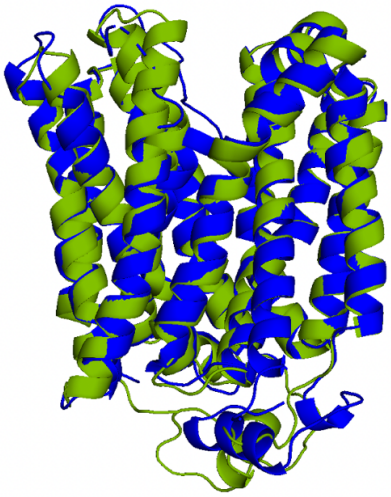
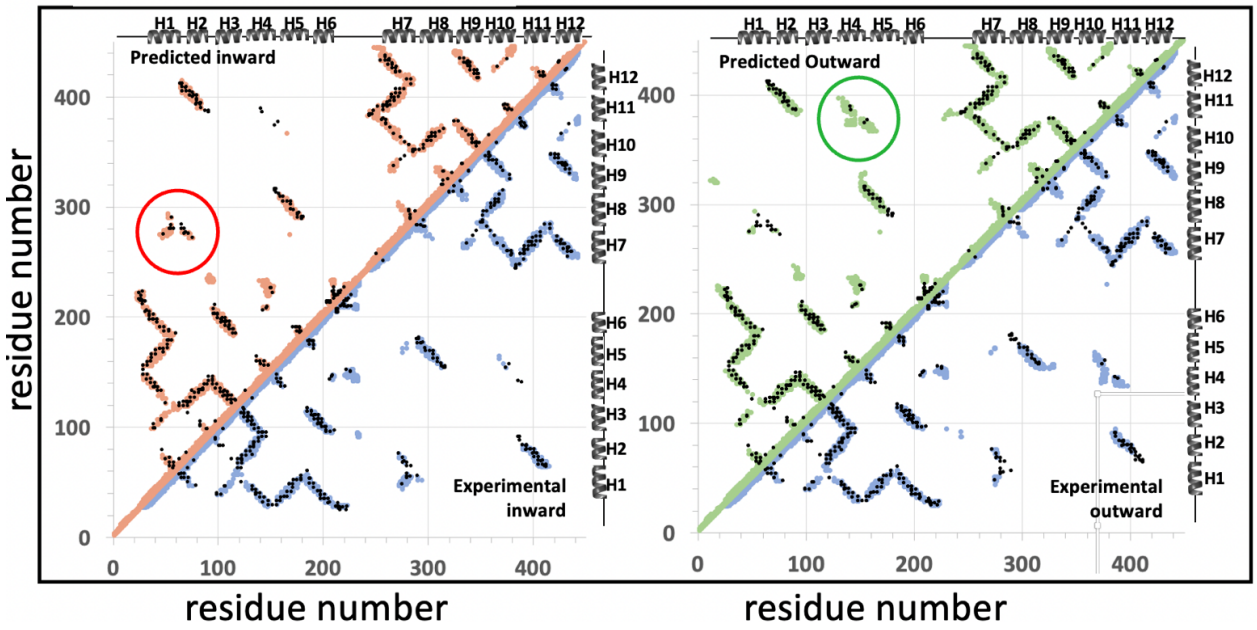
Memorization Bias Impacts Modeling of Alternative Conformational States of Symmetric Solute Carrier Membrane Proteins with Methods from Deep Learning

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.





*Corresponding authors.

G.V.T. Swapna orcid:0000-0001-9342-2525 email: swapna.gurla@gmail.com
Namita Dube orcid:0000-0002-2107-4899 email: duben@rpi.edu
Monica J. Roth orcid: 0000-0002-2219-0402 email: roth@rwjms.rutgers.edu
Gaetano T. Montelione orcid: 0000-0002-9440-3059 email: monteg3@rpi.edu

ABSTRACT
Conformational memorization impacts the performance of AlphaFold in modeling the dynamic structures of proteins. The Solute Carrier (SLC) superfamily of integral membrane proteins transport a wide array of small molecules across plasma and organelle membranes, and function as important drug transporters and as viral receptors. They populate different conformational states during the solute transport process, including outward-open, intermediate (occluded), and inward-open conformational states. For some SLC proteins this structural "flipping" corresponds to swapping between conformations of their N-terminal and C-terminal symmetry-related sub-structures. Conventional AlphaFold2, AlphaFold3, or Evolutionary Scale Modeling methods typically generate models for only one of these multiple conformational states of SLC proteins. While several AI-based protocols for modeling multiple conformational states of proteins have been described recently, these methods are often impacted by "memorization" of one of the alternative conformational states, and do not always provide both the inward- and outward-open conformations of SLC proteins. Here we assess the impact of memorization in modeling SLC proteins with AlphaFold, and describe a combined ESM – template-based-modeling process, based on a previously described template-based modeling method that relies on the internal pseudo-symmetry of many SLC proteins, to consistently model the alternate conformational states of SLC proteins. We also demonstrate how the resulting multi-state models can be validated experimentally 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. This simple, rapid, and robust 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.

Abbreviations: AF2 – AlphaFold2 Multimer; AF3 – AlphaFold3, 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-Distance Difference Test, a confidence score predicted from ML, TM – Template Modeling score to assess similarity between two protein structures.

INTRODUCTION

Proteins adopt multiple conformational states which are essential to their functions. While AlphaFold2/3 (AF2/3) (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 proteins, for systems that adopt multiple conformational states conventional AF2/3 and ESM calculations often deliver only one of the multiple states observed experimentally (Huang *et al.*, 2021; Del Alamo *et al.*, 2022; Saldano *et al.*, 2022; Bryant and Noé, 2024; Chakravarty *et al.*, 2024; Kalakoti and Wallner, 2024; Lazou *et al.*, 2024; Wayment-Steele *et al.*, 2024; Xie and Huang, 2024). Recent advances have been reported using modified AF2 protocols and "enhanced sampling" methods to model multiple conformational states of proteins, including integral membrane proteins (Sala *et al.*, 2023a). 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) chosen either randomly (AlphaFold-alt) (Del Alamo *et al.*, 2022; Sala *et al.*, 2023b) or by clustering homologous protein sequences (AF-cluster) (Wayment-Steele *et al.*, 2024), (iii) very shallow MSAs and even single protein sequences (Porter *et al.*, 2023; Chakravarty *et al.*, 2024) that allow knowledge inherent to the AI to dominate the modeling process, or (iv) using MSAs masked at multiple positions, as implemented in both (SPEACH-AF) (Stein and McHaourab, 2022) and AF-sample2 (Kalakoti and Wallner, 2024), to bias the prediction toward alternative conformational states. AF2 calculations using network dropouts (AF-sample) can also generate conformational diversity (Johansson-Akhe and Wallner, 2022; Wallner, 2023a; b; Benavides and Montelione, 2024; Huang and Montelione, 2024). Despite these advances, challenges remain in reliably modeling alternative conformational states that are observed experimentally. In particular, assessments report various enhanced sampling methods to be successful in modeling multiple conformational states for 50% (or less) of experimentally-available alternative conformer pairs (Bryant and Noé, 2024; Chakravarty *et al.*, 2024; Lazou *et al.*, 2024). This observation suggests that at least some cases of successful modeling result from some kind of memorization by the AI, rather than its inherent "learning" of protein structure principles (Chakravarty *et al.*, 2024; Lazou *et al.*, 2024). Such memorization can bias the AI-based modeling towards conformational state(s) used in the training process and prevent accurate modeling of alternative conformational states. To the degree that memorization biases the successful prediction of alternative conformational states, more robust methods leveraging the tools of AI-based modeling are required.

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. SLC proteins can also function as receptors for viral entry into the cell (Sarangi *et al.*, 2007). 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).

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 fold but non-identical conformations, enabling the protein to adopt multiple conformational states essential for its function. MFS-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 a 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 states (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). These "inward-open" and "outward-open" conformational states are sometimes called inward-facing and outward-facing states in the literature. Crystal structures have been solved for inward-open, occluded, and outward-open states of several 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. 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).

Computational methods significantly impact our understanding of SLC protein structure-function relationships and can guide experimental design. However, as they are medium-sized integral membrane proteins, molecular dynamics simulations are quite challenging, requiring powerful computing resources, accurate potential energy functions, and appropriate simulation of membrane-mimicking environments. The evolving AI-based enhanced sampling methods outlined above can sometimes provide models of multiple conformational states of SLC proteins, but they are not always successful (Del Alamo *et al.*, 2022; Saldano *et al.*, 2022; Bryant and Noé, 2024; Chakravarty *et al.*, 2024; Kalakoti and Wallner, 2024; Wayment-Steele *et al.*, 2024; Xie and Huang, 2024). These observations suggest the need for more robust methods for addressing this important class of membrane protein transporters.

Importantly, multiple conformational state modeling of proteins can be guided by evolutionary covariance (EC) analysis of functionally-preserved direct contacts, which can 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). Of special significance for SLC proteins, is their unique pseudo-symmetrical transport mechanisms, which provides the basis for classical method of modeling the inward-open (or outward-open) conformations of some SLC proteins from knowledge of their outward (or inward) open conformations by swapping the pseudo-symmetric structures of the N- and C-terminal halves, and then using the resulting virtual structure as a template to model the alternative conformational state (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). Although it is logical to combine the two concepts of EC-based contact information with swapping of pseudo-symmetric structures, it has not yet been implemented as a general strategy for modeling SLC proteins.

Here we describe a simple and robust approach for modeling alternative conformational states of pseudo-symmetric SLC proteins using a combined ESM – template-based-modeling process inspired by the methods of Forrest and others (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). In this approach, templates for alternative conformational states are generated from a "flipped virtual

sequence" using ESMFold (Lin *et al.*, 2023), and template-based modeling is then performed using either AF2 (Jumper *et al.*, 2021) or, where training bias impacts the AF2 structure prediction, with the template-based modeling software MODELLER (Webb and Sali, 2016). First, an 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 against EC-based contact maps and by comparison against atomic coordinates determined by cryoEM or X-ray crystallography. For two additional SLC proteins, *Zea mays* CMP-sialic acid transporter 1 (SLC35A1) and *Saccharomyces cerevisiae* GDP-mannose sugar transporter 1 (SLC35D subfamily), only the outward-open forms are available as experimental structures. In these cases, AF modeling was found to be biased towards these states, and the alternative inward-open forms were modeled with an ESM-MODELLER process, and then validated by comparison against EC-based contact maps. For SLC35F2, although neither inward nor outward-open experimental structures are available, the outward-open is strongly preferred when using conventional AF2. However, the inward-open conformational state could be modeled using either the ESM– AF2 or ESM – MODELLER processes. Both the inward- and outward-open structures were validated against EC-based contact maps. For other SLC proteins where experimental structures are available for only one conformational state, significant bias towards one state or the other was also observed using AF2. In these cases, the ESM-MODELLER approach was successful in modeling both inward and outward-open states, which were validated by comparisons against EC-based contact maps.

