

# AlphaFold2 transmembrane protein structure prediction shines

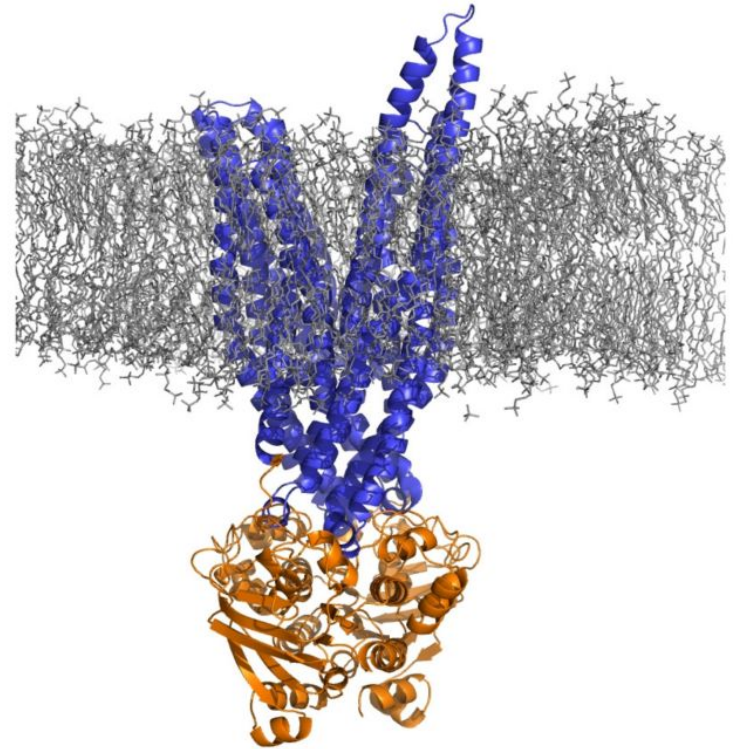
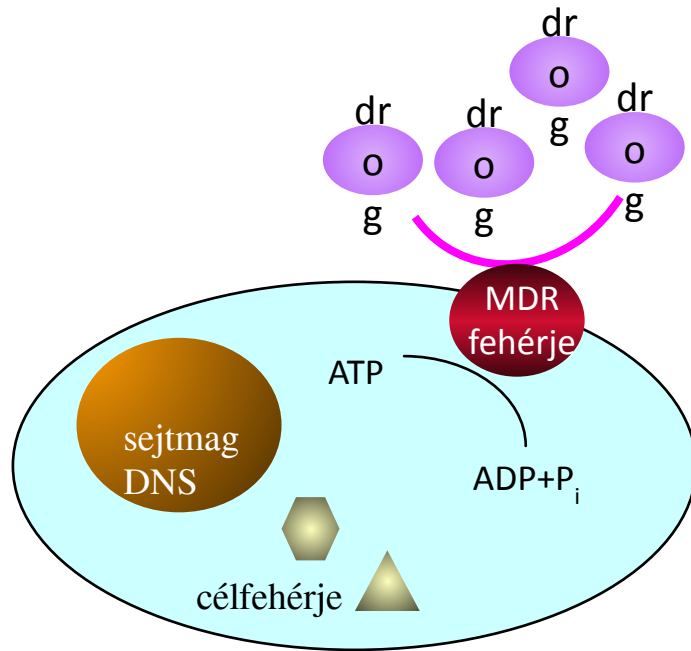
Tamás Hegedűs

[tamas.hegedus@hegelab.org](mailto:tamas.hegedus@hegelab.org)

Dept. of Biophysics and Radiation Biology  
Semmelweis University



# Multidrug transport – membrane ABC proteins

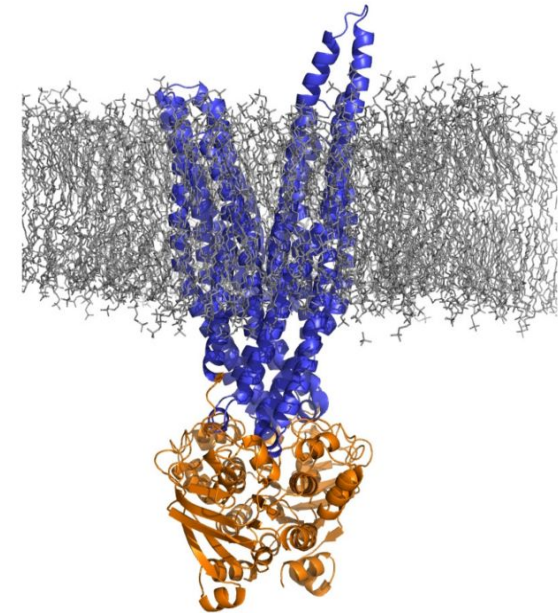
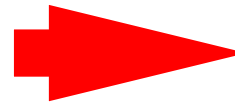


