DOI:10.14753/SE.2025.3101

SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3101.

GÁSPÁRNÉ CSIZMADIA GEORGINA

Celluláris és molekuláris biofizika

című program

Programvezető: Dr. Kellermayer Miklós, egyetemi tanár

Témavezető: Dr. Hegedűs Tamás, tudományos főmunkatárs

DOI:10.14753/SE.2025.3101

MEMMORFS, MEMBLOBS AND CONFTORS: COMPUTATIONAL STUDIES OF MEMBRANE PROTEINS

PhD thesis

Georgina Gáspárné Csizmadia

Theoretical and Translational Medicine Doctoral School
Semmelweis University

Supervisor: Tamás Hegedűs, D.Sc
Official reviewers: Gábor Turu, Ph.D
Zoltán Gáspári, Ph.D

Head of the Complex Examination Committee: Alán Alpár, D.Sc

Members of the Complex Examination Committee:
Mihály Kovács, D.Sc
László Cervenák, Ph.D

Budapest
2024

DOI:10.14753/SE.2025.3101

TABLE OF CONTENTS

DOI:10.14753/SE.2025.3101

LIST OF ABBREVIATIONS

ABC	ATP-binding cassette
ABCB1/MDR1	P-glycoprotein 1, permeability glycoprotein, also known as multidrug resistance protein 1 or ATP-binding cassette sub-family B member 1
ABCC1/MRP1	ATP binding cassette subfamily C member 1 or Multidrug resistance-associated protein 1
ATP	Adenosine triphosphate
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator or ABCC7
CG	coarse-grained
CH	coupling helix
COG	center of geometry
conftors	CONFormational vecTORs; vectors that represents relevant parts of a protein structure defined by an expert of the protein
cryo-EM	cryo-electron microscopy
CT	cytoplasmic tail
DDM	n-Dodecyl β-D-Maltopyranoside
EPR	Electron paramagnetic resonance
GPCR	G protein-coupled receptor
HCV	Hepatitis C virus
ICD	intracellular domain
IDR	intrinsically disordered region
ITAM	Immunoreceptor tyrosine-based activation motif
MAP	membrane associated protein
McjD	Microcin-J25 export ATP-binding/permease protein McjD
MD	Molecular dynamics
mMDR1	mouse P-glycoprotein 1
MemMoRF	Membrane associable molecular recognition feature
MoRF	Molecular Recognition Feature
MsbA	ATP-dependent lipid A-core flippase
NBD	nucleotide binding domain
NMR	Nuclear magnetic resonance
Pgp	P-glycoprotein 1, permeability glycoprotein, also known as multidrug resistance protein 1 or ATP-binding cassette sub-family B member 1
PLN	Phospholamban
PTM	post-translational modification
RCI	random coil index
RMSD	Root-mean-square deviation
SARS-CoV E	Envelope protein of Severe-acute-respiratory-syndrome–related coronavirus
SDS	Sodium dodecyl sulfate
SERCA	sarcoplasmic/endoplasmic reticulum Ca2+-ATPase
SMALP	Styrene maleic acid lipid particle
SSPop	secondary structure propensity
TH	transmembrane helix
TM	transmembrane
TMD	transmembrane domain
TMP	transmembrane protein
WAH	helix in Walker-A motif

DOI:10.14753/SE.2025.3101

1. INTRODUCTION

A protein called cystic fibrosis transmembrane conductance regulator (CFTR) connected all my projects with distinct objectives. CFTR is a member of the ATP-binding cassette (ABC) superfamily of transmembrane proteins. These proteins transport substrates through the membrane or regulate channel function. Therefore, they play important roles in every living organism and hold significant pathological relevance (1-3). The most widely studied disorder connected to ABC proteins is cystic fibrosis, which is caused by mutations in CFTR (4, 5). CFTR is the only known member of the superfamily functioning as a phosphorylation-regulated, ATP-gated anion channel. Regulation by phosphorylation occurs at sites in the regulatory R domain, which is also unique to CFTR (6, 7). Despite these special features of CFTR, it has a Pgp-like (previously called Type I) structural organization (8).

1.1. Evaluating Pgp-like ABC protein structures

Pgp-like exporter structures have two transmembrane (TM) domains, each usually formed by six helices (Figure 1). The TM helices continue in the intracellular domains (ICDs) in the cytosol. These intracellular domains contain coupling helices, which interact with the nucleotide binding domains (NBDs) (9).

During the transport cycle, the ABC proteins can be in different, characteristic conformations. In the absence of ATP, the NBDs are partially or completely separated and the TM domains exhibit a cavity facing the cytoplasm (10). These conformations are called “bottom-closed, inward-facing” and “bottom-open, inward-facing”, respectively. The NBDs form a close interaction with each other through the LSGGQ signature and the Walker A motifs when two ATPs are sandwiched between them. This leads to structural changes in the ICDs and transmembrane domains (TMDs). As a result, the cavity closes at the cytosolic side and opens to the extracellular space, which leads to the export of the substrate (“bottom-closed, outward facing” conformation). “Occluded” structures have interacting NBDs without an open cavity to the extracellular side. These are suggested to be an intermediate state between the “inward-facing” and “outward-facing” conformations. The cycle is believed to end with ATP hydrolysis, resulting in an inward-facing conformation (11, 12).

DOI:10.14753/SE.2025.3101

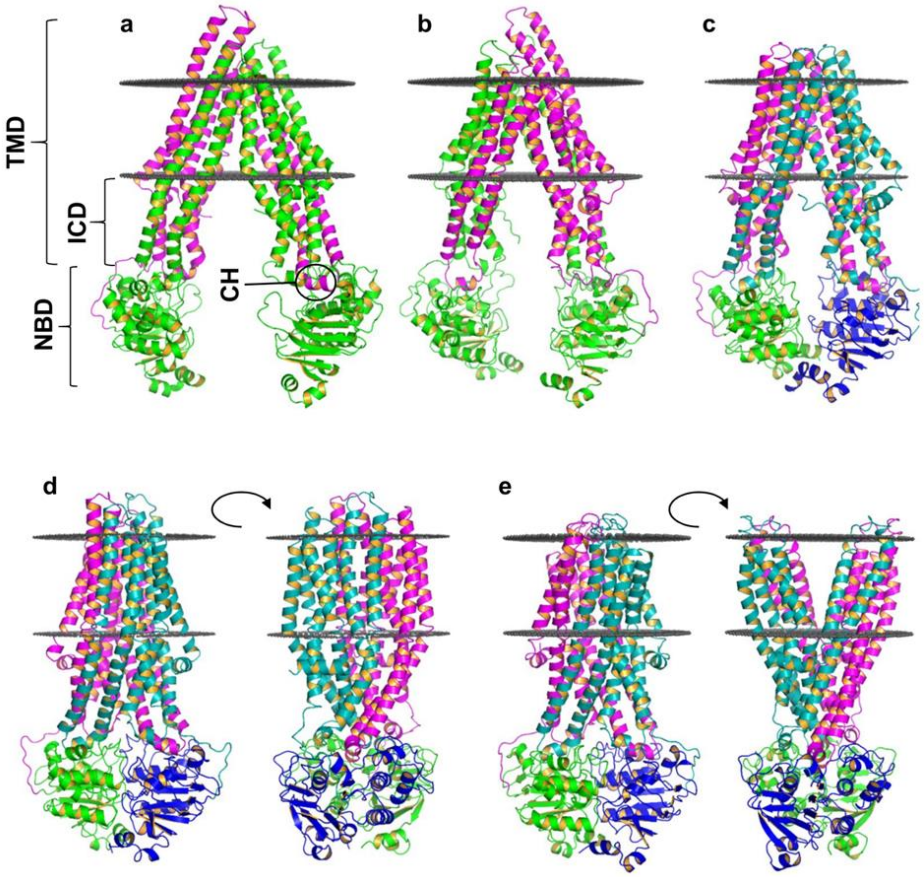

Figure 1 *Type I ABC exporter conformations.* ***a***, ***b****) "bottom-open, inward" facing mMDR1 structures (PDBIDs: 5KPJ, 3G5U) (13, 14). The NBDs are separated (open) at different extents.* ***c****) a "bottom-closed, inward-facing" TM287/288 structure (PDBID: 3QF4 (15).* ***d****) a "bottom-closed, top-closed" (occluded) McjD (PDBID: 4PL0) (11).* ***e****) a "bottom-closed, outward-facing" MsbA structure (PDBID: 3B60) (16). TMD: transmembrane domain, ICD: intracellular domain or intracellular "loop", NBD: nucleotide binding domain, CH: coupling helix (17).*

There are several models of the catalytic cycle, which can be divided into two categories based on whether they predict the complete or partial separation of the NBDs. The most notable model of the former category is the ATP switch model, which has two alternatives called the tweezers-like model and the processive clamp model (18). These models promote the separation of NBDs by more than 20 Å after the hydrolysis of the ATPs in both sites. In contrast, the constant-contact, the alternating sites and the nucleotide occlusion models propose that the NBDs remain in constant contact with each other through at least one nucleotide binding site (18, 19). This means that one site

remains closed while the other opens after ATP hydrolysis. The open site closes upon ATP binding, and the previously closed site opens (18).

Models like the constant-contact model question the validity of the experimentally determined "bottom-open, inward facing" conformations (18). There are several conformations in the "bottom-open, inward facing" category, where there is a large distance between the NBDs. This may be the result of crystallization artifacts, like crystal contacts or the lack of a membrane environment. Structures with widely separated NBDs were also unstable in molecular dynamics simulations (10, 20). However, MsbA (ATP-dependent lipid A-core flippase) with widely bottom-open NBDs recently was found to be present in *Escherichia coli* cells (21). There are much fewer structures of "bottom-closed, outward-facing" conformations in RCSB PDB than inward facing structures. Some of the outward-facing structures exhibit a widely open cavity to the extracellular space. These conformations were suggested to be the results of the lacking lateral pressure from the membrane bilayer during crystallization (22).

Differentiating the various conformations is crucial to study and understand the working mechanism of these proteins. Validation and quality assessment of experimental structures is crucial for all types of membrane proteins. This is especially true in the case of ABC proteins, as these undergo significant structural changes during their transport cycle.

There are existing tools and databases for evaluating structures. OneDep (23) is a tool to curate, validate and deposit structures to PDB. It uses metrics like clashes between atoms and Wilson B-value, which are important basis for curating 3D structures (24-26). However, higher level information (e.g. domain-domain orientation) on structure validity is not included. The CoDNaS conformational database provides a higher level comparison (e.g. RMSD, TM-score) of different structures of the same protein (27). Using a different approach, a molecular dynamics study was performed to compare and assess structures of ABCB1/MDR1, where TMDs and NBDs were compared between the structures before and during the simulations (28). Since using MD for numerous structures is highly resource intensive, a less demanding approach would be more suitable to catch up to the increasing number of structures.

1.2. Transmembrane regions of membrane proteins

We investigated multiple structures of ABC proteins, solved by cryo-electron microscopy (cryo-EM). Compared to X-ray diffraction, cryo-EM enables biomolecule structure determination in an environment that is more similar to their native one. However, in the first step the target proteins have to be separated from other proteins and solubilized. Detergents are used to solubilize TM proteins from their native membrane. After the solubilization of the protein, the detergent prevents its aggregation and precipitation by covering its TM region and acting as a membrane mimetic. Zwitterionic and anionic detergents (e.g. SDS) have very high extraction efficiency. However, these partially denature the soluble regions of the protein. In contrast, mild detergents, such as DDM, have minimal effects on both the TM and soluble parts, but they provide much lower yield. Amphipols are emerged as an alternative to these detergents, while nanodisc and SMALPs (Styrene maleic acid lipid particles) are developed to provide a native-like lipid environment around the TM protein (29).

The output of cryo-EM is a 3D Coulomb potential map (cryo-EM map). Importantly, many of these density maps contain more information than just the protein structure. For example, the identification of density blobs of ligands is solved (30). Intrinsically disordered regions (IDRs) also have densities in cryo-EM maps, though extracting and converting them to structural ensemble is challenging. We noticed that the membrane mimetic is also visible in many densities associated with structure determination of transmembrane proteins.

Knowing the transmembrane regions of membrane proteins is crucial to understand their working mechanism, and to develop new therapies targeting them, yet experimental data on the membrane embedment is sparse. There are experiments, where membrane embedded proteins are tagged at various positions around putative transmembrane helices and the accessibility of the tag is tested, but these provide only low-resolution data (31, 32). Structure determination by Nuclear magnetic resonance (NMR) can be done in the presence of membrane mimetics, but due to the limitation of the method it is used mostly for small proteins or short regions and membrane interaction sites are usually not directly tested (33). In the case of crystallography, we rarely see the lipids in the structure, but when we do, they may have attached to non-physiological sites.

This lack of sufficient experimental data led to the development of *in silico* transmembrane region prediction methods. The three most widely used are TMDET from PDBTM database (34) (https://tmdet.enzim.hu/, https://pdbtm.unitmp.org/), PPM from OPM database (35) (https://opm.phar.umich.edu/) and MEMPROTMD (36), (http://memprotmd.bioch.ox.ac.uk/). TMDET is a geometrical approach based on an objective function that measures the amino acid hydrophobicity and combines it with structural features. PPM calculates and minimizes the transfer free energy of protein structures at different membrane thicknesses, positions and protein tilting. TMDET and PPM define the membrane as a slab with two parallel planes, while MEMPROTMD builds a bilayer around the protein using molecular dynamics simulations. However, the only experimental data these methods directly use is protein structure.

1.3. Disordered regions in membrane proteins

CFTR has a two hundred amino acids long regulatory region called R domain. R domain is an IDR, which blocks the interaction of NBDs by intercalating between them. Upon phosphorylation, the R domain moves away and the NBDs can interact with each other when ATP binds to the active site, which leads to the opening of the chloride channel (37). We proposed in our group that the large number of positively charged amino acids in the R domain may result in anchoring to the negatively charged cytosolic surface of the plasma membrane. Therefore, we also started to investigate proteins with disordered regions interacting with lipid bilayers.

IDRs do not exhibit well-defined structures. Therefore, they are represented by conformational ensembles. They are often enriched in polar or charged residues and depleted of hydrophobic residues (38, 39). IDRs can bind to various partners with high specificity and low affinity (40). The high specificity is a result of the combination of their electrostatic interactions due to their structural flexibility (41). The low affinity binding is caused by the high entropic penalty due to the reduced conformational space upon binding. The consequence of this issue is a high enthalpic gain (42). These grant them exceptional flexibility regarding binding to partners, which is the base of various physiologically relevant processes, such us signaling, gene and cell cycle regulation and chaperoning (42-44).

41% of membrane proteins contain at least one long IDR. Experimental studying of IDRs in transmembrane proteins (TMPs) is challenging, as it requires to deal with both

the membrane environment and IDRs. Prediction of disordered regions in transmembrane proteins is also hard because they have different amino acid composition profile than IDRs in soluble proteins (45). Nearby large hydrophobic regions (the TM domains) make predictions even less accurate (46)

IDRs in TM proteins also have more post-translational modification (PTM) sites, especially phosphorylation. Phosphorylation mainly happens at intracellular IDRs. This provides tunable regulation by controlling disorder-order transitions as regulatory switches (45). Transition between ordered and disordered states often happens during binding to partners, too. Disorder to order transition of a segment in the IDR upon binding to protein partners is common. These segments are called Molecular Recognition Feature (MoRFs) (47-49). MoRFs often overlap with other motifs like PTM sites (44). However, IDRs can also bind to biological membranes. Similarly to interactions with proteins, the physicochemical properties of membranes affects the IDRs conformational state (45). Therefore, we named the IDRs that become ordered upon membrane binding MemMoRFs.

It is challenging to clarify the structural features of MemMoRFs in detail. As a result, literature focusing on this subject is scarce. Information about these regions have not been collected and categorized, although databases of IDRs and IDR-protein interactions are available. There is a database exclusive to protein interacting MoRFs in membrane proteins, but there is no database about IDR-membrane interactions.

DOI:10.14753/SE.2025.3101

2. OBJECTIVES

The first aim of my thesis was defining metrics to assess the differences between the conformational states of Pgp-like exporter structures in a way which considers the geometry of the whole structure. Therefore, my first objective was to define conformational vectors (conftors, Figure 4) to represent functionally relevant parts of Pgp-like protein structures. My next objective was to choose a TMD predictor by comparing the most popular ones and by calculating membrane solvation energy at different membrane positions.

My second aim was to present a tool to provide high resolution experimental data on transmembrane regions based on cryo-EM maps. The first objective regarding this project was to develop a tool called MemBlob for calculating transmembrane regions based on the Coulomb densities of the membrane and make the tool available as a web application. The second objective was to compare these results to regions based on TM predictors.

The third aim was to facilitate the investigation of membrane interacting intrinsically disordered regions (IDRs). My objective was to gather knowledge about MemMoRFs and organize them with various information into a manually curated database.

Each of the above aims is discussed in its own subsection called Conftors, MemBlob, and MemMoRF, as they are based on different publications (17, 65, 88).

3. METHODS

Custom Python (50, 51) scripts were used for general tasks like downloading and managing multiple files.

