on the initial state that is used as a template.

We then apply the standardized protocol to VcCNT, a CNT from *V. cholerae*, whose structure is available only in an inward-facing conformation (Johnson et al., 2012). VcCNT has also been proposed to transport its substrate and ions via an elevator-like movement. Although not related by sequence, or even TM topology (Johnson et al., 2012), several features of VcCNT are similar to those of Glt$_{Ph}$, namely that it appears to comprise scaffold/oligomerization and transport domains, as well as helical hairpin elements that dip into, but do not cross, the membrane (**Figure 1**). In addition, comparison of the model obtained for VcCNT in an outward-open state to the inward-open X-ray structure, suggests that a similar elevator-like mechanism also occurs in the CNT transporter family. Our results therefore show that the new protocol may be useful for predicting conformational mechanisms in other neuronal transporters whose structures that are undoubtedly to follow. Such information will be important in the design of effective therapies targeted to neuronal transporters.

Materials and Methods

The eight steps (a–h) of the standardized repeat-swap homology modeling protocol, illustrated in **Figure 2**, are as follows:

(a) *Identification of structural repeat units.* The starting point of this methodology is the identification of the residues that constitute each of the inverted repeats in the known structure (**Figure 3A**). The underlying assumption is that these repeat

FIGURE 1 | Structure of the Glt$_{Ph}$ and VcCNT transporters. (A) Cartoon representation of the Glt$_{Ph}$ (top) and VcCNT (bottom) protomer structures shown from the extracellular side of the membrane (left). HP, helical hairpin; IH, interfacial helix; EH, extracellular helix. In both proteins, the transport domain is colored orange. The scaffold domain is colored cyan and blue, with the blue regions indicating helices involved in oligomerization. The helices forming each domain are detailed in the topology diagrams to the right. **(B)** Trimeric structures of Glt$_{Ph}$ (left) and VcCNT (right) shown from the extracellular side of the membrane.

units adopt different internal and relative conformations in order to create the asymmetry of an outward- or inward-open state; exchanging between these two conformations is therefore what enables the major conformational change required for TM transport. To date, no reproducible strategy has been developed to identify structural repeats reliably, mainly because the asymmetry and low sequence identity between the domains can stymie even the most sensitive structure alignment programs. Thus, the repeats may be identified in a number of, essentially manual, ways. Two tools are available to aid this process: SymD (Tai et al., 2014) and CE-symm (Myers-Turnbull et al., 2014), which take as input a single protein structure file in Protein Data Bank (PDB) format, and then report internal pseudo-symmetry within the structure (**Figure 3B**). Progressively dividing the structure into smaller domains can be a useful strategy for identifying smaller repeated elements such as those in Glt$_{Ph}$ (**Figure 3A**).

FIGURE 2 | Repeat-swap homology modeling protocol. To build a molecular model of membrane transporters with internal structural pseudo-symmetry, the crystal structure of a protomer of the protein (e.g., Glt_{Ph}) is required as an input. In step (a) the protomer must be carefully analyzed in order to identify the internal repeats. This information is often provided by the authors that report the structure, however, additional insights may be obtained using SymD and CE-symm. In preparation for the modeling, (b,c) an initial sequence alignment is built from a structural alignment of the repeat units (e.g., RU1 and RU2) using a structure alignment program such as TM-Align. In addition, (d) a template structure is built by changing the order of the repeat units in the PDB file and including the regions of the protein that do not belong to the repeat units. At this point, (e) an initial swapped model of the protein can be built, using Modeler. (f) The input alignment is then carefully refined to optimize the agreement with the secondary structure definition and the evolutionary conservation pattern, obtained with DSSP and ConSurf, respectively. Once a model is built using the optimal alignment, the scaffold and transport domains are identified, where possible. (g) $C\alpha$-$C\alpha$ distance restraints can then be defined for each of these domains according to the crystal structure values, and the model is rebuilt using those restraints. Finally, (h) substrates are added to the binding site, if known, and the final repeat-swapped model can be compared with the crystal structure.

(b) *Structural and sequence alignment of the repeat units.* After structural superimposition of the repeat fragments using a structure alignment algorithm (e.g., TM-Align Zhang and Skolnick, 2005), a sequence alignment between the repeats can be extracted, based on the pairs of residues that are close in space in the structure alignment. Such a pairwise sequence alignment is typically reported by the structure alignment program.

(c) *Preparing the initial full-length sequence alignment.* The initial sequence alignment of the full-length protein to the template is constructed by duplicating the sequence alignment obtained in step b. Specifically, the first repeat of the protein is modeled based on – and is therefore aligned to – the second repeat, while the second repeat is modeled based on the first repeat (see **Figure 3C**).