# Structure prediction from sequence

Christian B. Anfinsen, Nobel prize in 1972

>MDR1\_HUMAN | P08183

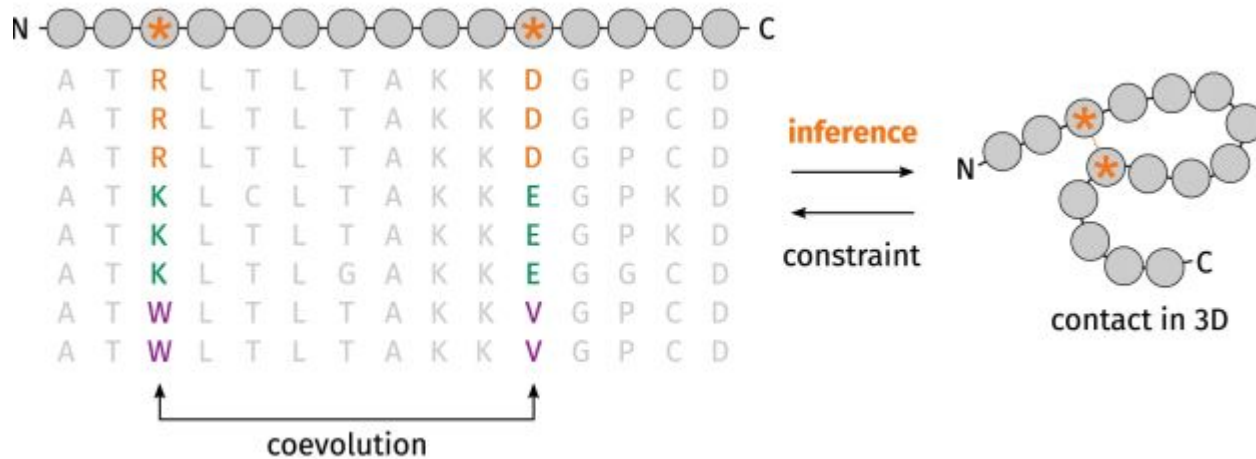
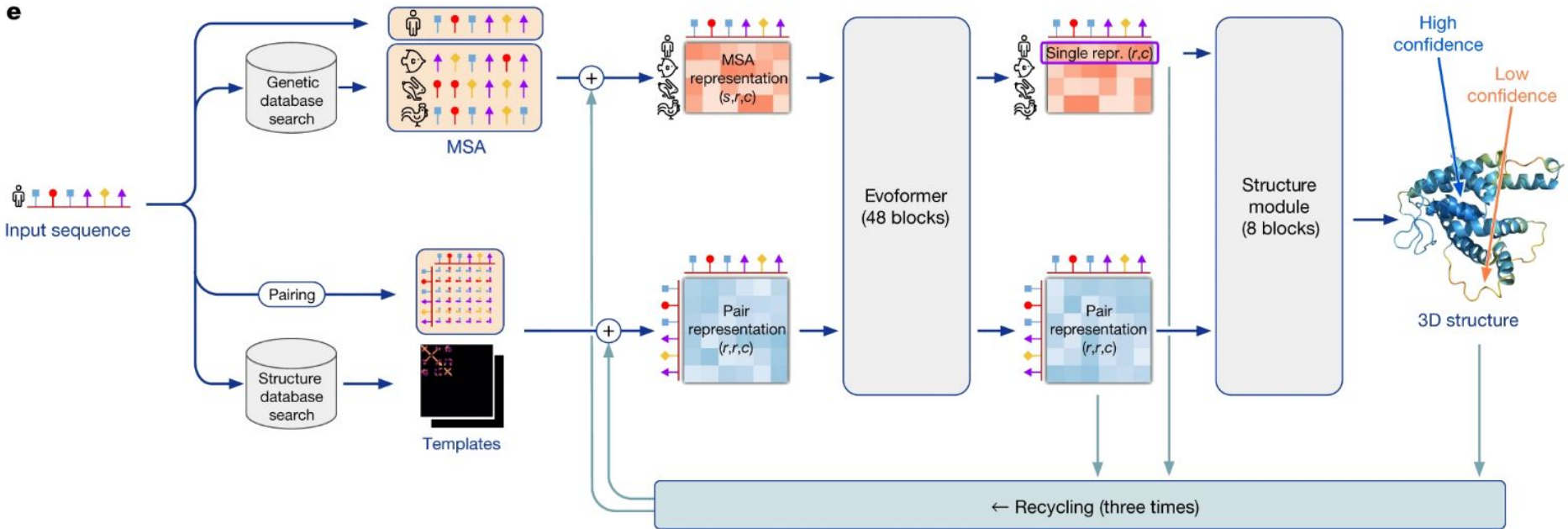
```
MDLEGDRNGGAKKKNFVKLNKSEKDKKEKKPTVSVFSMFRYSNWLDKLYMVVGTAAII
HGAGLPLMMLVFGEMTDIFANAGNLEDLMSNITNRSDINDTGFFMNLEEDMTRYAYYYSG
IGAGVLVAAYIQVSFWCLAAGRQIHKIRKQFFHAIMRQEIGWFDVHDVVELNTRRLTDDVS
KINEGIGDKIGMFFQSMATFFTGFI VGFTRGWKLTLVILAI SPVLGLSAAVWAKILSSFT
DKELLAYAKAGAVAEVLAAIRTVIAFGGQKKELEERYNKNLEEAKRIGIKKAITANISIG
AAFLLIYASYALAFWYGTTLVLSGEYSIGQVLTVFFSVLIGAFSVGQASPSIEAFANARG
AAYEI FKI IDNKPSIDSYSKSGHKPDNIKGNLEFRNVHFSYPSRKEVKILKGLNLKVQSG
QTVALVGN SGC GKSTTVQLMQRLYDPT EGMVSV DGD IRTINVRFLREIIGVVSQEPVLF
ATTIAENIRYGRENVTMDEIEKAVKEANAYDFIMKLP HKFDTLVGERGAQLSGGQKQRIA
IARALVRNP KILLLDEATSALDTESEAVVQVALDKARKGR TTI VIAHRLSTVRNADVIAG
FDDGVIVEKGNHDELMKEKGIYFKLVTMQTAGNEVELENAADESKSEIDALEMSNDSRS
SLIRKRSTRRSVRGSQAQDRKLS TKEALDESIPPVSFWRIMKLNLT EWPYFVVGVFCAII
NGGLQPAFAIIFSKIIGVFTRIDDPETKRQNSNLFSLFLALGII SFITFFLQGF TFGKA
GEILTKRLRYMVFRSMLRQDVSWFDDPKNTTGAL TTRLANDAAQVKGAIGSRLAVITQNI
ANLGTGIIISFIYGWQLTLLLLLAIVPIIAIAGVVEMKMLSGQALKDKKELEGS GKIATEA
IENFRTVVSLTQE QKFEHMYAQS LQVPYRNSLRKAHIFGITFSFTQAMMYFSYAGCFRFG
AYLVAHKLMSFEDVLLVFS AVVFGAMAVGQVSSFAPDYAKAKISAAHIIMIIEKTPLIDS
YSTEGLMPNTLEGNVTFGEVVFNYPTRPDI PVLQGLSLEVKKGQTLALVGS SGC GKSTVV
QLLERFYDPLAGKVLLDGKEIKRLNVQWLR AHLGIVSQEPILFDCSIAENIAYGDN SRVV
SQEEIVRAAKEANI HAFIESLPNKYSTKVGDKGTQLSGGQKQRIAIARALVRQPHILLLD
EATSALDTESEKVVQEALDKAREGR TCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQL
LAQKGIYFSMVS VQAGTKRQ
```