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* (Hopf *et al.*, 2012) 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, AlphaFold3, ESMfold, and MODELLER modeling. AlphaFold2 (AF2) (Jumper *et al.*, 2021) modeling was performed using ColabFold v1.5.5 server (Mirdita *et al.*, 2022) with *AlphaFold2.ipynb* scripts. The standard AF2 modeling in this study used no templates (unless specified), default multiple sequence alignments (MSAs), recycle of 12, and random dropouts, although other protocols were also assessed. The Amber-relaxed top-ranked model was taken as the final predicted structure. AlphaFold3 (AF3) modeling was performed using the Google Deep Mind / Isomorphic Lab server (https://alphafoldserver.com/welcome) (Abramson *et al.*, 2024) with no structural templates. 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 (masking_rate = 0.15), stochastic_mode="LM" no dropout), and recycle of 12. The model with the highest 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. For each

run, 20 models were generated, and the one with the lowest DOPE (Discrete Optimized Protein Energy score) was selected as the representative structure.

AlphaFold-alt. Enhanced sampling using shallow MSAs 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 with four A100 Nvidia HGX GPUs. In each AF-alt run, 480 models were generated using randomly sampled shallow MSAs (16-32 sequences), with 30 models created per MSA depth. Each run took < 3 hrs. For each model, disordered N- and C-terminal regions were removed, and the average pLDDT (<pLDDT>) was computed for the remaining residues.

AF_Sample and AF_Sample2. Massive sampling was carried out using *AF_Sample* and *AF_Sample2* (Johansson-Akhe and Wallner, 2022; Wallner, 2023a; Kalakoti and Wallner, 2024), executed on a local cluster of four A100 Nvidia HGX GPU processors, following protocols described elsewhere (Benavides and Montelione, 2024). *AF_Sample* inferences used *AF-Multimer* model weights v2.1.2, v2.2.0, and v2.3.2,with no templates. Runs with *v2.1.2* used 21 max_recycles, *v2.2.0* used the default of 3, and *v2.3.2* used 9. *AF_Sample2* inferences used the same model weight variations but were run with 3 max_recycles and no templates.

Hydrogen atoms were added to files generated by *AF_Sample, AF_Sample2,* and *AF_Alt* using a custom script which employs the Amber force field, analogous to the method employed by the original AF2 manuscript (Jumper et al, 2021). These scripts are provided at https://github.rpi.edu/RPIBioinformatics/FilteringAF2_scripts. Each of these enhanced sampling methods can be quite aggressive in generating conformational diversity and also models that are not physically reasonable: e.g. incorrect amino acid chirality, non-native cis peptide bonds, and other biophysically incorrect features, particularly in the not-well-packed residue segments of the modeled proteins. The most egregious of these physically unreasonable models were identified and removed. The resulting relaxed models were used for further analysis.

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

The challenge we encountered arises from the fact that conventional AF modeling generally provides only one of the multiple conformations of SLC proteins when only one of these states was available as an experimental structure at the time of training. Even enhanced sampling methods successfully generate alternative conformational states for only some multistate proteins (Del Alamo *et al.*, 2022; Bryant and Noé, 2024; Chakravarty *et al.*, 2024; Kalakoti and Wallner, 2024; Lazou *et al.*, 2024; Wayment-Steele *et al.*, 2024; Xie and Huang, 2024). These observations motivate the need for robust and consistent methods for modeling alternative conformational states (outward-open vs inward-open) of SLC proteins, at the very least for use as reference states for assessing the evolving deep learning methods for generating alternative conformational states of proteins.

Bias of AF2 in modeling alternative conformational states. AF2 bias in modeling alternative conformational states is documented for several SLC proteins used in this study in **Supplementary Table**

S1. In the first two systems (DgoT and ZnT8), both inward- and outward-open states were available in the PDB at the time of AF2 training. AF2 with full MSAs (with or without dropouts) is biased towards predicting only the inward-open state. AF2 inference with a single sequence consistently fails to generate a reasonable model. However, using shallow MSAs (4 to 16 sequences) alternative states similar to the experimentally-determined alternative state conformations are delivered, but only as a small fraction of the generated models. For the next 5 systems listed in **Supplementary Table S1**, only one state (inward- or outward-open) was available in the PDB at the time of AF2 training. For this set, AF2 with full MSAs (with or without dropouts) is biased towards predicting only the state available for training (in 4 cases the outward-open state, in 1 case the inward-open state). Again, AF2 inference with only a single sequence consistently fails. Using shallow MSAs the alternative state is delivered as one or more of the generated models in only one of the 5 cases; i.e. for SLC19A1. For the last 3 cases summarized in **Supplementary Table S1**, no homologous structures were available in the PDB at the time of training. AF2 with full MSAs (with or without dropouts) again delivers a single dominant state. AF2 inference with only a single sequence consistently fails. Using shallow MSAs (4 to 16 sequences) the alternative state is delivered as at least one of the generated models for SLC19A2, but only the outward-open state is generated for SLC35F2 or SLC35F3. These results document the challenges AF2 faces in modeling alternative conformational states when one or more state is potentially available in the training data. Overall it was much harder to generate alternative conformational states (i.e. both inward- and outward-open states) for these SLC proteins using shallow MSAs than we expected from published studies.

ESM-AF2/MODELLER protocol. To address these challenges, we reasoned that it might be possible to generate structural templates using alternative deep learning methods, and use these templates to guide the modeling process along the lines that have been demonstrated so successfully by Forrest et al (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). The ESM-AF2/ ESM-MODELLER 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, application of this method using conventional modeling methods can be challenging if the sequence similarity in these two halves of the protein sequence is low, making it difficult to determine the correct alignment for template-based modeling. In the ESM-AF2/MODELLER process, the N-terminal (blue) and C-terminal (purple) segments of protein sequences are first swapped to create a *virtual flipped sequence*. Note that this flipped sequence has no homologs with which to generate a multiple-sequence alignment. The 3D structure of this virtual sequence is then modeled using *ESMfold*, a large-language model-based method that requires no templates and only a single input sequence. The resulting *virtual structure* serves as a structural template for modeling the original protein sequence using template-based modeling with *AF2* (if no state-specific bias is observed) or with MODELLER.

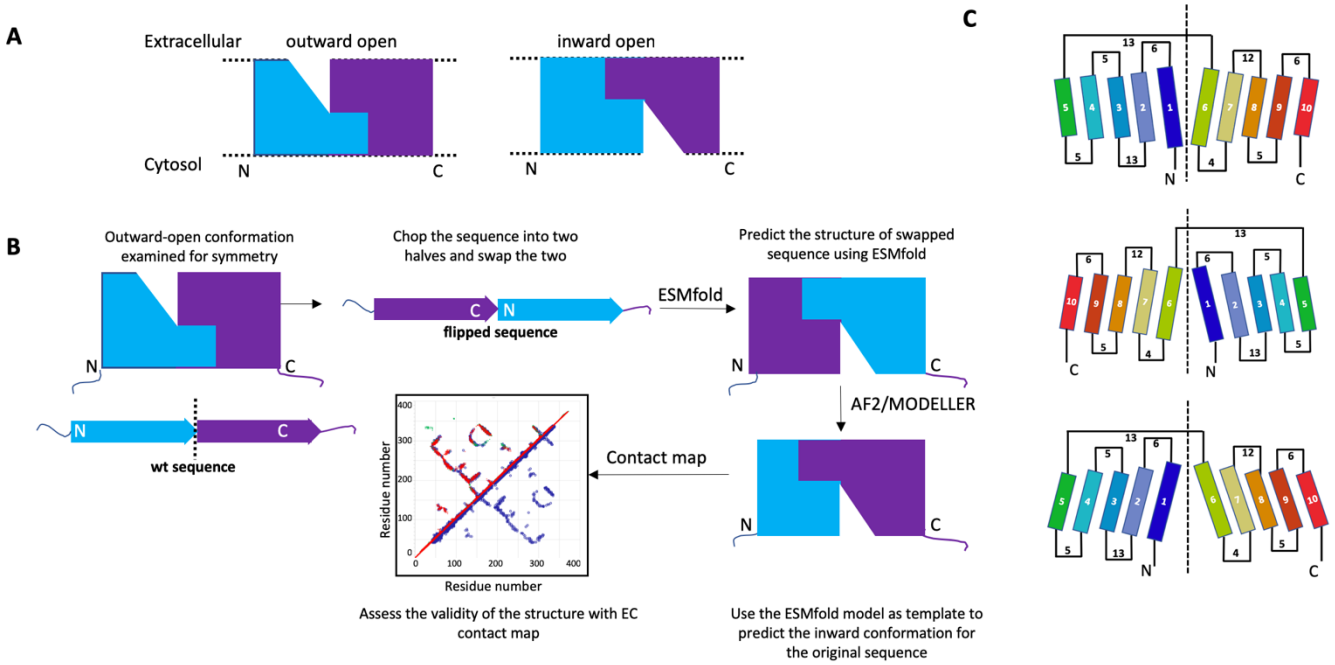


Fig. 1. The ESM-AF2/MODELLER protocol for modeling alternative conformational states of pseudo-symmetric SLC proteins. (A) cartoon representation of inward/outward-open conformers representing the pseudo-symmetry of the helices, with pseudo-symmetry halves indicated in blue and purple. (B) Protocol to 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-open state, the middle image is the *ESMfold* virtual protein structure generated from a *virtual flipped protein sequence*, and the bottom image the inward-open state generated by comparative modeling using the virtual protein structure as a modeling template.