(d) *Preparing the template coordinate file.* The "swapped" template used in this procedure is obtained by simply changing the order of the repeats in the PDB-formatted coordinate file: in this way, the first part of the coordinate file contains the residues from the second repeat, and is followed by those that constitute the first repeat. To model fragments of the protein that are not part of the structural repeats and therefore have no counterpart (e.g., helices 4a and 4b in Glt_{Ph}, **Figures 3A,B**, or other peripheral TM helices), the template can be the same fragment in the X-ray structure (**Figure 3C**). Note that a characteristic of the repeat-swap modeling process is that the model obtained is oriented "upside-down" with respect to the template (for example, in Glt_{Ph}, the N-terminal end of TM1 is modeled to have the same coordinate space as the N-terminal end of TM4c in the template); as a consequence an additional step is required to reorient the additional or peripheral template fragments to the position in space that they need to be in the model. In practice, structurally aligning the X-ray
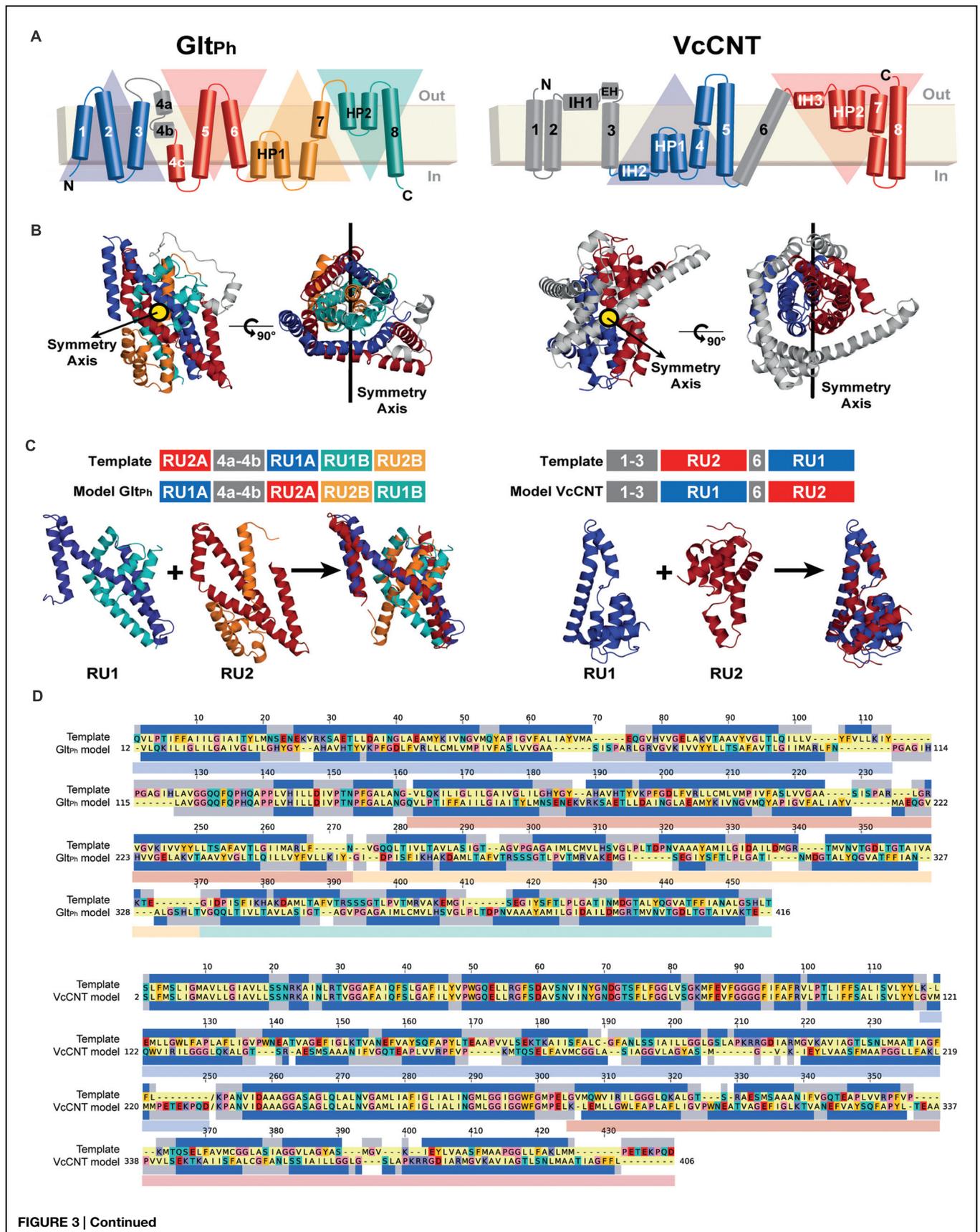
FIGURE 3 | Continued

Repeat-swap modeling of the Glt$_{Ph}$ and VcCNT transporters in alternate conformations. **(A)** A schematic representation of the topology of each protein is shown colored according to the structural repeats. In Glt$_{Ph}$, the helices on the blue and cyan triangular backgrounds comprise repeat unit 1 (RU1), while repeat unit 2 (RU2) is composed of the helices on the red and orange triangular backgrounds. In VcCNT, the blue helices form RU1 and the red helices comprise RU2. **(B)** In both cases, RU1 is related to RU2 by two-fold pseudo-symmetry, with the symmetry axis (black line) parallel to the membrane. **(C)** A structural alignment, built with TM-Align, of the repeats of each protein is shown in cartoon representation with the helices colored according to the topology. The initial sequence alignment used to build a swapped-repeat model was generated based on this structural alignment. A schematic of the sequence alignment is shown above, with peripheral helices in the model and template shown in gray. **(D)** Refined sequence alignment between model and template sequences. The percentage of identical sequences is 21.2 and 46.4% for Glt$_{Ph}$ and VcCNT, respectively, for the full length alignment, including peripheral elements; within the repeats, 14.0 and 7.3% of the residues are identical. Note that the template sequence is the input X-ray protein structure with the order of the repeats reversed. The sequence range of each repeat is displayed with transparent rectangles under the target sequences colored according to the triangular backgrounds in **(A)**. The secondary structure (helix) assignment obtained with DSSP for the X-ray structure is indicated by dark blue rectangles.

structure onto an initial swapped model using the same repeats (e.g., repeat 1 in the X-ray onto repeat 1 in the model) is a straightforward way to reorient the peripheral segments; these must then be added to the template coordinate file in the new orientation, for inclusion in a subsequent round of modeling. The approximate orientation of these peripheral segments will be corrected by adding restraints in a later step of the protocol (step g). Once all the pieces are assembled in order, the residues in the new "swapped" template coordinate file need to be renumbered.

