Modeling Alternative Conformational States of Pseudo-Symmetric Solute Carrier Transporters using Methods from Machine Learning

G.V.T. Swapna[1,2,*], Namita Dube[1], Monica J. Roth[2*], and Gaetano T. Montelione[1,*]

[1]Dept. of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Sciences, Rensselaer Polytechnic Institute, Troy, New York, 12180 USA.

[2]Department of Pharmacology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway NJ 08854 USA.

ABSTRACT

The Solute Carrier (SLC) superfamily of integral membrane proteins function to transport a wide array of solutes across the plasma and organelle membranes. SLC proteins also function as important drug transporters and as viral receptors. Despite being classified as a single superfamily, SLC proteins do not share a single common fold classification; however, most belong to multi-pass transmembrane helical protein fold families. SLC proteins populate different conformational states during the solute transport process, including outward open, intermediate (occluded), and inward open conformational states. For some SLC fold families this structural "flipping" corresponds to swapping between conformations of their N-terminal and C-terminal symmetry-related sub-structures. Conventional AlphaFold2 or Evolutionary Scale Modeling methods typically generate models for only one of these multiple conformational states of SLC proteins. Here we describe a fast and simple approach for modeling multiple conformational states of SLC proteins using a combined ESM - AF2 process. The resulting multi-state models are validated by comparison with sequence-based evolutionary co-variance data (ECs) that encode information about contacts present in the various conformational states adopted by the protein. We also explored the impact of mutations on conformational distributions of SLC proteins modeled by AlphaFold2 using both conventional and enhanced sampling methods. This approach for modeling conformational landscapes of pseudo-symmetric SLC proteins is demonstrated for several integral membrane protein transporters, including SLC35F2 the receptor of a feline leukemia virus envelope protein required for viral entry into eukaryotic cells.

*Corresponding authors.

Abbreviations: AF2 – AlphaFold2 Multimer; EC - Evolutionary Covariance; ESM- Evolutionary-Scale Modeling, LDDT – local-distance difference test; MD – molecular dynamics; ML – machine learning; mmCIF - macromolecular crystallographic information file; MSA – multiple sequence alignment; PDB - protein data bank; pLDDT - predicted local model confidence score predicted from ML

INTRODUCTION

The Solute Carrier (SLC) superfamily of integral membrane proteins function to transport a wide array of solutes across the plasma and organelle membranes. The superfamily includes more than 66 SLC protein families (https://www.bioparadigms.org/slc/intro.htm), each including many individual proteins. SLC proteins transport a wide array of molecules, including sugars, amino acids, vitamins, nucleotides, metals, inorganic ions, organic anions, oligopeptides, and drugs (Hediger *et al.*, 2013; Colas *et al.*, 2016; Bai *et al.*, 2017; Pizzagalli *et al.*, 2021). Some are orphan transporters with no known substrate. They constitute a major portion of all human transporter-related proteins and play key roles in human health and disease (Fredriksson *et al.*, 2008; Bai *et al.*, 2017; Pizzagalli *et al.*, 2021).

SLC proteins also have important roles beyond their basic functions as physiological solute transporters. For example, Sarangi *et al* have reported that both feline and human SCL35F2 proteins bind a variant of a feline leukemia virus envelope protein (FeLV Env A5) and function as retroviral receptors (Sarangi *et al.*, 2007). SLC35F2 also transports the anticancer drug YM155 (Winter *et al.*, 2014) and analogs (West *et al.*, 2023) into cancer cells. YM155's mode of action includes induction of DNA damage toxicity (Winter *et al.*, 2014) and downregulation of the apoptosis inhibitor protein survivin through binding to the interleukin enhancer-binding factor 3/NF110 transcription factor (Nakamura *et al.*, 2012). Winter et al have shown an absolute dependency of YM155 transport on expression of SLC35F2 (Winter *et al.*, 2014). Nyquist *et al* have also reported synergy between high-androgen therapy for prostate cancer and YM155 that is mediated by SLC35F2 (Nyquist *et al.*, 2017a; Nyquist *et al.*, 2017b). Additionally SLC35F2 is reported to interact with SYVN1of TRIM 59, promoting p53 degradation (Che *et al.*, 2023). These studies highlight the pleiotropic functions of SLC35F2, illustrating the broad biological significance of some SLC proteins.

Despite being classified as a single superfamily, the various SLC fold families do not share a single common fold classification and are not all phylogenetically related. For example, the two most common SLC fold families, the major facilitator superfamily (MFS) fold, which constitute the largest class of SLC proteins, and the LeuT fold, another important class of SLCs, are topologically and structurally distinct (Bai *et al.*, 2017). However, despite these differences, many SLC transporters have a characteristic structural architecture with pseudo two-fold symmetry, where the two halves of the protein structure are related by a two-fold symmetry axis in the plane of the membrane bilayer (Forrest, 2013; Bai *et al.*, 2017). These halves have a similar folds but non-identical conformations, enabling the protein to adopt multiple conformational states essential for its function. MSF-fold SLC proteins have a "6+6" topology comprised of two "inverted pseudo-repeat" 6-helical bundles with antiparallel orientations related by a pseudosymmetry axis, while the strikingly similar but topologically distinct LeuT-fold membrane proteins feature two 5-helical bundles with "inverted pseudo-repeat" sequences that form structures related to one another by pseudosymmetry axis (Bai *et al.*, 2017). Some (but not all) other SLC proteins also have folds with internal structural pseudosymmetry (Bai *et al.*, 2017).

SLC proteins populate different conformational states during the transport process, including "outward open", with a surface cavity directed one way, intermediate (i.e., occluded, with no surface cavity), and "inward open" with a surface cavity directed to the opposite side of the membrane (Colas *et al.*, 2016; Bai *et al.*, 2017). Crystal structures have been solved for inward-open, occluded, and outward-open states of MFS and LeuT SLC proteins; for a few SLC proteins both inward and outward open states have been determined by X-ray crystallography or cryoEM (Leano *et al.*, 2019; Xue *et al.*, 2020; Killer *et al.*, 2021; Wang *et al.*, 2021; Lu *et al.*, 2023). This conformational "flipping" confers an "airlock" or "revolving door" function, which underlies their mechanisms of symporter or antiporter solute transport (Forrest, 2013; Colas *et al.*, 2016; Bai *et al.*, 2017). The switch between outward and inward open states results from swapping of the conformations of the N-terminal and C-terminal symmetry-related sub-structures, in which the N-terminal helical bundle switches to adopt the conformation of the C-terminal helical bundle, while simultaneously the C-terminal helical bundle switches into the original conformation of the N-terminal helical bundle. This pseudo-symmetrical transport mechanism provides the basis for modeling the inward open (or outward open) conformations of some SLC proteins from knowledge of their outward (or inward) open conformations (Crisman *et al.*, 2009; Kowalczyk *et al.*, 2011; Radestock and Forrest, 2011; Liao *et al.*, 2012; Mancusso *et al.*, 2012; Schushan *et al.*, 2012; Forrest, 2013; Kim *et al.*, 2019).

These dynamic structural and biophysical properties confer to SLC proteins their functions as gates for symporter and antiporter transport of biochemically-important solutes and biomolecules (Bai *et al.*, 2017; Pizzagalli *et al.*, 2021). Both experimental and computational studies of SLC proteins have provided important insights into the role of these conformational dynamics in solute transport. However, as they are medium-sized integral membrane proteins, molecular dynamics simulations are quite challenging, requiring powerful computing resources and appropriate simulation of membrane-mimicking environments. While AlphaFold2 (AF2) (Jumper *et al.*, 2021), Evolutionary Scale Modeling (ESM) (Lin *et al.*, 2023), and related machine-learning methods (Baek *et al.*, 2021; Ahdritz *et al.*, 2024) can provide accurate structural models of integral membrane proteins, for systems that adopt multiple conformational states like these SLCs, conventional AF2 calculations generally identify only one of the multiple states observed experimentally (Huang *et al.*, 2021; Del Alamo *et al.*, 2022; Saldano *et al.*, 2022; Chakravarty *et al.*, 2023; Kalakoti and Wallner, 2024; Wayment-Steele *et al.*, 2024; Xie and Huang, 2024). Multiple conformational state modeling of SLC proteins can also be guided by evolutionary covariance (EC) analysis of functionally-preserved direct contacts, and thus provide information about contacts present in the two (or more) states adopted by the protein structure (Hopf *et al.*, 2012; Morcos *et al.*, 2013; Sutto *et al.*, 2015; Toth-Petroczy *et al.*, 2016; Zheng *et al.*, 2018; Huang *et al.*, 2019; Schafer and Porter, 2023).

Recently significant advances have been reported using modified AF2 protocols to accurately model multiple conformational states of proteins, including integral membrane proteins. Promising approaches use a conventional AF2 platform with curated input such as (i) state-annotated conformational templates (Heo and Feig, 2022), (ii) shallow multiple sequence alignments (MSAs) that are chosen either randomly (AlphaFold-alt) (Del Alamo *et al.*, 2022; Sala *et al.*, 2023) or by clustering homologous protein sequences (AF-cluster) (Wayment-Steele *et al.*, 2024), (iii) single protein sequences (Porter *et al.*, 2023), or (iv)

MSAs masked at multiple positions (SPEACH-AF) (Stein and McHaourab, 2022) to switch the prediction toward alternative conformational states. AF2 calculations using network dropouts (AFsample) can also generate conformational diversity (Wallner, 2023b; a; Kalakoti and Wallner, 2024). These evolving AI-based enhanced sampling methods can sometimes, but not always, provide models of multiple conformational states of SLC proteins (Del Alamo *et al.*, 2022; Chakravarty *et al.*, 2023; Kalakoti and Wallner, 2024; Wayment-Steele *et al.*, 2024; Xie and Huang, 2024).