# AlphaFold2

machine learning, deep learning, AI

Jumper et al. Nat 2021



# AF2 – IT parts

## sequence searches

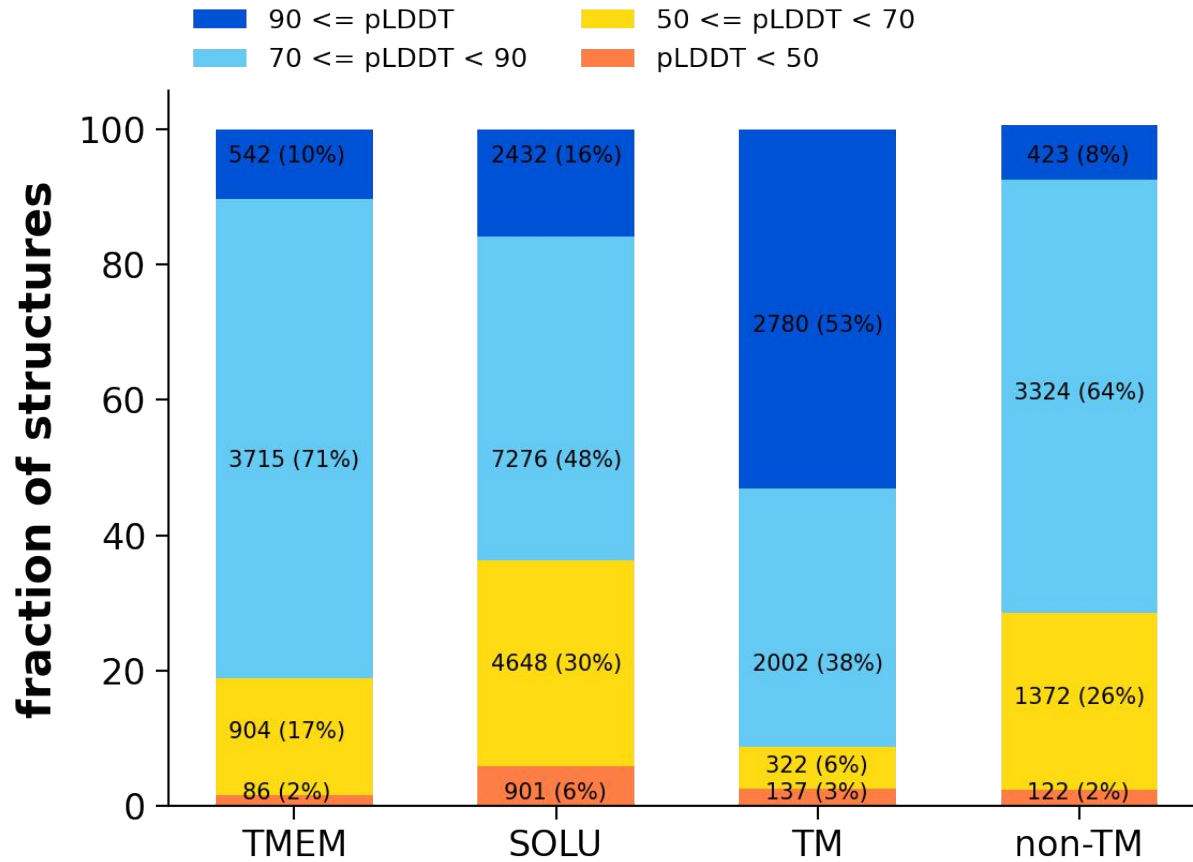
- 2.3 TB disk space for sequence database
  - I/O limited, SSD
  - shrinking to 550 GB and keeping in RAM
- RAM: at least 90 GB
- Parsing the output in RAM
  - it may be over ~100 GB
  - `open(filename).read()`