In this protocol, *ESMfold* is used to model a *virtual template structure*. **Supplementary Table S2** summarizes tests carried out to compare the effectiveness of *ESMFold* and *AF2* in generating these *virtual template structures*. In all cases, when using single sequences as input *ESMfold* provided a structural template with backbone structure matching the expected alternative conformational state, while AF2 was not able to generate reasonable structures for any of the *virtual flipped sequence*. AF2 could sometimes successfully generate a virtual structure with the alternative conformation when using shallow MSAs as input; however, this is not useful for our protocol which uses a single *virtual flipped sequence*. In our experience *AF2* is less consistent and robust than *ESMfold* in generating a good quality structural template from the *virtual flipped sequence*.

Validating the ESM-AF2 modeling protocol. As an initial test case of the ESM-AF2 method for modeling alternative conformational states of SLC proteins, we selected human ZnT8 (SLC30A8), a 2 x 320-residue homodimeric integral membrane Zn-transporter, for which structures have been 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. The contact maps for the inward- and outward-open states demonstrate that key differences involve interactions between helices H1, H2 and H3 with helices H5 and H6 (**Figure 2**). Conventional AF2-colab calculations using the standard protocol outlined in the Methods section provided a structure with the inward-open conformation, matching 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 generate the outward-open conformation and compared it with the experimentally determined outward-open cryoEM structure. The computed outward-open ZnT8 model showed excellent agreement with the experimental 6xpf-B (Cα RMSD = 1.09 Å) (**Figure 2B**). We compared residue-residue contact maps for the experimental and ESM-AF2 outward-open models against each other and with an EC-based contact map derived from multiple-sequence alignments of ZnT8 homologs (**Figure 2C,D**). The AF2-modeled inward-open structure, has a contact map that is nearly identical to that of the experimental inward-open structure (**Figure 2C**); the outward-open structure computed using the ESM-AF2 protocol is also essentially identical to the experimental outward-open 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, aligning precisely with the corresponding contacts in the computed models (circled in **Figures 2C,D**). Hence, the ESM-AF2 protocol successfully modeled both conformations of Znt8, as validated by comparison with experimental EC-derived contacts. A second test case for the ESM-AF2 modeling protocol using an SLC protein with both inward and outward-open experimental structures is presented for the *E. coli* D-galactonate:proton symporter (DgoT) in **Supplementary Figure S1**. Again, the ESM-AF2 protocol successfully modeled both inward- and outward-open states, consistent with experimental EC-derived contacts.

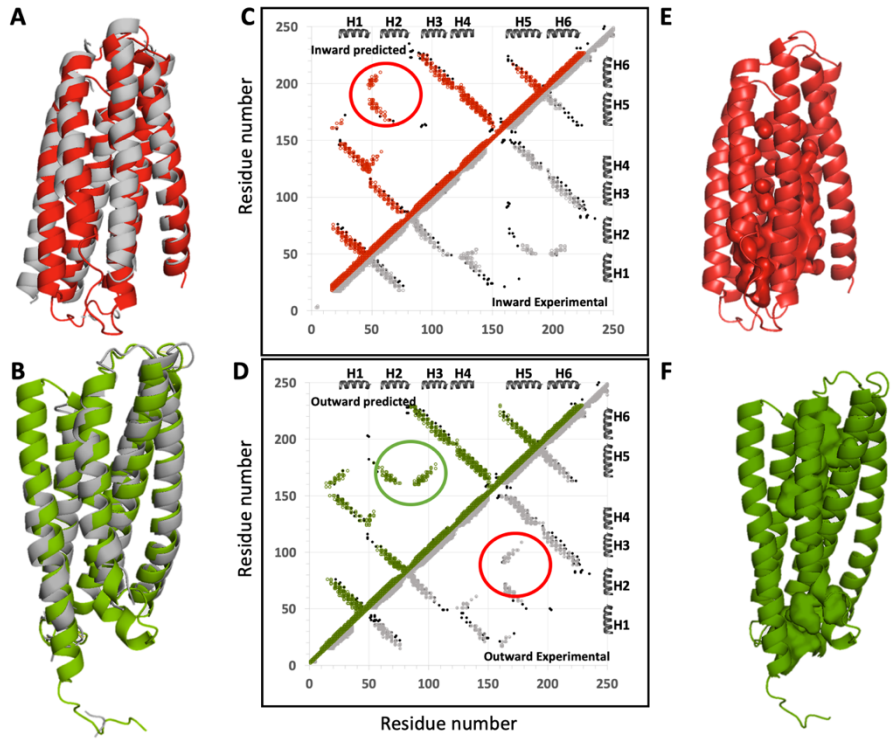


Fig. 2. Validation of ESM-AF2 protocol using an SLC protein with both outward- and inward-open experimental structures. The cryo-EM structure of human ZnT8 WT in the absence of zinc has two chains, with one subunit in an inward-open conformation and the other in an outward-open conformation (PDB id: 6xpf chain A and B respectively). (A) Superposition of the AF2-predicted (red) and experimental (grey) inward-open structures (Cα RMSD = 2.00 Å. (B) Superposition of the ESM-AF2 outward-open model (green) with the experimental structure (grey) (Cα RMSD = 1.09 Å). (C) Comparison of the EC-based contact map of ZnT8 (black points) with contacts in the experimental (grey) and predicted (red) inward-open models. (D) Comparison of the EC-based contact map of ZnT8 (black) with contacts in the experimental (grey) and predicted (green) outward-open 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. Surface pockets for (E) inward-open and (F) outward-open states are represented as space-filled voids using the server https://kvfinder-web.cnpem.br/

Modeling alternative conformations of SLC proteins when a single experimental state is avaiable. In the two cases above, we chose SLC 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 Cα RMSD metrics) and against EC based contact maps, which are based on experimental primary sequence data. However, for most SLC proteins, experimental

structures are only available for one (or neither) 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 322-residue *Zea mays* CMP-sialic acid transporter 1 SLC35A1 [PDB id 6i1r-A (Nji *et al*., 2019)], a SLC35A subfamily member and in **Figure 4** for the 337-residue *Saccharomyces cerevisiae* GDP-mannose sugar transporter 1 Vrg4 (PDB id 5oge (Parker and Newstead, 2017)), an SLC35D subfamily member. For both proteins, only outward-open X-ray crystal structures determined at 3.22 Å and 2.80 Å resolution, respectively, are available. In both of these cases, the ESM-AF2 protocol was not successful in providing models of the inward-open state that could be validated by patterns in the EC-based contact map unique to each conformer. However, using the ESM-MODELLER protocol, in which the outward-open state is modeled with AF2, and the inward-open state is modeled using a "flipped-sequence" as input to *ESMfold*, providing a virtual template that is then used with a conventional template-based modeling approaches, both outward- and inward-open states were generated. In both cases, the EC-based contact maps could be largely explained by the combined contact maps of these 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 validate the ESM-MODELLER process for cases where, due to the impact of memorization of conformational states available at the time of training on the AF inference, the ESM-AF2 method fails.

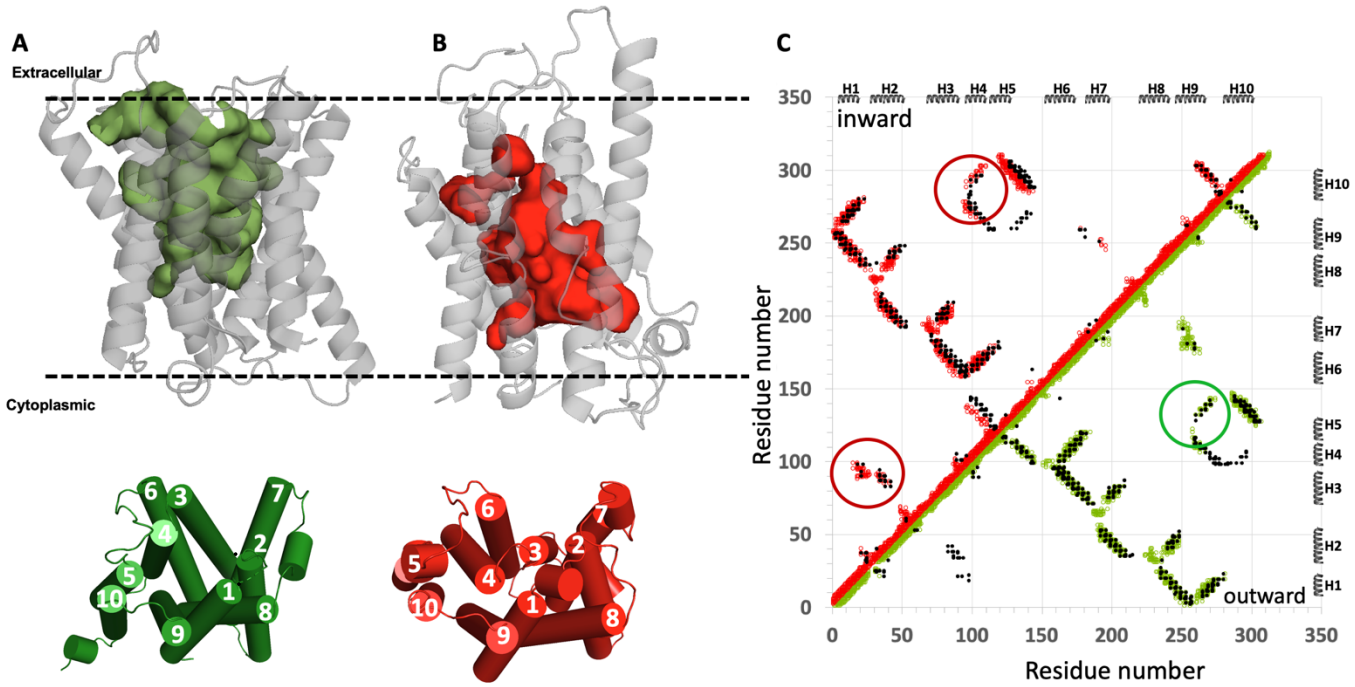


Fig. 3. ESM-MODELLER modeling of the inward-open conformation of the *Zea mays CMP-sialic acid transporter 1*. (A) The experimental outward-open structure (PDB id 6i1r-A). (B) The inward-open structure modeled using ESM-MODELLER. In each of panels A and B the top images are ribbon representations of 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 the 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/.

