Collection, prediction and publication of ABC transmembrane protein structures

Erzsébet Suhajda[1,2], Hedvig Tordai[1], Ian Sillitoe[3], Sreenath Nair[4], Mihály Váradi[4], Tamás Hegedűs[1,5]

[1] *Semmelweis University, Department of Biophysics and Radiation Biology, Budapest, Hungary*

[2] *Budapest University of Technology and Economics, Faculty of Electrical Engineering and Informatics, Budapest, Hungary*

[3] *University College London, Department of Structural and Molecular Biology, London, UK*

[4] *European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK*

[5] *Eötvös Loránd Research Network, ELKH-SE Biophysical Virology Research Group, Budapest, Hungary*

Introduction

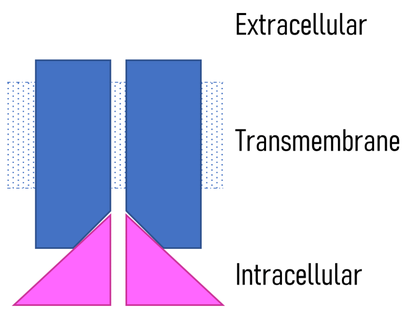
The number of resolved protein structures skyrocketed in the last few years thanks to novel computational methods, like AlphaFold and also experimental advancements in the field, like cryo-electron microscopy. Although this increment is a welcome change, collecting the structures of a specific protein family is a challenging task despite existing general and domain-specific databases. Here, we demonstrate and assess this with the ABC (ATP-binding casette) transmembrane (TM) protein superfamily.

Objectives

- Collect, classify and publish every available ABC TM protein structure from structural databases
- Run AlphaFold predictions on human ABC proteins functioning as dimers and connect them to large databases via 3D-Beacons
- Develop a web application to expose our thorough collection of ABC TM protein structures

Contact
abc3d.hegelab.org
suhajda@hegelab.org

Identification and classification of ABC proteins

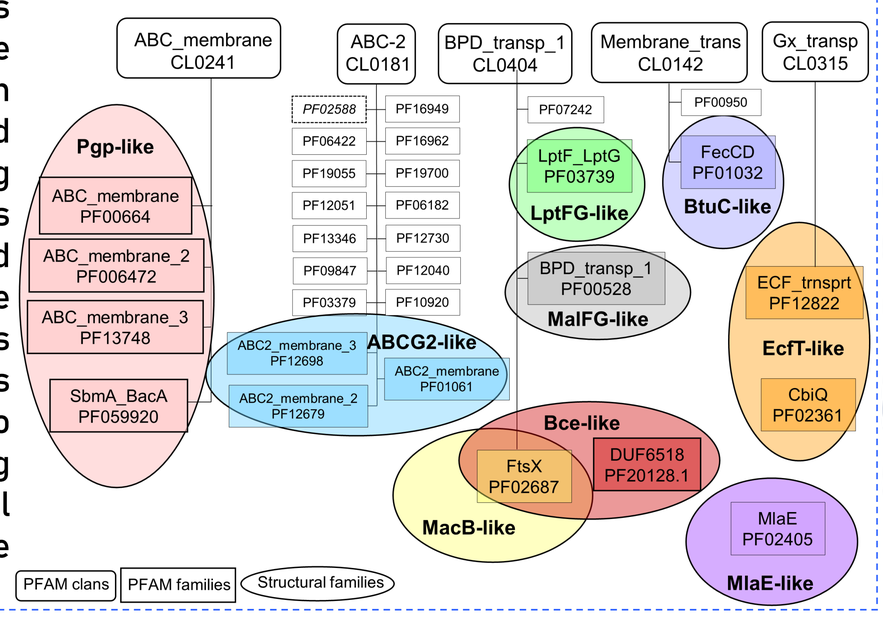


1. General structure of ABC proteins.

ABC proteins consist of two transmembrane domains (TMD) shown in blue and two nucleotide-binding domains (NBDs) shown in pink. Due to their heterogeneous transport functions, their TMDs are not conserved, while NBDs are highly conserved.