# AF2 – IT parts

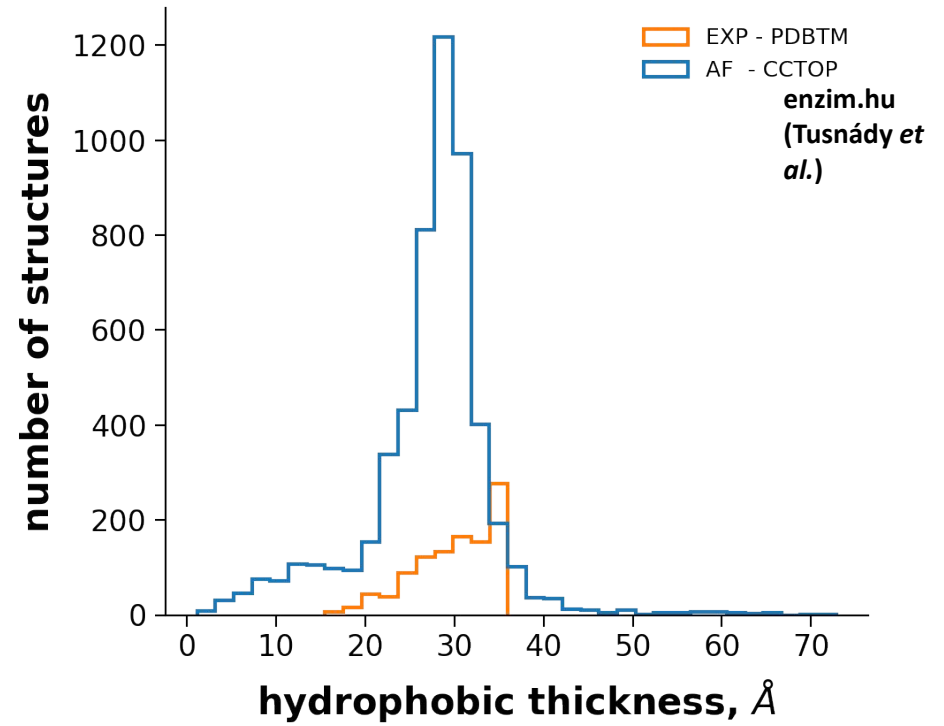
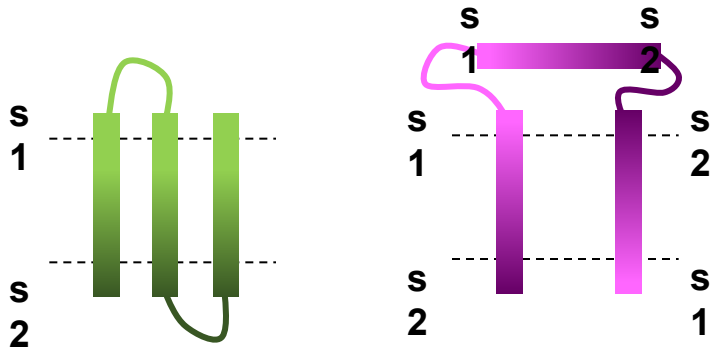
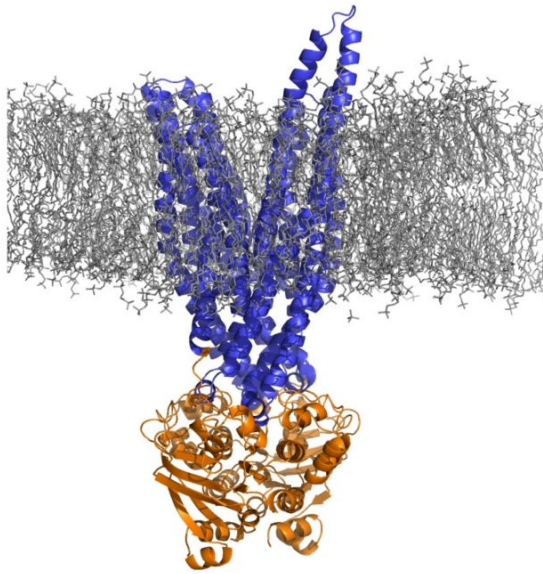
## DL models' running and structure building

- GPU
  - 1,500 amino acid long / 24 GB GPU RAM
  - multi-GPU
  - max. 2 GB limit for a TensorFlow data struct  
bug solved with 2 lines of code