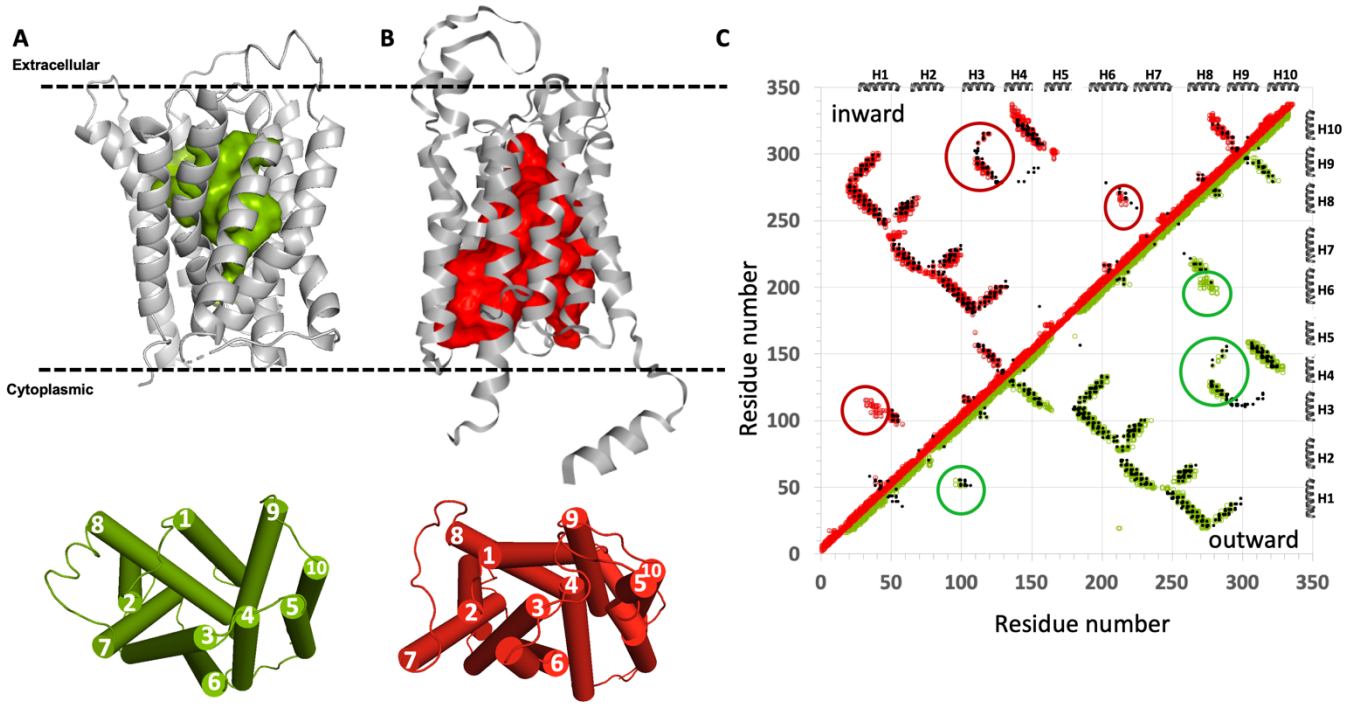


Fig. 4. ESM-MODELLER modeling of the inward-open conformation of the *S. cerevisiae* GDP-mannose sugar transporter 1, Vrg4. (A) The experimental outward-open structure (PDB id 5oge). (B) The inward-open structure modeled using ESM-AF2. In each of panels A and B the top images are ribbon representations 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. Of particular interest are SLC proteins for which no experimental structures are available for either the inward- or outward-open states. 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- or outward-open conformations. Conventional AF2 modeling was carried out using the AF2-multimer colab server (Mirdita *et al.*, 2022) executed both with the standard protocol without structural templates described in the Methods section and also with various other protocols using templates of distant homologues and multiple seeds. Modeling was also attempted using AF3 (Abramson *et al.*, 2024) which does not support template-directed modeling. Only the outward-open conformational state of SLC35 was returned by AF3. Hence, even without a state-specific structure in their training sets, AF2 (and AF3) are biased towards the outward-open state of SLC35F2.

For SLC35F2 we also explored using various protocols with shallow MSAs, dropouts, and the combination of dropouts with MSA masking to generate alternative conformational states. *AF-alt* was used to generate 480 models, and *AF_Sample* and *AF_Sample2* were used to generate 3,000 models each. These enhanced sampling methods are very GPU intensive and require long run times. For this particular protein, for which no experimental structures were available in the PDB at the time of AF2 training, all three of these methods generated exclusively outward-open states (**Supplementary Figure S2A-C**). These results for SLC35F2 illustrate the common case where even enhanced sampling methods fail to generate reliable models of multiple alternative conformational states. Interestingly, when *AF-Sample* was run on virtual flipped sequence of SLC35F2, exclusively inward-open conformational states for the flipped sequence were generated.

Having established the reliability, consistency, and limitations of the ESM-AF2 protocol, AF2 was used to model the outward-open conformation of SLC35F2, and both ESM-AF2 and ESM-Modeller were used to model its inward-open conformation (**Figure 5**). For EMS-AF2, the top-ranked model was outward-open, but other top-scoring models were inward-open. The contact maps of the resulting inward-open models generated by the two methods were then compared with their EC-based contact maps. The ESM-AF2 inward-open structure explains a few more EC-based contacts than the ESM-Modeller protocol, particularly for predicted contacts between helices H6 and H9 (*cf.* contact maps of **Figures 5D and E**). 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 variability of SLC35F2.

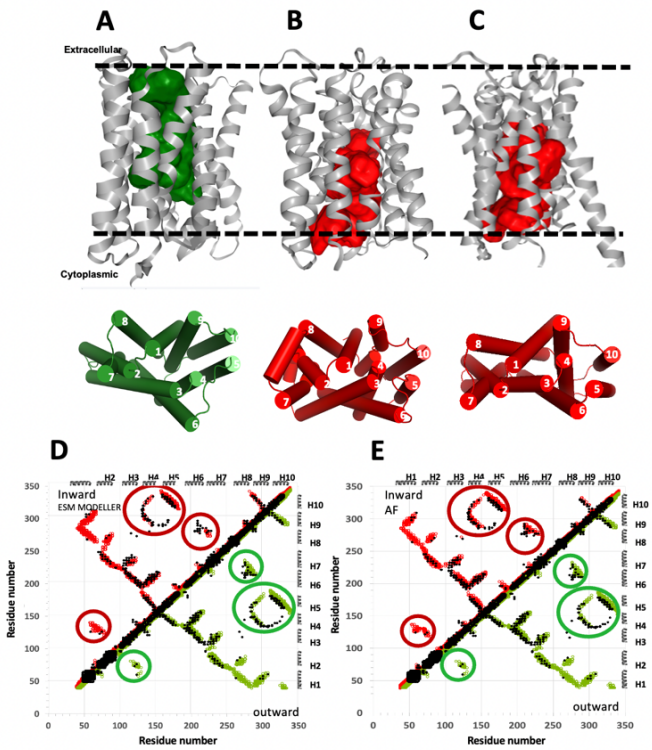


Fig. 5. Modeling of the outward- and inward-open conformations of human SLC35F2. (A) The outward-open structure modeled with AF2. (B) The inward-open structure modeled using ESM-MODELLER. (C) The inward-open structure modeled using ESM-AF2. In each of panels A-C the top images are ribbon representations 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. (D) Contact maps of inward-open (red circles) and outward-open (green circles) models of panels A and B, and (E) contact maps of inward-open (red circles) and outward-open (green circles) models of panel A and C, superimposed on the EC contact map (black dots). In panels A, B and C, surface pockets are represented as space-filled voids using the server https://kvfinder-web.cnpem.br/.

Modeling alternative conformational states of other SLC proteins. We selected 4 additional SLC proteins for modeling with the ESM-AF2 and ESM-MODELLER protocol. These results are summarized in **Supplementary Figures S3 – S6**. In all of these cases for which structure of one conformational state were available in the PDB at the time of AF2 training, bias toward this state was observed when using AF2 alone or even when using AF2 with a template for the alternative state generated with ESM using a flipped sequence; i.e. the ESM-AF2

protocols described here fail to generate the alternative conformational state when one conformational state was available in the PDB at the time of AF2 training. However, the ESM-MODELLER protocol, which avoids the bias of conformational state modeling due to "memorization" often observed using AF2, provided models of both inward-open and outward-open states, with excellent agreement (< 1 - 2 Å rmsd) to experimental models where available, and in concordance with EC-predicted contact maps.

In carrying out the studies described above, we also assessed an array of protocols using the *ESMfold* models generated from a virtual flipped sequence as a template for modeling of the alternative conformational state followed by either AF2 or conventional template-based modeling. In this process, a shallow MSA was used so that the template structural information dominates the modeling process. AF2 modeling was done using single-sequence inference, and also with shallow MSAs (8, 16, or 32), recycle of 12, and with dropout. All 5 top-scoring models were assessed for representatives of the alternative conformational state. The original (e.g., outward-open) and final (e.g., inward-open) structures were validated by comparison against the EC-based contact map that will generally include predicted contacts for both conformational states. While the ESM-AF2 protocol could sometimes model the alternative conformational state, it was not successful in all cases (**Supplementary Table S3**). In all of these cases, the template-based modeling step of **Fig. 1,** with the virtual flipped ESMfold structure as a template, could be performed successfully using MODELLER (Sali and Blundell, 1993; Webb and Sali, 2016). Template-based modeling could also be done using SwissModel (Waterhouse *et al.*, 2018) or other template-based modeling methods.

DISCUSSION

We were very surprised to observe significant weaknesses of various published protocols using AF2 for modeling alternative conformations of pseudo-symmetric SLC transporters. However, where conventional AF2 modeling (or even AF2 modeling with enhanced sampling) provides only one (either inward- or outward-open) conformational state; the alternative state can then be modeled by the templated-based 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 to generate a structural template for the alternative state, which can be quite difficult to generate. In the ESM-AF2 (or ESM-MODELLER) approach, we use *ESMfold* to generate from a virtual flipped sequence a virtual protein structure, which is then used as a structure modeling template. Memorization bias does not significantly impact this process. This allowed us to reliably model alternative conformational states of several SLC transporters that were difficult to model using the traditional approach. Importantly, 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 approach is simple to implement and runs fast using publicly-available servers. However, despite the successful examples demonstrated in this study, the ESM-AF2 protocol for modeling alternative conformational states of pseudo-symmetric SLC proteins has some limitations. In particular, where structures of only one of the alternative states was available in the PDB at the time of AF2 training, a significant bias towards this state was observed when AF2 was used either directly or as part of the ESM-AF2 modeling process. Although this bias is overcome using the ESM-MODELLER protocol, is it

somewhat disappointing to have to sometimes resort to older template-based modeling methods in place of AI-based methods like AF2. Another shortcoming is that neither protocol can 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 SLC sequence families for which only shallow MSAs are available. While 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/MODELLER protocols sometimes also generate such occluded states, these states were not explored in this study.