(e) *Constructing an initial structural model*. The initial full-length alignment, with the sequence of the additional or peripheral fragments correctly included, is used as a guide to build a preliminary homology model based on the "swapped" template.

(f) *Refinement of the sequence alignment*. The initial sequence alignment typically requires refinement, due to the sequence and structural divergence of the repeats, which can make them difficult to align. A common step is to remove gaps within secondary structural elements, where the secondary structure is assigned according to DSSP (Kabsch and Sander, 1983) for the known structure (**Figure 3D**). In addition, the conservation patterns obtained from the ConSurf web server (Ashkenazy et al., 2010) can be mapped onto the initial model (**Figures 4B,D**), and used to position the most conserved residues so that they are packed inside the protein, while sequence-variant positions face the exterior, mimicking the arrangement in the input X-ray structure. The model built with the refined alignment is useful for identifying the scaffold and transport domains in the case of a two-domain elevator-like mechanism; superimposing the preliminary model onto the known structure using only helices that form the oligomerization interface (**Figure 1**) can provide a first indication of the mobile elements of the protein.

(g) *Refinement of the model by introducing distance restraints*. Additional distance restraints between Cα atoms taken from the known structure can be used in the repeat-swap modeling process. These restraints are intended to conserve the native helical arrangement and intramolecular packing within domains that move relative to one another, reducing the deviations that arise due to the sequence divergence of the repeats. They also help to fix the orientation of any peripheral segments (see blue helices in **Figure 4C**). However, identification of suitable restraints

requires definition of the scaffold and transport domains, as described above, and should be applied cautiously if these boundaries are not clear.

(h) *Introduction of the ions and/or substrate*. In the final step, any known ions and/or substrates are introduced into the binding site (**Figure 4**). First, the template substrate position is estimated by superposing the binding site of the known structure on that of the initial model. Then, the model building is repeated using this template substrate, while applying additional distance restraints between substrate atoms and binding site atoms.

The details of the software and parameter choices used are as follows:

Model building – Models were built using Modeller v9.13 (Šali and Blundell, 1993). After each of the refinements of the alignment in step f of the protocol, 200 iterations of model building were performed. For steps g and h, 2000 iterations of model building were performed, in order to increase the chances that the sampling creates a model that satisfies all of the input restraints.

Model assessment – Each set of models was evaluated using MolPDF and ProQM scores (Ray et al., 2010), as well as Procheck analysis (Laskowski et al., 1993). The MolPDF score describes how well the model satisfies the input restraints, including those created from the template, the alignment, and any additional applied restraints, and is therefore an arbitrary value dependent on those features; the model with the smallest MolPDF score best satisfies all the restraints. The ProQM score measures the degree to which a set of coordinates is consistent with a number of required features, including TM segments predicted using TOPCONS (Bernsel et al., 2009), the distance to the membrane center for residues in α-helical segments predicted with ZPRED (Granseth et al., 2006), secondary structure elements predicted by PSIPRED (Jones, 1999), burial of conserved residues, and exposure of variant residues. The ProQM score is assigned as a 21-residue window-average to each residue, and is then summed to give a total score per model, with values ranging from 0 to 1, where values of 0.7 are typical of membrane protein structures solved by X-ray crystallography. Procheck assesses the degree to which the features of the model are consistent with those of known protein structures in terms of bond distances, angles, dihedrals and overlapping atoms. In this work, Procheck is used

FIGURE 4 | Refinement of repeat-swapped outward-facing models of Glt$_{Ph}$ and VcCNT compared with input inward-facing structures. Models of Glt$_{Ph}$ **(A,B)** and VcCNT **(C,D)** built using initial or refined alignments, both before and after applying restraints within the scaffold and transport domains. Models are shown as cartoon helices, colored with separate colors for each helix (see **Figure 5**) in **(A,C)**, and using ConSurf coloring in **(B,D)**. Substrates are shown as spheres, with their pathways illustrated using arrows. ConSurf conservation patterns are mapped to the helices using colors from cyan (most variant residues) to dark pink (most conserved residues); yellow indicates positions with insufficient sequence information to allow a conservation assignment. As seen for the X-ray structures (*right*), conserved residues cluster on the inside of the protein, and at oligomerization interfaces, while variant residues are found primarily on the surface. This observation, and the patterns seen in the template structure were used to guide the refinement of the alignments. The position of the proteins in the membrane was determined with the OPM server.

to identify the fraction of backbone groups that lie outside the favored regions of the Ramachandran plot.

The final model was selected as that in step h with the lowest MolPDF score, and/or the highest Procheck and global ProQM scores.