# TM protein structure prediction by AF2



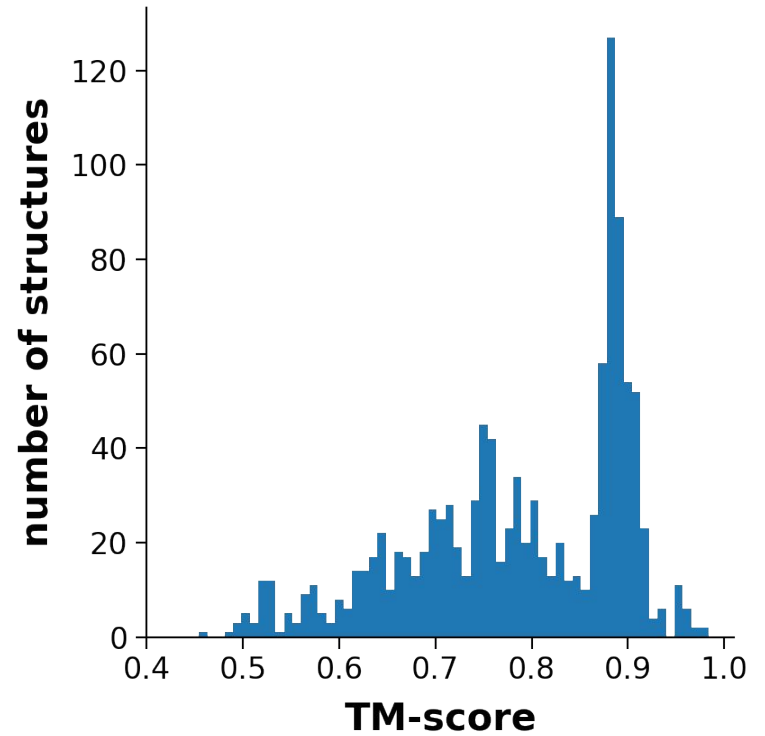
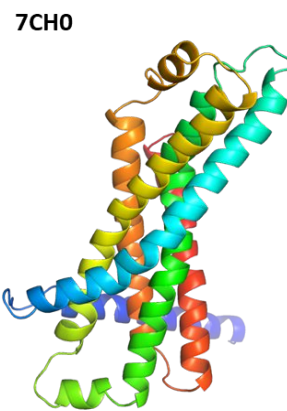
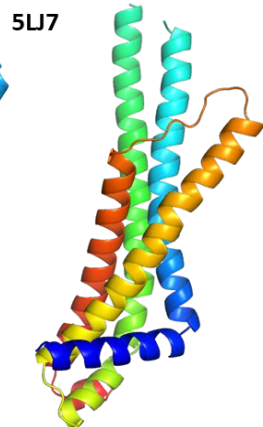
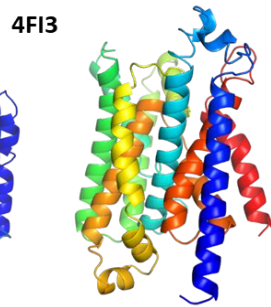
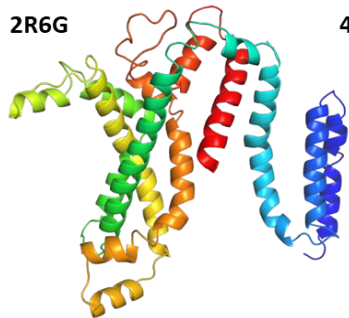
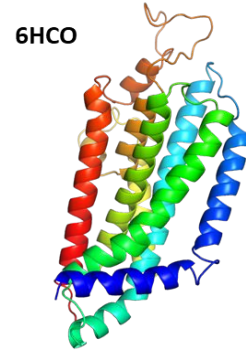
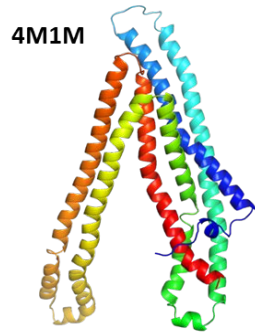
# TM protein structure prediction by AF2