Here we describe a simple and robust approach for modeling alternative conformational states of pseudo-symmetric SLC proteins using a combined ESM - AF2 process. Using these models to guide site-directed mutagenesis, we also explore the impact of point mutations on SLC35F2 conformational distributions using an enhanced AF2 sampling method (Del Alamo *et al.*, 2022). The ESM-AF2 approach was used to model the inward- / outward-open forms of two SLC proteins, human ZnT8 (SLC30A8, a Zn transporter) and *Escherichia coli* D-galactonate:proton symporter (SLC17, a MFS superfamily transporter) for which experimental structures of both outward- and inward-open states are available, and the resulting models of alternative conformations were validated by comparison against these cryoEM or X-ray crystal structures. These models were also validated against EC-based contact maps. For two additional SLC proteins, *Saccharomyces cerevisiae* GDP-mannose sugar transporter 1 (SLC35D subfamily) and CMP-sialic acid transporter 1 (which SLC family SLC35A1) only the outward-open forms are available as X-ray crystal structures. Here, the alternative inward-open forms were modeled with the ESM-AF2 process, and then validated by comparison against EC-based contact maps. For SLC35F2, neither inward nor outward open experimental structures are available. The outward-open form was modeled using conventional AF2, and the inward-open conformational state was then modeled using the ESM-AF2 process. Both the inward- and outward-open structures were then validated against EC-based contact maps. Studies of the conformational diversity of SLC35F2 using AF-alt (Del Alamo *et al.*, 2022) exhibit a strong preference for the outward-open state. Using the EC-predicted contacts as a guide, clusters of mutations were identified that bias the distribution towards either the inward- or outward-open states.

METHODS

Evolutionary covariance (EC) - based contact predictions. EC-based contact predictions were performed using evolutionary covariance analysis with *NeBcon* (Neural-network and Bayes-classifier based contact prediction) https://seq2fun.dcmb.med.umich.edu//NeBcon/, a hierarchical algorithm for sequence-based protein contact map prediction (He *et al.*, 2017), with a probability threshold of 0.7. A second server, *EVcouplings server* https://evcouplings.org/ was also used to confirm these contact predictions.

Contact maps for experimental and predicted structures were obtained from *CMview* (Vehlow *et al.*, 2011), an interactive contact map visualization and analysis tool. Contact maps were generated for interresidue Cα distances of < 10.0 Å. The contact lists generated from protein structure models were then imported into excel spreadsheets for overlay and comparison with the EC-based predicted contacts.

AlphaFold2, ESMfold, and Modeller modeling. AlphaFold2 (Jumper *et al.*, 2021) modeling was performed using Colabfold v1.5.5 server (Mirdita *et al.*, 2022) with *AlphaFold2.ipynb* scripts, with no templates, default multiple sequence alignments (MSAs), recycle of 12, and with random dropouts. The Amber-relaxed top-ranked model was taken as the final predicted structure. Evolutionary Scale Modeling (ESMfold) (Lin *et al.*, 2023) models were generated using the *ESMFold_advanced.ipynb* colab script. Models were generated with random masking of input sequences (defined by masking_rate of 0.15), stochastic_mode="LM" that uses no dropout, and recycle of 12. The model with the maximum pTM score was selected as the final model. A locally installed version of MODELLER 10.4 (Sali and Blundell, 1993; Webb and Sali, 2016) was used for conventional template-based modeling. 20 models were predicted for each run and the model with the best DOPE (Discrete Optimized Protein Energy score) was selected as the representative structure.

AlphaFold-alt. Enhanced sampling with AlphaFold-alt (AF-alt) was carried out as described by Meiler and co-workers (Del Alamo *et al.*, 2022), using scripts kindly provided by Dr. Davide Sala and executed on a local cluster of 4 A100 Nvidia HGX GPU processors. In each AF-alt run, 480 models were made using randomly-generated shallow MSAs of 16-32 sequences. 30 models were generated for each MSA depth of 16 to 32 sequences. Each run was < 3 hrs. No structural templates were used. For each model, disordered N- and C-terminal regions were removed and the average pLDDT score (<pLDDT>) was then computed for all of the remaining residues.

Statistical methods. Backbone root-mean-squared deviation (RMSD) and global distance test (GDT) scores for structural comparisons were performed using the methods of Zemla implemented on their public server http://linum.proteinmodel.org/ (Zemla *et al.*, 2005).

RESULTS

EMS-AF2 modeling protocol. The challenge we addressed arises from the fact that conventional AF2 modeling will generally provide only one of the multiple conformations that are observed experimentally (Huang *et al.*, 2021; Del Alamo *et al.*, 2022; Saldano *et al.*, 2022; Chakravarty *et al.*, 2023; Kalakoti and Wallner, 2024; Wayment-Steele *et al.*, 2024; Xie and Huang, 2024), motivating a need for complementary methods for generating alternative conformational states. The ESM-AF2 process for modeling alternative conformational states of SLC transporters that have structural pseudo-symmetry is outlined in **Figure 1**. It is based conceptually on methods used for other pseudo-symmetric SLC proteins (Forrest, 2013; Kim *et al.*, 2019), in which the pseudo-symmetric halves of the transporter are first identified as an N-terminal protein sequence (blue in Figure 1) and C-terminal protein sequence (purple in **Figure 1**), and the N-terminal protein sequence is then modeled using the C-terminal segment as a structural template, and the C-terminal protein sequence is modeled using the N-terminal segment as a structural model. However, this method can be challenging if the sequence similarity in these two halves of the protein sequence is low. In the ESM-AF2 process, the N-terminal (blue) and C-terminal (purple) segments of protein sequences are first swapped to create a *virtual flipped sequence*. The entire structure of this virtual

sequence is then modeled using ESM fold, a large-language model based method that requires no templates and only a single protein sequence. The resulting *virtual protein structure* is then used as a structural template to model the original protein sequence using AF2.

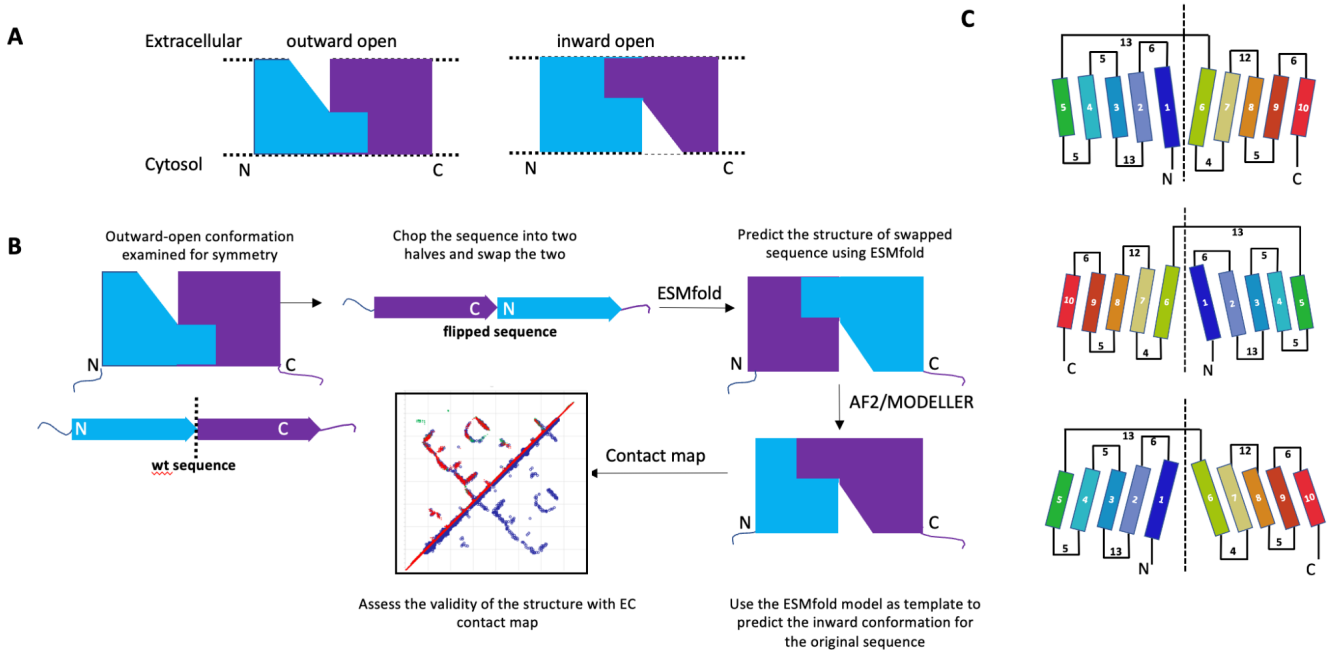

Fig. 1. The ESM-AF2 protocol for modeling alternative conformational states of pseudo-symmetric SLC proteins. (A) cartoon representation of inward/outward open conformers representing the pseudo-symmetry of the helices, with pseudo-symmetry halves indicated in blue and purple. (B) Protocol to model inward/outward open conformers for symmetric helical transmembrane proteins (C) Topology diagrams showing the conformational flip of a representative 10-helical SLC protein (SLC35F2). The vertical dotted line represents the symmetry axis of the pseudo-symmetric halves of the SLC protein. Numbers represent the number of residues in the membrane-external loops. The top image represents the outward-facing state, the middle image is the ESMfold virtual protein structure generated from a virtual flipped protein sequence, and the bottom image the inward-facing state generated by comparative modeling using the virtual protein structure as a modeling template.