Domain definitions – The repeats were defined as comprising residues 12-108, 335-416 for repeat unit 1 (RU1) and 150-334 for repeat unit 2 (RU2) in Glt$_{Ph}$, and 119-229 for RU1 and 294-416 for RU2 in VcCNT. The scaffold/oligomerization and transport domains were defined as comprising residues 12-77 plus 129-218 and 78-128 plus 219-416, respectively, for Glt$_{Ph}$, and residues 2-137 plus 241-311 and 138-229 plus 312-416, respectively, in VcCNT.

Distance restraints – All distance restraints applied in steps g and h were represented as a Gaussian with a standard deviation of 0.1 Å. Distance restraints between Cα atoms were applied to

FIGURE 5 | Crystal structures in an inward-facing conformation compared with the final repeat-swapped models of an outward-facing conformation. (A,B) The X-ray crystal structures of a protomer of **(A)** Glt$_{Ph}$ and **(B)** VcCNT in an inward-facing conformation (*left*) are compared with the repeat-swapped models of an extracellular-facing conformation (*right*). Structures are viewed along the plane of the membrane, with the extracellular side at the top. The Bendix plugin to VMD was used to represent the helices (Dahl et al., 2012). To quantify the conformational change of the two transporters, the rotation axis, rotation angle, and the displacement of the helices were evaluated. The black line shows the rotation axis, which lies approximately perpendicular to the symmetry axis in each transporter (see **Figure 3B**). The displacement values suggest an elevator-like movement of the transport domain containing the substrates (*spheres*), and this displacement is associated with a rotation of 43.5 and 11.8°, for Glt$_{Ph}$ and VcCNT, respectively. For reference, the conformational change in the transport domain between two X-ray structures of Glt$_{Ph}$ (PDB IDs 3KBC and 1XFH) involves a rotation of 39.6° and a displacement by x = 3.9, y = 2.8, z = 16.4 Å. **(C,D)** Displacement per residue determined as the distance between Cα atoms in the crystal structure and the model after fitting using the scaffold domains (*black*), for Glt$_{Ph}$ **(C)** and VcCNT **(D)**. In the case of Glt$_{Ph}$, the conformational change between known inward-facing and outward-facing substrate-bound structures is shown (*red*).

preserve the internal structure of the scaffold/oligomerization and transport domains, independently, and were assigned according to the input crystal structure for all pairs of Cα atoms < 60 Å apart within each domain.

The distance restraints applied in step h between substrates and binding site atoms were defined as follows:

For Glt$_{Ph}$, restraints were applied between the two Na$^+$ ions or any non-hydrogen ("heavy") atom of the aspartate substrate, and any heavy atom of the protein within 5 Å in the known/template structure, PDB (Berman et al., 2000) identifier, ID: 3KBC (Reyes et al., 2009).

For VcCNT, restraints were applied between the Na$^+$ ion and the O atoms in the backbone of N149, V152 and I184, the hydroxyl group of S183, and a crystallographic water molecule. Restraints were also applied between any heavy atoms of the uridine and any heavy-atom of the protein within 3.5 Å in the template structure, PDB ID: 3TIJ (Johnson et al., 2012).

Other programs – Structural alignments were obtained using TM-align (Zhang and Skolnick, 2005) and Jalview was used for editing and visualization of alignments (Waterhouse et al., 2009). Secondary structural elements were assigned with DSSP (Edgar, 2010) and conservation patterns were obtained using the ConSurf server with default parameters (Ashkenazy et al., 2010). Molecular figures were generated using PyMol v1.6 (Schrödinger, Inc.). The position of the proteins in the membrane

was determined for the X-ray structures with the OPM server (Lomize et al., 2006), and the orientation of the models was defined after superposition of the scaffold domain onto that of the corresponding X-ray structure.

Results and Discussion

Crisman et al. (2009) previously applied an *ad hoc* repeat-swap protocol to construct an inward-facing substrate-bound occluded model of Glt$_{Ph}$ (Protein Model Database (Castrignanò et al., 2006), PMDB identifier: PM0075966) based on an available structure of Glt$_{Ph}$ in an outward-facing substrate-bound occluded conformation (Boudker et al., 2007, PDB ID: 2NWL). The predicted inward-facing state captured the major conformational change revealed in a structure published around the same time (Reyes et al., 2009), thereby demonstrating that the conformational change is a result of asymmetry exchange. Here, we propose a more standardized and generalized protocol for repeat-swap homology modeling (see Materials and Methods), and calibrate the method by building an outward-facing occluded model of Glt$_{Ph}$ based on the available inward-facing structure (PDB ID: 3KBC), thereby providing a measure of the expected accuracy of the method for that case by comparison with a known structure in the same state (PDB ID: 1XFH). We then apply the

FIGURE 6 | The repeat-swapped model of VcCNT contains an extracellular substrate pathway that is not present in the template structure. The trimeric VcCNT X-ray structure **(A,C)** and model **(B,D)** are viewed from the extracellular side, colored according to **Figure 1** and rendered either as cartoon helices **(A,B)** or using the van der Waals surface **(C,D)**. The substrate (*spheres, arrows*) is visible from the extracellular side in the model **(D)**, but not in the inward-facing crystal structure **(C)**.

updated protocol to VcCNT, revealing that the mechanism of this secondary transporter is also likely to involve an elevator-like conformational change.