3.1. Conftors and electrostatics calculations

Collecting protein structural information

A list of Pgp-like ABC protein structures was collected from the UniProt database (52) (https://www.uniprot.org/), using the advanced search for Type I ABC proteins with 3D structure (UniProt, March 2017). I downloaded the coordinates of the structures from the OPM database (53) (https://opm.phar.umich.edu/) . This way, the coordinate files included the slabs of a predicted membrane bilayer represented by DUMMY atoms. Some of the structures were not included in the OPM database. I downloaded these from RCSB PDB (54) (https://www.rcsb.org/) and calculated their membrane regions using the PPM server. Then I extracted the transmembrane helix boundaries from files from OPM and PPM. Other notable regions, like coupling helices, were identified manually. The method used for structure determination was extracted from the downloaded coordinate files. The conformational state of each structure was determined by visual inspection.

Calculating Conftors

Python scripts were used with NumPy and MDAnalysis packages (55, 56) to calculate the length and angle of conftors.

Comparing membrane localizations

We selected the membrane predictors PPM, TMDET and MEMPROTMD to assign TM regions to our target proteins. PPM was chosen as a reference, because this is the most widely used predictor of them. I investigated the differences between their predictions on membrane positions along the z axis and on the tilting angle of the structures. I compared the values from TMDET, MEMPROTMD, and coarse-grained (CG) simulations (17) performed by my colleagues to the values from PPM.

The angle between the membrane normal and the bisector between the THX1 and THX2 conftors (Table 1, Figure 4, TH4–5 and TH10–11) was used to describe protein tilting. For PPM structures, the membrane normal was the vector normal of the plane defined by the DUMMY atoms in PPM entries. In the case of CG structures, the

membrane normal was calculated by subtracting the center of geometry of PO4 and NC3 beads in the upper leaflet from that in the lower leaflet.

The membrane position along the z axis was characterized by the center of geometry (COG) of PO4 and NC3 beads. We used the COG of the DUMMY membrane atoms from OPM as reference.

In the case of TMDET, the downloadable structures did not include any membrane atoms. However, they were arranged in such a manner that the center of geometry of the predicted membrane coincided with the origin, and the normal vector of the membrane was perpendicular to the XY plane.

The structures were aligned. In the case of the CG structures, the back-bone beads were aligned to the Cα atoms of the all-atom structures.

Electrostatics calculations

I ran PDB2PQR (57) with PARSE force field (58) and I set the pH to 7.0. In those cases, where PDB2PQR failed to run, I made the necessary structural preparation with VMD's Automatic PSF Builder (59) using default values. I modified the input template created by PDB2PQR for APBS (60) to add 150 mM Na^+ and Cl^- ions with charge +1 and -1, and radius 0.95 Å and 1.81 Å, respectively. For membrane solvation calculations, I ran APBSmem (61) with the parameters listed in (17), Table S4. In addition, grid dimensions and fine grid size for x and y coordinates were collected from the input file of APBS. The z coordinate of the fine grid size was $-2\times z_{min}+40$ Å, where z_{min} is the smallest z coordinate from the PQR file. The medium grid size was 2 times larger and the coarse grid size was 5 times larger than the fine grid size. Grid dimension was 161 for each axis. The membrane thicknesses from PPM were applied and the flooding algorithm of APBSmem was used as the membrane filling method.

Visualization

I used PyMOL (The PyMOL Molecular Graphics System, Version 1.8.4 Schrödinger, LLC) and its APBS Electrostatics plugin to visualize 3D structures and surface electrostatics. To standardize the orientation of the structures, I selected TM287/288 (PDBID: 3QF4) (15) as a reference and rotated and translated other structures to structurally align them. This structure has inward-facing TM domains and bottom-closed NBDs, corresponding to an intermediate 3D structure. I used Python Matplotlib (62) to plot conftors.

3.2. MemBlob

Collecting structures, Cryo-EM maps and TMDET translation matrices

I downloaded the structures of transmembrane proteins that has a resolution higher than 4 Å from RCSB PDB. I downloaded the Cryo-EM maps from EMDB (July, 2018) (63). I collected the corresponding TMDET translation matrices from PDBTM. When PDBTM did not contain the structure, I submitted the PDB file to TMDET.

Calculating MemBlob boundaries

The calculations were done using self-written Python scripts (Figure 2). The input is an EM density map and the corresponding model of protein structure in a PDB format. From this model, we generated a theoretical density map of the protein at 6 Å resolution using VMD MDFF plugin (64). This was subtracted from the EMD map, to eliminate the densities corresponding to the protein structure.

In some cases, what remains after the subtraction is not just a membrane blob, but other densities may be present, such as disordered regions or ligands. To avoid these extra densities, we started our search for the membrane boundaries from the inside of the membrane blob. First, we converted the map to 3D points and corresponding density values. Then, we translated the coordinates of the map by a matrix from TMDET, which sets the origin into the predicted center of the bilayer. This way we made sure that the origin was inside the membrane blob. That was an important step for ensuring the reliability of our calculations, as we searched for the first minimum density value starting from the origin (after the simplifying steps described below). We used TMDET, since it is a reliable TM predictor and can be easily integrated into such a pipeline we developed. To convert the membrane blob to transmembrane regions, we made a two dimensional projection from it.

First, we generated x-y sections of the blob along the z-axis at 2 Å intervals, starting from the origin in both directions. Then, we cut these sections into angle slices of 10°, and summed the density values in each slice. This way we got an array of density values for every z-angle pair. This array is smoothed and presented in two dimensions. We searched for the first minimum density values from zero towards both positive and negative z directions in each slice. The membrane blob's boundaries were identified as these minimum density values. We projected these boundaries to the all-atom model to pair the atoms with their localization. Surface atoms that are beyond the determined

DOI:10.14753/SE.2025.3101

membrane boundaries were defined as water accessible. The interface region thickness was set to 8 Å. Atoms closer to the origin than the interface region were assigned to the hydrophobic core.

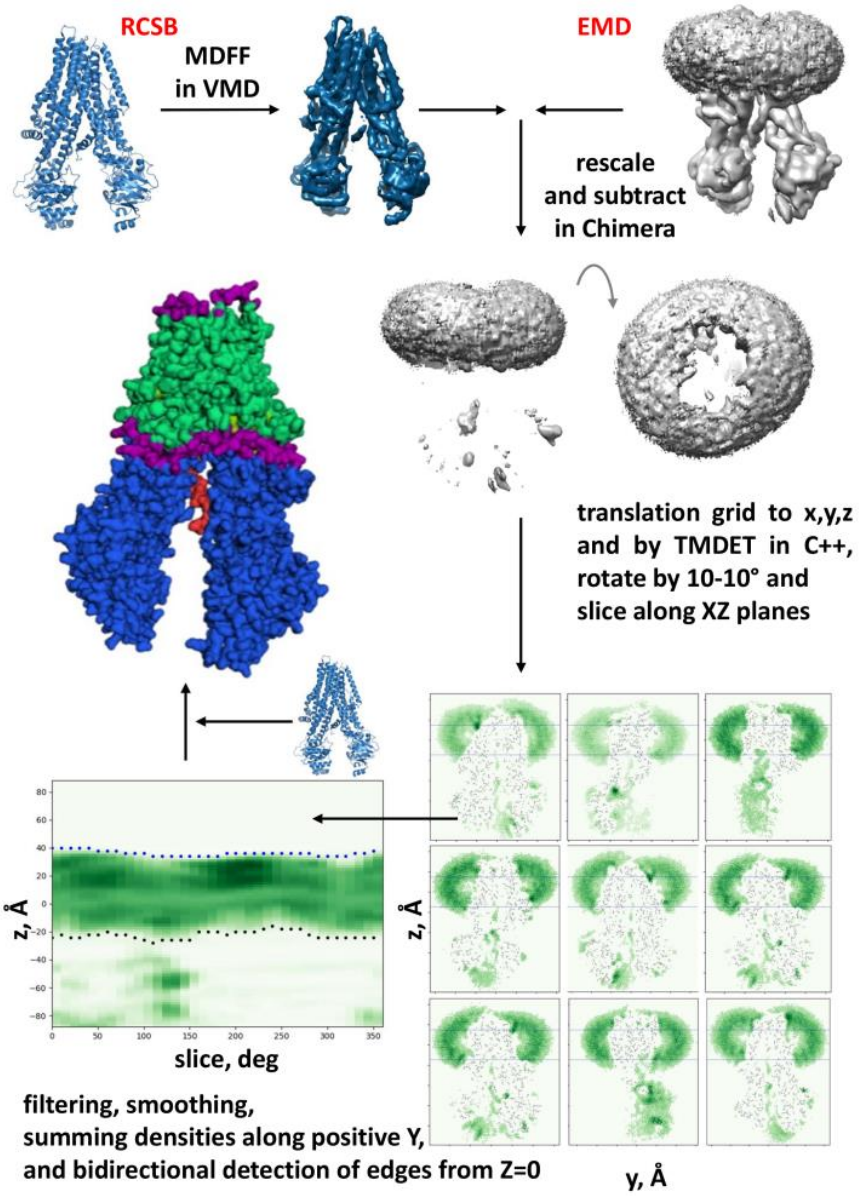

Figure 2 *MemBlob pipeline (65).*

Comparing the MemBlob method to TMDET

First, we needed to convert our boundaries, which conform to the shape of the lipid environment, into parallel slabs. To do this, we averaged the z-coordinates of the boundaries on each side. We then determined the thickness of the hydrophobic core by

calculating the distance between these boundaries and subtracting the combined thickness of the two interface regions. The center of the calculated membrane slabs and their thickness were compared to the outputs of TMDET.

3.3. MemMoRF

Collecting structures and structural information

First, I collected the NMR structures of TMPs (transmembrane proteins) and MAPs (membrane associated proteins). TM protein structures were identified by using the Membrane Protein Browser of RCSB PDB. MAPs were searched using the "peripheral membrane protein" keyword for subcellular location in the UniProt database and their structures in RCSB PDB via UniProt accession number. The BMRB IDs (66) (https://bmrb.io/) were extracted from the CIF or PDB files and used to download the NMR-STAR files that contain chemical shifts.

Identification of MemMoRFs

I did an excessive literature study including all articles published with the collected structures (approximately 500 articles). I searched for non-TM regions that were able to bind to membrane mimetics and were disordered. However, structures of TMPs and MAPs are generally solved in the presence of a membrane mimetic. It is rarely investigated whether a membrane attached region is disordered in solution or not. Therefore, I used the collected chemical shifts to calculate secondary structure propensity (SSPop; using δ2D) (67) and flexibility (1-S2; using Random Coil Index, RCI) (68) to recognize potential IDRs. If the residues within a membrane binding region populated "coil" secondary structure and were flexible, I considered it as an IDR. I used the minimum thresholds 0.15 and 0.5 for δ2D and RCI, respectively.

I also looked up if DisProt (https://disprot.org/), DIBS (https://dibs.enzim.ttk.mta.hu/), MFIB (https://mfib.enzim.ttk.mta.hu/) and PFAM (http://pfam.xfam.org/) databases (69-72) annotated an investigated region as disordered. This was especially important in those cases, where the lack of data prevented the calculation of SSPop and flexibility. I also checked the missing or invisible residues in X-ray and cryo-EM structures of TMPs or MAPs and considered them as IDRs.

The automatic and manual searching and arranging of these regions was challenging, since in many cases the numbering of residues in articles differs from the

numbering in structural and sequence databases, and sometimes even from the published structural files. I used CIF type coordinate files instead of PDB files to convert the unique numbering of the regions to the numbering of the whole protein automatically, because they usually contain the used numberings and the position of the region in the whole protein. In these instances, where the CIF file did not contain the necessary information regarding the position of the sequence in the whole protein, I identified the sequence position manually.

DOI:10.14753/SE.2025.3101

4. RESULTS

4.1 Conftors and electrostatics calculations

To facilitate the evaluation of Pgp-like ABC protein structures, I defined measures that can quantitatively describe their 3D structural features. We named these measures as conftors (conformational vectors).

Comparison of membrane predictors

The location of the hydrophobic bilayer core and the tilting of the protein in the bilayer are important features of TM proteins. However, atomic resolution experimental information on these data is limited. Therefore, I selected a TM predictor to derive this information. In order to choose the best one, I compared three TM predictors, including PPM (OPM), TMDET and MEMPROTMD (Figure 3).

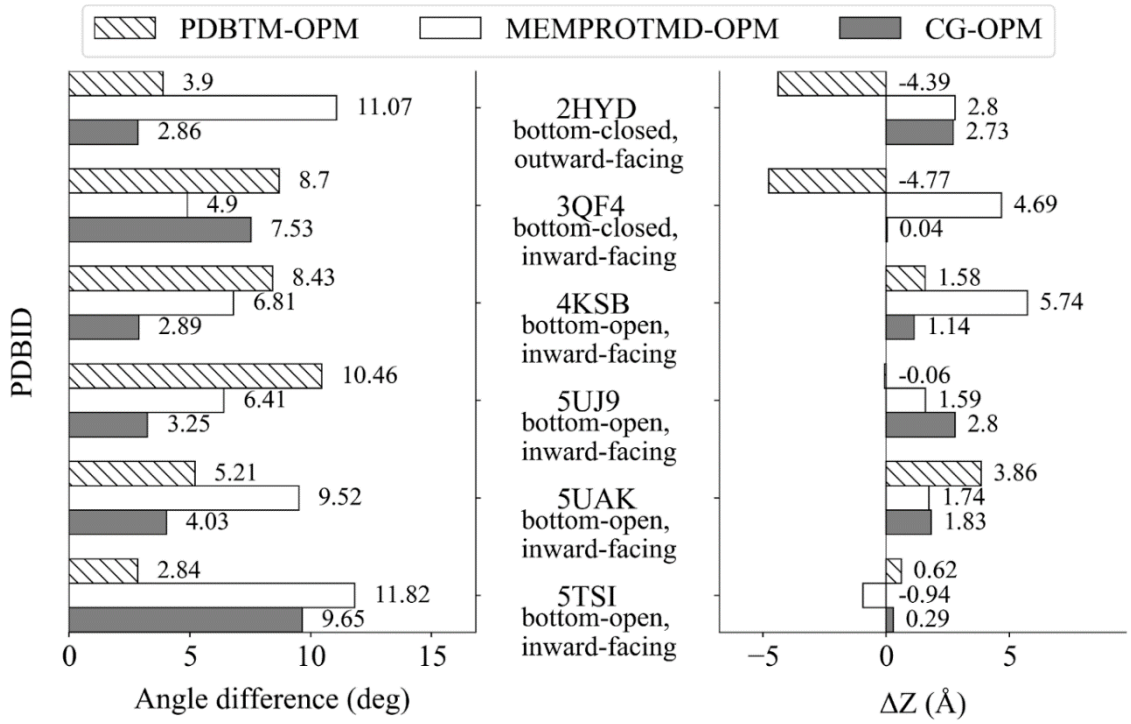

Figure 3 *Comparing TM region prediction methods. TMDET, MEMPROTMD, and our CG simulations are compared to PPM (OPM). Largest differences in the tilting angle of proteins in the membrane (left) and the z-positioning of the membrane bilayer around the proteins (17).*

I collected 62 structures of 17 Pgp-like ABC protein exporters for performing the comparison. I calculated the tilting angle and bilayer location of these proteins using data

from these predictors and our coarse-grained (CG) simulations. The values of TMDET, MEMPROTMD and our CG simulations were compared to the values from PPM.

Protein tilting is the angle between the membrane normal and the principal axis of the protein, which axis is the bisector between the THX1 and THX2 conftors (Table 1, Figure 4). These conftors represent the two crossover (X) transmembrane helices (TH) in each half of the protein. The largest differences in tilting were 10-12°, which corresponds to a few amino acids difference in membrane immersion.