In this study, an ESMfold structure model generated from a virtual flipped sequence was used as a custom template for AF2 with low MSA (16 - 32), recycle of 12, and with dropout. A shallow MSA is used so that the template information dominates the modeling process. This "comparative modeling" step can also be performed using Modeller (Sali and Blundell, 1993; Webb and Sali, 2016), SwissModel (Waterhouse *et al.*, 2018), or other template-based modeling method. Finally, the original (e.g., outward open) and final (e.g., inward open) structures are validated by comparison against the EC-based contact map that will generally include predicted contacts for both conformational states.

Validating the EMS-AF2 modeling protocol. As an initial test case of the EMS-AF2 method for modeling alternative conformational states of SLC proteins, we selected human ZnT8 (SLC30A8), a 2 x 320-residue homodimeric integral membrane protein Zn-transporter, a representative SLC protein for which structures have determined by cryoEM (Xue *et al.*, 2020) (PDB ids: 6xpd, 6xpde, and 6xpf, at resolutions of 3.9 Å, 4.1 Å, and 5.1 Å, respectively). ZnT8 (PDB id: 6xpf) has two subunits; in the

absence of Zn, chain-A is in an inward-open conformation and chain-B in an outward-open conformation. Conventional AF2-colab calculations provided a structure with inward-open conformation, in agreement with the cryoEM inward-open structure 6xpf-A (Cα RMSD 2.00 Å). (**Figure 2A**). We then used the ESM-AF2 modeling protocol outlined in **Figure 1** to model the outward-open conformational state, and compared the resulting model with the experimentally-determined outward-open cryoEM structure. The computed outward-open conformation of ZnT8 has excellent agreement with experimental outward-open structure 6xpf-B, with backbone Cα RMSD of 1.09 Å (**Figure 2B**). We also compared residue-residue contact maps for the experimental and EMS-AF2 outward open models with each other and with the EC-based contact map generated from multiple-sequence alignments of ZnT8 homologs (**Figure 2C,D**). The computed inward facing structure, modeled with AF2, has a contact map that is nearly identical to that of the experimental inward-facing structure (**Figure 2C**); the outward-facing structure computed from the inward-facing structure using the ESM-AF2 protocol is also essentially identical to the experimental outward-facing structure (**Figure 2D**). While many ECs are common to both the outward- and inward-open conformations, the ECs contain information about both states, and several are unique to each conformation; i.e. there are 6 unique ECs for outward-open and 6 unique ECs for inward-facing states. These several ECs unique to the outward/inward open conformations superimpose on top of the corresponding unique contacts in the outward and inward-open computed models, respectively (circled in **Figures 2C,D**). Hence, the ESM-AF2 protocol successfully modeled the outward-open conformation of Znt8, as validated by comparison with EC-derived contacts. A second test case for the ESM-AF2 modeling protocol using an SLC protein with both inward and outward-facing experimental structures is presented for the *E. coli* D-galactonate:proton symporter in **Supplementary Information**.

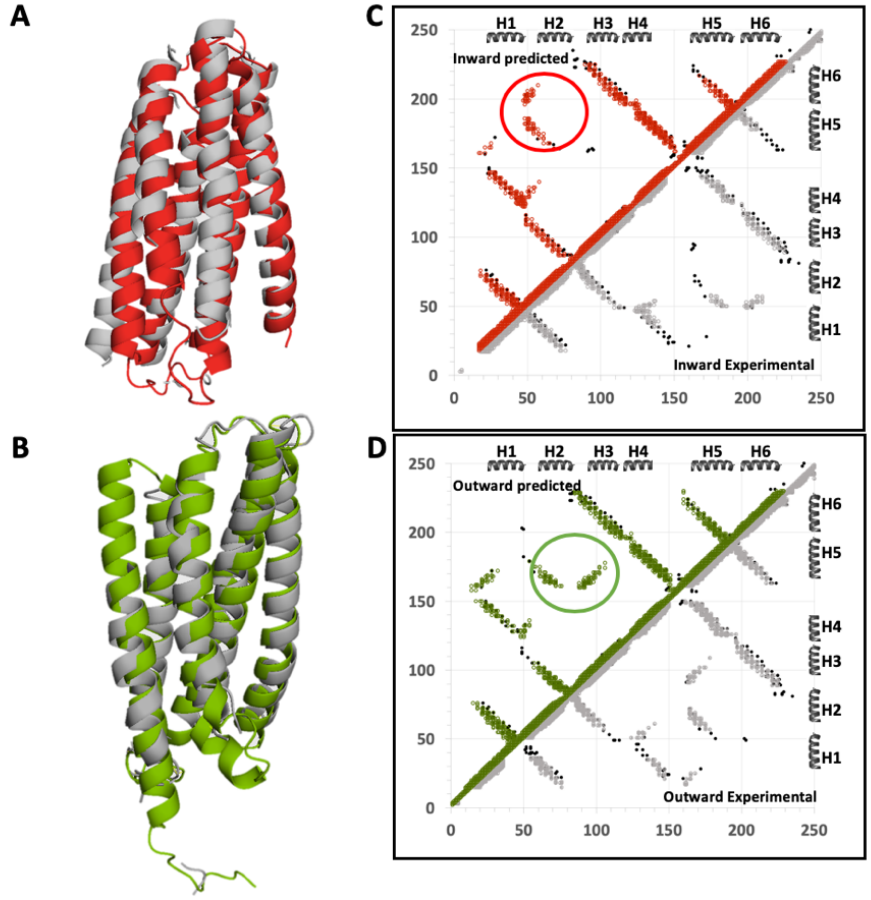

Fig. 2. Validation of ESM-AF2 protocol using and SLC protein with both outward- and inward-open experimental structures. The experimentally-determined cryo-EM structure of human ZnT8 WT in the absence of zinc has two chains, with one subunit in an inward-facing conformation and the other in an outward-facing conformation (PDB id: 6xpf chain A and B respectively). (A) Superposition of the AF2-predicted (red) and experimental (grey) inward-open structures, with backbone RMSD of 2.00 Å. (B) Superposition of outward-open model generated using the ESM-AF2 protocol (green) with the experimental (grey) outward-open structure, with backbone RMSD of 1.09 Å. (C) Comparison of the EC-based contact map of ZnT8 (points shown in black) with contacts in the experimental (grey points) and predicted (red points) inward-facing models. (D) Comparison of the EC-based contact map of ZnT8 (points shown in black) with contacts in the experimental (grey points) and predicted (green points) outward-facing models. In panels C and D, major differences in the contact patterns of inward-open and outward-open states, supported by ECs unique to each state, are circled.

Modeling inward-open forms of SLC proteins. In the case above (and in the SI), we chose SLCs proteins for which experimental structures of both outward and inward open conformations are available, and validated the ESM-AF2 modeling protocol against both the experimental atomic coordinates (using RMSD and GDT metrics) and against contact maps predicted from EC analysis which are based on experimental primary sequence data. However, for most SLC proteins, experimental structures are only available for one (or none) of the two states. We next modeled inward-open structures for two integral membrane proteins for which only the outward-open state is experimentally available. The results are shown in **Figure 3** for the 337-residue *Saccharomyces cerevisiae* GDP-mannose sugar transporter 1 Vrg4 (PDB id 5oge (Parker and Newstead, 2017)), an SLC35D subfamily member, and in **Figure 4** for the 322-residue *Zea mays* CMP-sialic acid transporter 1 [PDB id 6i1r-A (Nji *et al*., 2019)], an SLC35A subfamily member. For both proteins, only outward-open forms determined at 3.22 Å and 2.80 Å resolution, respectively, are available as X-ray crystal structures. The inward-open forms modeled with the ESM-AF2 process were validated by comparison against the EC-based contact map. In both cases, the EC-based contact maps could be largely explained by the combined contact maps of the outward- and inward-

open conformations, although some sporadic predicted ECs at the edge of the cutoff value used for identifying ECs were also present. These results further validate the ESM-AF2 process.