An Outward-Facing Model of Glt$_{Ph}$ Built Using the Inward-Facing Structure as a Template

As mentioned above, structures of Glt$_{Ph}$ are available in several different conformations, including outward- and inward-open, outward-occluded, intermediate, and apo forms (Yernool et al., 2004; Boudker et al., 2007; Reyes et al., 2009; Boudker and Verdon, 2010). Here, we wished to compare predictions of the major conformational change between outward-occluded and inward-occluded states, with structures of the end points at a similar resolution. Thus, we elected to use the inward-occluded substrate-bound structure of Glt$_{Ph}$ at 3.51 Å resolution (PDB ID: 3KBC, Reyes et al., 2009) as a template to model the outward-occluded substrate-bound state, which could then be compared with a known structure also at 3.50 Å resolution (PDB ID: 1XFH, Yernool et al., 2004). This we did by following the protocol described in the Section "Materials and Methods."

As mentioned above, a few steps in the repeat-swap homology modeling protocol do not yet have reliable automated procedures. Specifically, we defined the inverted-topology repeats

in Glt$_{Ph}$ according to Crisman et al. (2009; **Figure 3A**). We note that this differs slightly from the definition of Reyes et al. (2009), who defined TM3 and TM6 as part of RU1B and RU2B, respectively. We elected to define them this way because the structure superpositions led to more reasonable initial sequence alignments, requiring fewer cycles of refinement. The final refined sequence alignment is shown in **Figure 3D**.

For assigning the domains that move relative to one another (transport and scaffold domains) we used the definition of Reyes et al. (2009; **Figure 1**, see Materials and Methods). In an important update to the protocol, restraints were imposed between all pairs of Cα atoms within each of these domains separately during model building. The final model contains aspartate and two sodium ions (**Figure 4A**).

Accuracy of the Repeat-Swapped Model of Glt$_{Ph}$

The model of the outward-occluded state of Glt$_{Ph}$ (**Figure 5A**) was compared with the inward-occluded (template) X-ray crystal structure by structurally aligning the scaffold/oligomerization domains. This comparison revealed a conformational change consisting of a ∼44° rotation of the transport domain about an axis running approximately perpendicular to the symmetry axis, and a ∼10 Å translocation (normal to the membrane), relative to the scaffold/oligomerization domain. As a result, the substrate binding site is exposed to the opposite side of the membrane. This domain movement translates to positional changes of between 5 and 23 Å for the individual Cα atoms in the transport domain (**Figure 5C**, black line). The helical hairpins HP1 and HP2, which are at the interface with the scaffold domain, undergo the largest movements, whereas the helices facing the lipids (TM3 and TM6) undergo the smallest changes.

These predicted changes recapitulate those observed in a comparison of the template inward-occluded structure (PDB ID: 3KBC), with an outward-occluded substrate-bound structure of similar resolution (PDB ID: 1XFH), which reveals a rotation of ∼40° and a translation perpendicular to the membrane of ∼16 Å, corresponding to positional changes of between 10 and 25 Å per Cα atom (**Figure 5C**, red line). Thus, the degree of rotation and direction of translocation are very similar. Moreover, this comparison allows a quantitative analysis of the accuracy of the repeat-swap method. Two factors determine the accuracy of the model: first, the degree to which the two repeats are structurally similar, which in turn is related to their sequence similarity; and second, the degree of symmetry that relates the two states.

With respect to the first factor, the similarity between the structures of the repeats is related to the expected accuracy of a typical homology model, which decreases as a function of decreasing sequence identity (Forrest et al., 2006). In Glt$_{Ph}$, the sequence identities between repeats are only ∼9–14% (**Figure 3**), leading to an expected accuracy of ∼1.5–4 Å in the Cα positions in the TM segments (Forrest et al., 2006). However, with the new protocol we have also captured information about the known structural features of each domain, by incorporating restraints within the oligomerization/scaffold and transport domains

separately, based on intra-domain distances in the known structure (see Materials and Methods). During the modeling, these intra-domain distance restraints will be satisfied as long as they do not lead to significant violations of the other restraints obtained from the use of the other repeat as a template (for which there are a larger number of distance restraints). Applying these new intra-domain restraints in the case of Glt_{Ph} leads to a significant improvement in the accuracy of the individual domains. Specifically, after applying these restraints the root mean squared deviation (RMSD) between the model and the crystal structure of the outward-facing conformation decreases from 4.8 Å and 6.7 Å, to 2.5 Å and 2.9 Å over all Cα atoms in the scaffold/oligomerization and transport domains, respectively. This improvement occurs without altering the overall elevator-like movement; before incorporating the intradomain restraints, the rotation and translocation were ~45° and ~11 Å, respectively (cf. ~44° and ~10 Å for the final model described above).

The above results show that the error in the new model is only slightly higher than the difference between the two known structures in different conformations (PDB IDs: 1XFH and 3KBC), which is 1.2 and 3.4 Å for the scaffold/oligomerization and transport domains, respectively. Using a length-independent measure for the same comparison, the template modeling score (TM-score) increases from 0.62 and 0.64 to 0.83 and 0.86, for the models built without and with intra-domain restraints, respectively. (A TM-score of 1.0 indicates that the Cα atoms have identical positions.) For comparison, the TM-scores when comparing the inward- and outward-facing crystal structures domains are 0.91 and 0.73, for the scaffold/oligomerization and transport domains, respectively. The remarkable accuracy of this modeling procedure is achieved in spite of the extremely low (<15%) sequence identity, and even while the overall conformational change is captured (**Figure 5**).