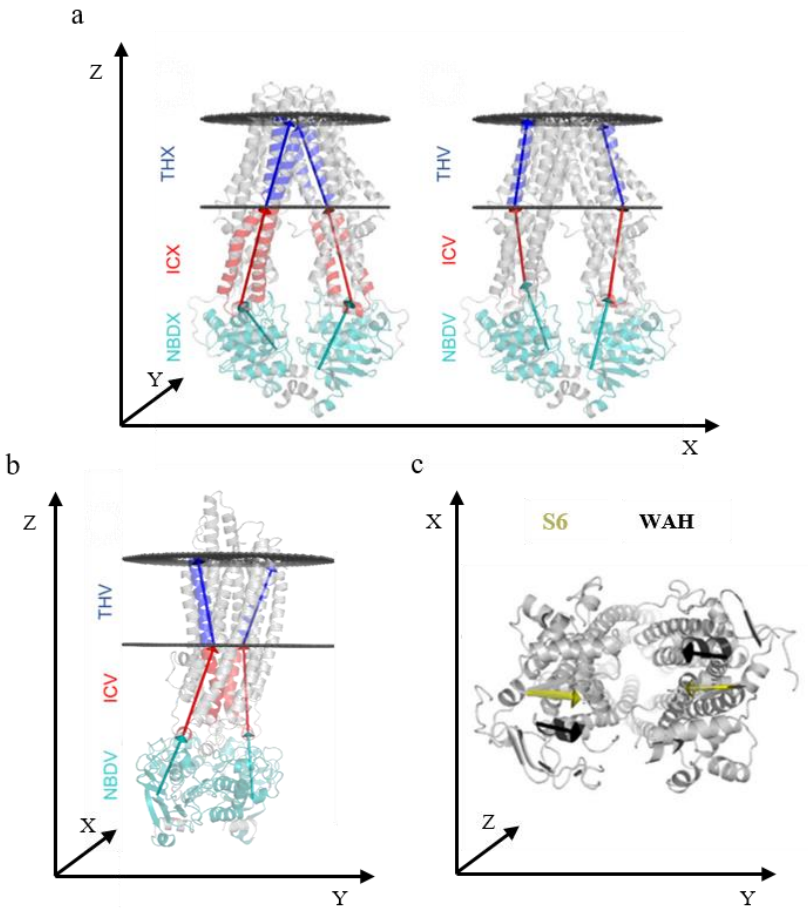

Figure 4 *ABC conftors in 3D.* ***a)****,****b)*** *Vectors point towards the extracellular space. Conftors for the TMDs, ICDs and NBDs are blue, red and teal, respectively.* ***c)*** *Conftors representing the Walker A helices (WAH) and strand S6 are black and yellow with black line, respectively (17) (modified).*

The bilayer location around the protein was characterized by the center of geometry (COG) of the DUMMY membrane atoms from PPM or of PO4 and NC3 beads for CG structures. In terms of their placement along the z-axis of the membrane, the

structures 2HYD, 3QF4, 4KSB, and 5UAK exhibited differences greater than 3 Å, but in no case did these differences exceed 6 Å. To illustrate these differences, I use the 5UAK CFTR structure, where I demonstrate a nearly 4 Å difference between TMDET and PPM (Figure 5).

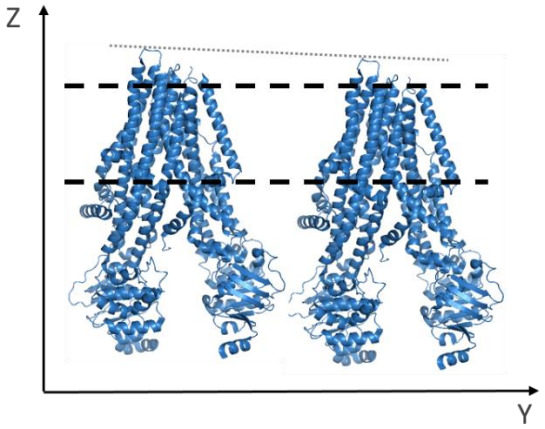

Figure 5 *A difference of 3.86 Å in membrane location (PDBID: 5UAK).*

In order to select which predictor should be used, I calculated the membrane solvation energy of these structures (Figure 6) while immersing them differently into the bilayer. This was achieved by moving the bilayer by 1 Å steps up and down around the center of the putative transmembrane region.

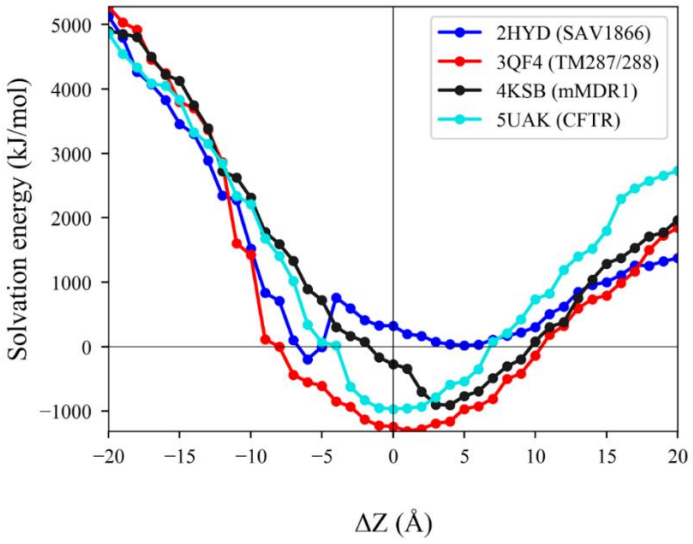

Figure 6 *Membrane insertion assessed by APBS calculations. Zero Å equals to the middle of the membrane leaflet defined by PPM. For each calculation, the middle point of the bilayer was moved by 1 Å steps (17).*

These calculations showed that the membrane solvation energy of 3QF4 and 5UAK aligns with PPM, whereas MEMPROTMD and PDBTM provided better membrane insertion for the 4KSB and 2HYD structures, respectively. In conclusion, the assessed membrane prediction methods produced similar membrane positions. I used the PPM method in this study, because that is the most well-know.

Application of conftors

My aim was to define measures that can highlight the differences in intra- and inter-domain arrangements (Table 1, Figure 4). In the case of ABC proteins, I compared their opening level by defining three set of vectors (THV1-2, THX1-2 and THC3-9, Table 2) The angle between THV conftors was able to separate the occluded (bottom-closed, top closed) conformations from the other conformations. However, bottom-open, inward facing and occluded conformational states were distinguished from the others by the angle between the THX conftors (Table 2).

I also defined conftors representing the intracellular domains (ICDs, Figure 4). ICV conftors differenciated between the inward-facing, the outward-facing and the occluded conformations (Table 2).

While the openness of NBDs can be easily defined based on the 3D structure, their rotation and orientation are harder to inspect. Therefore, I also defined conftors describing the orientation of NBDs. Among these, the length of the NBDX1&NBDX2_ext conftor can differentiate bottom-open inward facing and bottom-closed inward facing conformations from the occluded and outward facing ones (Table 2). I also specified conftors that represents the S6 strands and the helix of the Walker A motif (conftor S6 and WAH, respectively, Figure 4). These conftors can be used to monitor the relative orientation of NBDs.

Conftors can also be used for visual differentiation of bottom-open, inward facing, occluded and bottom-closed, outward facing conformations by projecting them to 2D (Figure 7). Though inward facing conformations are generally easy to separate from other conformations, occluded and outward facing conformations are harder to be distinguished from each other by visual inspection of the all-atom structures. However, THX conftors projected to the YZ plane and THC conftors simplify the visual differentiation of these conformations.

Table 1 *ABC conftors. COG: center of geometry, ext: extracellular membrane plane, int: intracellular membrane plane*

ICC3	$\overrightarrow{COG(CH1_{start}, CH1_{end})TH3_{int}}$
ICC9	$\overrightarrow{COG(CH3_{start}, CH3_{end})TH9_{int}}$
ICV1	$\overrightarrow{COG(CH1_{start}, CH1_{end})TH2_{int}}$
ICV2	$\overrightarrow{COG(CH3_{start}, CH3_{end})TH8_{int}}$
ICX1	$\overrightarrow{COG(CH2_{start}, CH2_{end})COG(TH4_{int}, TH5_{int})}$
ICX2	$\overrightarrow{COG(CH4_{start}, CH4_{end})COG(TH10_{int}, TH11_{int})}$
NBD1_S6	$\overrightarrow{NBD1S6_{start}NBD1S6_{end}}$
NBD2_S6	$\overrightarrow{NBD2S6_{start}NBD2S6_{end}}$
NBDC3	$\overrightarrow{COG(NBD1S8_{start}, NBD1S9_{end})COG(CH1_{start}, CH1_{end})}$
NBDC9	$\overrightarrow{COG(NBD2S8_{start}, NBD2S9_{end})COG(CH3_{start}, CH3_{end})}$
NBDV1	$\overrightarrow{COG(NBD1S8_{start}, NBD1S9_{end})COG(CH1_{start}, CH1_{end})}$
NBDV2	$\overrightarrow{COG(NBD2S8_{start}, NBD2S9_{end})COG(CH3_{start}, CH3_{end})}$
NBDX1	$\overrightarrow{COG(NBD2S8_{start}, NBD2S9_{end})COG(CH2_{start}, CH2_{end})}$
NBDX2	$\overrightarrow{COG(NBD1S8_{start}, NBD1S9_{end})COG(CH4_{start}, CH4_{end})}$
THC3	$\overrightarrow{TH3_{int}TH3_{ext}}$
THC9	$\overrightarrow{TH9_{int}TH9_{ext}}$
THV1	$\overrightarrow{TH2_{int}TH2_{ext}}$
THV2	$\overrightarrow{TH8_{int}TH8_{ext}}$
THX1	$\overrightarrow{COG(TH4_{int}, TH5_{int})COG(TH4_{ext}, TH5_{ext})}$
THX2	$\overrightarrow{COG(TH10_{int}, TH11_{int})COG(TH10_{ext}, TH11_{ext})}$
WAH1	$\overrightarrow{WAH1_{end}WAH1_{start}}$
WAH1-SIG2	$\overrightarrow{WAH1_{start}SIG2_{start}}$
WAH2	$\overrightarrow{WAH2_{end}WAH2_{start}}$
WAH2-SIG1	$\overrightarrow{WAH2_{start}SIG1_{start}}$

DOI:10.14753/SE.2025.3101

Table 2 *Angles between conftors and lengths of conftors. Bold: mentioned in text (17).*

	bottom-open, inward-facing (n=27)	**bottom-closed, inward-facing (n=6)**	**bottom-closed, top-closed (n=3)**	**bottom-closed, outward-facing (n=4)**
	Angle between conftors, deg			
TH1&TH3	32.79 (±0.96)	39.78 (±1.45)	34.85 (±7.46)	50.92 (±4.61)
TH7&TH9	36.48 (±1.14)	34.53 (±1.66)	29.64 (±8.59)	48.33 (±5.21)
THX1&THX2	**46.13 (±1.78)**	39.73 (±2.35)	**27.25 (±5.91)**	35.04 (±2.38)
ICX1&ICX2	50.67 (±1.77)	44.58 (±2.17)	52.31 (±5.46)	54.02 (±2.81)
NBDX1&NBDX2	74.68 (±2.02)	75.36 (±4.87)	85.32 (±3.63)	82.84 (±0.72)
THV1&THV2	37.85 (±1.31)	36.76 (±2.91)	**22.79 (±4.49)**	40.35 (±4.49)
ICV1&ICV2	**42.70 (±1.24)**	37.96 (±3.32)	**53.39 (±2.62)**	**60.12 (±2.01)**
NBDV1&NBDV2	26.73 (±1.88)	27.39 (±4.56)	34.23 (±3.22)	27.74 (±2.57)
WAH1&WAH2	104.21 (±3.77)	123.09 (±5.20)	104.20 (±9.10)	110.49 (±3.59)
NBD_S91&NBD_S92	164.53 (±1.60)	172.13 (±2.60)	173.36 (±3.14)	174.83 (±1.17)
THX1&ICX1	13.64 (±1.05)	19.88 (±3.12)	18.85 (±5.16)	13.76 (±2.09)
THX2&ICX2	14.52 (±1.05)	16.30 (±2.91)	23.07 (±3.41)	15.65 (±3.32)
ICX1&NBDX1	37.68 (±1.55)	37.71 (±2.25)	28.23 (±7.81)	30.91 (±3.89)
ICX2&NBDX2	37.53 (±1.25)	40.41 (±1.86)	37.89 (±1.38)	34.82 (±3.06)
THV1&ICV1	18.88 (±1.59)	17.47 (±0.91)	21.14 (±7.78)	16.67 (±4.26)
THV2&ICV2	19.24 (±1.52)	20.70 (±1.28)	30.94 (±11.66)	15.63 (±2.91)
ICV1&NBDV1	14.77 (±0.99)	12.39 (±1.02)	13.78 (±3.30)	15.52 (±0.76)
ICV2&NBDV2	14.38 (±1.37)	13.36 (±1.58)	15.41 (±7.04)	17.60 (±2.45)
	Distances between conftor ends, Å			
THX1&THX2_ext	30.97 (±0.59)	29.68 (±1.76)	35.09 (±2.25)	40.63 (±3.12)
ICX1&ICX2_ext	35.39 (±0.68)	35.91 (±1.27)	33.36 (±2.93)	32.50 (±1.10)
NBDX1&NBDX2_ext	49.46 (±1.66)	38.93 (±1.66)	**28.57 (±0.35)**	**28.70 (±0.95)**
NBDX1&NBDX2_int	47.05 (±2.12)	27.84 (±4.09)	39.50 (±0.55)	40.04 (±1.38)
THV1&THV2_ext	32.64 (±0.67)	33.35 (±1.14)	28.23 (±2.26)	40.47 (±2.26)
ICV1&ICV2_ext	41.77 (±0.86)	37.49 (±0.85)	31.71 (±3.67)	29.69 (±1.49)
NBDV1&NBDV2_ext	43.96 (±1.40)	32.07 (±2.21)	28.30 (±2.23)	30.31 (±0.83)
WAH1&WAH2_int	59.70 (±1.64)	44.80 (±1.47)	40.83 (±0.90)	42.84 (±0.82)
NBD_S91&NBD_S92_ext	42.33 (±2.04)	29.50 (±2.07)	20.81 (±0.37)	20.75 (±0.05)

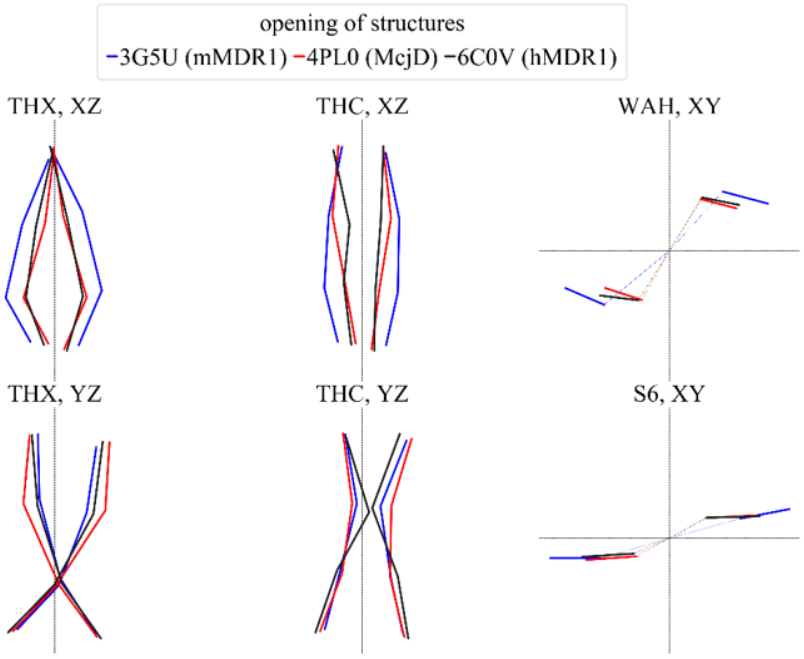

Figure 7 *2D projection of conftors. "Bottom-open, inward-facing" conformation (PDBID: 3G5U) (13) can be visually differentiated from the "occluded" (PDBID: 4PL0) and "bottom-closed, outward-facing" (PDBID: 6C0V) (73) conformations by THX and THC conftors (17) (modified).*

I use CFTR as an example of applying conftors for visually inspection of structures. The apo cryo-EM structure of CFTR (PDBID: 5U71) (74) is similar to other inward-facing Pgp-like ABC structures. This structure does not exhibit an open chloride pathway. Even an ATP bound and phosphorylated CFTR structure (PDBID 5W81) is closed for chloride current (75). Therefore, Das *et al.* (76) generated conformations with open chloride channel by MD simulation using cross-linking, labeling, mutagenesis and single channel measurements as experimental constraints during pore construction. However, projecting the conftors of these models to 2D revealed that they were very different from known ABC protein structures (Figure 8).

These type of differences is challenging to recognize in 3D structures, but the WAH and S9 conftors showed that the NBD rotations in the open Das-model were largely different from other CFTR structural models. For the closed Das-model, THX conftors revealed differences in their TM helix conformations when compared to other closed CFTR models. Furthermore, there are not any known ABC structure, in which the relative orientation of the NBDs and the TMDs is similar to those observed in the closed Das-

model, shown by conftor WAH. The differences in the transmembrane regions may also exist *in vivo*, since the channel gating may need different conformations compared to active transport.