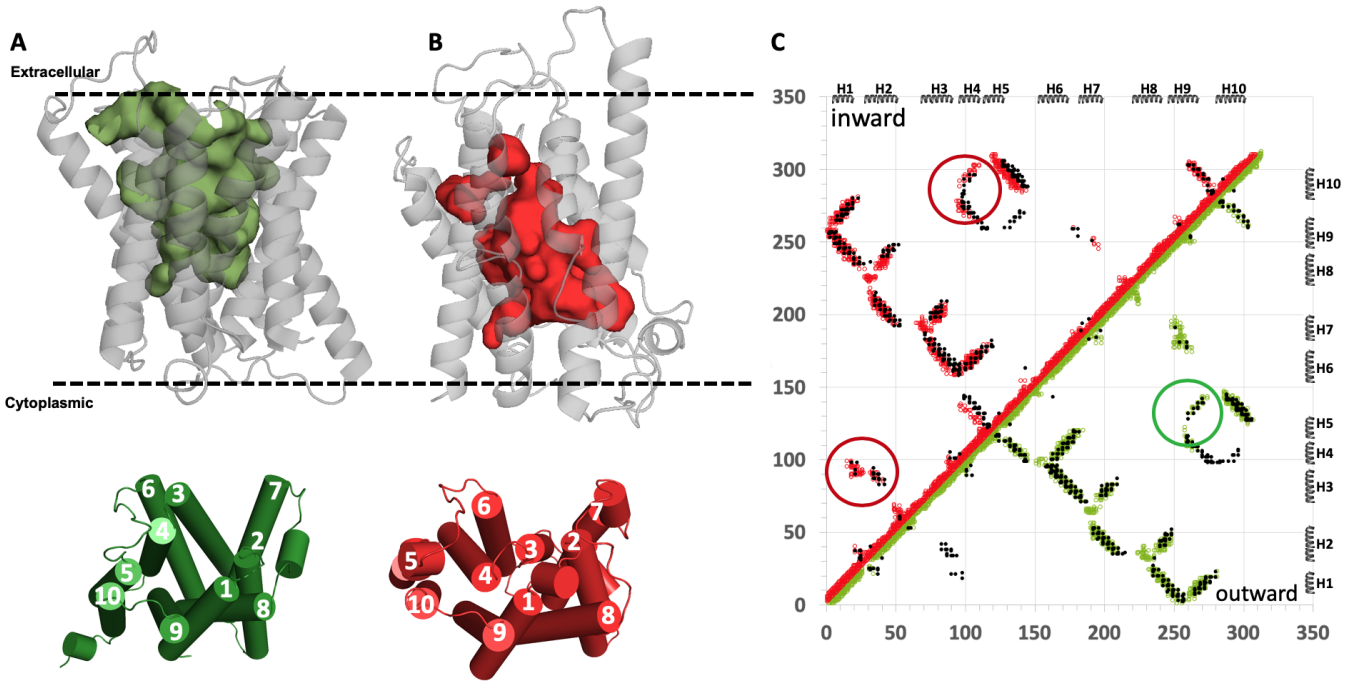

Fig. 3. ESM-AF2 modeling of the inward-open conformation of the *Zea mays CMP-sialic acid transporter 1*. (A) The experimental outward-open structure (PDB id 6i1r-A). (B) The inward-open structure modeled using ESM-MODELLER. In each of panels A and B the top images are ribbon representations of the protein structure with surface exposed cavities shown in either green (outward-open) or red (inward-open), and the bottom images are cylinder representations of these structural states with helices numbered 1 - 10. The dashed horizontal lines in panels A and B denote the approximate locations of the membrane boundaries. (C) The combined contact maps of the two resulting models are consistent with the experimental EC-based contact map. Green contacts are those present in the experimental outward-open model, and red contacts are those present in the predicted inward-open model. EC-based contacts are shown as black dots. The EC-based contacts circled in green are unique to the outward-open conformation, and those circled in red are unique to the inward-open conformation. At the thresholds chosen for ECs several predicted contacts are not explained by the combination of two conformational states. In panels A and B (top), surface pockets are represented as space-filled voids using the server https://kvfinder-web.cnpem.br/.

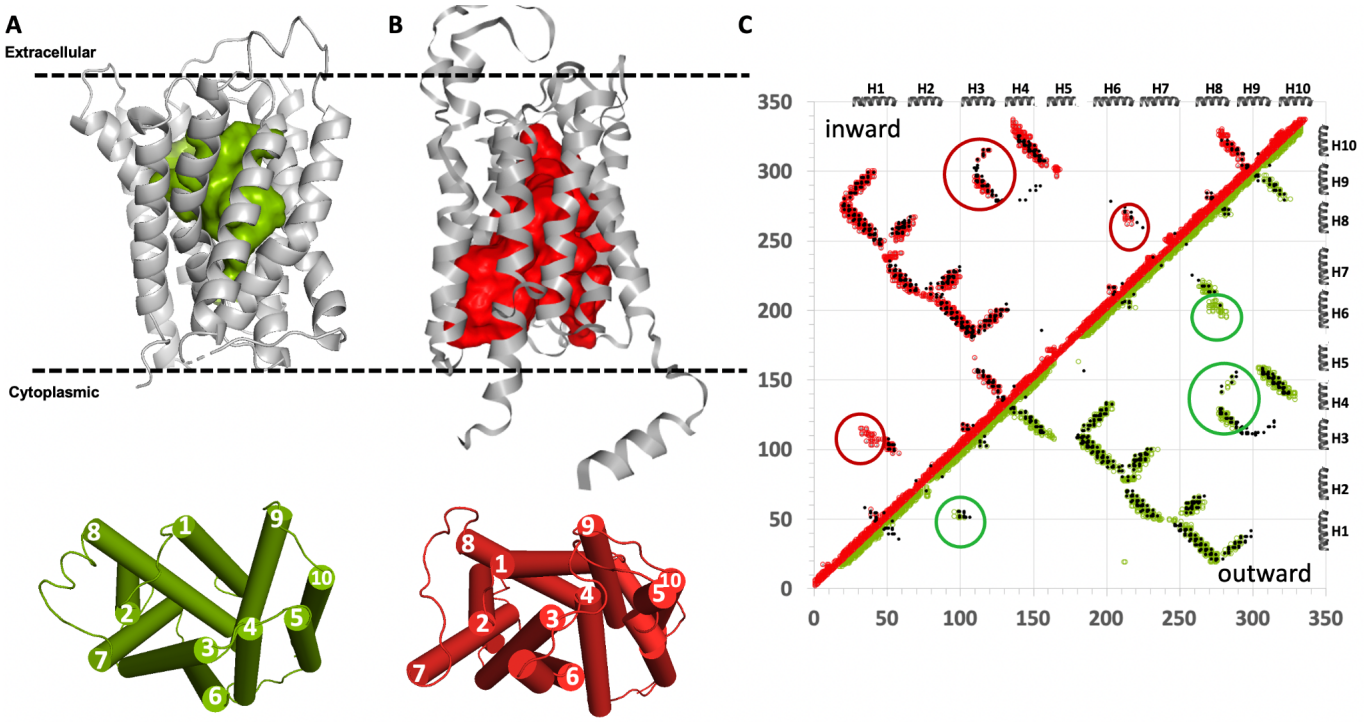

Fig. 4. ESM-AF2 modeling of the inward-open conformation of the *S. cerevisiae* GDP-mannose sugar transporter 1, Vrg4. (A) The experimental outward-open structure (PDB id 5oge). (B) The inward-open structure modeled using ESM-AF2. In each of panels A and B the top images are ribbon representations of the protein structure with surface exposed cavities shown in either green (outward-open) or red (inward-open), and the bottom images are cylinder representations of these structural states with helices numbered 1 - 10. The dashed horizontal lines in panels A and B denote the approximate locations of the membrane boundaries. (C) The combined contact maps of the two resulting models are consistent with the EC-based contact map. EC-based contacts are shown as black dots, inward-open contacts as red circles and outward-open contacts as green circles. The EC-based contacts circled in green are unique to the outward-open conformation, and those circled in red are unique to the inward-open conformation. At the thresholds chosen for ECs several predicted contacts are not explained by the combination of two conformational states. In panels A and B, surface pockets are represented as space-filled voids using the server https://kvfinder-web.cnpem.br/.

Modeling alternative conformations of SLC35F2 with ESM-AF2. SLC35F2 has < 12% sequence identity with the SLC35 subfamily members of known structure; in particular there is no good experimental structure that can be used as a template for comparative modeling of its inward-open conformation. Having established the reliability, consistency, and limitations of the EMS-AF2 protocol, AF2 was used to model the outward-open conformation of SLC35F2, and ESM-AF2 was used to model its inward-open conformation (**Figure 5**). The contact maps of these two conformations were then compared with its EC-based contact map. The excellent agreement between the EC-based contact map and combined contact maps of the computed outward- and inward-open structures validate the accuracy of the ESM-AF2 protocol for modeling this conformational heterogeneity of SLC35F2.

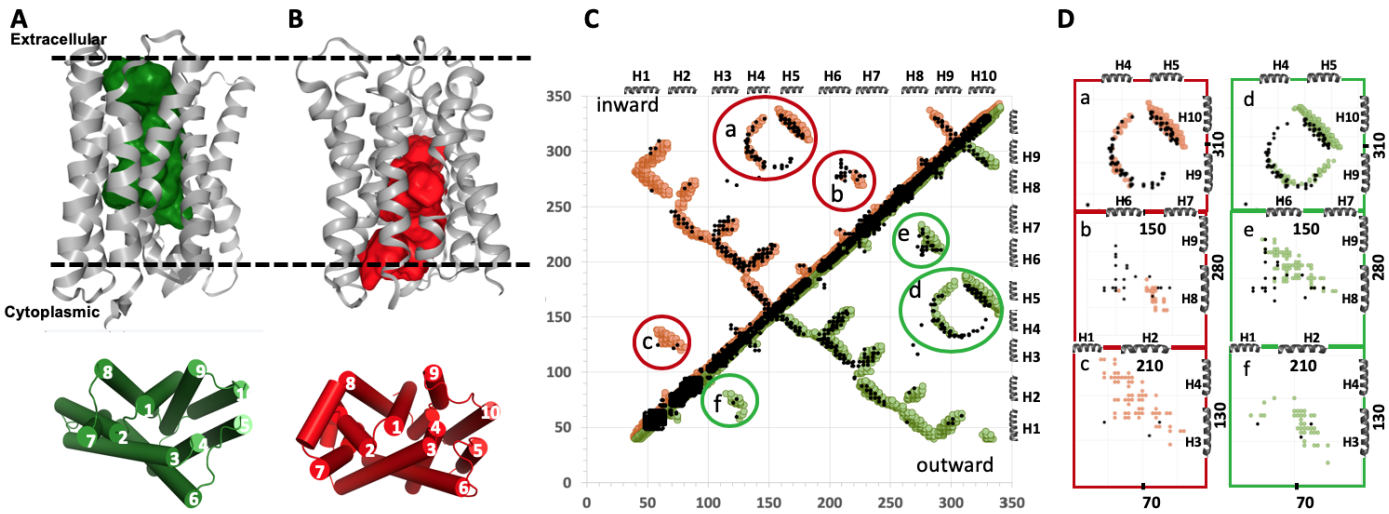

Fig. 5. AF2 / EMS-AF2 modeling of the outward- and inward-open conformations of human SLC35F2. (A) The outward-open structure modeled with AF2. (B) The inward-open structure modeled using ESM-AF2. In each of panels A and B the top images are ribbon representations of the protein structure with surface exposed cavities shown in either green (outward-open) or red (inward-open), and the bottom images are cylinder representations of these structural states with helices numbered 1 - 10. The dashed horizontal lines in panels A and B denote the approximate locations of the membrane boundaries. (C) Contact maps of the outward-open (green circles) and inward-open (red circles) structures superposed on the EC contact map (black dots). (D) Expanded regions (labeled a through f) of panel C, focusing on key distinguishing contacts and ECs between helices H4 and H10 in the inward-open conformation (subpanel D.a) and between helices H5 and H9 in the outward-open conformation (subpanel D.d). Also shown in panel D, subpanels a, b, c, e and f are other key contacts and ECs distinguishing the two states. In panels A and B (top), surface pockets are represented as space-filled voids using the server https://kvfinder-web.cnpem.br/.