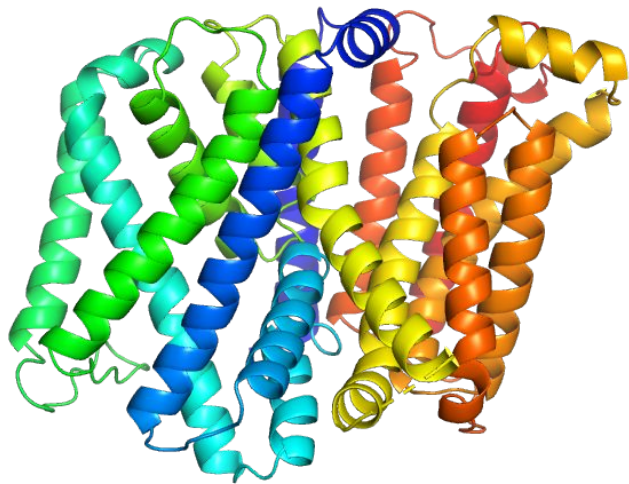
# ABC protein folds

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

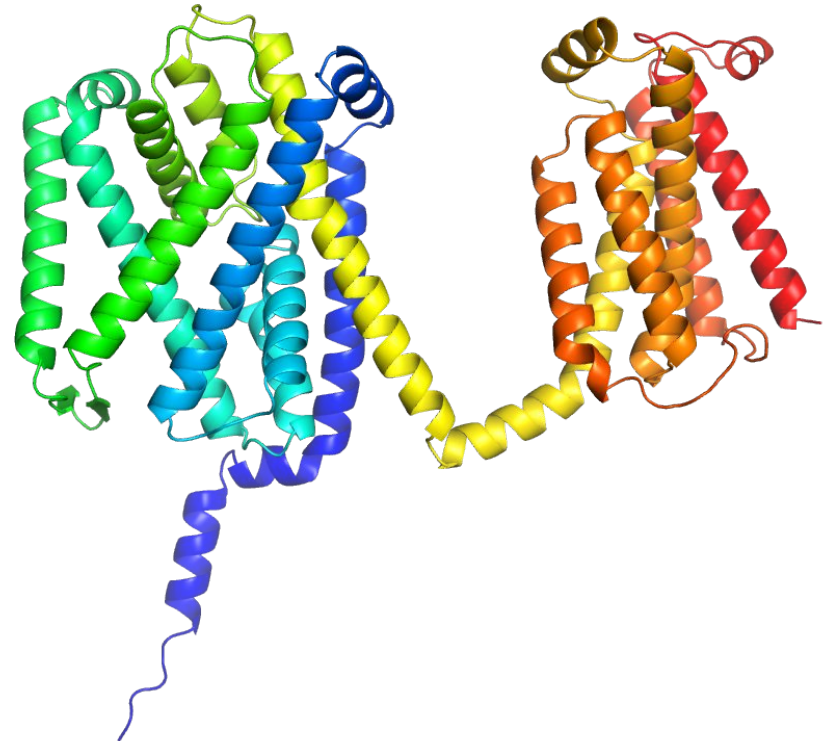


# Prediction of new TM folds

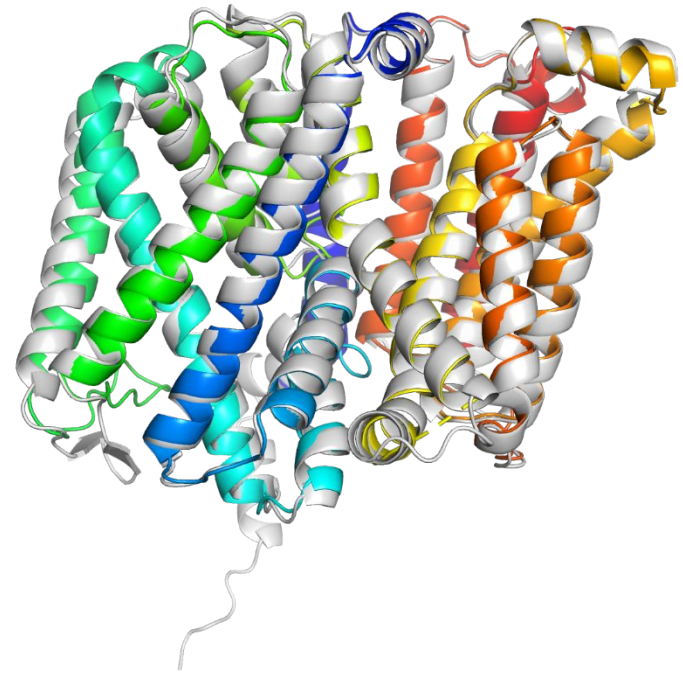
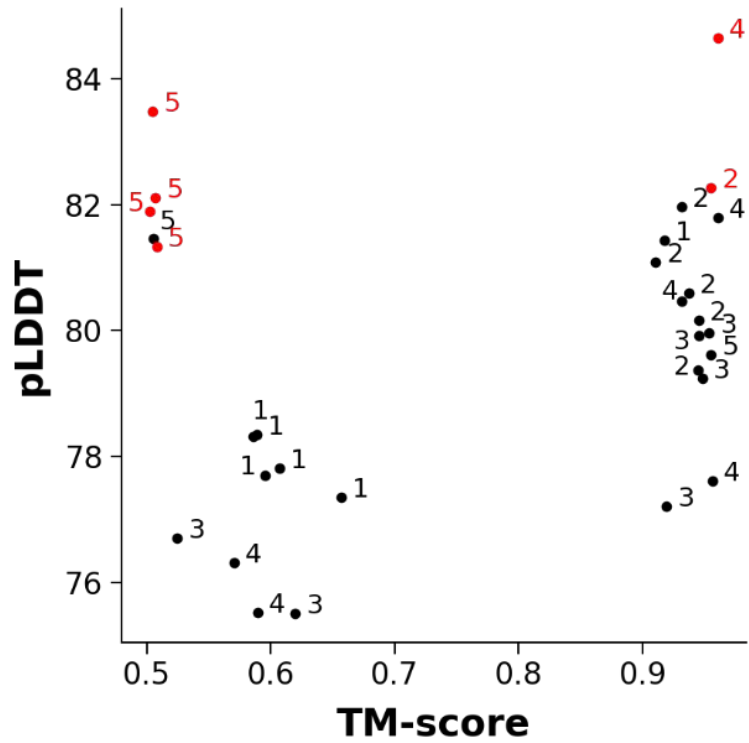
MprF (PDBID: 7DUW)