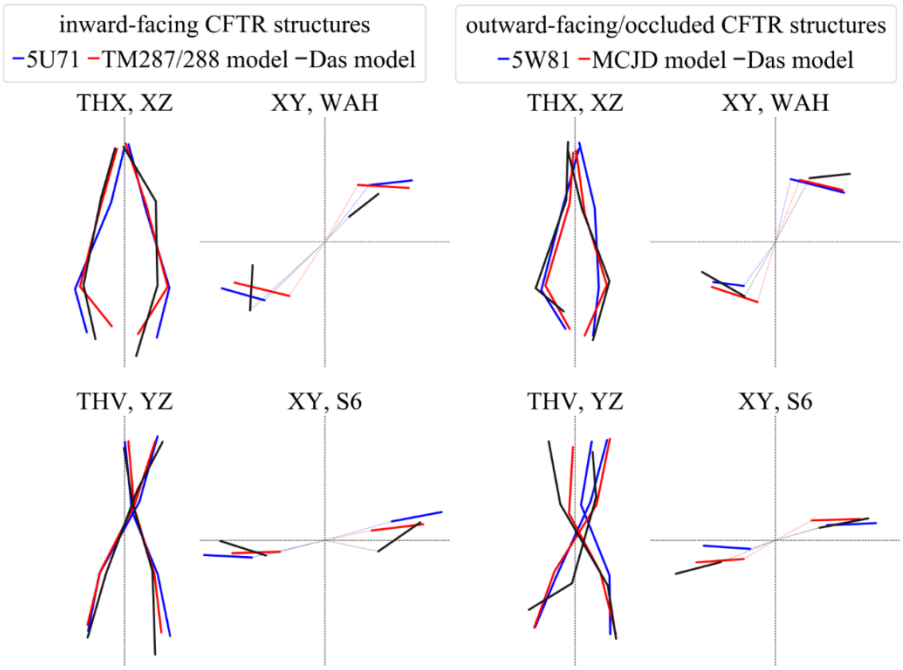

Figure 8 *Conftors emphasize key similarities and differences among various CFTR structural models. The CFTR cryo-EM structures are represented by PDB IDs 5U71 and 5W81. Homology models based on TM287/288 and McjD were generated by Corradi et al., while CFTR models from Das et al. are available at http://troll.med.unc.edu/cftr/ (17).*

Previously we performed MD simulations of the inward-facing CFTR cryo-EM structure (PDBID: 5UAK) (74, 77), where NBDs were closing during the simulations. To gain a better understanding of these events, we used conftors to analyze the trajectories (Figure 9). The bottom of the NBDs (the side distant from the membrane bilayer) got closer to each other, shown by the increase in the NBDX1/2 angle. I used the NBDX1/2_int distance conftor for further investigation. This is the distance between the COGs of the S8 and S9 strands in both NBDs (Figure 4), which are at the intracellular side (bottom) of the NBDs. NBDX1/2_int kept fluctuating till ~18 ns of our simulation then this distance stabilized. This observation suggested that the NBDs became stabilized slower than other parts of the protein. As CFTR has a nonfunctional degenerate ATP-binding site (site-1), I aimed to monitor whether the NBDs became closed in a

physiologically relevant mode. Therefore, I used another conftor to follow the distance between Walker A and Signature motifs. This conftor showed that the two NBDs associate asymmetrically, as motifs in the nonfunctional degenerate ATP-binding site-1 (WAH1-SIG2) were more separated than the motifs in the canonical site-2 (WAH2-SIG1).

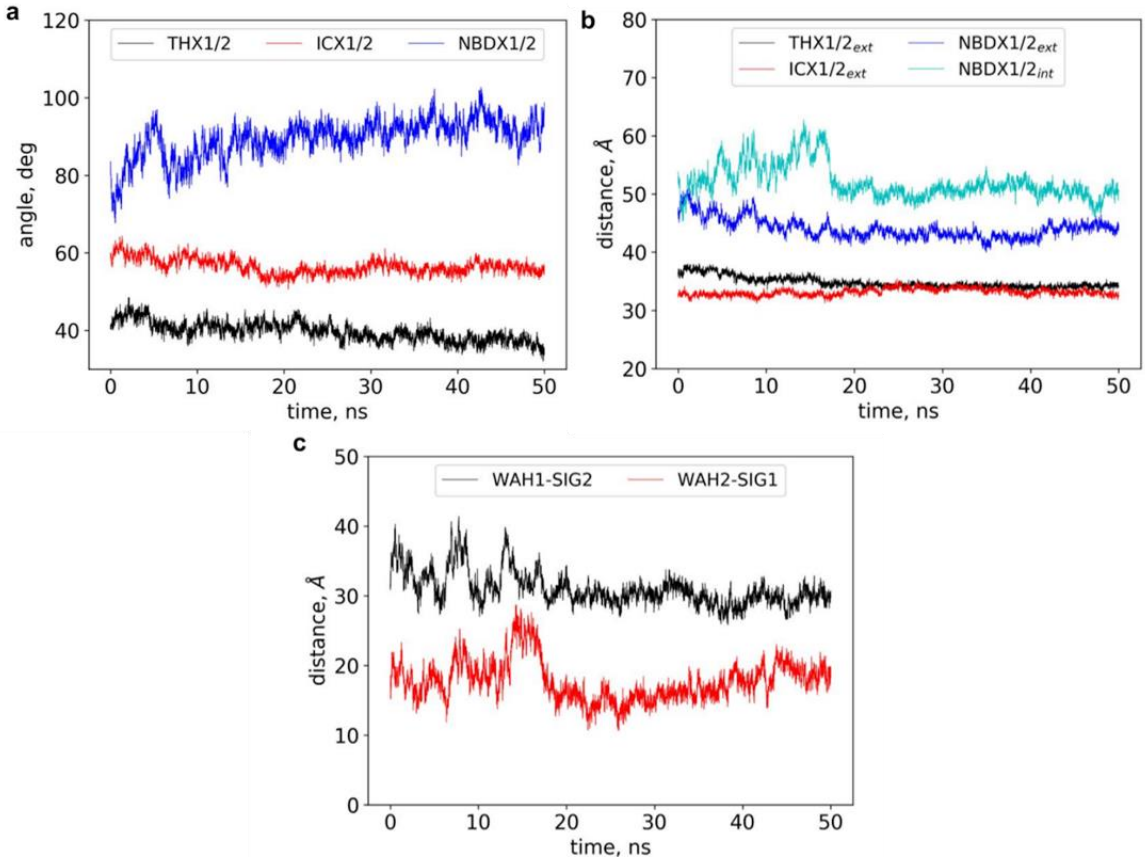

Figure 9 *MD analysis using conftors. The trajectory of an MD simulation with the CFTR bottom-open, inward-facing structure (PDBID: 5UAK) was used to calculate the angles between conftors (**a**) and lengths of conftors (**b**, **c**) (17) (modified).*

Electrostatics calculations

Though conftors can be used to compare and differentiate structures, they do not provide information about the overall quality of a structure on their own. Therefore, I also used other measures describing physicochemical properties to evaluate structures. I utilized and calculated the membrane solvation energy for each ABC protein structure (Figure 10a) to assess their stability in the membrane. The calculation was done for 36 structures. One of the inward-facing structures (PDBID: 3WME) (78) exhibited high

membrane solvation energy, because the structural file contained the coordinates of only half of the protein. Both of the two outward-facing conformations with widely open TMDs (PDBIDs 3B60 and 2HYD) (9, 16) also exhibited high solvation energies. However, the solvation energy of the two other bottom-closed, outward facing structures was small (PDBIDs 6C0V and 6BHU) (73, 79). These structures are less open to the extracellular space than the formers.

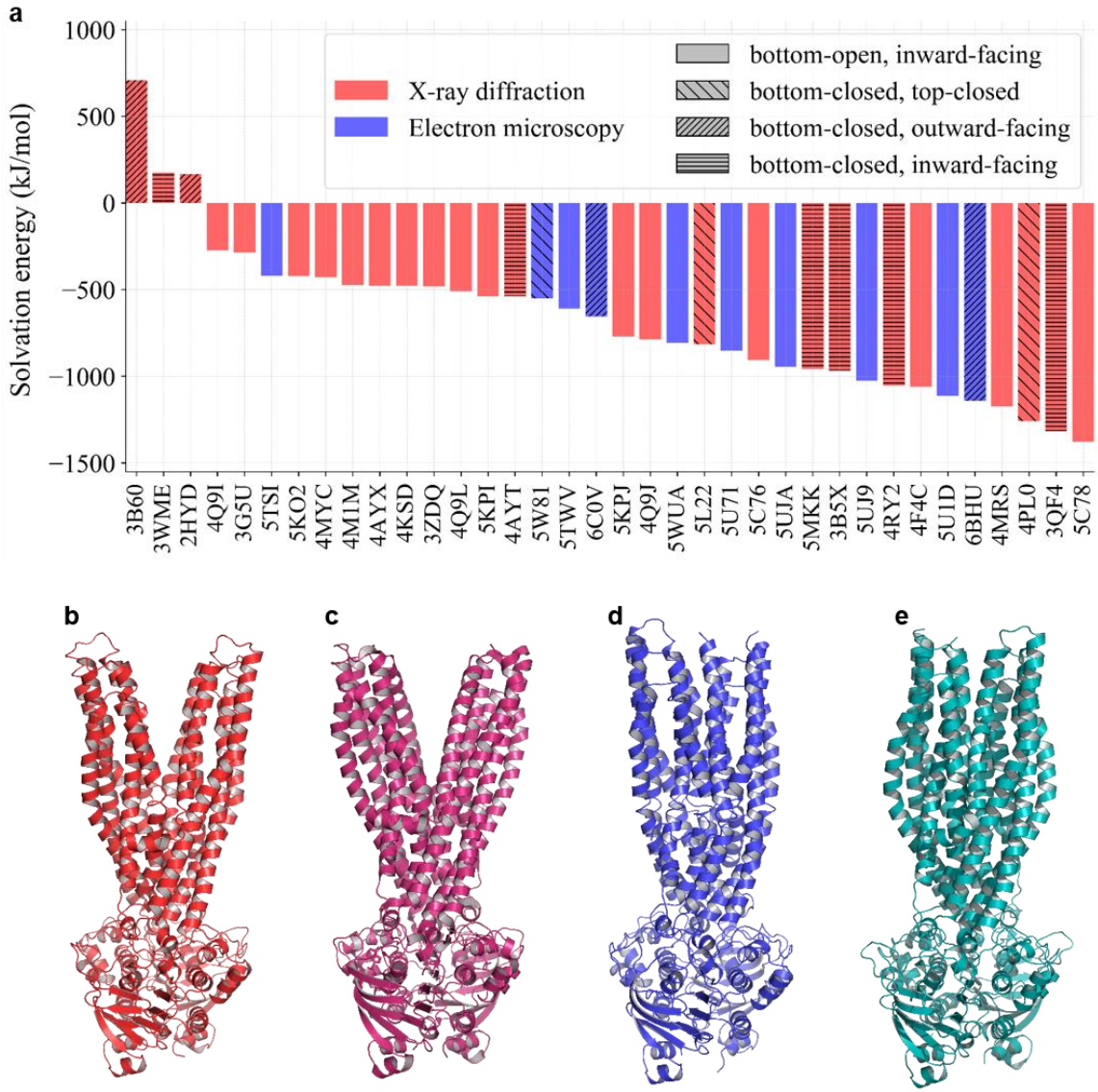

***Figure 10 a**) Membrane solvation energies. Variations in the degree of opening are observed for Sav1866 (PDBID: 2HYD) (**b**), MsbA (PDBID: 3B60) (**c**) and Pgp/ABCB1 (PDBID: 6C0V) (**d**). The wide opening of Sav1866 has been questioned and an alternative ATP-bound conformation has been suggested (Protein Model Database: PM0075213) (**e**) (17) (modified).*

I also used protein surface electrostatics to characterize the properties of TMD conformations. According to the positive-inside rule, positive amino acids are located at the intracellular ends of TM helices because of the negatively charged lipids in the intracellular bilayer leaflet, and form a positively charged ring in most of the transmembrane proteins (e.g. calcium ATPases) (Figure 11a) (80-82). The quantitative description of this ring is challenging and currently we can study this phenomenon in individual cases without automation. I also observed such a positively charged ring in the case of ABCC structures. However, at the site where the L0/Lasso motive interacts with ABCC1/MRP1 and ABCC7/CFTR, a hydrophobic spot breaks the ring (Figure 11b,c). This indicates the binding location of a protein partner, which is the intramolecular N-terminal region in this case.

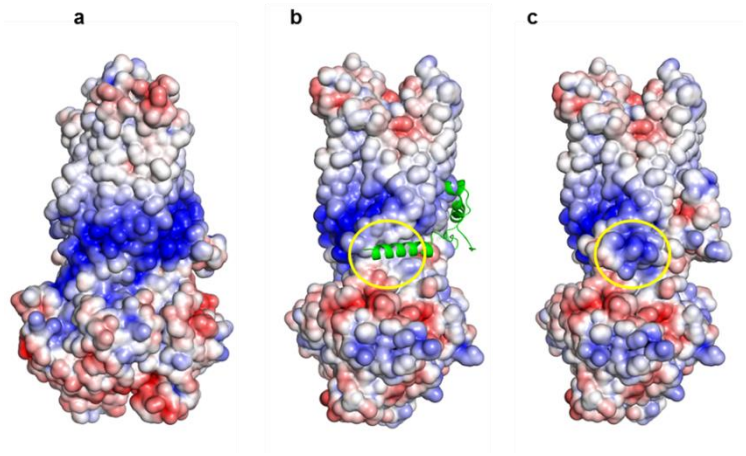

Figure 11 *The intracellular positive blue ring is broken by a hydrophobic patch corresponding to a protein-protein interaction site.* ***a****) A positively charged ring is located at the intracellular side of the membrane. ABCC1 (PDBID: 6BHU) is used as an example.* ***b****) Calculating surface electrostatics after the removal of the N-terminal L0/lasso region (green cartoon) shows a disruption of the positive ring by a hydrophobic patch (yellow circle).* ***c****) If the N-terminal L0/lasso region is present then the patch is covered and a significant level of positive charge is observed (yellow circle) (17).*

4.2. MemBlob

Although the position of the transmembrane region is an important property of transmembrane proteins, only low amount of experimental information is available at high resolution. Since we noticed that cryo-EM density maps of transmembrane proteins contain the densities of the membrane mimetics in numerous cases, we made a pipeline

called MemBlob to convert these densities to membrane boundaries. I tested the MemBlob pipeline for 92 TM protein structures determined by cryo-EM with at least 4 Å resolution.

Approximately 30% of the collected maps lacked well-defined densities for the membrane. One of the possible reasons of this was the absence of a well formed lipid environment during structure determination. In Figure 12, two maps of KATP channel (EMD-7073 and EMD-7338) demonstrate this issue.

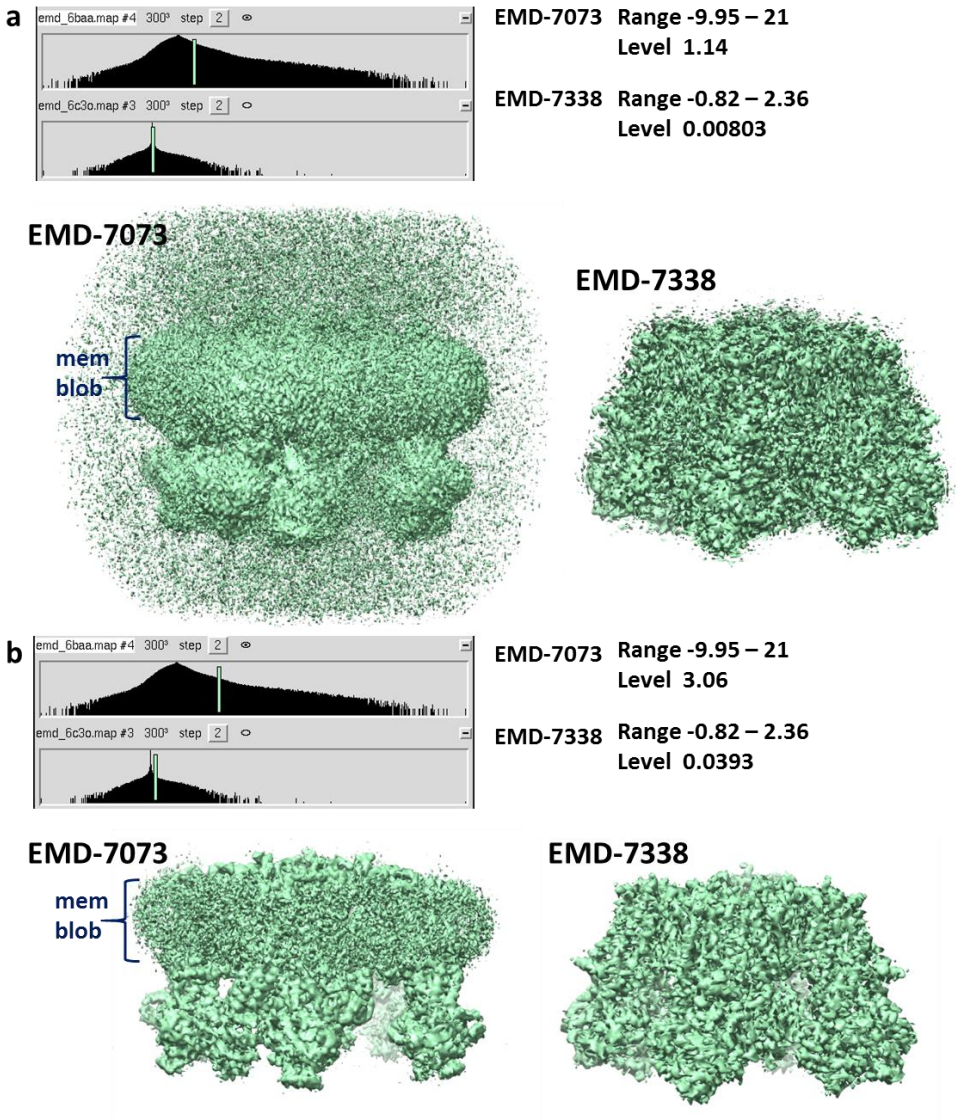

Figure 12 *Many EM density maps lack a membrane blob. Densities in EMD-7073 (PDBID: 6BAA, micelle) (83) and EMD-7338 (PDBID: 6C3O, amphipol) (84) are shown with a lower (**a**) and a higher (**b**) cutoff levels, using Chimera. The protein is ABCC8/Kir6.2 complex in both cases (65).*

I used Chimera to visualize densities with lower (a) and higher (b) cutoff levels, at which a blob was visible in EMD-7073 solved in a micelle, but not in the case of EMD-7338. Though EMD-7338 was determined in the presence of amphipol, densities of the membrane mimetic are missing. This is the case for 14%, 22% and 20% of structures solved in micelle, nanodisc and amphipol, respectively. In 8% and 12% of density maps determined in the presence of micelle and amphipol, respectively, the signal to noise ratio was so low that the membrane blob boundaries cannot be detected.