The ability of AF2 to model protein structures not included in its training has been demonstrated in various CASP blind assessments (Kryshtafovych *et al.*, 2021; 2023). Conventional AF2 was also reported to be successful in accurate modeling of protein structures determined by NMR methods which were not included in its training data, and for which no structures of homologous proteins were available at the time of training (Li *et al.*, 2023). However, for proteins adopting multiple conformational states, AI training carried out with data that includes only one state can bias the predictor and limit its ability to model the alternative state.

Recently Porter and co-workers have demonstrated that, at least for fold-flipping proteins which have significant structural differences between conformational states, AF2 modeling with enhanced sampling is often biased toward the conformational state reported in the PDB and potentially used in the AF2 training, and often is not able to predict conformational states not represented in the AF2 training data. Combining >280,000 models from several implementations of AF2 and AF3, only a 35% success rate was achieved in modeling alternative states of fold switchers for which one state was available for AF training (Chakravarty *et al.*, 2024; Chakravarty *et al.*, 2025). In a related study of cryptic sites in proteins, Lazou al were able to use AF2 to generate both open and closed conformations for only 6 of 16 proteins studied (Lazou *et al.*, 2024), attributing this low success rate to bias due to training memorization. Bryant and Noé also have explored this question by training a structure prediction network, *Cfold*, on a conformational split of the PDB that excludes alternative conformations for protein structure pairs solved in two conformational states. While > 50% of experimentally-known nonredundant alternative protein conformations evaluated were predicted with high accuracy (TM-score > 0.8), for the remaining pairs *Cfold* failed to correctly model the alternative conformational state that was not included in the training data (Bryant and Noé, 2024) and was biased toward the conformational state used in training. These results indicate that while in some cases, the network has learned enough to model alternative conformational states not included in the training data (Roney and Ovchinnikov, 2022), in other cases success may in fact rely on some kind of memorization; i.e. both factors can be at play. It has been suggested that *ESMfold* may be less sensitive to this memorization bias (Xie and Huang, 2024). Consistent with these observations, we also observed a bias toward previously reported conformational states when modeling with either AF2 or with the ESM-AF2 modeling protocol outlined here. This bias was not suppressed by using single-sequences or very shallow MSAs (8-16 sequences) in the AF2 modeling. Nor was it overcome in the cases tested using enhanced sampling with AFSample (Wallner, 2023a) or AFSample2 (Kalakoti and Wallner, 2024). For these SLC proteins, this bias is overcome using the ESM-MODELLER protocol. However, where no memorization bias is involved, the ESM-AF2 protocol is preferable as template-guided AF2 has more accurate properties than conventional template-

based modeling method. More significantly, the successful generation of multistate models of SCL35F2 using ESM-AF2, despite the absence of structural templates, demonstrates the suppression of conformational bias and suggests a general protocol using a retrained AF network that excludes homologous structures from training data to improve inference.

While previous studies have demonstrated that training memorization bias impacts AF2 modeling of alternative states of "fold flip" proteins that have significant structural differences between the states (Chakravarty *et al.*, 2024), the *Cfold* study cited above, using a structure prediction network trained on a split version of the PDB excluding alternative conformational states of multistate proteins (Bryant and Noé, 2024) demonstrated less impact of memorization and high success rates in modeling distinct conformational states with smaller structural differences. The SLC proteins studied here have very similar overall structures and contact maps for the two states, yet the bias from conformational states available for the training process still strongly impacts the reliability of alternative conformational modeling by AF2.

Conclusions. In this work we document bias in modeling multiple conformational states of SLC proteins that challenges the view that multiple conformational state modeling of this important class of integral membrane proteins is a largely solved problem. We describe, validate, and compare hybrid ESM-AF2 and ESM-MODELLER protocols for modeling alternative conformational states of pseudo-symmetric SLC proteins. The approach overcomes one shortcoming of conventional AF2 structure calculations which generally provide only one of the multiple conformational states observed experimentally. We observed that while AF2 generally does an excellent job of modeling one of the conformational states, there is a significant bias of AF towards conformational states available in the PDB at the time of its training. This bias can be overcome using the ESM-AF2 or ESM-MODELLER protocols. 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. The method is simple to use, rapid to run, and can be implemented using public domain servers. Overall, the current study validates the ESM-AF2/MODELLER protocol for modeling conformational heterogeneity of pseudo-symmetric SLC transporters, one of the most extensive class of transporters in the human proteome.

RESOURCE AVAILABILITY

Lead contact

Corresponding authors are Monica J. Roth and Gaetano T. Montelione. Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Gaetano T. Montelione (monteg3@rpi.edu).

Materials availability

This study did not generate new unique reagents or materials.

Data and code availability

All scripts and key data generated in this study are available at https://github.rpi.edu/RPIBioinformatics/SLCModeling.

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.

Star Methods

Key Resources – Star Methods Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Deposited data includes: 1) Data obtained using *ESMfold* modeling using the flipped-sequence as input, for all the proteins reported in this manuscript 2) Data generated using *AlphaFold2* with wild-type sequences 3) Data generated for wild-type sequence using *AlphaFold2* with *ESMfold* model of alternate conformer as template 4) Data generated for wild-type sequence using *MODELLER* with *ESMfold* model of alternate conformer as template 5) Data generated for wild-type sequence using *AFsample* and *AFsample2.*	This paper	Key data generated in this study are available at https://github.rpi.edu/RPIBioinformatics/SLCModeling
Software and Algorithms		
AlphaFold2 – Colab Server ver 1.5.5	(Mirdita *et al.*, 2022)	https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb?authuser=1#scrollTo=kOblAo-xetgx
AlphaFold3	(Abramson *et al.*, 2024)	https://alphafoldserver.com/welcome
AlphaFold_Sample	(Wallner, 2023b; a)	http://wallnerlab.org/AFsample.
AlphaFold_Sample2	(Kalakoti and Wallner, 2024)	https://wallnerlab.org/AFsample2
ESM_Fold server ver XX	(Lin *et al.*, 2023)	https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/ESMFold_advanced.ipynb
EVcouplings server ver XX	(Hopf *et al.*, 2019)	https://evcouplings.org/
MODELLER ver 10.4	(Webb and Sali, 2016)	https://salilab.org/modeller/
NeBcon (Neural-network and Bayes-classifier based contact prediction)	(He *et al.*, 2017)	https://seq2fun.dcmb.med.umich.edu//NeBcon/
PyMOL Molecular Graphics System PyMOL (Schrödinger, LLC. (2020). *The PyMOL Molecular Graphics System, Version 2.4*)	(Schrödinger and DeLano, 2020)	https://github.com/schrodinger/pymol-open-source

CMView 1.1.1 Protein contact map visualization and analysis	(Vehlow *et al.*, 2011)	https://www.bioinformatics.org/cmview/download.html

SUPPLEMENTARY INFORMATION

Table S1. Assessment of modeling alternative states of SLC proteins with AF2.
Table S2. Assessment of successful generation of alternative states of SLC proteins from flipped sequence using *ESMfold* vs *AlphaFold2*.
Table S3. Modeling of both outward-open and inward-open states of pseudo-symmetric SLC proteins.

Fig. S1. Validation of ESM-AF2 protocol using an SLC protein with both outward- and inward-open experimental structures.
Fig. S2. Conformational states of SLC35F2 modeled using various enhanced or massive sampling protocols.
Fig. S3. Thiamine transporter 1 (SLC19A2).
Fig. S4. Aromatic amino acid exporter YddG.
Fig. S5. Reduced folate transporter (SLC19A1).
Fig. S6. Chloroquine resistance transporter I.

REFERENCES

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. **Nature** *630*, 493-500. https://doi.org/10.1038/s41586-024-07487-w.

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.

Benavides, T.L., and Montelione, G.T. (2024). Integrative Modeling of Protein-Polypeptide Complexes by Bayesian Model Selection using AlphaFold and NMR Chemical Shift Perturbation Data. **bioRxiv**. https://doi.org/10.1101/2024.09.19.613999.

Bryant, P., and Noé, F. (2024). Structure prediction of alternative protein conformations. **Nature Communications** *15*, 7328. https://doi.org/10.1038/s41467-024-51507-2.

Chakravarty, D., Lee, M., and Porter, L.L. (2025). Proteins with alternative folds reveal blind spots in AlphaFold-based protein structure prediction. **Current Opinion in Structural Biology** *90*, 102973. https://doi.org/https://doi.org/10.1016/j.sbi.2024.102973.

Chakravarty, D., Schafer, J.W., Chen, E.A., Thole, J.F., Ronish, L.A., Lee, M., and Porter, L.L. (2024). AlphaFold predictions of fold-switched conformations are driven by structure memorization. **Nat Commun** *15*, 7296. https://doi.org/10.1038/s41467-024-51801-z.

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.

Hopf, T.A., Green, A.G., Schubert, B., Mersmann, S., Scharfe, C.P.I., Ingraham, J.B., Toth-Petroczy, A., Brock, K., Riesselman, A.J., Palmedo, P., et al. (2019). The EVcouplings Python framework for coevolutionary sequence analysis. **Bioinformatics** *35*, 1582-1584. https://doi.org/10.1093/bioinformatics/bty862.

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., and Montelione, G.T. (2024). Hidden structural states of proteins revealed by conformer selection with AlphaFold-NMR. **bioRxiv**, 2024.2006.2026.600902. https://doi.org/10.1101/2024.06.26.600902.