As mentioned, the second factor determining repeat-swap model accuracy is the particular conformational state of the template. Thus, if the input structure is semi-occluded, then the output model will also be semi-occluded. In the case of the swapped model of Glt_{Ph}, the rotation of the transport domain is very similar to that observed in comparing the two structures, but the translocation of that domain implied by the model is slightly (~7 Å) smaller (**Figure 5C**, compare red and black lines) demonstrating that the modeled conformational change is less extreme than expected based on the crystal structures. In other words, the (template) inward-facing structure is less "inward" than the known outward-facing structure is "outward," resulting in a model that is also less "outward." Thus, the major value of the repeat-swap modeling is to identify the elements involved in the conformational change, and the directionality of their movement, which are clearly both well captured by this approach.

A Model of the Outward-Facing Conformation of VcCNT

In Johnson et al. (2012), a structure of a CNT from *V. cholerae* (VcCNT) was reported at 2.4 Å resolution, with a substrate binding site that is exposed toward the presumed cytoplasm.

VcCNT is related to transporters of the human solute carrier SLC28 family. To date, all other structures reported for this transporter fold reflect the same state of the transporter (Johnson et al., 2014) and therefore the mechanism of conformational change is unknown. It has been proposed, based on visual analysis of the structure and binding site accessibility data, that CNTs undergo a similar elevator-like movement (Johnson et al., 2012), with two helical hairpins moving above and below TM helix 6, which serves as a hydrophobic barrier. In particular, the overall "design principle" of the CNTs appears to be similar to that of Glt_{Ph}, with a trimeric quaternary structure, helical hairpins, and clear transport- and scaffold-like boundaries (Johnson et al., 2012, **Figure 1**). However, at the level of individual helices, a structural comparison revealed few commonalities (Johnson et al., 2012). Nevertheless, the VcCNT fold clearly contains inverted-topology repeats, which together assemble into the transport domain (**Figures 3A,B**). Internal symmetry within the scaffold/oligomerization domain of VcCNT is less clear, except for the interfacial helices IH2 and IH3, and therefore, if additional internal symmetry ever existed, it appears to have been lost during evolution, perhaps because trimerization provides sufficient scaffolding without compromising the transport mechanism. As in Glt_{Ph}, the scaffold domain of VcCNT also contains within it the helices involved in forming the physiologically relevant trimer interface (**Figure 1B**).

To pursue the hypothesis that VcCNT undergoes an elevator movement, we built a repeat-swapped structural model using the inward-facing structure as a template, and using the definition of repeats shown in **Figures 3A,C**, by applying our updated protocol (**Figure 2**, see Materials and Methods). The final alignment after initial model-building and refinement (**Figures 4C,D**) is shown in **Figure 3D**.

We defined repeat units 1 and 2 as comprising the transport domain, with the exception of interfacial helices IH2 and IH3, which were assigned to the scaffold domain, together with the peripheral trimerization helices (**Figures 1A and 3A–C**). Along with restraints to position a uridine substrate and a sodium ion in binding sites formed by the same residues as in the template, we also imposed restraints within each domain to build a refined model that matches as many aspects of the known structure as possible, while capturing the same overall conformational change. The final model of VcCNT (**Figures 5B,D**) is freely available from the PMDB (https://bioinformatics.cineca.it/PMDB/) with the identifier PM0080188, and the model generated without restraints is available upon request.

The repeat-swapped model of VcCNT adopts a conformation in which the substrates are visible via an aqueous pathway connected to the extracellular solution (**Figure 6**). The conformational change predicted by the VcCNT model also involves an elevator-like movement of the hairpin domains, similar to that seen for Glt_{Ph} as suggested by Johnson et al. (2012). These changes arise due to a ~12° rotation of the transport domain around a similar axis to that found in Glt_{Ph}, as well as lateral and displacements along the membrane normal of ~3–4 Å (**Figure 5B**). Note that very similar changes are found

when the intradomain restraints were not included; specifically, the transport domain rotates by ~15°, and is also translocated by 3–4 Å. Interestingly, the extent of this conformational change is more modest than that of Glt_{ph}, with Cα displacements of between 4 and 10 Å, after superposition on the scaffold helices (**Figure 5D**).

It is entirely possible that VcCNT undergoes more extreme movements than can be predicted based on the available structural data. Nevertheless, as noted above, the directionality and composition of the moving domains predicted by the repeat-swap model are likely to be correct, based on our observations using Glt_{ph}. Therefore, this model of the outward-facing state of VcCNT should provide a useful starting point for experimental studies of the conformational change required for transport in CNTs, by illustrating which elements become buried or exposed, and which elements move relative to one another.

Conclusion

Repeat-swap modeling has proven to be helpful for analyzing the conformational changes in various homologs of neuronal transporters, including Na^+/Cl^--dependent transporters of the NSS family (Forrest et al., 2008), EAATs (Crisman et al., 2009), and vesicular monoamine transporters (Yaffe et al., 2013). Although the models generated may not be high-resolution enough for drug design, the identification of alternate conformations and conformational mechanisms is essential for the development of treatments against such transporters, as it provides insight into changes in the transporter structure that may affect inhibitor binding. For example, inhibitors of serotonin uptake such as ibogaine and cocaine bind to different conformations of SERT (Tavoulari et al., 2009), and similar conclusions have been evoked for reserpine and tetrabenazine binding to the vesicular monoamine transporter VMAT2 (Schuldiner et al., 1995). Therefore, molecular models based on repeat-swap modeling can be extremely useful to help design further experiments such as cross-linked constructs for crystallography, whose structures can then be used for drug design.