Interestingly, the absence of a membrane blob not necessary inhibited the MemBlob pipeline. In one instance (PDBID: 5H3O) (85), the pipeline detected reasonable TM regions in the absence of membrane densities. In this case, densities of lipid molecules intercalated between the TM helices provided the data for successful TM region calculation.

Comparing MemBlob results to TMDET

We aimed to compare the TM boundaries determined by our MemBlob pipeline to those predicted by one of the reliable TM predictors, TMDET (figure 13). We chose TMDET, as we utilized it to improve the reliability of our calculations.

Since TMDET, similar to other TM predictors, output a membrane slab with two parallel planes, we calculated the average thickness of the MemBlob lipid boundaries (Figure 13 a, b). To achieve this, we averaged the z-coordinates of the boundaries for both sides and subtracted one from the other. We also decreased this distance by twice of the average interface region thickness (2 x 8 Å).

There were only four cases, where the membrane center calculated by us differs from the TMDET center by more than 5 Å (Figure 13c). These structures included PDBID: 5U70 (dz = -7.4 Å), PDBID: 5UAK (dz = 6.9 Å), PDBID: 5UJ9 (dz = -5.1 Å), and PDBID: 5H3O (dz = -6.1 Å) (74, 85-87). In the latter case the difference was most likely caused by the lack of membrane densities as discussed above.

Plotting the distribution of membrane thickness as calculated by MemBlob and TMDET revealed that the MemBlob pipeline often identifies a thicker membrane environment than TMDET (Figure 13d). The average membrane interface thickness, set at 8 Å in our calculations, does not explain the observed increase in hydrophobic core thickness, as the uncertainty in the interface region is not comparable to the difference in bilayer thickness values.

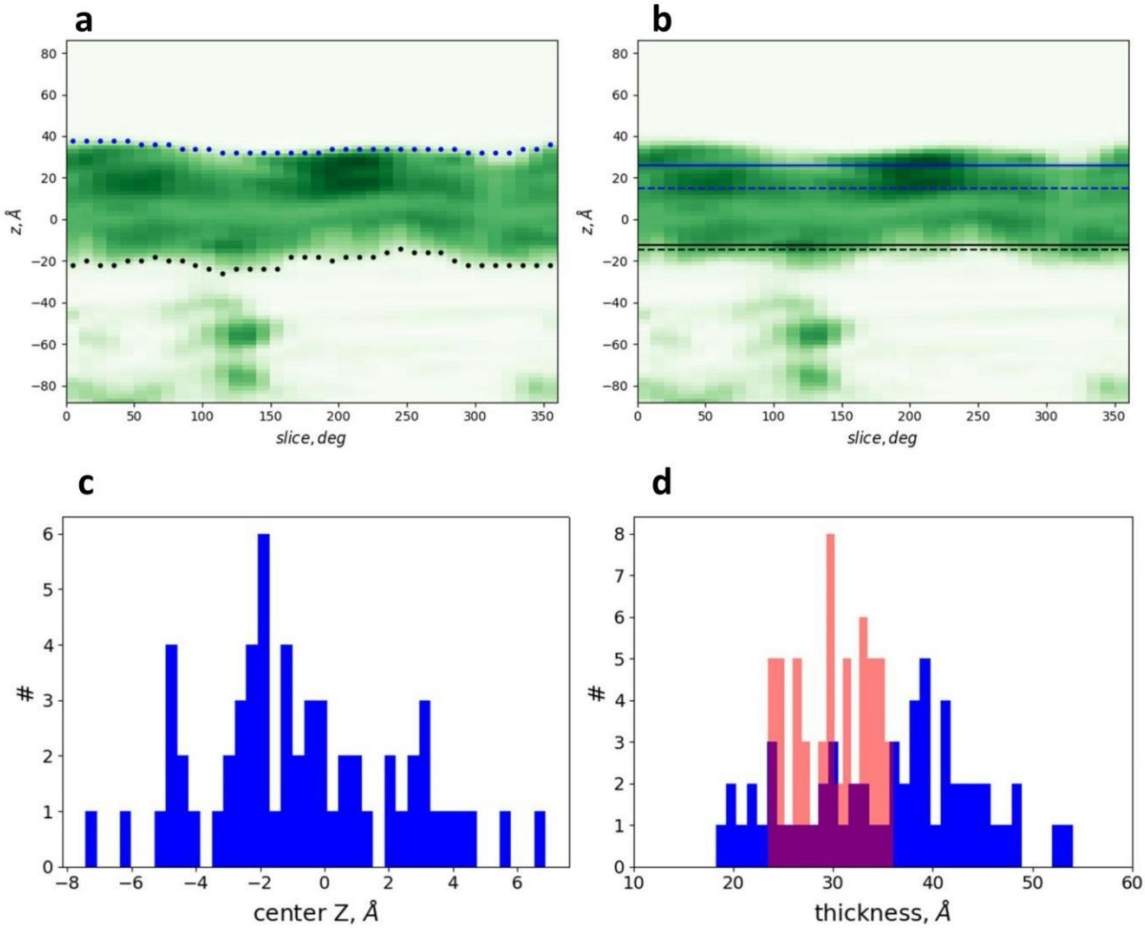

Figure 13 *Comparing membrane boundaries from MemBlob and TMDET. I use hCFTR as an example (PDBID: 5UAK). The MemBlob determines edges that follows the shape of the membrane blob (**a**). Averaging the z-coordinates of these edges (solid lines) and subtracting the average interface region thickness (8 Å) from both sides allows the comparison with the membrane slabs of predictors (TMDET, dashed lines) (**b**). Distribution of the bilayer center of proteins with a resolution at least 4 Å (**c**). Distribution of membrane thickness, calculated by MemBlob (blue) and TMDET (coral). (**d**) (65).*

In many cases we observed deep embedment of the protein into the membrane. Thus, short regions that were predicted to be extracellular, were indicated intramembranous by our pipeline. For example, the extracellular regions of the ABCC8/Kir6.2 KATP channel structure are located in a pit formed by the edge of the membrane blob (Figure 14).

DOI:10.14753/SE.2025.3101

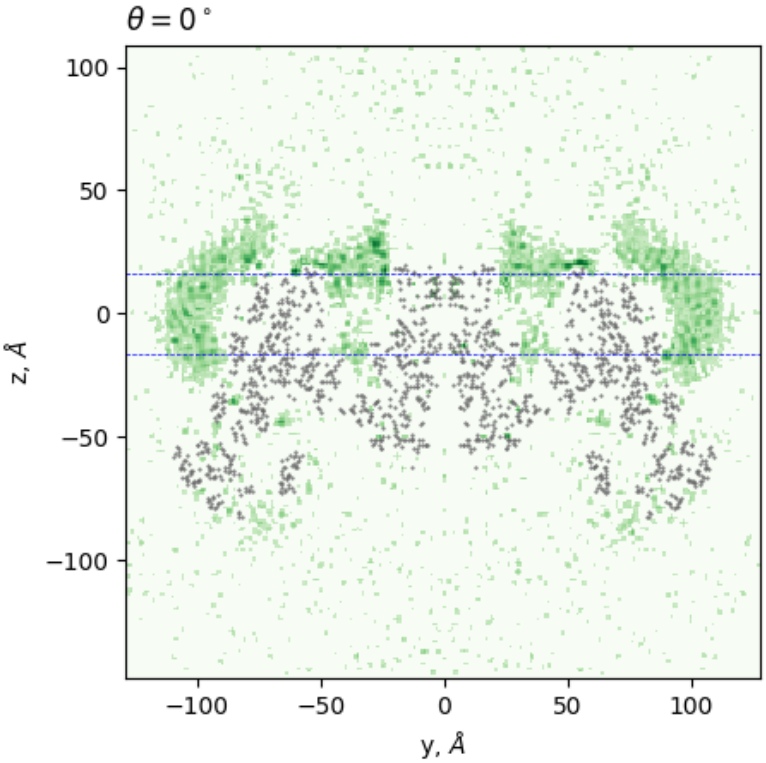

Figure 14 *Some experimental membrane blobs are wider than expected. This is shown on a ABCC8/Kir6.2 structure (PDBID: 6BAA, grey dots) (83). Remaining densities after the subtraction of the protein density from the EM map (EMD-7073) are green, protein densities are grey. Dotted blue lines represent the membrane slab predicted by TMDET (65).*

I also investigated whether the embedment is influenced by the type of the membrane mimetics applied during structure determination (Figure 15). I found that nanodisc produced mostly similar membrane thickness with TMDET, while this is true for only half of the structures solved in amphipols. Our pipeline calculated membrane regions thicker than TMDET thickness for sixty percent of the structures solved in micelle. This could be because the micelle adheres to the hydrophobic regions in a spherical shape, and the size of the sphere depends on the composition of the detergent.

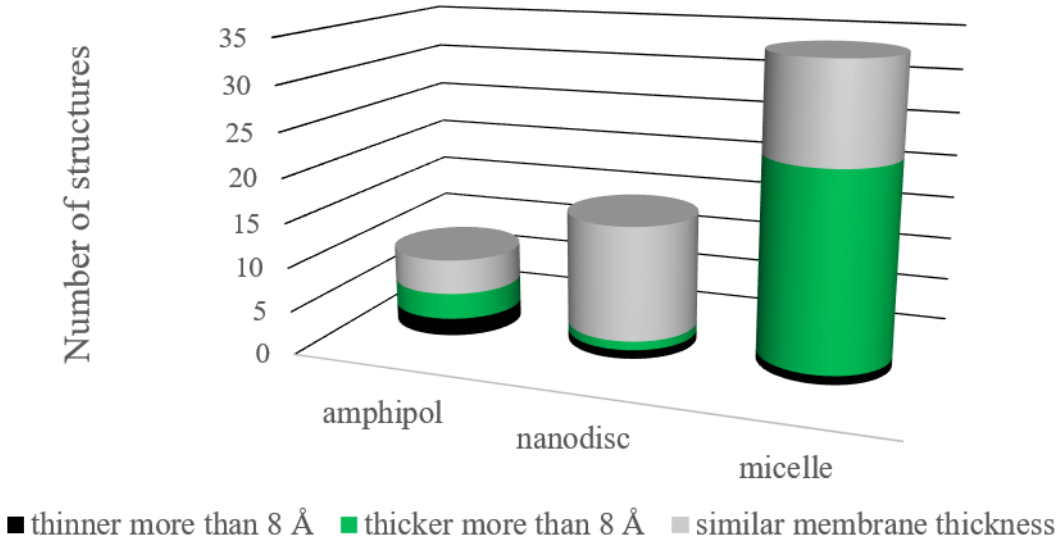

Figure 15 *Type of the used membrane mimetics and the thickness of their blob.*

Webserver

We developed a web application available at memblob.hegelab.org that allows users to utilize our pipeline using a graphical interface. The required inputs are a Cryo-EM density file in MRC format and the corresponding all-atom structure in PDB format. Providing a TMDET XML file is optional. The downloadable results include images of y-z cross-sections at every 10° and the plot of the densities. An interactive structural model is included in the result page for visualization and to help the checking and manual adjustment of the membrane boundaries. This manual correction is done by selecting expressions and simple commands listed in Table S1 in (65).

4.3. MemMoRF

Collection of MemMoRFs

There are several known cases, when IDRs (Intrinsically Disordered Regions) bind to biological membranes. However, this type of interaction is rarely studied in detail. Furthermore, there is a debate in the field, whether membrane binding facilitates the folding of the IDR (45). As investigations of protein-lipid interactions at atomistic level are most widely done by NMR, I collected NMR structures of transmembrane proteins (TMPs) and membrane associated proteins (MAPs). TMP structures were collected using the Membrane Protein Browser of RCSB PDB. I searched for MAPs using the “peripheral membrane protein” keyword for subcellular location in the UniProt database and their

structures solved by NMR in RCSB PDB using the UniProt accession number of the matches.

I downloaded the associated chemical shifts from BMRB, which were used to calculate SSPop and flexibility for identification of IDRs. In many cases, there were not any available chemical shift data or the files did not contain enough residues for the calculations. Therefore, SSPop and flexibility data (1-S2) were calculated only for 41% of the collected proteins. I considered a region as an IDR if it was flexible (1-S2 ≥ 0.15) and was in "coil" structure (SSPop ≥ 0.5). I also looked up disorder annotations in DisProt, DIBS, MFIB and PFAM databases. In addition, I searched for IDRs that can bind to membranes from the articles published along the structures. Finally, I checked collected X-ray and cryo-EM structures for unresolved residues that resulted in the identification of eleven additional MemMoRFs.This way, I collected 206 TMPs and 332 MAPs (Figure 16).

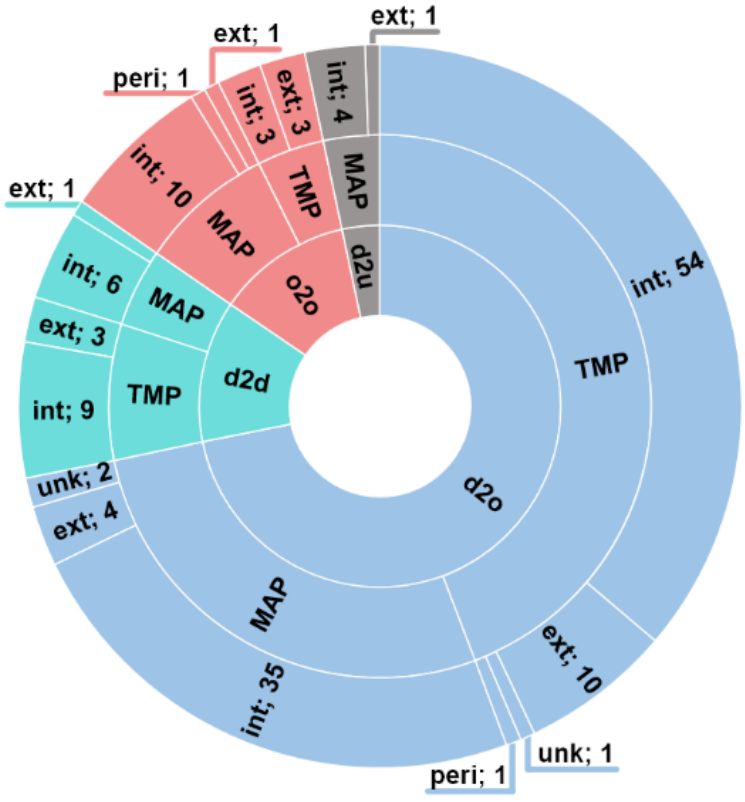

Figure 16 *Distribution of MemMoRF types in the database. d2o: disorder-to-order, d2d: disorder-to-disorder, o2o: order-to-order, d2u: disorder-to-unknown, TMP: transmembrane protein, MAP: membrane associated protein, int: intracellular side, ext: extracellular side, peri: periplasmic side, unk: unknown location (88).*

Among the collected proteins, 107 proteins included 149 membrane interacting regions. 131 are disordered in aqueous solution and 92 regions of them were not annotated in the DisProt database. 107 of the IDRs undergo disorder to order transition upon binding to membrane mimetics. I found 19 IDRs that maintain their flexibility in their membrane bound form. Therefore, I considered these segments as a new category of membrane interacting IDRs and labeled their transition as 'disorder-to-disorder'. I could not identify the membrane bound state of the IDR in five cases, these entries were categorized as 'disorder-to-unknown. In addition to the 107 IDRs, I also found 18 stable segments, which exhibit secondary structure in both soluble and membrane bound forms (labeled as 'bistable helix' or 'order-to-order'). Among all of the membrane interacting regions, 84 are in TMPs and the remaining 65 are in MAPs. I observed 121 intracellular, 23 extracellular and two periplasmic localizations.

Validation process of MemMoRFs

As an example of the IDR and MemMoRF validation process, I discuss a MemMoRF in integrin β3 in detail (Figure 17).