AF2



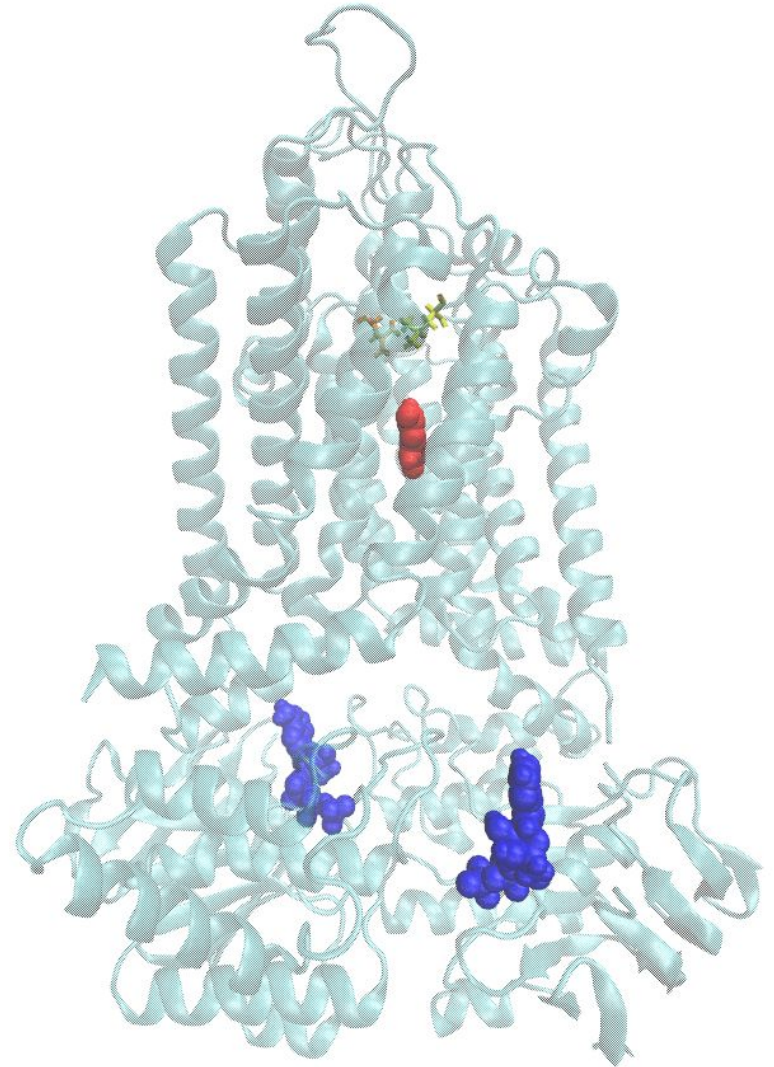
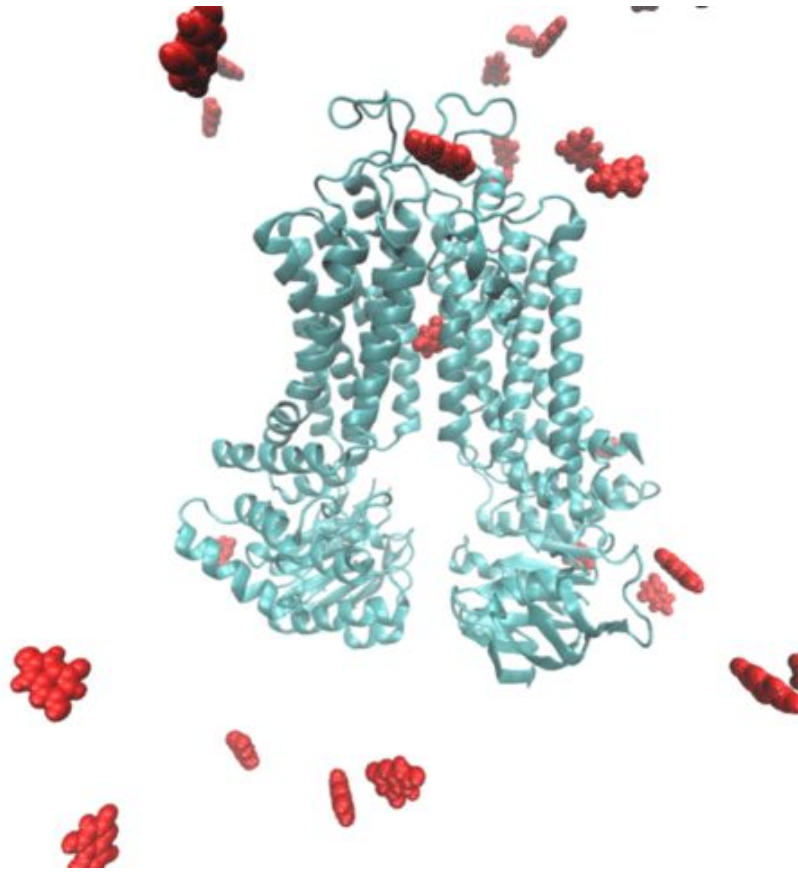
# Prediction of new TM folds