Having established models of the outward- and inward-open conformation states, we next explored using AF2 modeling calculations to design point mutations that can destabilize the outward-open structure relative to the inward-open structure, and *visa versa*. Contact maps generated by ESM-AF2 for the wild-type SCL35F2 sequence (**Figure 5C,D**) were used together with the models of inward- and outward-facing states to design mutations predicted to stabilize or destabilize these states. ESMfold predictions, which are relatively fast, were then performed for these various mutant sequences. Examples of the mutant designs include Mutant 1 ([E216G, V220G, S224G, F228G]-SLC35F2; contacts in subpanels **Figure 5D.b** and **e**) designed to prefer the inward-open conformation by destabilizing the outward-open conformation, and Mutant 2 ([A57G, Y129G, L134G]-SLC35F2; contacts in subpanels **Figure 5D.c** and **f**) designed to prefer the outward-open conformation by destabilizing the inward-open conformation. The locations of these mutations in the outward-open model of SLC35F2 are shown in **Figure 6A**. Using AF2 modeling [no template, shallow (16 sequences) MSA, recycle of 12, and dropouts], Mutant 1 ([E216G, V220G, S224G, F228G]-SLC35F2) returns an inward-open structure (**Figure 6B**), while similar AF2 modeling of Mutant 2 ([A57G, Y129G, L134G]-SLC35F2) results in an outward-open model. The Cα contact maps generated for Mutant 1 (shown in red in **Figure 6B**) are in good agreement with the contact map (shown in blue in **Figure 6B**) for the inward-open model predicted using the flipped-template method, indicating the effect of mutations in inducing conformational change.

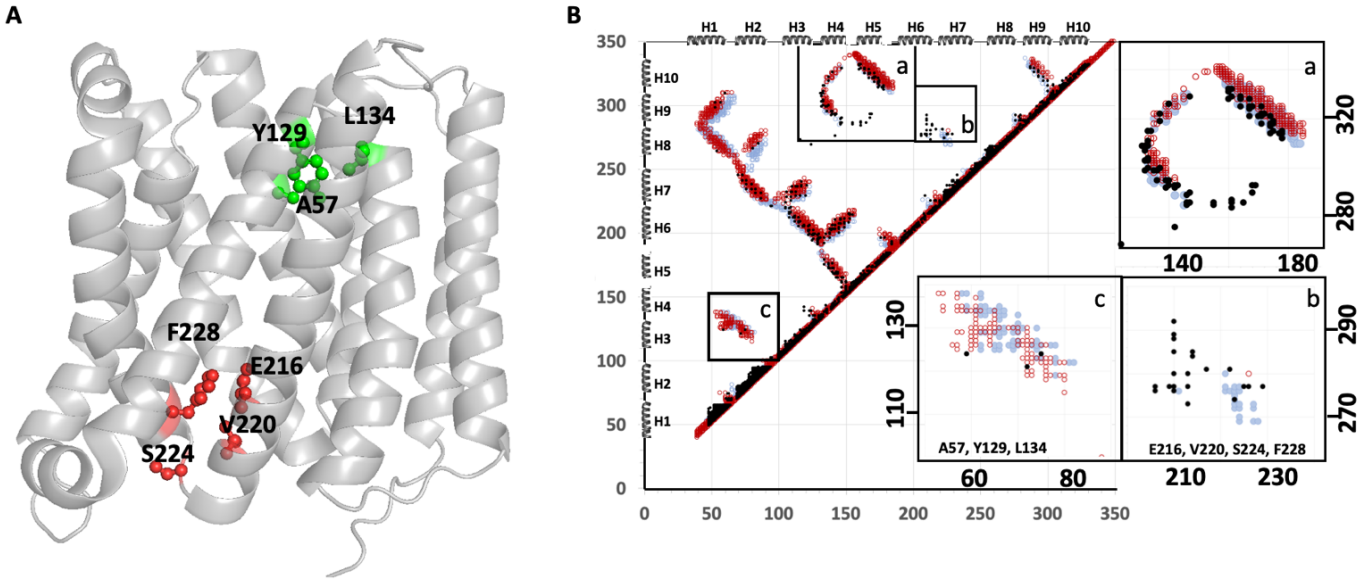

Fig. 6. (A) Mutation sites that shift the states from outward-open to inward-open (red residues) and inward-open to outward-open (green residues) are shown mapped on the outward-open model generated by AF2. (B) C_α contact map of wild-type (wt) SLC35F2 inward-open model generated by ESM-AF2 protocol (blue circles) and inward-open model of ([E216G, V220G, S224G, F228G]-SLC35F2 (Mutant 1) generated by standard AF2 modeling (red circles). Subpanels a, b, and c illustrate ECs (black dots) and key inter-residue contacts in the inward-facing models of wt (blue circles) and Mutant 1 SLC35F2 (red circles) used to design mutations, as discussed in the text.

AF-alt modeling of SLC35F2 conformational distributions. An alternative approach for modeling the multiple conformational states of SLC proteins is to employ recently described enhanced sampling protocols using AF2 with shallow MSAs or other forms of modified input information (Del Alamo *et al.*, 2022; Stein and McHaourab, 2022; Stein and McHaourab, 2024; Wayment-Steele *et al.*, 2024). In particular, AF-alt has been used successfully to generate both inward-open and outward-open conformations for SLC protein ZnT8 and other integral membrane proteins (Del Alamo *et al.*, 2022). We applied the AF-alt algorithm to SLC35F2 in order to further explore its potential for modeling the preferences of wild-type (wt) and mutant sequences for inward vs outward open states. We focused our analysis only on models with overall <pLDDT> > 70 (blue dots in **Figure 7**), a score characteristic of reliable models (Jumper *et al.*, 2021). For wt SLC35F2, AF-alt did not successfully model an inward-facing state; rather it predicts only the outward-open conformational state, with no models of inward-open conformations with <pLDDT> > 70 (**Figure 7A**). We next assessed if AF-alt can provide information on the conformational states accessible to the two mutant SLC35F2 proteins. For Mutant 1 ([E216G, V220G, S224G, F228G]-SLC35F2), AF-alt generates a distribution of high-scoring (<pLDDT> > 70) models with both inward- and outward-open conformations; for Mutant 2 ([A57G, Y129G, L134G]-SLC35F2), the distribution of high-scoring (<pLDDT> > 70) models is similar to that of wt-SLC25F2, with a small shift toward the open-out conformational state. The plots of **Figure 7** also indicate that the outward-open structure of Mutant 1 is distorted compared to the wt structure, with a slightly higher GDT-TS with the inward-open structure (i.e. GTD-TS inward-open of about ~70% rather than ~ 62%).

Although these results are not quantitative, and can only be used to identify trends, they demonstrate the potential for developing AF-alt -like enhanced sampling protocols to model the conformational preferences of SLC proteins and to design mutations that can shift from one conformational state to another.