The current study provides an updated and more systematic protocol for repeat-swap modeling of secondary transporters with inverted-repeats. Our analysis of the Glt_{ph} model reveals that the model accuracy can be dramatically improved by inclusion of restraints extracted from the input structure. Those restraints can be readily defined if there is a distinct boundary between the mobile and static segments of the transporter. In the cases of Glt_{ph} and VcCNT this boundary became clear after producing an initial repeat-swapped model of each protein, and comparing that model with the input crystal structure by superposition of the helices forming the oligomerization interface. With the benefit of this insight, this boundary becomes relatively clear merely from analysis of the architecture viewed from above the membrane (**Figure 1B**).

The inward-facing crystal structure of VcCNT and the swapped model in an outward-facing conformation that we present here provide a molecular-level description of the conformational changes in CNTs. We expect that this insight will prove to be important for understanding the mechanism of TM transport of nucleosides and of nucleoside-derived drugs, such as anticancer drugs like gemcitabine. More broadly, our model strongly suggests that the remarkable elevator-like movements identified in the glutamate transporter family are not unique, and that transporters in the SLC28 CNT family also undergo elevator-like movements during their alternating access mechanism. This raises the possibility that other transporter families may use a similar elevator-like mechanism. For example, the sodium-coupled dicarboxylate transporter VcINDY (Mancusso et al., 2012), which is a homolog of the neuronal sodium-coupled citrate transporter NaCT, contains inverted-topology repeats with reentrant helical hairpins similar to those observed in Glt_{ph} and VcCNT. It will be of interest to examine which other neuronal transporters use elevator mechanisms for TM transport of their substrates.

Acknowledgments

References

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R., and Iwata, S. (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* 301, 610–615. doi: 10.1126/science.1088196

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 38, W529–W533. doi: 10.1093/nar/gkq399

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235

Bernsel, A., Viklund, H., Hennerdal, A., and Elofsson, A. (2009). TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res.* 37, W465–W468. doi: 10.1093/nar/gkp363

Bolla, J. R., Su, C. C., and Yu, E. W. (2012). Biomolecular membrane protein crystallization. *Philos. Mag. (Abingdon)* 92, 2648–2661. doi: 10.1080/14786435.2012.670734

Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K., and Gouaux, E. (2007). Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature* 445, 387–393. doi: 10.1038/nature05455

Boudker, O., and Verdon, G. (2010). Structural perspectives on secondary active transporters. *Trends Pharmacol. Sci.* 31, 418–426. doi: 10.1016/j.tips.2010.06.004

Castrignanò, T., De Meo, P. D., Cozzetto, D., Talamo, I. G., and Tramontano, A. (2006). The PMDB protein model database. *Nucleic Acids Res.* 34, D306–D309. doi: 10.1093/nar/gkj105

Chen, Y.-J., Pornillos, O., Lieu, S., Ma, C., Chen, A. P., and Chang, G. (2007). X-ray structure of EmrE supports dual topology model. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18999–19004. doi: 10.1073/pnas.0709387104

Crisman, T., Qu, S., Kanner, B. I., and Forrest, L. R. (2009). Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20752–20757. doi: 10.1073/pnas.0908570106

Dahl, A. C. E., Chavent, M., and Sansom, M. S. P. (2012). Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics* 28, 2193–2194. doi: 10.1093/bioinformatics/bts357

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. doi: 10.1038/nature09406

Edgar, R. C. (2010). Quality measures for protein alignment benchmarks. *Nucleic Acids Res.* 38, 2145–2153. doi: 10.1093/nar/gkp1196

Fleishman, S. J., Harrington, S. E., Enosh, A., Halperin, D., Tate, C. G., and Ben-Tal, N. (2006). Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *J. Mol. Biol.* 364, 54–67. doi: 10.1016/j.jmb.2006.08.072

Forrest, L. R. (2015). Structural symmetry in membrane proteins. *Annu. Rev. Biophys.* 44, 311–337. doi: 10.1146/annurev-biophys-051013-023008

Forrest, L. R., Kraemer, R., and Ziegler, C. (2011). The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta* 1807, 167–188. doi: 10.1016/j.bbabio.2010.10.014

Forrest, L. R., and Rudnick, G. (2009). The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology* 24, 377–386. doi: 10.1152/physiol.00030.2009

Forrest, L. R., Tang, C. L., and Honig, B. (2006). On the accuracy of homology modeling and alignment methods applied to membrane proteins. *Biophys. J.* 91, 508–517. doi: 10.1529/biophysj.106.082313

Forrest, L. R., Zhang, Y.-W., Jacobs, M. T., Gesmonde, J., Xie, L., Honig, B., et al. (2008). A mechanism for alternating access in neurotransmitter transporters. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10338–10343. doi: 10.1073/pnas.0804659105