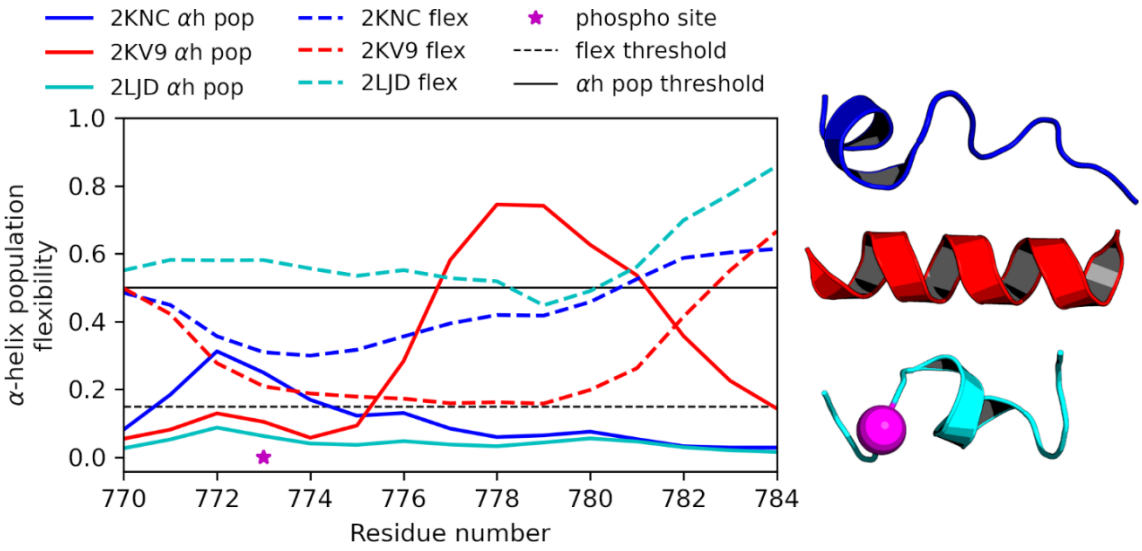

Figure 17 *Experimental NMR data used for MemMoRF identification. Blue: in organic solvent; red: in DPC; cyan: phosphorylated in DPC; magenta: phosphorylation site; αh pop: α-helix population calculated by δ2D; flex: 1-S2 calculated by RCI, αh pop threshold: 0.5, flex threshold: 0.15 (88).*

Integrin β3 plays a role in angiogenesis and tumor growth (89, 90). Its cytoplasmic tails (CT) were used to develop cell permeable peptides for angiogenesis inhibition (91). I found that these peptides overlap with a MemMoRF (a.a. 770-784; Uniprot accession number: P05106). This region is disordered in solution, indicated by its low helical propensity (<0.5) and high flexibility (>0.15) (PDBID: 2KNC and BMRBID: 16496) (92). In addition, this segment is also annotated to be disordered in the DisProt and PFAM databases. The original publication containing the NMR experiments demonstrated this region to interact with membrane mimetics (93). In the presence of a lipid environment (PDBID: 2KV9 and BMRBID: 16771) (93), the flexibility of the peptide decreases and its helical propensity increased. These observations indicated that the region exhibited disorder to order transition upon lipid interactions, thus it was qualified for a MemMoRF label.

Regulation of membrane interaction of IDRs

Phosphorylation of the above discussed MemMoRF in integrin β3 promotes disorder even in the presence of a membrane mimetic (PDB ID: 2LJD and BMRB ID: 17930). The available experimental data allows a detailed description of phosphorylation dependent protein and membrane interactions of this MemMoRF that potentially can lead to a better understanding of the function of this region and improve the therapeutic use of integrin β CT peptides.

Another example of order to disorder rearrangement upon phosphorylation is the detachment of the intracellular domain of T cell receptor CD3-**ζ** chain from the membrane (Uniprot ID: P20963) (Figure 18a). This domain contains three ITAMs (ITAM1-3, immunoreceptor tyrosine-based activation motif), which sequence patterns include two tyrosine residues, which have been shown to serve as phosphorylation sites (94). I found that both ITAM2 and ITAM3 contained a MemMoRF (a.a. 111-126 and 142-156, respectively), which underwent disorder to order transition upon membrane binding. When phosphorylated, the ITAM MemMoRFs are released from the membrane and unfold, which leads to the activation of the receptor (95).

DOI:10.14753/SE.2025.3101

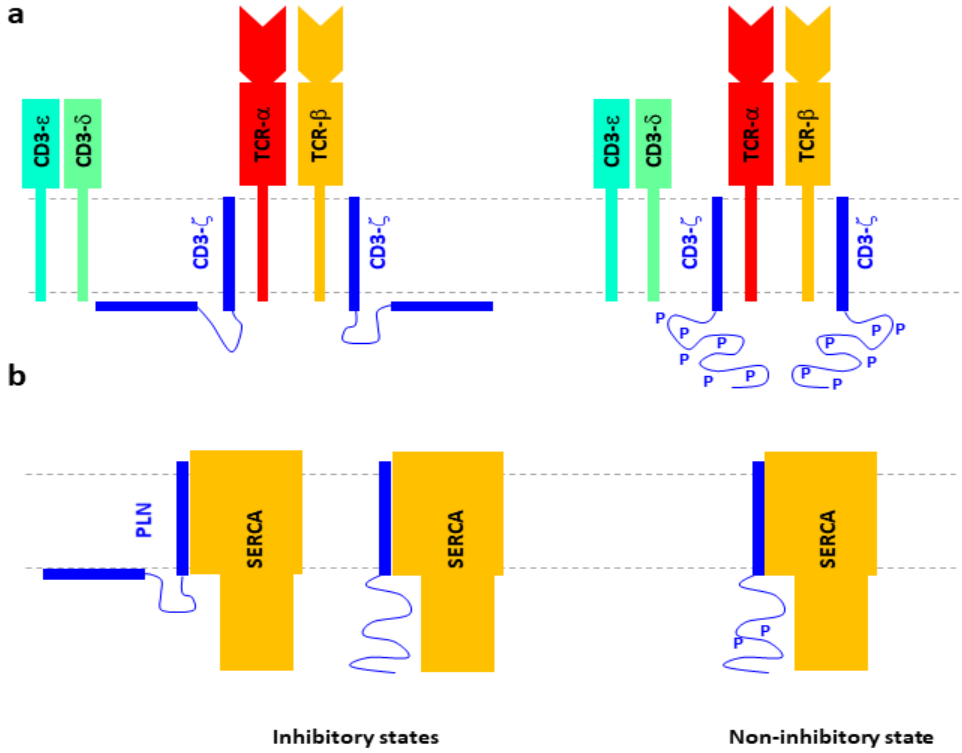

Figure 18 *MemMoRFs in phosphorylation dependent regulation. MemMoRF containing regulatory domains or proteins are blue.* ***a****) The intracellular domain of T cell receptor CD3-****ζ*** *chain contains two MemMoRFs (a.a. 111-126 and 142-156), which undergo disorder to order transition upon membrane binding. Upon phosphorylation, they detach from the membrane and unfold, which leads to the activation of the receptor.* ***b****) The N terminus of PLN contains a MemMoRF (a.a. 2-15). Membrane binding is prevented by phosphorylation near this MemMoRF. This phosphorylated form is non-inhibitory (88) (modified).*

Another example for phosphorylation regulated MemMoRF interactions is the regulation of SERCA (sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase) by phospholamban (PLN) (Figure 18b). The cytoplasmic N terminus of PLN affects the inhibition of SERCA exerted by the TM region of PLN. The N terminal IDR includes a MemMoRF (a.a. 2-15). When membrane bound, PLN inhibits SERCA function. According to the dynamic nature of IDRs, the MemMoRF does not permanently bound to the membrane. However, this unfolded and membrane detached state is also inhibitory. The MemMoRF unfolds ‘permanently’ upon phosphorylation at Ser16 or Thr17, which phosphorylated form is non-inhibitory (96, 97).

DOI:10.14753/SE.2025.3101

Helical content in NMR structures

In the absence of a membrane mimetic, the helical propensity for a.a. 771-776 (PDBID: 2KNC and BMRBID: 16496) in integrin β3 is lower than the commonly used SSPop threshold (0.5). However, all of the deposited structures present a small helix in this region. In the presence of a membrane mimetic, the increased helical propensity indicates an α-helix in a smaller part of the sequence (a.a. 776-781) than expected based on the accompanying structural ensemble (Figure 17). Interestingly, hydrogen-deuterium exchange experiments revealed transient helical contents and increasing disorder towards the C-terminus published in the same paper as the NMR structures (93). Moreover, the NMR ensemble of the phosphorylated CT also features a stable small helix between residues 775-780.

Table 3 *Protein structures containing more stable and longer helices than our calculations suggest. The MemMoRF database includes 110 structures with calculated SSPop and flexibility values.*

Protein	Structure (PMID)
Sodium/potassium-transporting ATPase subunit gamma	2mkv
Myc box-dependent-interacting protein 1	2rmy, 2rnd
Cardiac phospholamban	1fjp
Amyloid-beta precursor protein	2lfm, 2llm, 1aml, 2lpq, 2mj1
Insulin receptor	2mfr
Sec-independent protein translocase protein TatA	2mn6
Receptor tyrosine-protein kinase erbB-2	2n2a
Epidermal growth factor receptor	1z9i, 2n5s
Stannin	1zza
Integrin alpha-IIb	2knc
Voltage-gated potassium channe	2kyh

Based on my preliminary investigation, I observed that 16 other structural ensembles exhibited longer and more stable helices compared to what was initially anticipated based on our calculations (Table 3 and www.memmorf.hegelab.org). Though further investigation is necessary to assess the prevalence of α-helical content

overrepresentation in the case of NMR structures, it is important to note that making a decision about disorder should not rely solely on atomistic NMR structures.

Webserver

We included all identified membrane interacting regions into our MemMoRF database at https://memmorf.hegelab.org, even those that do not undergo disorder-to-order transition upon membrane interaction. Although IDRs that remains disordered upon membrane binding or bistable membrane interacting regions are not MemMoRFs, we inserted them into the database, since knowledge on this kind of regions can also facilitate research in the field. All MemMoRF containing proteins are listed in a table format in the 'Browse' page. The table can be sorted and filtered using dropdown lists or queries. There is also a general search field at the top of the page, which is independent from the columns.

Each protein entry starts with the protein name, accession code, gene name, organism and keywords imported from UniProt. This is followed by a list of MemMoRFs containing their boundaries, the type of transition upon membrane binding, localization and supporting statements from the literature.

Flexibility and SSPop calculations are plotted for each sets of chemical shift data when available. Additionally, Wimley-White hydrophobicity plots and IUPred short predictions are also included (98, 99). The graph can be saved as well as manipulated by toggling the visibility of the calculated data or by zooming in or out. Corresponding NMR experiments can be selected from the right of the graph, where the PDBID of the NMR ensemble, the BMRB ID, the used membrane mimetic and the number of the set are shown. Sequence annotations are also shown at the bottom of the graph. This incorporates MemMoRF regions, TM domain, PFAM domains (69), short linear motifs from ELM database (100), PTMs from PhosphoSitePlus (101), IDRs from DisProt, DIBS and MFIB (70-72). Coverage by PDB entries are also indicated. Corresponding structures are shown and can be manipulated by LiteMol (102). MemMoRFs and TM helices are color coded on the structure to help visual inspection of them. Disease phenotypes, disease causing and polymorphic variations and DrugBank records (103-105) are included to help estimate the pathological roles of the MemMoRFs in the entry.

DOI:10.14753/SE.2025.3101

Finally, links to the IntAct and STRING entries (106, 107) of the protein and a STRING network graph are included at the bottom of the page to help evaluate protein-protein interactions.

5. DISCUSSION

5.1. Conftors and electrostatics calculations

The increasing number of membrane protein structures make it necessary to develop new 3D-bioinformatics methods to investigate them. This is especially true after the publication of the AlphaFold database (108, 109). One of the most important aspects of TM protein structure related tasks is structure comparison. In most of the cases new experimental structures are compared to old ones or to AlphaFold predictions. However, these comparisons lack quantitative and standardized measures distinguishing geometric and physicochemical properties of structures. TM structure analysis could benefit from application of conftors and electrostatic calculations. Conftors can highlight structural inaccuracies, contribute to understanding conformational changes related to function, and to facilitate the analysis of MD simulations. My colleagues investigated the channel opening and folding of CFTR TMDs (110). They concluded that the kinked TH8 observed in several CFTR cryo-EM structures could hinder the channel opening in MD simulations and might be attributed to experimental conditions. This altered TH8 conformation was also highlighted by the THV conftors (Figure 8).

Solvation energies determined from electrostatic calculations can also outline structures and structural regions that need additional and detailed investigation to validate them. Interestingly, membrane solvation energy calculations were not resulted in high energy values for inward-facing ABC structures exhibiting widely separated NBDs. This was in agreement with previous EPR studies of MsbA (111, 112). After the publication of our work, existence of the widely inward-open MsbA conformations were also confirmed in cells (21). However, the widely open bottom-closed, outward facing conformation may be the results of the lack of a membrane bilayer, which would restrict the TMDs, under crystallization conditions.

Lastly, surface electrostatics can also be used to identify protein-protein interaction sites in TM proteins.

In summary, validating and comparing structures using standardized metrics help to describe protein structures more accurately. These processes yield more information on functionally important conformations, which promotes the understanding of the working mechanism of important mutations, ultimately aiding structure-based drug design. These types of metrics are becoming extremely important after predicting millions

of protein structures by AI-based methods. It is important to note that outliers (e.g. values of bond angles or solvation energies) may sign important regions as well, and not necessarily just indicate a problem with the structure. My colleagues recently used some of these ABC conftors to classify novel ABC protein structures (https://abc3d.hegelab.org) (113) and are exploring the possibility how to define them semi automatically and use them in PDBe-KB.

The presented and similar metrics can also be used to analyze other kind of proteins outside of the ABC superfamily. However, in the lack of an automatic algorithm, an expert of the studied protein family should be involved in the definition of the vectors.

5.2. MemBlob

In spite of the increasing number of experimental and AI-based TM protein structures, determining their transmembrane regions remain elusive. Nowadays the state of the art method of membrane protein structure determination is Cryo-EM. Interestingly, many cryo-EM densities contain the density of the membrane environment used during structure determination experiments as well. We exploited this phenomenon and built our MemBlob pipeline, which is the first that calculates membrane localization from experimental data.

Importantly, the MemBlob pipeline differs from prediction methods in the following issues. First, it uses experimental data for determining TM regions and the prediction method (TMDET) is used only for orienting the protein along the z axis. Second, the output of our pipeline is not a slab with parallel edges, but a volume with boundaries that follows the shape of the lipid environment. While MEMPROTMD uses MD simulations to predict a more realistic configuration of the membrane around the protein when compared to slab models, it utilizes only the protein structure as input experimental data (36).

During the analysis of the output set of MemBlob, I found that detergent micelles tend to form thicker blobs than other membrane mimetics. However, one third of the blobs formed by micelles was not thicker. This is in agreement with the observation made by Thongin *et al.,* that the size of the micelle formed by the detergent varies according to the composition of the detergent (29). Although our method has the limitation that the thickness of the hydrophobic core regions can be affected by the type of membrane mimetic used for the cryo-EM structure determination, this could be corrected in the

future by the length of the used detergent. However, it can be assumed that nanodisc-produced membrane thicknesses closely mimic the natural environment, particularly when the nanodisc diameter is significantly larger than the protein itself.

The deep embedment of the protein into the membrane that we observed cannot be predicted by *in silico* methods. Most likely this deep embedment occurs under physiological conditions as well, since the extracellular parts of most ABC proteins are known to highly resistant to proteases. Thus the potential role of this embedment is to protect the protein from extracellular effects (e.g. proteases).

Our method gained attention in the last years in various scientific communities. My abstract was selected for oral presentation at the Gordon Research Conference on Mechanisms of Membrane Transport (Colby-Sawyer College in New London, New Hampshire, United States, 2019). The increasing number of cryo-EM studies should lead to the development of fully automatic membrane detection, of which our pipeline can be a base. My colleagues and G. Tusnády have started to integrate various transmembrane protein associated tools, including MemBlob, into a framework. Importantly, developers of membrane region predictors can apply our calculated, checked, and deposited results as a true positive experimental set. Baaklini *et al.* used TM regions determined by MemBlob during their investigation of chaperone binding sites of CFTR (114).

5.3. MemMoRFs

The increasing number of membrane protein structures, especially NMR structures, also allowed me to collect membrane binding regions and categorized them based on how membrane interaction affects their dynamic properties. Most of the regions undergo disorder to order transition with increased α-helical propensity, as indicated by related NMR data. Other segments remain unordered upon membrane binding, however, it is important to note that their conformational space becomes reduced, as shown for MoRFs binding to protein partners (45). I also included regions that were stable helices both in solution and membrane bound (e.g. kinase suppressor of Ras 1). These helices are not classical MemMoRFs and are in a dynamic equilibrium between lipid and water phases possessing helical conformation.