Huang, Y.J., Zhang, N., Bersch, B., Fidelis, K., Inouye, M., Ishida, Y., Kryshtafovych, A., Kobayashi, N., Kuroda, Y., Liu, G., et al. (2021). Assessment of prediction methods for protein structures determined by NMR in CASP14: Impact of AlphaFold2. **Proteins** *89*, 1959-1976. https://doi.org/10.1002/prot.26246.

Johansson-Akhe, I., and Wallner, B. (2022). Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. **Front Bioinform** *2*, 959160. https://doi.org/10.3389/fbinf.2022.959160.

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.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. (2021). Critical assessment of methods of protein structure prediction (CASP)-Round XIV. **Proteins** *89*, 1607-1617. https://doi.org/10.1002/prot.26237.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. (2023). Critical assessment of methods of protein structure prediction (CASP)-Round XV. **Proteins** *91*, 1539-1549. https://doi.org/10.1002/prot.26617.

Lazou, M., Khan, O., Nguyen, T., Padhorny, D., Kozakov, D., Joseph-McCarthy, D., and Vajda, S. (2024). Predicting multiple conformations of ligand binding sites in proteins suggests that AlphaFold2 may remember too much. **Proc Natl Acad Sci U S A** *121*, e2412719121. https://doi.org/10.1073/pnas.2412719121.

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.

Li, E.H., Spaman, L.E., Tejero, R., Janet Huang, Y., Ramelot, T.A., Fraga, K.J., Prestegard, J.H., Kennedy, M.A., and Montelione, G.T. (2023). Blind assessment of monomeric AlphaFold2 protein structure models with experimental NMR data. **J Magn Reson** *352*, 107481. https://doi.org/10.1016/j.jmr.2023.107481.

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.

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.

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.

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.

Roney, J.P., and Ovchinnikov, S. (2022). State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold. **Phys Rev Lett** *129*, 238101. https://doi.org/10.1103/PhysRevLett.129.238101.

Sala, D., Engelberger, F., McHaourab, H.S., and Meiler, J. (2023a). Modeling conformational states of proteins with AlphaFold. **Curr Opin Struct Biol** *81*, 102645. https://doi.org/10.1016/j.sbi.2023.102645.

Sala, D., Hildebrand, P.W., and Meiler, J. (2023b). 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.

Schrödinger, L., and DeLano, W. (2020). PyMOL.

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.

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** *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.

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

Table S1. Assessment of modeling alternative states of SLC proteins with AF2.

Name	Uniprot id	PDB ID	AlphaFold2 full MSA		AlphaFold2 single-sequence		AlphaFold2 low MSA 4-8		AlphaFold2 low MSA 8-16	
			dropout	No dropout	dropout	No dropout	dropout	No dropout	dropout	No dropout
***E. coli* D-galactonate:proton symporter SLC17 (DgoT)**	P0AA76	6e9n (I) 6e9o (O)	inward	inward	failed	failed	occluded	occluded	**outward**	**outward**
proton-coupled zinc antiporter SLC30A8 ZnT8	Q8IWU4	6xpf_B (O) 6xpf_A (I)	inward	inward	failed	failed	outward	**outward**	inward	inward
aromatic amino acid exporter YddG	D7A5Q8	5i20 (O)	outward	outward	failed	failed	failed	failed	occluded	occluded
reduced folate transporter SLC19A1	P41440	7xtk (I)	inward	inward	failed	failed	**inward**	failed	outward	outward
chloroquine resistance transporter I	W7FI62	6ukj (O)	outward	outward	failed	failed	outward	outward	outward	outward
***Zea mays* CMP-sialic acid transporter 1 SLC35A1**	B4FZ94	6i1r (O)	outward	outward	failed	failed	outward	outward	outward	outward
***S. cerevisiae* GDP-mannose sugar transporter 1, SLC35D Vrg4.**	P40107	5oge(O)	outward	outward	failed	failed	outward	outward	outward	outward
thiamine transporter 1 SLC19A2	O60779	n/a	inward	inward	failed	failed	**outward**	**outward**	**outward**	**outward**
solute carrier family 35 member F2	Q8IXU6	n/a	outward	outward	failed	failed	failed	outward/ occluded	outward	occluded
solute carrier family 35 member F3	Q8IY50	n/a	inward	inward	failed	failed	inward	inward	inward	inward

The outcomes “inward”, “outward”, and “occluded” indicate the conformational state observed, and the outcome “failed” indicates that a reasonable structure consistent with EC contact maps and/or experimental structures was not obtained. Results in which the alternative conformational state was generated are indicated in bold font.

Table S2. Assessment of successful generation of alternative states of SLC proteins from flipped sequence using *ESMfold* vs *AlphaFold2*.

Name	Uniprot id	ESMfold	AlphaFold2 single sequence	AlphaFold2 MSA 8 - 16
E. coli **D-galactonate:proton symporter SLC17 (DgoT)**	P0AA76	modeled	failed	modeled
proton-coupled zinc antiporter SLC30A8 ZnT8	Q8IWU4	modeled	failed	failed
aromatic amino acid exporter YddG	D7A5Q8	modeled	failed	modeled
reduced folate transporter SLC19A1	P41440	modeled	failed	modeled
chloroquine resistance transporter I	W7FI62	modeled	failed	failed
Zea mays **CMP-sialic acid transporter 1 SLC35A1**	B4FZ94	modeled	failed	modeled
S. cerevisiae **GDP-mannose sugar transporter 1, SLC35D Vrg4.**	P40107	modeled	failed	modeled
thiamine transporter 1 SLC19A2	O60779	modeled	failed	modeled
solute carrier family 35 member F3	Q8IY50	modeled	failed	modeled

For each of these 9 SLC proteins, a "flipped sequence" was used to generate a template for modeling the alternative conformation state, as outlined in main text **Fig. 1**, using either *ESMfold* or *AF2*. The outcome "modeled" indicates that the result was a folded protein with a structure generally similar to the alternative inward- or outward-open conformational state; while "failed" indicates that a reasonable structure was not obtained. Both methods were initially performed using a single-sequence as input; in this case the *ESMfold* models were alternative conformational states as expected, but the AF2 modeling based on the "flipped sequence" failed for all 9 proteins. The *AF2* modeling from the "flipped sequence" was then repeated with MSA max of 8 to 16, in which case 7 of the 9 systems provided reasonable structures for the alternative state.

Table S3. Modeling of both outward-open and inward-open states of pseudo-symmetric SLC proteins

Name	Uniprot id	Experimental structure (O: Outward open) (I: Inward open)	AF2 prediction (rmsd Å)	AF Database (rmsd Å)	ESM-AF (rmsd Å)	ESM-MODELLER (rmsd Å)	
***E. coli* D-galactonate:proton symporter SLC17 (DgoT)**	P0AA76	6e9n (I) 6e9o (O)	0.76 -	0.84	0.76	3.8	**Supplementary Figure S2**
proton-coupled zinc antiporter SLC30A8 ZnT8	Q8IWU4	6xpf_B (O) 6xpf_A (I)	1.09 -	1.12	2.00	-	**Figure 2B**
aromatic amino acid exporter YddG	D7A5Q8	5i20 (O) n/a (I)	0.68 -	0.43	0.45	modeled	**Supplementary Figure S4**
reduced folate transporter SLC19A1	P41440	n/a (O) 7xtk (I)	- 0.64	0.58	0.59	modeled	**Supplementary Figure S5**
chloroquine resistance transporter I	W7FI62	6ukj (O) n/a (I)	0.89 -	2.76	0.77	modeled	**Supplementary Figure S6**
***Zea mays* CMP-sialic acid transporter 1 SLC35A1**	B4FZ94	6i1r (O) n/a (I)	0.69 -	0.26	0.65	modeled	**Figure 3**
***S. cerevisiae* GDP-mannose sugar transporter 1, SLC35D Vrg4.**	P40107	5oge (O) n/a (I)	0.56 -	0.46	0.52	modeled	**Figure 4**
thiamine transporter 1 SLC19A2	O60779	n/a (O) n/a (I)	modeled -	n/a n/a	modeled	modeled	**Supplementary Figure S3**

For each SLC protein, conventional AF2 modeling provided either an inward-open (I) or outward-open (O) state for which the backbone root-mean-squared deviation to an available X-ray crystal or cryoEM structure is listed. Shading designates where both states (dk grey), neither state (med grey), and one state (lt gray) were available in the PDB at time of AF2 training. The alternative outward-open or inward-open conformation was then generated with the ESM-AF2 or ESM-MODELLER protocol. For these models of alternative conformational states, backbone rmsd's are reported where experimental structures are available. Models for each protein available in the AlphaFold2 data base are indicated along with backbone rmsd's to the most similar experimental structure. The outcome "modeled" indicates that the result was a folded protein with a structure generally similar to the alternative inward- or outward-open conformational state, n/a – an experimental structure is not available.