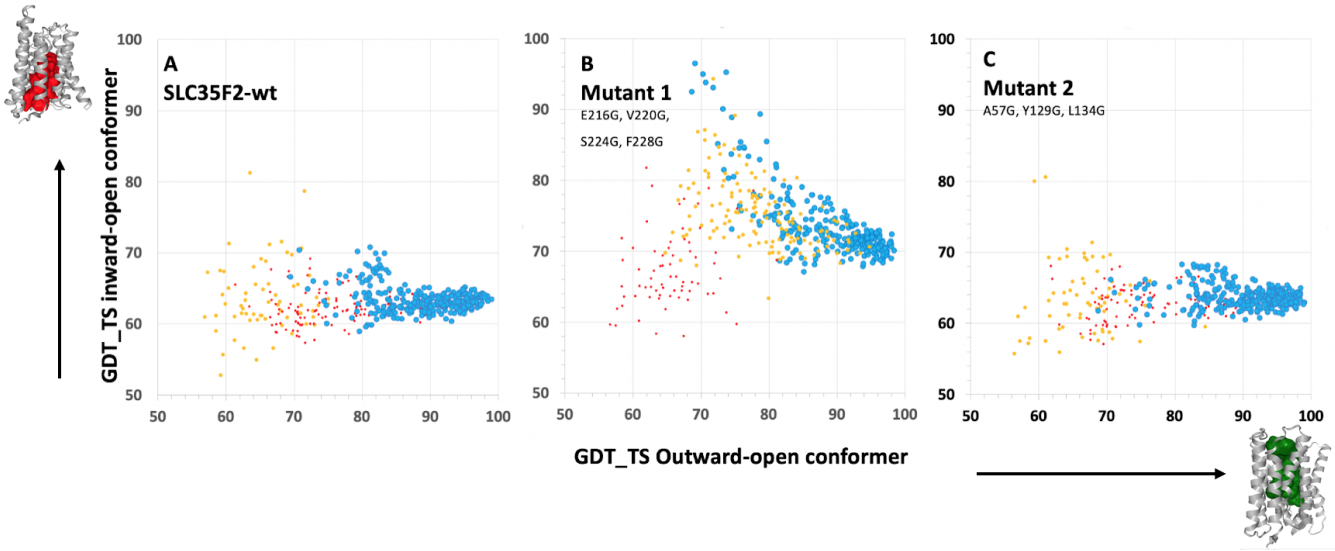

Fig. 7. Conformational distributions predicted by AF-alt. 480 models were predicted for each of the three sequences (A) wild-type SLC35F2, (B) Mutant 1 ([E216G, V220G, S224G, F228G]-SLC35F2), and (C) Mutant 2 ([A57G, Y129G, L134G]-SLC35F2), using shallow (16-32 sequence) MSAs. The resulting models are color coded by average residue-specific pLDDT (<pLDDT>) scores: < 50% - red, small dots; 50 - 70% - yellow, medium-sized dots; > 70% - blue, large dots. <pLDDT> scores were computed after eliminating the apparently flexible 36-residues from the N-terminal and 38-residues from the C-terminal regions. These N- and C-terminal segments are predicted to be disordered and have very low pLDDT scores.

DISCUSSION

We have developed and tested a hybrid AF2 / ESM-AF2 protocol for modeling alternative conformations of pseudo-symmetric SLC transporters. Generally, where AF2 provides only one (either inward- or outward-open) conformational state; the alternative state can then be modeled by the ESM-AF2 (or ESM-MODELLER) protocol. The ESM-AF2 protocol is inspired by a more traditional approach using comparative modeling of the pseudo-symmetric halves of SLC transporters (Crisman *et al.*, 2009; Kowalczyk *et al.*, 2011; Radestock and Forrest, 2011; Liao *et al.*, 2012; Mancusso *et al.*, 2012; Schushan *et al.*, 2012; Forrest, 2013; Kim *et al.*, 2019). This traditional approach requires an accurate sequence alignment between the two symmetric halves of SLC protein. However, in some cases it is difficult to determine the correct sequence alignment needed for accurate comparative modeling. In the ESM-AF2 (or ESM-MODELLER) approach, we use EMSfold to generate from a virtual flipped sequence a virtual protein structure, which is then used as a structure modeling template without the need for any sequence alignment between the two halves of the SLC protein. This allowed us to reliable model alternative conformational states of several SLC transporters that were difficult to model using the traditional approach. The resulting multi-state models are validated by comparison with sequence-based evolutionary

co-variance data (ECs) that encode information about contacts present in the various conformational states adopted by the protein.

The ESM-AF2 (or ESM-MODELLER) approach is simple to implement and runs fast using publicly-available AI-based servers. However, despite the successful examples demonstrated in this study, these ESM-AF2 (or ESM-MODELLER) protocols for modeling alternative conformational states of pseudo-symmetric SLC proteins have some limitations. They cannot be applied directly to homodimeric pseudo-symmetric SLC proteins, such as YiiP or EmrE (Fleishman *et al.*, 2006; Bai *et al.*, 2017). Coordinates of SLC proteins with large loops and other structural decorations require manual editing to eliminate these loops / decorations prior to applying the protocol. In addition, the validation of alternative state conformations by contact predictions relies on the quality of these contact predictions, and may not work well for smaller SLC sequence families.

Although we have focused our analysis on the outward and inward conformational states of SLC transporters, intermediate "occluded" states have also been captured in X-ray crystal and cryoEM structures. Although the ESM-AF2 protocol could potentially also generate such occluded states, this was not observed in the cases studied here. However, the AF-alt enhanced sampling protocol produced a range of conformational states, including outward-open, inward-open, and intermediate occluded conformational states (**Figure 7**). The significance of these intermediate conformational states in mechanisms of solute transport will require further experimental studies.

The idea of introducing mutations to induce conformational shifts is an important tool for biological studies. In our work on SLC35F2, we observed that while conventional AF2 returns only one of multiple states, carefully selected mutations can shift the resulting model between outward-facing and inward-facing states. Specifically, we observed that conventional AF2 can be used to model conformational switching induced by targeted structurally-clustered mutations defined by ECs and contacts present in the inward- and outward-open models. In this approach, key contacts were selected from regions of these contact maps that are unique to one conformational state and mutations were designed to disrupt or stabilize these interactions. Conformational shifting between states of SLC35F2 by these designed mutations was also observed using the AlphaFold-alt protocol (**Figure 7b**).

Previous studies have pointed out that AF2 may not reliably predict the structural effects of missense mutations (Buel and Walters, 2022) or the effects of mutations on protein stability (Buel and Walters, 2022; Pak *et al.*, 2023); generally speaking, AF2 does not provide quantitative information about how mutations may destabilize protein structures. However, more recent work demonstrates some accuracy in predicting effects of single-site mutants with AF2. McBride et al report strong correlations of local structural perturbations resulting from 1-3 mutations across 3,901 sets of experimental and AF2 models (McBride *et al.*, 2023). AlphaFold-cluster, using shallow MSAs of highly homologous sequence clusters was reported to provide reliable predictions of the effects of single-site mutations on conformational flipping between states of both KaiB and RafH (Wayment-Steele *et al.*, 2024). SPEACH_AF with model

relaxation and energetic analysis with Rosetta has also been used successfully to assess structural effects of single site mutations (Stein and McHaourab, 2024). While still evolving, the prudent application of AI-based modeling methods in designing specific mutations useful for structure-function studies is an important area for further investigations.

Conclusions. In this work we describe and validate a hybrid ESM-AF2 approach for modeling alternative conformational states of pseudo-symmetric SLC proteins. The approach overcomes the shortcoming of conventional AF2 structure calculations which generally provide only one of the multiple conformational states observed experimentally. The method is simple to use, rapid to run, and can be implemented using the public domain ESMfold (Lin *et al.*, 2023) and AF2-colab (Mirdita *et al.*, 2022) servers. In this approach, the resulting multi-state models are validated by comparison with sequence-based EC data that encode information about contacts present in the various conformational states adopted by the protein. We also explored the complementary use of conventional AF2, ESM-AF2, and AF-alt (Del Alamo *et al.*, 2022) to model the effects of mutations on conformational preferences of the SLC35F2 transporter. Experimental studies will be required to characterize the structures, dynamics, and functions of these SLC35F2 mutants. Overall, the current study validates the ESM-AF2 protocol for modeling conformational heterogeneity of pseudo-symmetric SLC transporters, one of the most extensive class of transporters in the human proteome.

ACKNOWLEDGEMENTS
We thank Dr. Davide Sala for providing scripts for running AF-alt, T.B. Acton, T. Benavides, A. De Falco, K. Fraga, A. Gaur, R. Greene-Cramer, Y.J. Huang, T.A. Ramelot, B. Shurina, L. Spaman, and R. Tejero for helpful discussions and comments on the manuscript, and S. Collen for computer system administration support. This work was supported financially by National Institutes of Health NIGMS grants R35 GM141818 (to G.T.M.) and R35 GM122518 (to M.J.R.), and by the Rensselaer Polytechnic Institute (RPI) Bio-computing and Bio-informatics Constellation Chair Fund. GTM also acknowledges access to the RPI Center for Computational Innovations (CCI) computing infrastructure.

AUTHOR CONTRIBUTIONS
GVTS, ND, MJR, and GTM jointly conceptualized the study and analyzed data. GVTS carried out bioinformatics analyses and generated graphics. All authors contributed in writing and editing the manuscript.

DECLARATION OF INTERESTS
GTM is a founder of Nexomics Biosciences, Inc. This does not represent a conflict of interest for this study.

SUPPLEMENTARY MATERIAL
Fig. S1. Example validation of ESM-AF2 protocol using SLC proteins with both outward- and inward-open experimental structures.

REFERENCES

Ahdritz, G., Bouatta, N., Floristean, C., Kadyan, S., Xia, Q., Gerecke, W., O'Donnell, T.J., Berenberg, D., Fisk, I., Zanichelli, N., et al. (2024). OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. **Nature Methods**. https://doi.org/10.1038/s41592-024-02272-z.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. **Science** *373*, 871-876. https://doi.org/10.1126/science.abj8754.

Bai, X., Moraes, T.F., and Reithmeier, R.A.F. (2017). Structural biology of solute carrier (SLC) membrane transport proteins. **Mol Membr Biol** *34*, 1-32. https://doi.org/10.1080/09687688.2018.1448123.

Buel, G.R., and Walters, K.J. (2022). Can AlphaFold2 predict the impact of missense mutations on structure? **Nat Struct Mol Biol** *29*, 1-2. https://doi.org/10.1038/s41594-021-00714-2.

Chakravarty, D., Schafer, J.W., Chen, E.A., Thole, J.R., and Porter, L.L. (2023). AlphaFold2 has more to learn about protein energy landscapes. **bioRxiv**, 2023.2012.2012.571380. https://doi.org/10.1101/2023.12.12.571380.

Che, B., Du, Y., Yuan, R., Xiao, H., Zhang, W., Shao, J., Lu, H., Yu, Y., Xiang, M., Hao, L., et al. (2023). SLC35F2-SYVN1-TRIM59 axis critically regulates ferroptosis of pancreatic cancer cells by inhibiting endogenous p53. **Oncogene** *42*, 3260-3273. https://doi.org/10.1038/s41388-023-02843-y.