Reducing the ensemble of an IDR to one conformation causes high entropic penalty, which means that binding requires an equally high enthalpic gain (42). This allows the IDR to bind to the membrane reversibly. This reversible membrane association

suggests that MemMoRFs can alternate between membrane-bound and -dissociated forms. When they are dissociated from the membrane bilayer, they become accessible for enzymes, such as kinases. In the case of gaining negatively charged phosphate group, the dissociated form is promoted because of charge repulsion with negatively charged lipids in the inner leaflet of the plasma membrane. This phenomenon occurs with the MemMoRF of integrin β3, thus this segment becomes available for protein-protein interactions. Importantly, many membrane proteins with different roles contain MemMoRFs. This observation indicates that MemMoRFs play crucial roles in various cellular processes. Some of the proteins utilizing MemMoRFs, like NOTCH1, take part in cell cycle regulation (115). Other proteins with MemMoRFs, such as Myc box-dependent-interacting protein 1 (116), α-synuclein (117), and pro-neuregulin-1 (118), participate in cellular trafficking and membrane shaping. I also identified MemMoRFs in transmembrane receptors like integrin β3 (119), PTEN tumor suppressor (120), and receptor tyrosine kinases, such as EGFR (121) and erbB-2 (122) and also in membrane-associated proteins, like tyrosine-protein kinase Src (123).

There are MemMoRFs in a high number of proteins with pathological relevance. For example, proteins associated with neurodegenerative diseases contain MemMoRFs, including prion protein, α-synuclein and amyloid-β precursor protein (124). Toxins like melittin (125) and the GPCR impairing toxin of *Gila monster* (126) also contain MemMoRFs. Furthermore, I identified MemMoRFs also in viral proteins, like capsid proteins of HIV1 (127) and HCV (128) and transmembrane proteins (e.g. the potassium channel of the Influenza A virus) (129). I observed MemMoRFs in the C-terminal region of SARS-CoV E protein, too. The E protein disturbs the function of the α-1 subunit of the Na^+/K^+-ATP-ase (130) potentially by competing with its regulator proteins called FXYD proteins. E protein also disrupts the Ca^{2+} transport by similar competition with the SERCA regulator PLN (131). FXYD (132) and PLN proteins are single pass transmembrane proteins, which also contain MemMoRFs.

These MemMoRF containing proteins are potential therapeutic targets. For example, SERCA2a activation in cardiac diseases could be achieved by increasing PLN phosphorylation (133). This could be attained by decreasing the propensity of membrane binding of the PLN MemMoRF, which would increase the availability of the phosphorylation site for the kinase.

DOI:10.14753/SE.2025.3101

Dobson & Tusnády used the MemMoRF database to assess how well IDR predictors can detect lipid-binding regions (134). Their conclusion was that none of the investigated predictors can detect MemMoRFs reliably. This highlights the importance of our manually curated database. As the database is entirely manually curated, automatic detection methods should be developed in the future to be able to catch up to the increasing number of new data on IDRs. My colleagues are working on a MemMoRF predictor in collaboration with Lukasz Kurgan (Virginia Commonwealth University) using his tools. In addition, they are also testing the exploitability of protein language models in membrane binding and MemMoRF predictors.

DOI:10.14753/SE.2025.3101

6. CONCLUSIONS

1. The examined membrane prediction methods provide similar results regarding the embedding of proteins into the membrane, suggesting that they likely to be accurate.

2. For Pgp-like ABC proteins, the description of the functional units of proteins with vectors is suitable for comparing structures, determining conformational states, and analyzing molecular dynamics simulations.

3. Binding locations can be identified through electrostatic calculations.

4. The position of TM regions can be determined at the atomic level from data representing the lipid environment found in Coulomb potential maps determined by cryo-electron microscopy.

5. The membrane mimetic used in structure determination affects the thickness of the transmembrane region calculated based on cryo-EM maps.

6. Some of the disordered regions become ordered during interaction with the membrane, and these regions are referred to as MemMoRFs.

7. Phosphorylation of MemMoRFs may impact the interaction with the membrane, influencing the transition between ordered and disordered states.

8. In NMR structures found in the PDB database, stable structures (helices) have been observed numerous times, which are not justified by flexibility values calculated from raw data. Therefore, to identify disordered regions, it is worthwhile to consider additional information beyond the NMR ensembles uploaded to databases.

DOI:10.14753/SE.2025.3101

7. SUMMARY

This thesis outlines three computational studies on membrane proteins, emphasizing the crucial need to comprehend their diverse conformations for deciphering functional mechanisms during the transport cycle. Despite the significance of studying transmembrane regions, sparse experimental data exists on membrane embedment. Intrinsically disordered regions in membrane proteins may impact function through membrane bilayer interaction, yet detailing these structural aspects poses a challenging gap in existing literature. My aim was to fill these gaps using computational methods.

First, I developed computational methods to assess the new structures of Pgp-like membrane proteins, which are constantly increasing in number. During this project, a comparison between the most popular transmembrane prediction methods showed that they produce similar results on the tilting and position of proteins in membranes. I showed that structure-specific conformational vectors called conftors are suitable to describe the differences in conformations, to compare structures, to highlight inaccuracies, and to analyze molecular dynamic simulations. Though the hereby described conftors are specific to the Pgp-like ABC subfamily, the idea can be used to any protein family. Electrostatic calculations showed that the widely open outward facing conformation of some Pgp-like proteins may be the result of the lack of a membrane bilayer under crystallization conditions. Additionally, I found that electrostatic calculations are suitable to highlight binding locations. and to assess the overall quality of structures.

Second, I aided the development of the MemBlob pipeline, which provides experiment-based transmembrane regions for transmembrane proteins solved by cryo-EM that is an entirely new addition to the field. The analysis of the resulting transmembrane regions showed that the thickness of the membrane blob could be affected by the membrane mimetics used during structure determination.

Finally, I collected, categorized and analyzed the data of membrane interacting intrinsically disordered regions (MemMoRFs) on which a database was built. This study showed that disordered regions tend to form helical structures during membrane interaction and that this interaction can be regulated by phosphorylation. We also found examples implying that helical structures tend to be overrepresented in NMR ensembles. Lastly, as several pathogenically relevant proteins contain MemMoRFs, these regions could be potential therapeutic targets.

8. REFERENCES

1. George AM. ABC transporters-40 years on: Springer; 2016.

2. Locher KP. Review. Structure and mechanism of ATP-binding cassette transporters. Philos Trans R Soc Lond B Biol Sci. 2009;364(1514):239-45.

3. Szakács G, Váradi A, Ozvegy-Laczka C, Sarkadi B. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). Drug Discov Today. 2008;13(9-10):379-93.

4. Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. Nat Rev Genet. 2015;16(1):45-56.

5. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science. 1989;245(4922):1066-73.

6. Hegedus T, Aleksandrov A, Mengos A, Cui L, Jensen TJ, Riordan JR. Role of individual R domain phosphorylation sites in CFTR regulation by protein kinase A. Biochim Biophys Acta. 2009;1788(6):1341-9.

7. Hegedus T, Serohijos AW, Dokholyan NV, He L, Riordan JR. Computational studies reveal phosphorylation-dependent changes in the unstructured R domain of CFTR. J Mol Biol. 2008;378(5):1052-63.

8. Csanády L, Vergani P, Gadsby DC. STRUCTURE, GATING, AND REGULATION OF THE CFTR ANION CHANNEL. Physiol Rev. 2019;99(1):707-38.

9. Dawson RJ, Locher KP. Structure of a bacterial multidrug ABC transporter. Nature. 2006;443(7108):180-5.

10. Locher KP. Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nat Struct Mol Biol. 2016;23(6):487-93.

11. Choudhury HG, Tong Z, Mathavan I, Li Y, Iwata S, Zirah S, Rebuffat S, van Veen HW, Beis K. Structure of an antibacterial peptide ATP-binding cassette transporter in a novel outward occluded state. Proc Natl Acad Sci U S A. 2014;111(25):9145-50.

12. Perez C, Gerber S, Boilevin J, Bucher M, Darbre T, Aebi M, Reymond JL, Locher KP. Structure and mechanism of an active lipid-linked oligosaccharide flippase. Nature. 2015;524(7566):433-8.

13. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL, Chang G. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science. 2009;323(5922):1718-22.

14. Esser L, Zhou F, Pluchino KM, Shiloach J, Ma J, Tang WK, Gutierrez C, Zhang A, Shukla S, Madigan JP, Zhou T, Kwong PD, Ambudkar SV, Gottesman MM, Xia D. Structures of the Multidrug Transporter P-glycoprotein Reveal Asymmetric ATP Binding and the Mechanism of Polyspecificity. J Biol Chem. 2017;292(2):446-61.

15. Hohl M, Briand C, Grütter MG, Seeger MA. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nat Struct Mol Biol. 2012;19(4):395-402.

16. Ward A, Reyes CL, Yu J, Roth CB, Chang G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proc Natl Acad Sci U S A. 2007;104(48):19005-10.

17. Csizmadia G, Farkas B, Spagina Z, Tordai H, Hegedűs T. Quantitative comparison of ABC membrane protein type I exporter structures in a standardized way. Comput Struct Biotechnol J. 2018;16:396-403.

18. George AM, Jones PM. Perspectives on the structure–function of ABC transporters: The Switch and Constant Contact Models. Progress in Biophysics and Molecular Biology. 2012;109(3):95-107.

19. Szöllősi D, Rose-Sperling D, Hellmich UA, Stockner T. Comparison of mechanistic transport cycle models of ABC exporters. Biochimica et Biophysica Acta (BBA) - Biomembranes. 2018;1860(4):818-32.

20. Gyimesi G, Ramachandran S, Kota P, Dokholyan NV, Sarkadi B, Hegedus T. ATP hydrolysis at one of the two sites in ABC transporters initiates transport related conformational transitions. Biochim Biophys Acta. 2011;1808(12):2954-64.

21. Galazzo L, Meier G, Januliene D, Parey K, De Vecchis D, Striednig B, Hilbi H, Schäfer LV, Kuprov I, Moeller A, Bordignon E, Seeger MA. The ABC transporter MsbA adopts the wide inward-open conformation in E. coli cells. Science Advances. 2022;8(41):eabn6845.

22. Stockner T, de Vries SJ, Bonvin AM, Ecker GF, Chiba P. Data-driven homology modelling of P-glycoprotein in the ATP-bound state indicates flexibility of the transmembrane domains. Febs j. 2009;276(4):964-72.

23. Young JY, Westbrook JD, Feng Z, Sala R, Peisach E, Oldfield TJ, Sen S, Gutmanas A, Armstrong DR, Berrisford JM, Chen L, Chen M, Di Costanzo L, Dimitropoulos D, Gao G, Ghosh S, Gore S, Guranovic V, Hendrickx PMS, Hudson BP, Igarashi R, Ikegawa Y, Kobayashi N, Lawson CL, Liang Y, Mading S, Mak L, Mir MS, Mukhopadhyay A, Patwardhan A, Persikova I, Rinaldi L, Sanz-Garcia E, Sekharan MR, Shao C, Swaminathan GJ, Tan L, Ulrich EL, van Ginkel G, Yamashita R, Yang H, Zhuravleva MA, Quesada M, Kleywegt GJ, Berman HM, Markley JL, Nakamura H, Velankar S, Burley SK. OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive. Structure. 2017;25(3):536-45.

24. Gore S, Sanz García E, Hendrickx PMS, Gutmanas A, Westbrook JD, Yang H, Feng Z, Baskaran K, Berrisford JM, Hudson BP, Ikegawa Y, Kobayashi N, Lawson CL, Mading S, Mak L, Mukhopadhyay A, Oldfield TJ, Patwardhan A, Peisach E, Sahni G, Sekharan MR, Sen S, Shao C, Smart OS, Ulrich EL, Yamashita R, Quesada M, Young JY, Nakamura H, Markley JL, Berman HM, Burley SK, Velankar S, Kleywegt GJ. Validation of Structures in the Protein Data Bank. Structure. 2017;25(12):1916-27.

25. Read RJ, Adams PD, Arendall WB, 3rd, Brunger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Lütteke T, Otwinowski Z, Perrakis A, Richardson JS, Sheffler WH, Smith JL, Tickle IJ, Vriend G, Zwart PH. A new generation of crystallographic validation tools for the protein data bank. Structure. 2011;19(10):1395-412.

26. Wilson AJC. Determination of Absolute from Relative X-Ray Intensity Data. Nature. 1942;150(3796):152-.

27. Monzon AM, Rohr CO, Fornasari MS, Parisi G. CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state. Database (Oxford). 2016;2016.

28. Condic-Jurkic K, Subramanian N, Mark AE, O'Mara ML. The reliability of molecular dynamics simulations of the multidrug transporter P-glycoprotein in a membrane environment. PLoS One. 2018;13(1):e0191882.

29. Thonghin N, Kargas V, Clews J, Ford RC. Cryo-electron microscopy of membrane proteins. Methods. 2018;147:176-86.

30. Kowiel M, Brzezinski D, Porebski PJ, Shabalin IG, Jaskolski M, Minor W. Automatic recognition of ligands in electron density by machine learning. Bioinformatics. 2019;35(3):452-61.

31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.

32. Zagotta WN, Gordon MT, Senning EN, Munari MA, Gordon SE. Measuring distances between TRPV1 and the plasma membrane using a noncanonical amino acid and transition metal ion FRET. J Gen Physiol. 2016;147(2):201-16.

33. Chang XB, Hou YX, Jensen TJ, Riordan JR. Mapping of cystic fibrosis transmembrane conductance regulator membrane topology by glycosylation site insertion. J Biol Chem. 1994;269(28):18572-5.

34. Tusnády GE, Dosztányi Z, Simon I. Transmembrane proteins in the Protein Data Bank: identification and classification. Bioinformatics. 2004;20(17):2964-72.

35. Lomize AL, Pogozheva ID, Mosberg HI. Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. J Chem Inf Model. 2011;51(4):930-46.

36. Stansfeld PJ, Goose JE, Caffrey M, Carpenter EP, Parker JL, Newstead S, Sansom MS. MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. Structure. 2015;23(7):1350-61.

37. Bozoky Z, Krzeminski M, Chong PA, Forman-Kay JD. Structural changes of CFTR R region upon phosphorylation: a plastic platform for intramolecular and intermolecular interactions. Febs j. 2013;280(18):4407-16.

38. Sormanni P, Piovesan D, Heller GT, Bonomi M, Kukic P, Camilloni C, Fuxreiter M, Dosztanyi Z, Pappu RV, Babu MM, Longhi S, Tompa P, Dunker AK, Uversky VN, Tosatto SC, Vendruscolo M. Simultaneous quantification of protein order and disorder. Nat Chem Biol. 2017;13(4):339-42.

39. Uversky VN. Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. J Biol Chem. 2016;291(13):6681-8.

40. Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput. 1998:473-84.

41. Mittag T, Kay LE, Forman-Kay JD. Protein dynamics and conformational disorder in molecular recognition. J Mol Recognit. 2010;23(2):105-16.

42. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005;6(3):197-208.

43. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. Intrinsic disorder and protein function. Biochemistry. 2002;41(21):6573-82.

44. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM. Classification of Intrinsically Disordered Regions and Proteins. Chemical Reviews. 2014;114(13):6589-631.

45. Appadurai R, Uversky VN, Srivastava A. The Structural and Functional Diversity of Intrinsically Disordered Regions in Transmembrane Proteins. The Journal of Membrane Biology. 2019;252(4):273-92.

46. Tusnády GE, Dobson L, Tompa P. Disordered regions in transmembrane proteins. Biochimica et Biophysica Acta (BBA) - Biomembranes. 2015;1848(11, Part A):2839-48.

47. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. Analysis of molecular recognition features (MoRFs). J Mol Biol. 2006;362(5):1043-59.

48. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry. 2005;44(37):12454-70.

49. Yang J, Gao M, Xiong J, Su Z, Huang Y. Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. Protein Sci. 2019;28(11):1952-65.

50. Van Rossum GaD, Fred L. Python 3 Reference Manual: CreateSpace; 2009.

51. Van Rossum GaDJ, Fred L. Python reference manual: Centrum voor Wiskunde en Informatica Amsterdam; 1995.

52. Consortium TU. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research. 2022;51(D1):D523-D31.

53. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 2012;40(Database issue):D370-6.

54. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Research. 2000;28(1):235-42.

55. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. Nature. 2020;585(7825):357-62.

56. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem. 2011;32(10):2319-27.

57. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 2007;35(Web Server issue):W522-5.

58. Tang CL, Alexov E, Pyle AM, Honig B. Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. J Mol Biol. 2007;366(5):1475-96.

59. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996;14(1):33-8.

60. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, Chun M, Li P, Gohara DW, Dolinsky T, Konecny R, Koes DR, Nielsen JE, Head-Gordon T, Geng W, Krasny R, Wei GW, Holst MJ, McCammon JA, Baker NA. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 2018;27(1):112-28.

61. Callenberg KM, Choudhary OP, de Forest GL, Gohara DW, Baker NA, Grabe M. APBSmem: a graphical interface for electrostatic calculations at the membrane. PLoS One. 2010;5(9).

62. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. 2007;9(3):90-5.

63. Lawson CL, Patwardhan A, Baker ML, Hryc C, Garcia ES, Hudson BP, Lagerstedt I, Ludtke SJ, Pintilie G, Sala R, Westbrook JD, Berman HM, Kleywegt GJ, Chiu W. EMDataBank unified data resource for 3DEM. Nucleic Acids Research. 2015;44(D1):D396-D403.