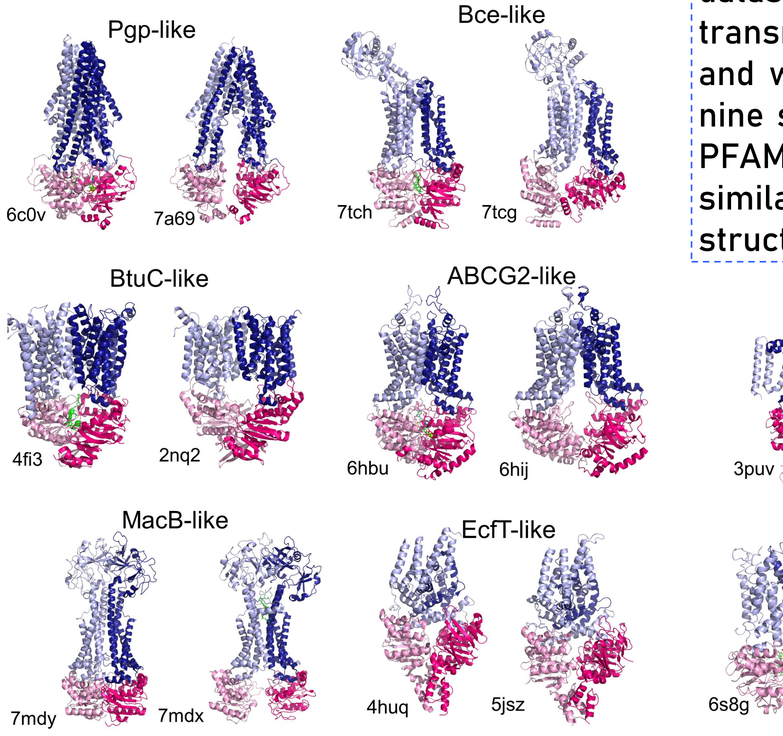
2. Connecting PFAM entries to structural classes

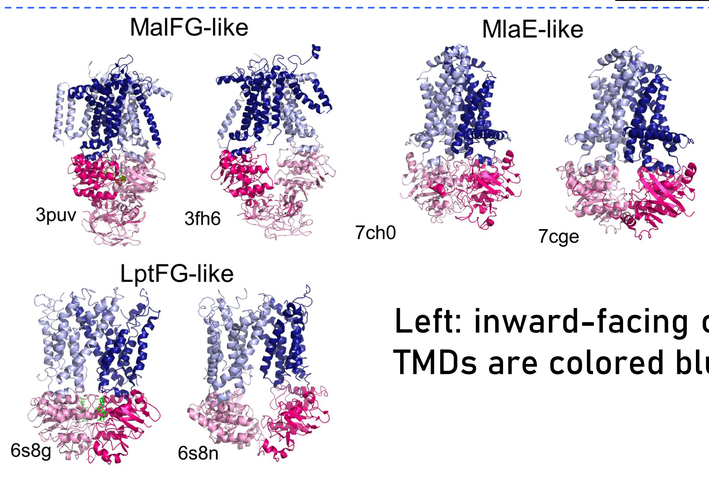
ABC protein PFAM (https://pfam.xfam.org) entries were identified and the sequence of every structure in RCSB and AFDB (AlphaFold Database) was searched using HMMER hmmsearch. Structures containing at least one TMD and two NBD PFAM hits in the datasets were selected as transmembrane ABC proteins and were further classified into nine structural subgroups using PFAM hits and structural similarity to reference structures.



3. Nine structural families of ABC proteins

Previously, ABC proteins were classified into numbered classes based on TMD folds. We propose to name each structural family after a well-known member instead (e.g. Pgp-like). We also aimed to group the structures according to their conformation exhibiting inward open or closed nature, using the distances between selected amino acids in Walker A and ABC signature sequences in opposite NBDs.





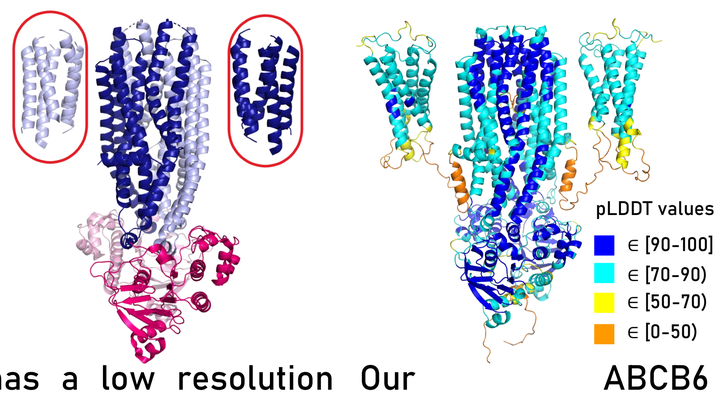
Two reference structures are shown for every structural family, along with their PDB IDs.

