SUPPLEMENTARY INFORMATION

Ins and outs of AlphaFold2 transmembrane protein structure predictions

Tamás Hegedűs, Markus Geisler, Gergely L. Lukács, Bianka Farkas

Fig. S1 Mean pLDDT distribution in various subsets of human AF2-2 predicted protein structures

![Graph showing mean pLDDT distribution in various subsets of human AF2-2 predicted protein structures.](image)

TMEM: transmembrane proteins, SOLU: soluble proteins, TM: transmembrane region of TM proteins, non-TM: non-transmembrane regions of TM proteins.
Fig. S2 Low thickness and low CCTOP reliability values may indicate structures forming complexes

(a) The TRPC4 (short transient receptor potential channel 4, ACC: Q9UBN4) exhibits low thickness value and some of its TM helices in the AF2 structure do not pass the membrane, but one forms a reentrant helix and one another is located parallel to the bilayer plane, based on CCTOP prediction with a reliability score of 63. Transmembrane helices based on CCTOP prediction are colored from blue to red from side1 to side2. End positions are labeled with spheres. (b) The structure of mouse Trpc4 (PDBID: 5z96) and its chain A is replaced with human AF2 TRPC4 structure (red, RMDS: 0.6Å). Light gray: chain B, yellow: chain C, gray: chain D, light blue: membrane bilayer predicted by OPM (https://opm.phar.umich.edu). Arrows indicated regions, which are not present in the experimental structure and do not immerse irrationally deep into the membrane region. Thus AlphaFold2 may have some idea about the presence of membranes around TM proteins.
Fig. S3 Structures used as reference ABC transmembrane domain folds

| fold class  | reference PDB |
|-------------|---------------|
| Pgp-like    | 4m1m          |
| ABCG2-like  | 6hco          |
| MalFG-like  | 2r6g          |
| BtuC-like   | 4fi3          |
| EcfT-like   | 4huq          |
| LptFG-like  | 5x5y          |
| MacB-like   | 5lj7          |
| MlaE-like   | 7ch0          |

One TMD structure was extracted from a structure from each fold class. These TMD structures were used as reference folds for classifying TMD domains using TMalign.
Fig. S4 The AF2 structure of human CFTR (ACC: P13569)

Side view and top view from the extracellular space. Blue and green: disordered regulatory domain and linker region, respectively; none of them enters the hydrophobic bilayer region which is indicated by dashed lines; red: dynamic regulatory insertion, which is located in the first nucleotide binding domain and is only partially resolved in isolated NBD1 or full-length structures.
Two ABCG2 (Q9UNQ0) TMD sequences were linked with a CFTR R domain segment and subjected to AlphaFold2 prediction without template usage. Side and top views from the extracellular space, the R domain segment is not shown. Grey: experimental structure (PDBID: 6hco), blue: AF2 model, red sticks: AF2 cysteine residues forming intramolecular and intermolecular disulfide bonds, located in the extracellular loop 3.
Fig. S6 Comparing MD simulations with the homology model and the AtABCg36 AF2 model (ACC: Q9XIE2).
(a) AF2 provided structures for regions that were difficult (non-conserved extracellular loops #3 and helices between NBD and the linker, labelled by purple and red, respectively) or impossible (N-terminal helices, blue) to homology model. (b) We misaligned the non-conserved β1 strand in the homology model, which segment is also registry shifted in ABCG8 experimental structure (Fig. 3a). TM6 was misaligned because the proceeding non-conserved EL3. (c) The homology model highly deviated from the initial structure more than the AF2 model based on RMSD (root mean square deviation). (d) The AF2 structure exhibited stability also on longer time scales (500 ns). (e) The homology model exhibited fluctuations in EL3 and β-strands of NBD1 higher than observed for the AF2 model in MD simulations. (f) Warmer color and larger thickness indicate higher B-factor, which was calculated by GROMACS rmsf tool for frames between 50 and 100 ns. The disordered regions a.a. 1-39 and 809-863 of AtABCG36, were removed before MD system building, since they exhibited lower quality score and could have possibly interfere with the simulation.
Table S1 Potentially erroneous AF-predicted human TM structures with hydrophobic thickness lower than 15Å or higher than 35 Å