Colas, C., Ung, P.M., and Schlessinger, A. (2016). SLC transporters: Structure, function, and drug discovery. **Medchem Comm** *7*, 1069-1081. https://doi.org/10.1039/C6MD00005C.

Crisman, T.J., 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. https://doi.org/10.1073/pnas.0908570106.

Del Alamo, D., Sala, D., McHaourab, H.S., and Meiler, J. (2022). Sampling alternative conformational states of transporters and receptors with AlphaFold2. **Elife** *11*. https://doi.org/10.7554/eLife.75751.

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. https://doi.org/10.1016/j.jmb.2006.08.072.

Forrest, L.R. (2013). Structural biology. (Pseudo-)symmetrical transport. **Science** *339*, 399-401. https://doi.org/10.1126/science.1228465.

Fredriksson, R., Nordstrom, K.J., Stephansson, O., Hagglund, M.G., and Schioth, H.B. (2008). The solute carrier (SLC) complement of the human genome: phylogenetic classification reveals four major families. **FEBS Lett** *582*, 3811-3816. https://doi.org/10.1016/j.febslet.2008.10.016.

He, B., Mortuza, S.M., Wang, Y., Shen, H.B., and Zhang, Y. (2017). NeBcon: protein contact map prediction using neural network training coupled with naïve Bayes classifiers. **Bioinformatics** *33*, 2296-2306. https://doi.org/10.1093/bioinformatics/btx164.

Hediger, M.A., Clemencon, B., Burrier, R.E., and Bruford, E.A. (2013). The ABCs of membrane transporters in health and disease (SLC series): introduction. **Mol Aspects Med** *34*, 95-107. https://doi.org/10.1016/j.mam.2012.12.009.

Heo, L., and Feig, M. (2022). Multi-state modeling of G-protein coupled receptors at experimental accuracy. **PROTEINS: Structure, Function and Bioinformatics** *90*, 1873-1885. https://doi.org/10.1002/prot.26382.

Hopf, T.A., Colwell, L.J., Sheridan, R., Rost, B., Sander, C., and Marks, D.S. (2012). Three-dimensional structures of membrane proteins from genomic sequencing. **Cell** *149*, 1607-1621. https://doi.org/10.1016/j.cell.2012.04.012.

Huang, Y.J., Brock, K.P., Ishida, Y., Swapna, G.V.T., Inouye, M., Marks, D.S., Sander, C., and Montelione, G.T. (2019). Combining evolutionary covariance and NMR data for protein structure determination. **Methods Enzymol** *614*, 363-392. https://doi.org/10.1016/bs.mie.2018.11.004.

Huang, Y.J., Zhang, N., Bersch, B., Fidelis, K., Inouye, M., Ishida, Y., Kryshtafovych, A., Kobayashi, N., Kuroda, Y., and Liu, G. (2021). Assessment of prediction methods for protein structures determined by NMR in CASP14: impact of AlphaFold2. **PROTEINS: Structure, Function and Bioinformatics**.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. **Nature** *596*, 583-589. https://doi.org/10.1038/s41586-021-03819-2.

Kalakoti, Y., and Wallner, B. (2024). AFsample2: Predicting multiple conformations and ensembles with AlphaFold2. **bioRxiv**, 2024.2005.2028.596195. https://doi.org/10.1101/2024.05.28.596195.

Killer, M., Wald, J., Pieprzyk, J., Marlovits, T.C., and Löw, C. (2021). Structural snapshots of human PepT1 and PepT2 reveal mechanistic insights into substrate and drug transport across epithelial membranes. **Science Advances** *7*, eabk3259. https://doi.org/doi:10.1126/sciadv.abk3259.

Kim, J., Tan, Y.Z., Wicht, K.J., Erramilli, S.K., Dhingra, S.K., Okombo, J., Vendome, J., Hagenah, L.M., Giacometti, S.I., Warren, A.L., et al. (2019). Structure and drug resistance of the Plasmodium falciparum transporter PfCRT. **Nature** *576*, 315-320. https://doi.org/10.1038/s41586-019-1795-x.

Kowalczyk, L., Ratera, M., Paladino, A., Bartoccioni, P., Errasti-Murugarren, E., Valencia, E., Portella, G., Bial, S., Zorzano, A., Fita, I., et al. (2011). Molecular basis of substrate-induced permeation by an amino acid antiporter. **Proc Natl Acad Sci U S A** *108*, 3935-3940. https://doi.org/10.1073/pnas.1018081108.

Leano, J.B., Batarni, S., Eriksen, J., Juge, N., Pak, J.E., Kimura-Someya, T., Robles-Colmenares, Y., Moriyama, Y., Stroud, R.M., and Edwards, R.H. (2019). Structures suggest a mechanism for energy coupling by a family of organic anion transporters. **PLoS Biol** *17*, e3000260. https://doi.org/10.1371/journal.pbio.3000260.

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. https://doi.org/10.1126/science.1215759.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. **Science** *379*, 1123-1130. https://doi.org/10.1126/science.ade2574.

Lu, Y., Zuo, P., Chen, H., Shan, H., Wang, W., Dai, Z., Xu, H., Chen, Y., Liang, L., Ding, D., et al. (2023). Structural insights into the conformational changes of BTR1/SLC4A11 in complex with PIP2. **Nature Communications** *14*, 6157. https://doi.org/10.1038/s41467-023-41924-0.

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. https://doi.org/10.1038/nature11542.

McBride, J.M., Polev, K., Abdirasulov, A., Reinharz, V., Grzybowski, B.A., and Tlusty, T. (2023). AlphaFold2 can predict single-mutation effects. **Phys Rev Lett** *131*, 218401. https://doi.org/10.1103/PhysRevLett.131.218401.

Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. **Nat Methods** *19*, 679-682. https://doi.org/10.1038/s41592-022-01488-1.

Morcos, F., Jana, B., Hwa, T., and Onuchic, J.N. (2013). Coevolutionary signals across protein lineages help capture multiple protein conformations. **Proc Natl Acad Sci U S A** *110*, 20533-20538. https://doi.org/10.1073/pnas.1315625110.

Nakamura, N., Yamauchi, T., Hiramoto, M., Yuri, M., Naito, M., Takeuchi, M., Yamanaka, K., Kita, A., Nakahara, T., Kinoyama, I., et al. (2012). Interleukin enhancer-binding factor 3/NF110 is a target of YM155, a suppressant of survivin. **Mol Cell Proteomics** *11*, M111 013243. https://doi.org/10.1074/mcp.M111.013243.

Nji, E., Gulati, A., Qureshi, A.A., Coincon, M., and Drew, D. (2019). Structural basis for the delivery of activated sialic acid into Golgi for sialyation. **Nat Struct Mol Biol** *26*, 415-423. https://doi.org/10.1038/s41594-019-0225-y.

Nyquist, M.D., Corella, A., Burns, J., Coleman, I., Gao, S., Tharakan, R., Riggan, L., Cai, C., Corey, E., Nelson, P.S., and Mostaghel, E.A. (2017a). Exploiting AR-regulated drug transport to induce sensitivity to the survivin inhibitor YM155. **Mol Cancer Res** *15*, 521-531. https://doi.org/10.1158/1541-7786.MCR-16-0315-T.

Nyquist, M.D., Prasad, B., and Mostaghel, E.A. (2017b). Harnessing solute carrier tansporters for precision oncology. **Molecules** *22*. https://doi.org/10.3390/molecules22040539.

Pak, M.A., Markhieva, K.A., Novikova, M.S., Petrov, D.S., Vorobyev, I.S., Maksimova, E.S., Kondrashov, F.A., and Ivankov, D.N. (2023). Using AlphaFold to predict the impact of single mutations on protein stability and function. **PLoS One** *18*, e0282689. https://doi.org/10.1371/journal.pone.0282689.

Parker, J.L., and Newstead, S. (2017). Structural basis of nucleotide sugar transport across the Golgi membrane. **Nature** *551*, 521-524. https://doi.org/10.1038/nature24464.

Pizzagalli, M.D., Bensimon, A., and Superti-Furga, G. (2021). A guide to plasma membrane solute carrier proteins. **FEBS J** *288*, 2784-2835. https://doi.org/10.1111/febs.15531.

Porter, L.L., Chakravarty, D., Schafer, J.W., and Chen, E.A. (2023). ColabFold predicts alternative protein structures from single sequences, coevolution unnecessary for AF-cluster. **bioRxiv**, 2023.2011.2021.567977. https://doi.org/10.1101/2023.11.21.567977.

Radestock, S., and Forrest, L.R. (2011). The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. **J Mol Biol** *407*, 698-715. https://doi.org/10.1016/j.jmb.2011.02.008.

Sala, D., Hildebrand, P.W., and Meiler, J. (2023). Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. **Front Mol Biosci** *10*, 1121962. https://doi.org/10.3389/fmolb.2023.1121962.

Saldano, T., Escobedo, N., Marchetti, J., Zea, D.J., Mac Donagh, J., Velez Rueda, A.J., Gonik, E., Garcia Melani, A., Novomisky Nechcoff, J., Salas, M.N., et al. (2022). Impact of protein conformational diversity on AlphaFold predictions. **Bioinformatics** *38*, 2742-2748. https://doi.org/10.1093/bioinformatics/btac202.

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

Sarangi, A., Bupp, K., and Roth, M.J. (2007). Identification of a retroviral receptor used by an envelope protein derived by peptide library screening. **Proc Natl Acad Sci U S A** *104*, 11032-11037. https://doi.org/10.1073/pnas.0704182104.