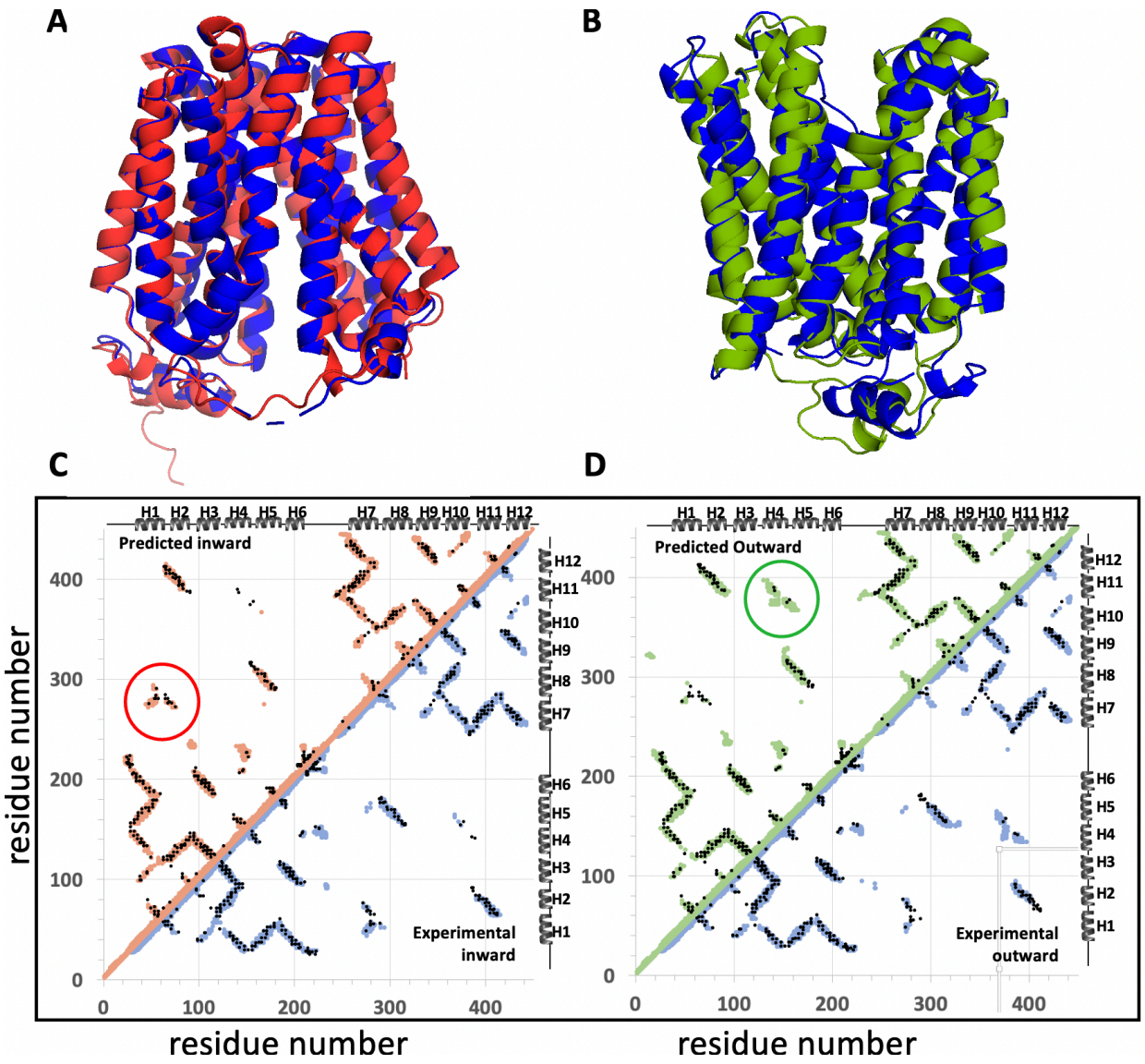


Fig. S1. Validation of ESM-AF2 protocol using an SLC protein with both outward- and inward-open experimental structures. (A) Superposition of the inward-open conformational state of *E. coli* D-galactonate:proton symporter (SLC17, DgoT) modeled using the ESM-AF2 protocol (red) with the experimentally-determined inward-open structure PDB id 6e9n (blue). (B) Superposition of the outward-open state modeled using the ESM-AF2 protocol (green) with the experimentally-determined inward-open structure PDB id 6e9o (blue). (C) Contact maps for the inward-open state of DgoT predicted by ESM-AF2 (red circles), observed in the cryoEM experimental structure (blue circles), and predicted by EC analysis (black dots). (D) Contact maps for the outward-open state of DgoT predicted by ESM-AF2 (green circles), observed in the cryoEM experimental structure (blue circles), and predicted by EC analysis (black dots). In this example, only a few EC-based contacts between helices H4 and H11 and helices H5 and H10 (circled in green in panel D) and between H1 and H7 (circled in red in panel C) distinguish the inward- from outward-open structures.

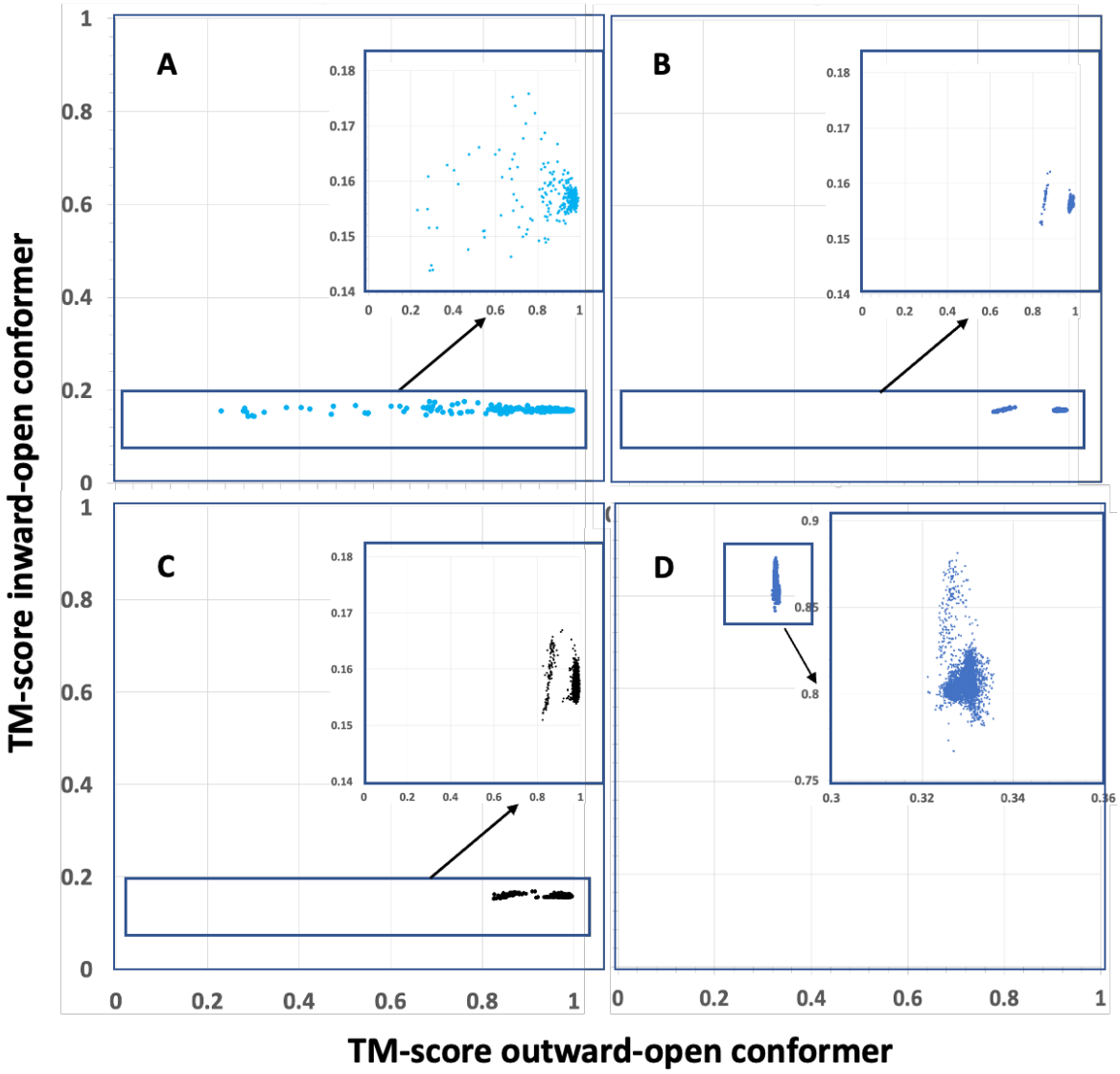


Fig. S2. Conformational states of SLC35F2 modeled using various enhanced or massive sampling protocols. (A) *AlphaFold-alt* (480 models), (B) *AlphaFold-sample* (3,000 models), and (C) *AlphaFold-sample2* (3,000 models); the resulting models were compared with the outward-open and inward-open models from the standard AF2 (for outward-open) and our ESM-AF2 protocol (for inward-open). None of these three methods modeled the alternative "inward-open" conformational state. (D) Massive sampling with *AF-sample2* (3,000 models) using a "flipped" protein sequence generated only inward-open conformations for SLC35F2 can provide an inward-open structure of SLC35F2. In each panel, the inset shows an expanded region of the plot. In all cases, no models were observed outside of the boxed regions. All models have <pLDDT > > 70.

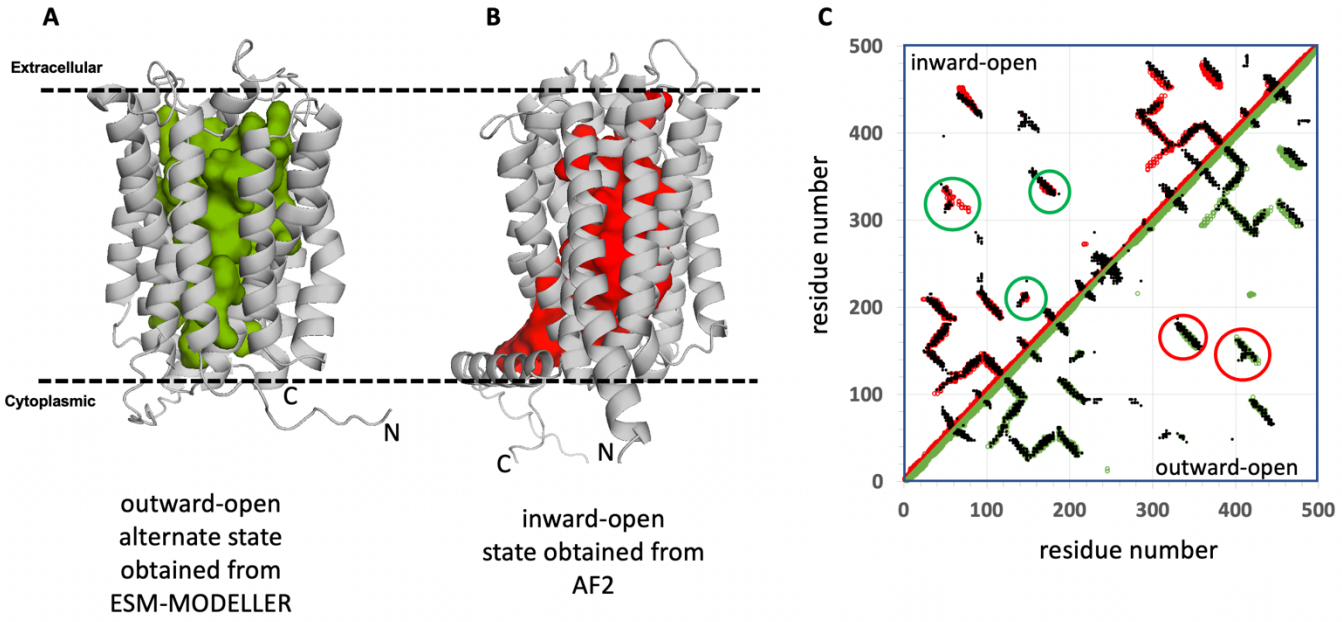


Fig. S3. Thiamine transporter 1 (SLC19A2). (A) The outward-open conformer generated using the ESM-MODELLER method. (B) AF2 models the inward-open state. (C) The outward-open conformer contact map (green, lower diagonal), and inward-open conformer contact map (red, upper diagonal), generated using Cα-Cα cutoff of 10 Å, are superposed on the symmetric EC-predicted contact map (black). ECs unique to each state are indicated with green and red circles, respectively.