Left: inward-facing closed, right: inward-facing open. TMDs are colored blue, NDBs are colored pink.

Running AlphaFold (AF) predictions on TM protein complexes

1. Human ABC dimers

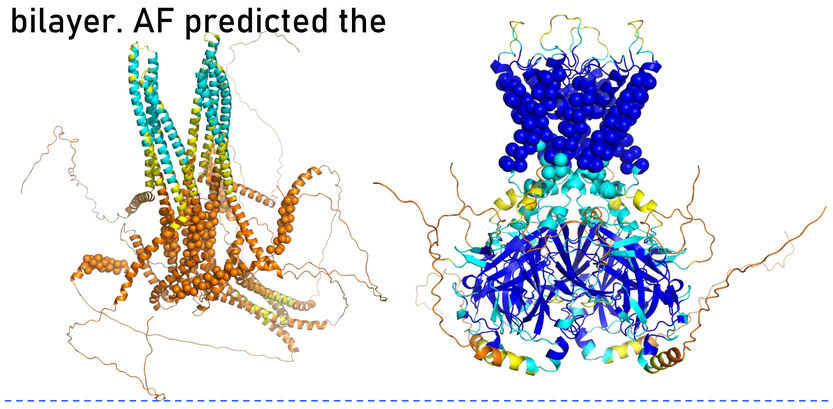
We ran AF predictions on dimeric human ABC proteins, since AFDB contains only monomeric structures. We predicted all dimeric ABC proteins without experimental structures and also those resolved experimentally to gain information on unresolved regions. E.g. TMD0s (circled) in ABCB6 has a low resolution (PDB ID: 7D7N, 5.2 Å), thus helices were observed but not sidechains.



Our ABCB6 homodimer prediction reveals TMD0s with a high confidence (pLDDT) value.

2. Limitations of AF?

We aimed to predict the complex of ABCB8 and mitoK (CCDC51), demonstrated to conduct potassium in the inner membrane of mitochondria. No reasonable structure was built. We also ran AF predictions of mitoK complex without ABCB8. Aside from overall low model confidence, the TM regions (shown as spheres) do not align with a putative membrane bilayer. AF predicted the structure of Kir6.2, potassium channel partner of ABCC8 ABCC9 excellently.



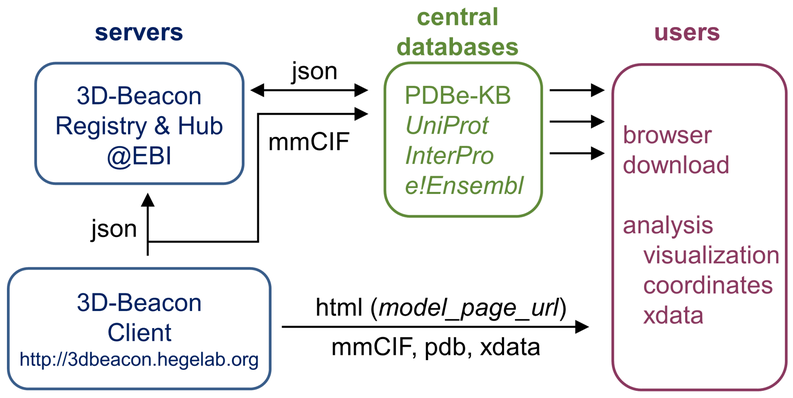
Exposing our collected and predicted structures

1. ABC3D webapplication

In our webapplication at *abc3d.hegelab.org*, structures can be browsed, searched and visualized. Users can choose to display Experimental, Computational, Human or All proteins and can further narrow down their search according to structural families and conformations (open or closed). Selected structures can be downloaded as .pdb files and .mmcif files are also available for bulk download.

2. 3D-Beacons Network

We connected our AF-predicted human dimer ABC proteins to the 3D-Beacons Network to reach the global scientific community. 3D-Beacons provides programmatic access to both theoretical and experimental protein structures and links them to central databases like PDBe-KB. At *3dbeacon.hegelab.org*, aside from the structures in .mmcif format, users can find information about methodology, publications and additional structure files.



Acknowledgements

NKFIH K127961

NKFIH K137610

NKFIH 2020-2.1.1-ED-2021-00179

CFF HEGEDU2010

Wigner Scientific Computing Laboratory (WSCLAB, the former Wigner GPU Laboratory)