Schafer, J.W., and Porter, L.L. (2023). Evolutionary selection of proteins with two folds. **Nat Commun** *14*, 5478. https://doi.org/10.1038/s41467-023-41237-2.

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. https://doi.org/10.1074/jbc.M111.336446.

Stein, R.A., and McHaourab, H.S. (2022). SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. **PLOS Computational Biology** *18*, e1010483. https://doi.org/10.1371/journal.pcbi.1010483.

Stein, R.A., and McHaourab, H.S. (2024). Rosetta energy analysis of AlphaFold2 models: Point putations and conformational ensembles. **bioRxiv**. https://doi.org/10.1101/2023.09.05.556364.

Sutto, L., Marsili, S., Valencia, A., and Gervasio, F.L. (2015). From residue coevolution to protein conformational ensembles and functional dynamics. **Proc Natl Acad Sci U S A** *112*, 13567-13572. https://doi.org/10.1073/pnas.1508584112.

Toth-Petroczy, A., Palmedo, P., Ingraham, J., Hopf, T.A., Berger, B., Sander, C., and Marks, D.S. (2016). Structured states of disordered proteins from genomic sequences. **Cell** *167*, 158-170 e112. https://doi.org/10.1016/j.cell.2016.09.010.

Vehlow, C., Stehr, H., Winkelmann, M., Duarte, J.M., Petzold, L., Dinse, J., and Lappe, M. (2011). CMView: interactive contact map visualization and analysis. **Bioinformatics** *27*, 1573-1574. https://doi.org/10.1093/bioinformatics/btr163.

Wallner, B. (2023a). AFsample: improving multimer prediction with AlphaFold using massive sampling. **Bioinformatics** *39*. https://doi.org/10.1093/bioinformatics/btad573.

Wallner, B. (2023b). Improved multimer prediction using massive sampling with AlphaFold in CASP15. **PROTEINS: Structure, Function and Bioinformatics** *91*, 1734-1746. https://doi.org/10.1002/prot.26562.

Wang, N., Jiang, X., Zhang, S., Zhu, A., Yuan, Y., Xu, H., Lei, J., and Yan, C. (2021). Structural basis of human monocarboxylate transporter 1 inhibition by anti-cancer drug candidates. **Cell** *184*, 370-383.e313. https://doi.org/10.1016/j.cell.2020.11.043.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic Acids Res** *46*, W296-w303. https://doi.org/10.1093/nar/gky427.

Wayment-Steele, H.K., Ojoawo, A., Otten, R., Apitz, J.M., Pitsawong, W., Homberger, M., Ovchinnikov, S., Colwell, L., and Kern, D. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. **Nature** *625*, 832-839. https://doi.org/10.1038/s41586-023-06832-9.

Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. **Curr Protoc Bioinformatics** *54*, 5.6.1-5.6.37. https://doi.org/10.1002/cpbi.3.

West, T.J., Bi, J., Martinez-Pena, F., Curtis, E.J., Gazaniga, N.R., Mischel, P.S., and Lairson, L.L. (2023). A cell type selective YM155 prodrug targets receptor-interacting protein kinase 2 to induce brain cancer cell death. **J Am Chem Soc**. https://doi.org/10.1021/jacs.2c11715.

Winter, G.E., Radic, B., Mayor-Ruiz, C., Blomen, V.A., Trefzer, C., Kandasamy, R.K., Huber, K.V.M., Gridling, M., Chen, D., Klampfl, T., et al. (2014). The solute carrier SLC35F2 enables YM155-

mediated DNA damage toxicity. **Nat Chem Biol** *10*, 768-773. https://doi.org/10.1038/nchembio.1590.
Xie, T., and Huang, J. (2024). Can protein structure prediction methods capture alternative conformations of membrane transporters? **J Chem Inf Model** *64*, 3524-3536. https://doi.org/10.1021/acs.jcim.3c01936.
Xue, J., Xie, T., Zeng, W., Jiang, Y., and Bai, X.C. (2020). Cryo-EM structures of human ZnT8 in both outward- and inward-facing conformations. **Elife** *9*. https://doi.org/10.7554/eLife.58823.
Zemla, A., Zhou, C.E., Slezak, T., Kuczmarski, T., Rama, D., Torres, C., Sawicka, D., and Barsky, D. (2005). AS2TS system for protein structure modeling and analysis. **Nucleic Acids Res** *33*, W111-115. https://doi.org/10.1093/nar/gki457.
Zheng, S., Sham, L.T., Rubino, F.A., Brock, K.P., Robins, W.P., Mekalanos, J.J., Marks, D.S., Bernhardt, T.G., and Kruse, A.C. (2018). Structure and mutagenic analysis of the lipid II flippase MurJ from Escherichia coli. **Proc Natl Acad Sci U S A** *115*, 6709-6714. https://doi.org/10.1073/pnas.1802192115.

SUPPLEMENTARY INFORMATION

As an additional test case, we selected the *E. coli* D-galactonate proton symporter (DgoT, an SLC17 family member), a 460-residue MFS superfamily symporter that transports D-galactonate while antiporting protons. The outward- (PDB id 6e9o) and inward- (PDB id 6e9n) open conformations of DgoT have been solved by X-ray crystallography at 3.50 Å and 2.92 Å resolution, respectively (Leano *et al.*, 2019). Conventional AF2-colab calculations provided a structure with outward open conformation, in excellent agreement with the X-ray crystal structure (C_α RMSD 0.85 Å). We then used the ESM- AF2 modeling protocol (**Figure 1)** to model the inward-open conformational state, and compared the resulting model with the inward-open subunit structure determined by cryoEM. The computed inward-open conformation has excellent agreement with experimental inward-open structure, with backbone C*α* RMSD of 1.35 Å (**Supplementary Figure S1A**). We also compared residue-residue contact maps for the experimental and ESM-AF2 inward-open models with each other and with the EC-based contact map, which are also in excellent agreement (**Supplementary Figure S1B**). Finally, in order assess the use of EC-based contact maps in distinguishing and validating outward- and inward-open models, we compared contact maps for the experimentally-determined outward-open conformation, the experimentally-determined inward-open conformations, and contacts predicted by EC analysis (**Supplementary Figure S1C**). In this case, many ECs are common to both the outward- and inward-open conformations, and only a few contacts distinguish outward- from inward-open states (circled in green in **Supplementary Figure S1C**); some of these distinguishing contacts for the outward-open conformation are also observed as EC-based contact predictions (black dots in the green circled region of **Supplementary Figure S1C**). Hence, in some cases the EC contact map is not complete enough to extensively validate accurate predictions of the alternative conformational states.

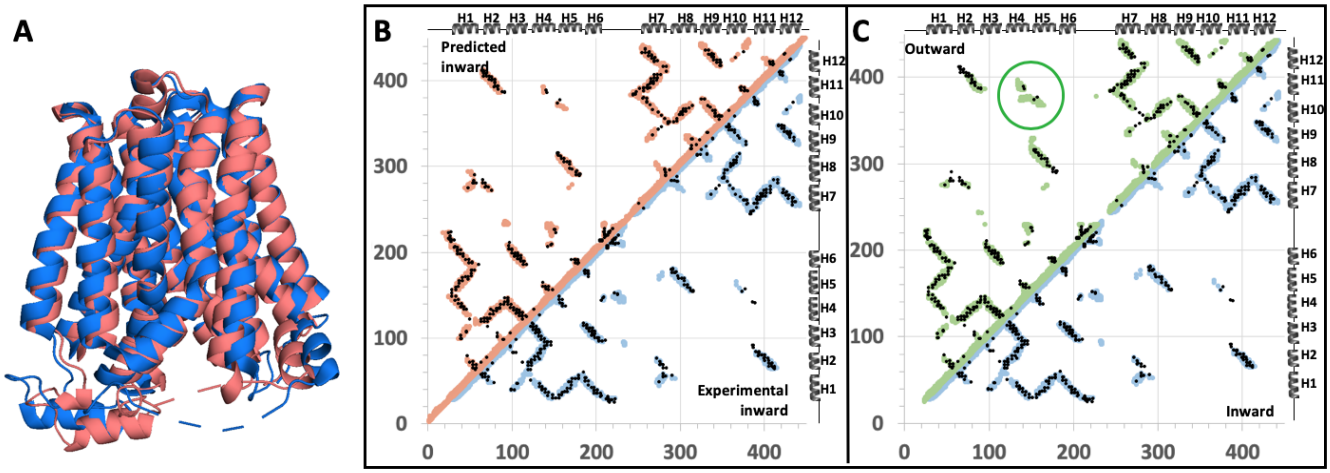

Fig. S1. Validation of ESM-AF2 protocol using SLC proteins with both outward- and inward-open experimental structures. (A) Superposition of the inward-open conformational state of *E. coli* D-galactonate:proton symporter (DgoT) modeled using the ESM-AF2 protocol from a template generated by ESMfold using a virtual flipped-sequence (red) with the experimentally-determined inward-open structure PDB id 6e9n (blue). (B) Contact maps for the inward-open conformational state of DgoT predicted by ESM-AF2 (red circles), observed in the cryoEM experimental structure (blue circles), and predicted by EC analysis (black dots). (C) Contact maps for the experimentally-determined outward-open conformational state of DgoT (green circles ; PDB id 6e9o) and experimentally-determined inward-open conformational state (blue circles; PDB id 6e9n) structures compared with contact predictions from EC analysis (black dots). In this example, only a few EC-based contacts between helices H4 and H11 and helices H5 and H10 distinguish the inward-open from outward-open structures.