| UniProt ACC | d < 15 Å 416 models 350 proteins |
|-------------|----------------------------------|
| Q01668, Q9NSA2, Q7Z4N2, Q9Y2P4, Q7ZW7M0, Q43462, Q8IWY9, Q9NUH8, Q8IZQ1, Q2QZL4, P21817, Q5B6D5, Q9NZV8, Q8WW34, Q9N623, Q94671, Q96TS4, Q13563, Q9H6D3, Q3KNW5, Q86UB9, Q9HS15, Q5K4L6, Q14534, Q9UKG4, Q9Y4D8, Q92521, Q75298, Q8IV84, Q13507, Q5BKT4, Q13571, A5PLL7, Q9UL62, P31513, A8MWL7, Q14735, Q9BR84, Q48651, Q95180, Q14880, Q9Y2E8, Q96PS6, Q7RTX7, Q5UK3, A6NKG5, Q92819, Q15858, Q8NBRO, Q9UND9, Q9UNB4, Q9NV58, Q9BDT3, Q9PO55, A5PLK6, Q15878, Q92575, Q9BQ86, Q9Y2U2, Q6PP77, Q29252, Q96BM0, Q14573, Q9Y69M7, P51811, P51795, Q95406, Q75T13, P22459, Q96L42, Q6IF36, Q9BSX9, Q8WUD6, P42858, P36269, P40879, Q9PD28, A8MWKO, P38435, Q53F2P, Q2VPB7, Q9UK17, Q9Y4A5, Q9BV8K, Q962F5, Q99456, Q6UX98, Q14524, Q6UVM3, Q14689, Q95279, Q6P1A2, Q7Z2W7, Q5GH77, Q12908, Q5WOB7, Q9Y210, Q14674, Q9H0A0, Q9NX78, Q00555, B2RXF0, Q9Y5Y9, Q94919, A14AF0, Q86VD9, Q6UNII, Q6ZRR5, Q7ZSM5, Q9U133, Q9UBU6, Q8NGG1, Q15056, Q6ZW7T, Q8TE9B, Q6ZN68, P30988, Q9Y2K9, Q57T71, Q4GON8, Q9BY78, Q9H2B4, P48995, Q9BWWV8, Q5VT9Y, Q9NRC1, Q96GC9, Q7Z449, Q6NTF9, Q8TEQ8, Q9H3K2, P46977, Q29604, Q2LD37, Q9NXE4, Q9UBY0, Q9JWIV, Q16799, Q96N66, Q92543, Q9UQ05, Q9NV0V, Q00975, Q96PR1, Q9Y672, Q5TEA6, O43511, A6NM10, P35499, P0C7U9, Q8IZY2, P6I25, P51787, Q13488, Q9NR6, Q6ZUK4, Q9BCVE, C9QIJ7, Q8NCM2, Q6NX76, Q8VVJ1, Q14667, Q9RPOX4, Q8NFU1, Q6PC67, P48751, Q6XZV5, Q8WVW5, Q95197, Q9U14, O00219, O43525, Q13936, Q9HB14, A6NKP2, Q6P4A7, Q8NM1, Q86WA9, Q96B96, Q7Z3C6, Q9NRR7, Q76024, Q7LBE3, Q9Y257, Q6NZC8, Q9UH99, Q9NR82, P50443, Q6TCH4, Q71RG4, Q8UI89, Q9UPG8, Q96HV5, Q06840, P15650, Q01118, Q93210, Q8IY26, A0A1BOGVZ9, Q9UKU0, Q92839, O43526, Q5T197, Q12791, Q9H0H0, H0YL14, O94911, Q5VW36, Q6P1M0, A6NFY4, P31213, Q9PO03, Q9NT52, Q725H4, Q8NVB4, Q9BV9G, Q96NB2, Q86UL3, Q9PO59, P35498, Q9UM55, Q8NW64, Q9G2Z3, Q74418, Q9H330, Q7L1W4, Q9BV10, Q9NY46, O60427, Q6ZVX9, Q9H6F2, Q99250, Q6ZM20, Q8TD71, Q13224, Q14656, Q9B8X9, Q03721, Q6UX04, P35610, Q8TE81, Q14649, Q13535, Q81U07, Q9ULQ1, Q1740, Q14524, Q9H490, Q9427, Q14003, Q9BRB3, Q9NB94, P58743, Q2T9J0, Q9UQDO, Q8TCJ2, P40305, Q86XZ9, Q9P7N7, Q10981, Q9NT68, Q9UGZ2, Q56HT5, Q6N22, Q59826, Q6UVW2, Q5JSX7, Q99942, Q6YX33, H3BN30, Q96SEO, Q6BCQ1, Q15392, Q15013, Q86JSJ2, Q7Z407, Q96C03, B6SEH9, Q16836, Q5STY1, Q75445, Q96AD5, Q9UKZ4, Q86X9, Q6VZ70, Q9U14, Q9UCL5, P60509, Q8IKX2, Q68CK6, P0DMW4, Q9H1A4, Q9NSK7, Q8WVF7, Q9PF26, Q8TCCG1, Q2TAAS, Q96GF1, P54829, O43914, Q8IYT2, P09110, Q97322, Q9H267, Q3SX5M, Q08AH3, Q9UNU1, A6NH66, P51841, Q75871, Q9PO53, Q902F9, Q962V1, Q9NN41, Q6GM87, Q86V5S, Q9S237, Q6GPH6, Q95976, Q6P1Q0, Q96M2O, Q9NNZC3, P61565, Q8NNF7, Q9UQ52, Q96ER9, Q96HA1, H3BR10, P0DMW5, Q96AX1, A8CG34, Q7RT59, Q16586, Q95YW7, Q5ITH9, Q9UKJ5, Q6DT37, Q96P6N, Q5VXU3, Q8TC41, Q5SWX8, Q07812, Q8N138, A2RU48, Q16795, Q06136, Q96958, Q96U76, Q9H689 |