Fowler, P. W., Orwick-Rydmark, M., Radestock, S., Solcan, N., Dijkman, P. M., Lyons, J. A., et al. (2015). Gating topology of the proton-coupled oligopeptide symporters. *Structure* 23, 290–301. doi: 10.1016/j.str.2014.12.012

Granseth, E., Viklund, H., and Elofsson, A. (2006). ZPRED: predicting the distance to the membrane center for residues in alpha-helical membrane proteins. *Bioinformatics* 22, e191–e196. doi: 10.1093/bioinformatics/btl206

Johnson, Z. L., Cheong, C. G., and Lee, S. Y. (2012). Crystal structure of a concentrative nucleoside transporter from *Vibrio cholerae* at 2.4 Å. *Nature* 483, 489–493. doi: 10.1038/nature10882

Johnson, Z. L., Lee, J.-H., Lee, K., Lee, M., Kwon, D.-Y., Hong, J., et al. (2014). Structural basis of nucleoside and nucleoside drug selectivity by concentrative nucleoside transporters. *Elife* 3, e03604. doi: 10.7554/eLife.03604

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202. doi: 10.1006/jmbi.1999.3091

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi: 10.1002/bip.360221211

Krishnamurthy, H., and Gouaux, E. (2012). X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* 459, 347–355. doi: 10.1038/nature10737

Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK – a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291. doi: 10.1107/S0021889892009944

Liao, J., Li, H., Zeng, W., Sauer, D. B., Belmares, R., and Jiang, Y. (2012). Structural insight into the ion-exchange mechanism of the sodium/calcium exchanger. *Science* 335, 686–690. doi: 10.1126/science.1215759

Lomize, M. A., Lomize, A. L., Pogozheva, I. D., and Mosberg, H. I. (2006). OPM: orientations of proteins in membranes database. *Bioinformatics* 22, 623–625. doi: 10.1093/bioinformatics/btk023

Mancusso, R., Gregorio, G. G., Liu, Q., and Wang, D.-N. (2012). Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter. *Nature* 491, 622–626. doi: 10.1038/nature11542

Myers-Turnbull, D., Bliven, S. E., Rose, P. W., Aziz, Z. K., Youkharibache, P., Bourne, P. E., et al. (2014). Systematic detection of internal symmetry in proteins using CE-Symm. *J. Mol. Biol.* 426, 2255–2268. doi: 10.1016/j.jmb.2014.03.010

Radestock, S., and Forrest, L. R. (2011). Outward-facing conformation of MFS transporters revealed by inverted-topology repeats. *J. Mol. Biol.* 407, 698–715. doi: 10.1016/j.jmb.2011.02.008

Ray, A., Lindahl, E., and Wallner, B. (2010). Model quality assessment for membrane proteins. *Bioinformatics* 26, 3067–3074. doi: 10.1093/bioinformatics/btq581

Reyes, N., Ginter, C., and Boudker, O. (2009). Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* 462, 880–885. doi: 10.1038/nature08616

Šali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/jmbi.1993.1626

Schuldiner, S., Shirvan, A., and Linial, M. (1995). Vesicular neurotransmitter transporters: from bacteria to humans. *Physiol. Rev.* 75, 369–392.

Schushan, M., Rimon, A., Haliloglu, T., Forrest, L. R., Padan, E., and Ben-Tal, N. (2012). A model-structure of a periplasm-facing state of the NhaA antiporter suggests the molecular underpinnings of pH-induced conformational changes. *J. Biol. Chem.* 287, 18249–18261. doi: 10.1074/jbc.M111.336446

Shimamura, T., Weyand, S., Beckstein, O., Rutherford, N. G., Hadden, J. M., Sharples, D., et al. (2010). Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* 328, 470–473. doi: 10.1126/science.1186303

Tai, C. H., Paul, R., Dukka, K. C., Shilling, J. D., and Lee, B. (2014). SymD webserver: a platform for detecting internally symmetric protein structures. *Nucleic Acids Res.* 42, W296–W300. doi: 10.1093/nar/gku364

Tavoulari, S., Forrest, L. R., and Rudnick, G. (2009). Fluoxetine (prozac) binding to serotonin transporter is modulated by chloride and conformational changes. *J. Neurosci.* 29, 9635–9643. doi: 10.1523/JNEUROSCI.0440-09.2009

Verdon, G., and Boudker, O. (2012). Crystal structure of an asymmetric trimer of a bacterial glutamate transporter homolog. *Nat. Struct. Mol. Biol.* 19, 355–357. doi: 10.1038/nsmb.2233

Verdon, G., Oh, S., Serio, R. N., and Boudker, O. (2014). Coupled ion binding and structural transitions along the transport cycle of glutamate transporters. *Elife* 3, e02283–e02283. doi: 10.7554/eLife.02283

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Weyand, S., Shimamura, T., Yajima, S., Suzuki, S. I., Mirza, O., Krusong, K., et al. (2008). Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* 322, 709–713. doi: 10.1126/science.1164440

Yaffe, D., Radestock, S., Shuster, Y., Forrest, L. R., and Schuldiner, S. (2013). Identification of molecular hinge points mediating alternating access in the vesicular monoamine transporter VMAT2. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1332–E1341. doi: 10.1073/pnas.1220497110

Yernool, D., Boudker, O., Jin, Y., and Gouaux, E. (2004). Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* 431, 811–818. doi: 10.1038/nature03018

Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33, 2302–2309. doi: 10.1093/nar/gki524