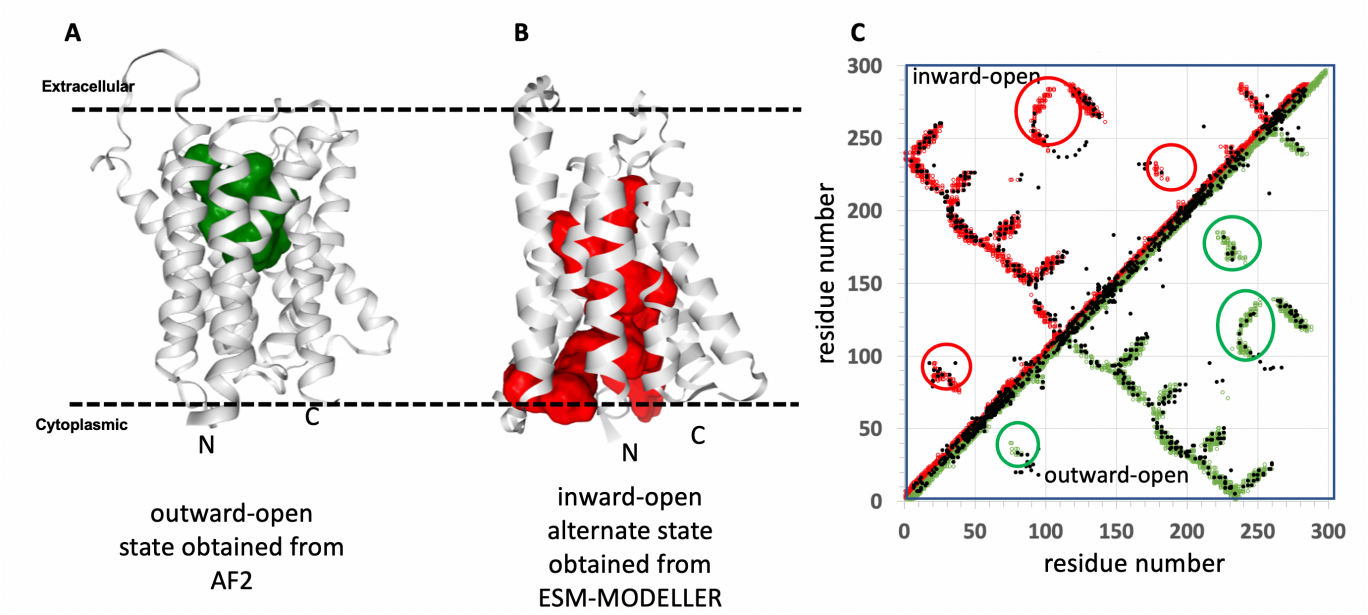


Fig. S4. Aromatic amino acid exporter YddG. (A) AF2 models the outward-open state. (B) The inward-open conformer was generated using the ESM-MODELLER method. (C) The outward-open conformer contact map (green, lower diagonal), and inward-open conformer contact map (red, upper diagonal), generated using Cα-Cα cutoff of 10 Å, are superposed on the symmetric EC-predicted contact map (black). ECs unique to each state are indicated with green and red circles, respectively.

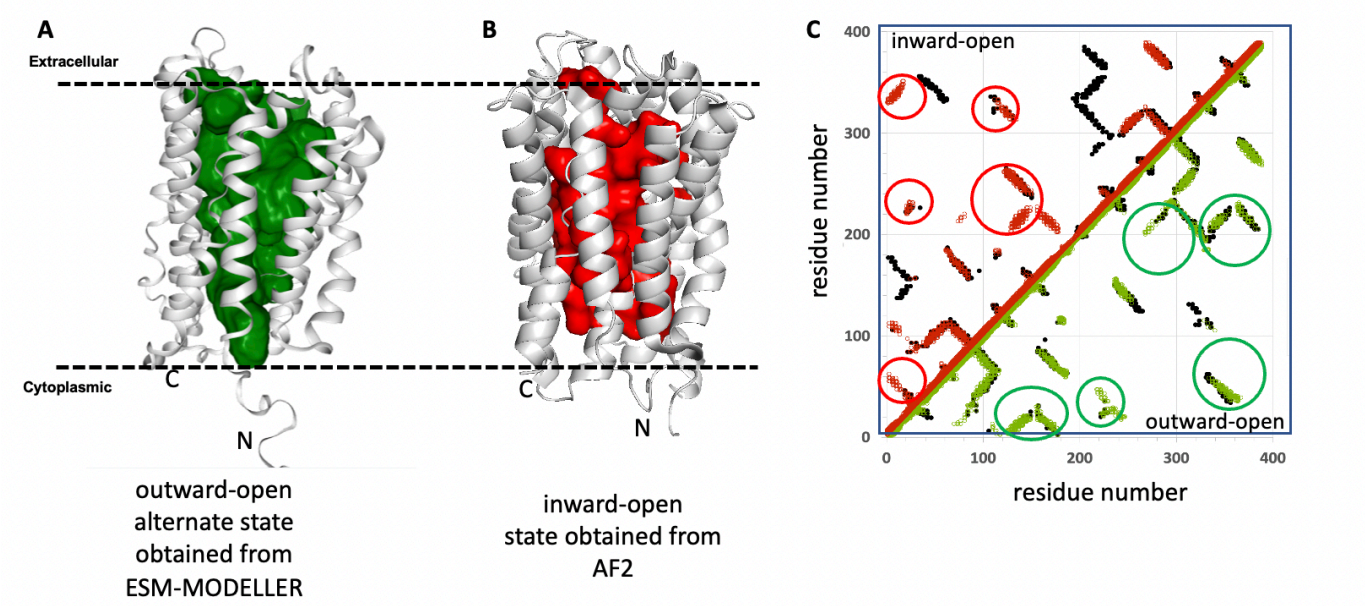


Fig. S5. Reduced folate transporter (SLC19A1). (A) The outward-open conformer was generated using the ESM-MODELLER method. (B) AF2 models the inward-open state. (C) The outward-open conformer contact map (green, lower diagonal), and inward-open conformer contact map (red, upper diagonal), generated using Cα-Cα cutoff of 10 Å, are superposed on the symmetric EC-predicted contact map (black). ECs unique to each state are indicated with green and red circles, respectively.

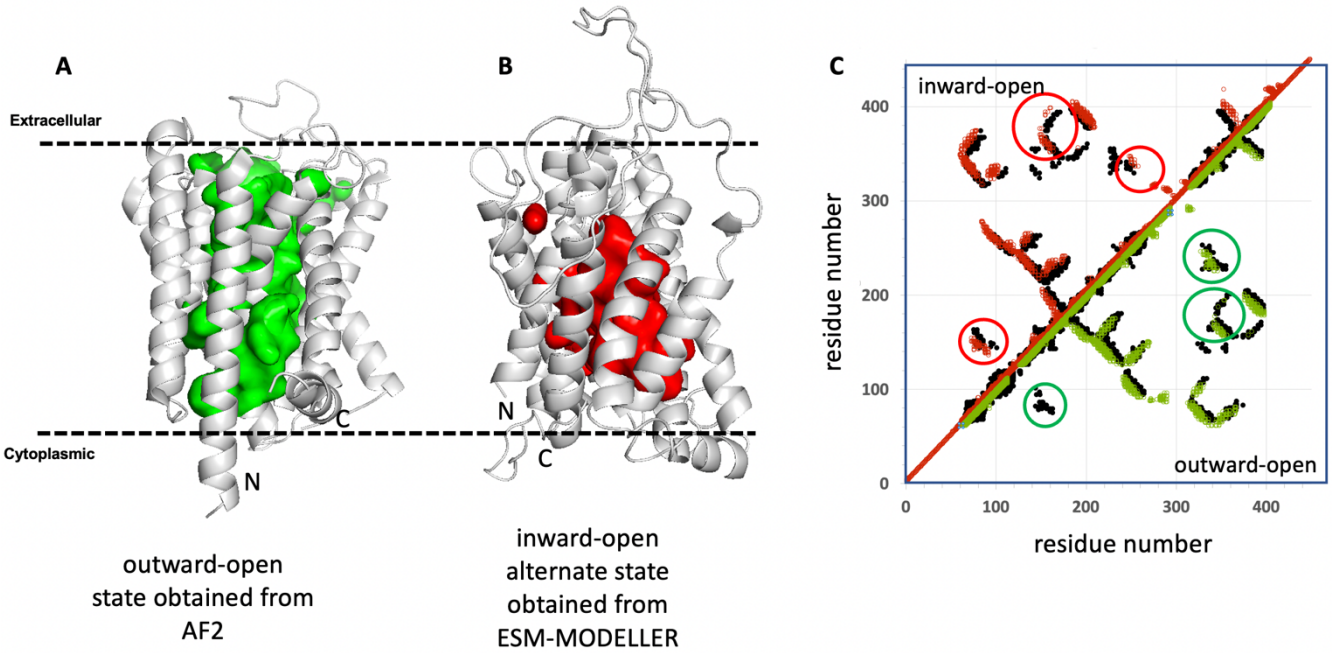


Fig. S6. Chloroquine resistance transporter I. (A) AF2 models the outward-open state. (B) The inward-open conformer was generated using the ESM-MODELLER method. (C) The outward-open conformer contact map (green, lower diagonal), and inward-open conformer contact map (red, upper diagonal), generated using Cα-Cα cutoff of 10 Å, are superposed on the symmetric EC-predicted contact map (black). ECs unique to each state are indicated with green and red circles, respectively.