| UniProt ACC | d > 35 Å 309 models 296 proteins |
|-------------|----------------------------------|
| Q8NF9P, Q86UQ4, P21817, Q9UKL4, P13569, Q2LD37, A6NH66, Q96HA4, Q8N539, Q07011, Q72419, Q9H4F1, E9PIF3, Q6UX15, P0D80, A6NGB0, Q14318, Q86T13, Q13449, Q13508, Q16620, Q9NY15, Q75900, I3L273, P34810, Q9UK23, Q9UMF0, P22748, Q8IZ0J, Q5STR5, Q6N022, Q8N2Q7, P23467, Q9BSX4, Q15389, Q8N158, Q9UKZ4, Q75581, Q14669, Q5V4V3, P14415, Q96RD9, P06731, Q72407, Q96YD3, Q5VT99, P00387, Q14BN4, E5RHQ5, Q96MH6, Q00198, Q71RC9, Q9ULX5, Q8JO25,
Table S2 PFAM entries used to detect transmembrane ABC domains in the 21 proteomes with AF predictions

| PFAM NAME                  | PFAM ACC   |
|----------------------------|------------|
| 12TM_1                     | PF09847.11 |
| ABC2_membrane              | PF01061.26 |
| ABC2_membrane_2            | PF12679.9  |
| ABC2_membrane_3            | PF12698.9  |
| ABC2_membrane_4            | PF12730.9  |
| ABC2_membrane_5<sup>a,b</sup> | PF13346.8 |
| ABC2_membrane_6            | PF06182.13 |
| ABC2_membrane_7            | PF19055.2  |
| ABC_export                 | PF16962.7  |
| ABC_membrane               | PF00664.25 |
| ABC_membrane_2             | PF06472.17 |
| ABC_membrane_3             | PF13748.8  |
| ABC_tran_2                 | PF16949.7  |
| BPD_transp_1               | PF00528.24 |
| BPD_transp_2<sup>a,c</sup> | PF02653.18 |
| CbiQ                       | PF02361.18 |
| CcmB<sup>a,e</sup>         | PF03379.15 |
| DUF2705                    | PF10920.10 |
| DUF3526                    | PF12040.10 |
| DUF3533<sup>a,c</sup>      | PF12051.10 |
| DUF6198                    | PF19700.1  |
| ECF_transprt               | PF12822.9  |
| FecCD                      | PF01032.20 |
| FtsX                       | PF02687.23 |
| Lptf_LptG                  | PF03739.16 |
| PDR_CDR                    | PF06422.14 |
| SbmA_BacA<sup>a,d</sup>    | PF05992.14 |
| YitT_membrane<sup>a,b</sup> | PF02588.17 |

<sup>a</sup> PFAM families without associated PDB structures  
<sup>b</sup> Our results indicated that YitT is not an ABC fold and incorrectly placed into the ABC-2 PFAM clan  
<sup>c</sup> The BPD_transp_2 HMM was identified in bacterial permeases and their AF2-predicted TM structures exhibit BtuC-like fold  
<sup>d</sup> AF2 structures of importers with SbmA_BacA HMM hits are Pgp-like  
<sup>e</sup> All ABC2_membrane_5, CcmB, and DUF3533 HMM hits include conformations corresponding to the ABCG2-like fold
Table S3 Opening of zfCFTR structures with kinked TM8 (PDBID: 5W81) and with straight TM8 in MD simulations

| system        | open frames in each trajectory, #                                      |
|---------------|------------------------------------------------------------------------|
| kinked TM8    | 427 in one and none in the other 21 trajectories*                       |
| straight TM8  | 0, 227, 174, 560, 441, 793**                                           |

* 22 simulations: 6x100 ns and 16x35ns
* 6 simulations: 6x100 ns