64. Trabuco LG, Villa E, Schreiner E, Harrison CB, Schulten K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. Methods. 2009;49(2):174-80.

65. Farkas B, Csizmadia G, Katona E, Tusnády GE, Hegedűs T. MemBlob database and server for identifying transmembrane regions using cryo-EM maps. Bioinformatics. 2020;36(8):2595-8.

66. Hoch JC, Baskaran K, Burr H, Chin J, Eghbalnia Hamid R, Fujiwara T, Gryk Michael R, Iwata T, Kojima C, Kurisu G, Maziuk D, Miyanoiri Y, Wedell Jonathan R, Wilburn C, Yao H, Yokochi M. Biological Magnetic Resonance Data Bank. Nucleic Acids Research. 2022;51(D1):D368-D76.

67. Camilloni C, De Simone A, Vranken WF, Vendruscolo M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. Biochemistry. 2012;51(11):2224-31.

68. Berjanskii MV, Wishart DS. A simple method to predict protein flexibility using secondary chemical shifts. J Am Chem Soc. 2005;127(43):14970-1.

69. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47(D1):D427-d32.

70. Fichó E, Reményi I, Simon I, Mészáros B. MFIB: a repository of protein complexes with mutual folding induced by binding. Bioinformatics. 2017;33(22):3682-4.

71. Hatos A, Hajdu-Soltész B, Monzon AM, Palopoli N, Álvarez L, Aykac-Fas B, Bassot C, Benítez GI, Bevilacqua M, Chasapi A, Chemes L, Davey NE, Davidović R, Dunker AK, Elofsson A, Gobeill J, Foutel NSG, Sudha G, Guharoy M, Horvath T, Iglesias V, Kajava AV, Kovacs OP, Lamb J, Lambrughi M, Lazar T, Leclercq JY, Leonardi E, Macedo-Ribeiro S, Macossay-Castillo M, Maiani E, Manso JA, Marino-Buslje C, Martínez-Pérez E, Mészáros B, Mičetić I, Minervini G, Murvai N, Necci M, Ouzounis CA, Pajkos M, Paladin L, Pancsa R, Papaleo E, Parisi G, Pasche E, Barbosa Pereira PJ, Promponas VJ, Pujols J, Quaglia F, Ruch P, Salvatore M, Schad E, Szabo B, Szaniszló T, Tamana S, Tantos A, Veljkovic N, Ventura S, Vranken W, Dosztányi Z,

Tompa P, Tosatto SCE, Piovesan D. DisProt: intrinsic protein disorder annotation in 2020. Nucleic Acids Res. 2020;48(D1):D269-d76.

72. Schad E, Fichó E, Pancsa R, Simon I, Dosztányi Z, Mészáros B. DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. Bioinformatics. 2018;34(3):535-7.

73. Kim Y, Chen J. Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. Science. 2018;359(6378):915-9.

74. Liu F, Zhang Z, Csanády L, Gadsby DC, Chen J. Molecular Structure of the Human CFTR Ion Channel. Cell. 2017;169(1):85-95.e8.

75. Zhang Z, Liu F, Chen J. Conformational Changes of CFTR upon Phosphorylation and ATP Binding. Cell. 2017;170(3):483-91.e8.

76. Das J, Aleksandrov AA, Cui L, He L, Riordan JR, Dokholyan NV. Transmembrane helical interactions in the CFTR channel pore. PLoS Comput Biol. 2017;13(6):e1005594.

77. Tordai H, Leveles I, Hegedűs T. Molecular dynamics of the cryo-EM CFTR structure. Biochem Biophys Res Commun. 2017;491(4):986-93.

78. Kodan A, Yamaguchi T, Nakatsu T, Sakiyama K, Hipolito CJ, Fujioka A, Hirokane R, Ikeguchi K, Watanabe B, Hiratake J, Kimura Y, Suga H, Ueda K, Kato H. Structural basis for gating mechanisms of a eukaryotic P-glycoprotein homolog. Proc Natl Acad Sci U S A. 2014;111(11):4049-54.

79. Johnson ZL, Chen J. ATP Binding Enables Substrate Release from Multidrug Resistance Protein 1. Cell. 2018;172(1-2):81-9.e10.

80. Elazar A, Weinstein JJ, Prilusky J, Fleishman SJ. Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. Proc Natl Acad Sci U S A. 2016;113(37):10340-5.

81. Penniston JT, Padányi R, Pászty K, Varga K, Hegedus L, Enyedi A. Apart from its known function, the plasma membrane Ca^{2+}ATPase can regulate Ca^{2+} signaling by controlling phosphatidylinositol 4,5-bisphosphate levels. J Cell Sci. 2014;127(Pt 1):72-84.

82. von Heijne G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. Nature. 1989;341(6241):456-8.

83. Martin GM, Kandasamy B, DiMaio F, Yoshioka C, Shyng SL. Anti-diabetic drug binding site in a mammalian K(ATP) channel revealed by Cryo-EM. Elife. 2017;6.

84. Lee KPK, Chen J, MacKinnon R. Molecular structure of human KATP in complex with ATP and ADP. Elife. 2017;6.

85. Li M, Zhou X, Wang S, Michailidis I, Gong Y, Su D, Li H, Li X, Yang J. Structure of a eukaryotic cyclic-nucleotide-gated channel. Nature. 2017;542(7639):60-5.

86. Hite RK, MacKinnon R. Structural Titration of Slo2.2, a Na(+)-Dependent K(+) Channel. Cell. 2017;168(3):390-9.e11.

87. Johnson ZL, Chen J. Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1. Cell. 2017;168(6):1075-85.e9.

88. Csizmadia G, Erdős G, Tordai H, Padányi R, Tosatto S, Dosztányi Z, Hegedűs T. The MemMoRF database for recognizing disordered protein regions interacting with cellular membranes. Nucleic Acids Res. 2021;49(D1):D355-d60.

89. Brooks PC, Montgomery AM, Rosenfeld M, Reisfeld RA, Hu T, Klier G, Cheresh DA. Integrin alpha v beta 3 antagonists promote tumor regression by inducing apoptosis of angiogenic blood vessels. Cell. 1994;79(7):1157-64.

90. Brooks PC, Strömblad S, Klemke R, Visscher D, Sarkar FH, Cheresh DA. Antiintegrin alpha v beta 3 blocks human breast cancer growth and angiogenesis in human skin. J Clin Invest. 1995;96(4):1815-22.

91. Cao Z, Suo X, Chu Y, Xu Z, Bao Y, Miao C, Deng W, Mao K, Gao J, Xu Z, Ma YQ. Peptides derived from the integrin β cytoplasmic tails inhibit angiogenesis. Cell Commun Signal. 2018;16(1):38.

92. Yang J, Ma YQ, Page RC, Misra S, Plow EF, Qin J. Structure of an integrin alphaIIb beta3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation. Proc Natl Acad Sci U S A. 2009;106(42):17729-34.

93. Metcalf DG, Moore DT, Wu Y, Kielec JM, Molnar K, Valentine KG, Wand AJ, Bennett JS, DeGrado WF. NMR analysis of the alphaIIb beta3 cytoplasmic interaction suggests a mechanism for integrin regulation. Proc Natl Acad Sci U S A. 2010;107(52):22481-6.

94. Aivazian D, Stern LJ. Phosphorylation of T cell receptor zeta is regulated by a lipid dependent folding transition. Nat Struct Biol. 2000;7(11):1023-6.

95. Shah K, Al-Haidari A, Sun J, Kazi JU. T cell receptor (TCR) signaling in health and disease. Signal Transduction and Targeted Therapy. 2021;6(1):412.

96. Gustavsson M, Verardi R, Mullen DG, Mote KR, Traaseth NJ, Gopinath T, Veglia G. Allosteric regulation of SERCA by phosphorylation-mediated conformational shift of phospholamban. Proc Natl Acad Sci U S A. 2013;110(43):17338-43.

97. Weber DK, Reddy UV, Wang S, Larsen EK, Gopinath T, Gustavsson MB, Cornea RL, Thomas DD, De Simone A, Veglia G. Structural basis for allosteric control of the SERCA-Phospholamban membrane complex by Ca2+ and phosphorylation. eLife. 2021;10:e66226.

98. Mészáros B, Erdos G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 2018;46(W1):W329-w37.

99. Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol. 1996;3(10):842-8.

100. Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J, Diakogianni A, Valverde JA, Bukirova D, Čalyševa J, Palopoli N, Davey NE, Chemes LB, Gibson TJ. ELM-the eukaryotic linear motif resource in 2020. Nucleic Acids Res. 2020;48(D1):D296-d306.

101. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015;43(Database issue):D512-20.

102. Sehnal D, Deshpande M, Vařeková RS, Mir S, Berka K, Midlik A, Pravda L, Velankar S, Koča J. LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. Nat Methods. 2017;14(12):1121-2.

103. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005;33(Database issue):D514-7.

104. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11.

105. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson

A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-d82.

106. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L, Lovering RC, Meldal B, Melidoni AN, Milagros M, Peluso D, Perfetto L, Porras P, Raghunath A, Ricard-Blum S, Roechert B, Stutz A, Tognolli M, van Roey K, Cesareni G, Hermjakob H. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 2014;42(Database issue):D358-63.

107. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-d13.

108. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.

109. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Žídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research. 2021;50(D1):D439-D44.

110. Hegedűs T, Geisler M, Lukács GL, Farkas B. Ins and outs of AlphaFold2 transmembrane protein structure predictions. Cell Mol Life Sci. 2022;79(1):73.

111. Mittal A, Böhm S, Grütter MG, Bordignon E, Seeger MA. Asymmetry in the homodimeric ABC transporter MsbA recognized by a DARPin. J Biol Chem. 2012;287(24):20395-406.

112. Zou P, Bortolus M, McHaourab HS. Conformational cycle of the ABC transporter MsbA in liposomes: detailed analysis using double electron-electron resonance spectroscopy. J Mol Biol. 2009;393(3):586-97.

113. Tordai H, Suhajda E, Sillitoe I, Nair S, Varadi M, Hegedus T. Comprehensive Collection and Prediction of ABC Transmembrane Protein Structures in the AI Era of Structural Biology. International Journal of Molecular Sciences. 2022;23(16):8877.

114. Baaklini I, Gonçalves CdC, Lukacs GL, Young JC. Selective Binding of HSC70 and its Co-Chaperones to Structural Hotspots on CFTR. Scientific Reports. 2020;10(1):4176.

115. Deatherage CL, Lu Z, Kroncke BM, Ma S, Smith JA, Voehler MW, McFeeters RL, Sanders CR. Structural and biochemical differences between the Notch and the amyloid precursor protein transmembrane domains. Sci Adv. 2017;3(4):e1602794.

116. Löw C, Weininger U, Lee H, Schweimer K, Neundorf I, Beck-Sickinger AG, Pastor RW, Balbach J. Structure and dynamics of helix-0 of the N-BAR domain in lipid micelles and bilayers. Biophys J. 2008;95(9):4315-23.

117. Ulmer TS, Bax A, Cole NB, Nussbaum RL. Structure and dynamics of micelle-bound human alpha-synuclein. J Biol Chem. 2005;280(10):9595-603.

118. Chukhlieb M, Raasakka A, Ruskamo S, Kursula P. The N-terminal cytoplasmic domain of neuregulin 1 type III is intrinsically disordered. Amino Acids. 2015;47(8):1567-77.

119. Vinogradova O, Vaynberg J, Kong X, Haas TA, Plow EF, Qin J. Membrane-mediated structural transitions at the cytoplasmic face during integrin activation. Proc Natl Acad Sci U S A. 2004;101(12):4094-9.

120. Nanda H, Heinrich F, Lösche M. Membrane association of the PTEN tumor suppressor: neutron scattering and MD simulations reveal the structure of protein-membrane complexes. Methods. 2015;77-78:136-46.

121. Choowongkomon K, Carlin CR, Sönnichsen FD. A structural model for the membrane-bound form of the juxtamembrane domain of the epidermal growth factor receptor. J Biol Chem. 2005;280(25):24043-52.

122. Bragin PE, Mineev KS, Bocharova OV, Volynsky PE, Bocharov EV, Arseniev AS. HER2 Transmembrane Domain Dimerization Coupled with Self-Association of

Membrane-Embedded Cytoplasmic Juxtamembrane Regions. J Mol Biol. 2016;428(1):52-61.

123. Pérez Y, Maffei M, Igea A, Amata I, Gairí M, Nebreda AR, Bernadó P, Pons M. Lipid binding by the Unique and SH3 domains of c-Src suggests a new regulatory mechanism. Sci Rep. 2013;3:1295.

124. Riek R, Eisenberg DS. The activities of amyloids from a structural perspective. Nature. 2016;539(7628):227-35.

125. Appadu A, Jelokhani-Niaraki M, DeBruin L. Conformational Changes and Association of Membrane-Interacting Peptides in Myelin Membrane Models: A Case of the C-Terminal Peptide of Proteolipid Protein and the Antimicrobial Peptide Melittin. J Phys Chem B. 2015;119(47):14821-30.

126. Neidigh JW, Fesinmeyer RM, Prickett KS, Andersen NH. Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. Biochemistry. 2001;40(44):13188-200.

127. Serrano S, Araujo A, Apellániz B, Bryson S, Carravilla P, de la Arada I, Huarte N, Rujas E, Pai EF, Arrondo JLR, Domene C, Jiménez MA, Nieva JL. Structure and immunogenicity of a peptide vaccine, including the complete HIV-1 gp41 2F5 epitope: implications for antibody recognition mechanism and immunogen design. J Biol Chem. 2014;289(10):6565-80.

128. Penin F, Brass V, Appel N, Ramboarina S, Montserret R, Ficheux D, Blum HE, Bartenschlager R, Moradpour D. Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. J Biol Chem. 2004;279(39):40835-43.

129. Schnell JR, Chou JJ. Structure and mechanism of the M2 proton channel of influenza A virus. Nature. 2008;451(7178):591-5.

130. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. Virol J. 2019;16(1):69.

131. Berta B, Tordai H, Lukács GL, Papp B, Enyedi Á, Padányi R, Hegedűs T. SARS-CoV-2 Envelope protein alters calcium signaling via SERCA interactions. bioRxiv. 2023:2023.06.13.544745.

132. Geering K. Function of FXYD proteins, regulators of Na, K-ATPase. J Bioenerg Biomembr. 2005;37(6):387-92.

133. Ablorh ND, Thomas DD. Phospholamban phosphorylation, mutation, and structural dynamics: a biophysical approach to understanding and treating cardiomyopathy. Biophys Rev. 2015;7(1):63-76.
134. Dobson L, Tusnády GE. MemDis: Predicting Disordered Regions in Transmembrane Proteins. Int J Mol Sci. 2021;22(22).

9. BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

9.1. Related to the dissertation

1. **Csizmadia G**, Erdős G, Tordai H, Padányi R, Tosatto S, Dosztányi Z, Hegedűs T. The MemMoRF database for recognizing disordered protein regions interacting with cellular membranes. Nucleic Acids Res. 2021 Jan 8;49(D1):D355-D360.
   IF: 19,160

2. Farkas B*, **Csizmadia G***, Katona E, Tusnády GE, Hegedűs T. MemBlob database and server for identifying transmembrane regions using cryo-EM maps. Bioinformatics. 2020 Apr 15;36(8):2595-2598.
   IF: 6,937

3. **Csizmadia G**, Farkas B, Spagina Z, Tordai H, Hegedűs T. Quantitative comparison of ABC membrane protein type I exporter structures in a standardized way. Comput Struct Biotechnol J. 2018 Oct 18;16:396-403.
   IF: 4,720

9.2. Independent of the dissertation

Kiss R, **Csizmadia G**, Solti K, Keresztes A, Zhu M, Pickhardt M, Mandelkow E, Tóth G. Structural Basis of Small Molecule Targetability of Monomeric Tau Protein. ACS Chem Neurosci. 2018 Dec 19;9(12):2997-3006.
IF: 3,861

DOI:10.14753/SE.2025.3101

10. ACKNOWLEDGEMENTS

Foremost, I extend my gratitude to my advisor, Tamás Hegedűs, whose mentorship has been invaluable throughout this journey. His profound knowledge, expertise, and patience have guided my work, and his encouragement to participate in scientific conferences has been instrumental from the outset.

I am also deeply thankful to Hedvig Tordai, whose guidance and critical thinking were indispensable whenever I encountered challenges. Her oversight in the gathering process of MemMoRFs and insightful contributions to this thesis are greatly appreciated. I extend my thanks to her and to my esteemed colleagues at the Institute – particularly Bianka Farkas, Erzsébet Besze, Krisztina Lór, and Rita Padányi – for fostering a productive work environment.

Special appreciation goes to Professor Miklós Kellermayer and the SE School of PhD Studies for their support of my research endeavors and publications.

I am grateful for the collaboration with Gábor Erdős and Zsuzsanna Dosztányi, whose creation of the MemMoRF webpage enhanced the scope of my work.

Lastly, I express heartfelt appreciation to my parents, in-laws, and husband for their unwavering support throughout my studies.