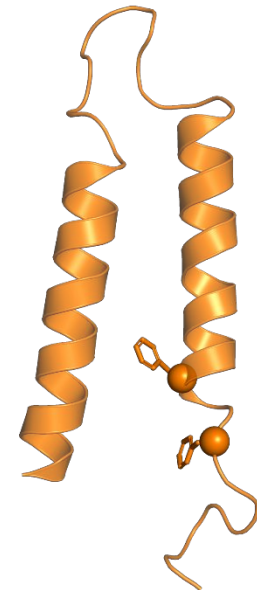
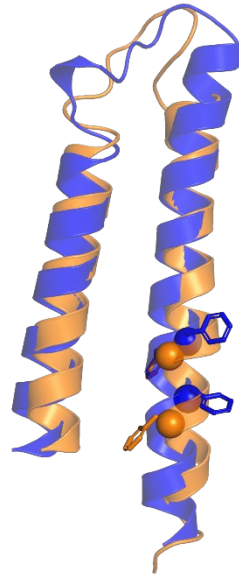
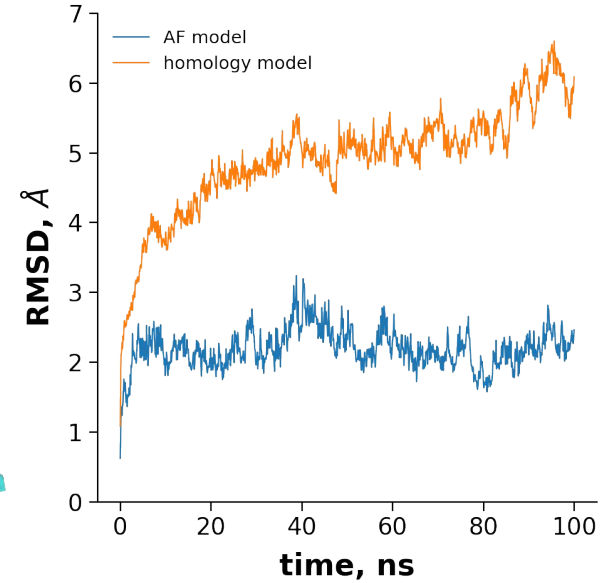
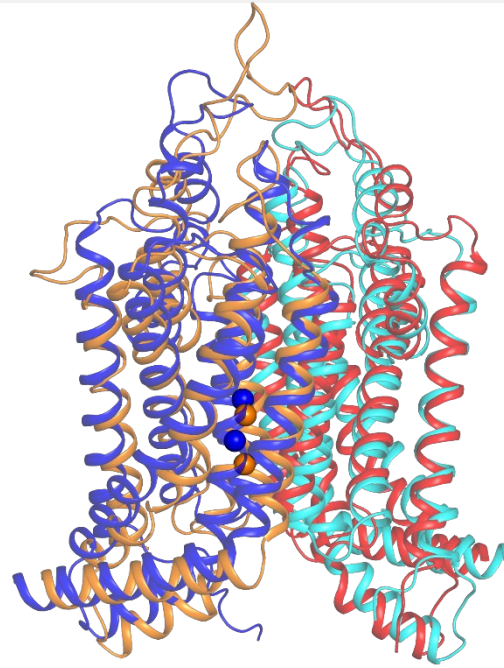
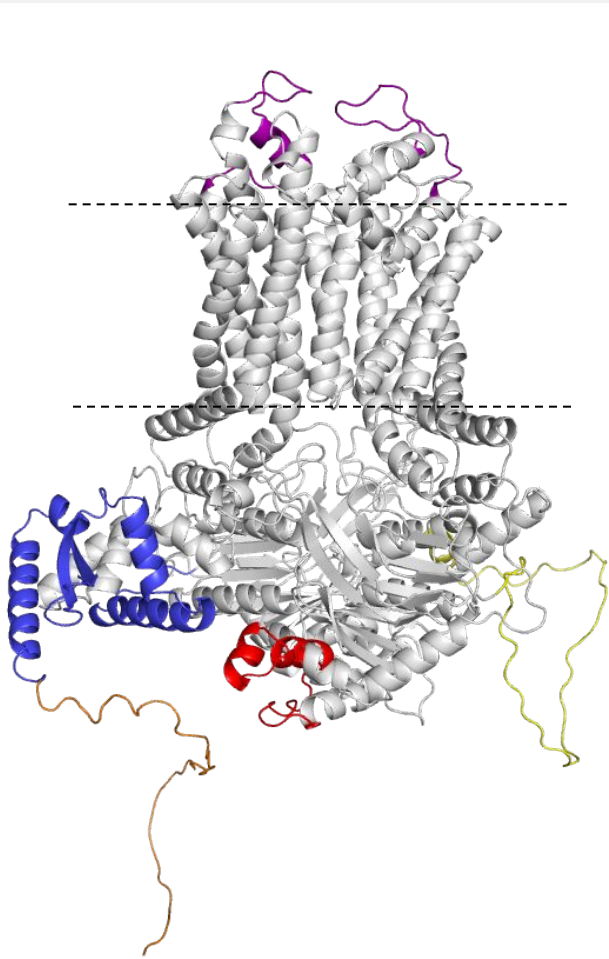
# Describing the transport process

Equilibrium MD and Metadynamics; substrate: uric acid

Nagy *et al.* CMLS 2021



# AF2 AtABCG36 is stable in MD simulations



# Summary

- AF2 performance is great in the case of TM proteins
- AF2-Multimer (released this Tuesday)
- The conformation of the output is „random”
- AF2 code needs some tuning (*kód foldozgatás*)
- The folding problem is **not** solved

Find our preprint and code at  
<http://alphafold.hegelab.org>

# Acknowledgements

hegelab.org



Csizmadia Gina+ReGina

Farkas Bianka

Tordai Hedvig

Lór Krisztina

Padányi Rita

Erzsébet Suhajda

- Gergely L. Lukács, McGill, Montréal
- Markus Geisler, University of Fribourg

- Grubmüller laboratory, MPI, Göttingen
- MTA Wigner, GPU Laboratórium
- Boblem Kft
- NIIF/KIFÜ HPC

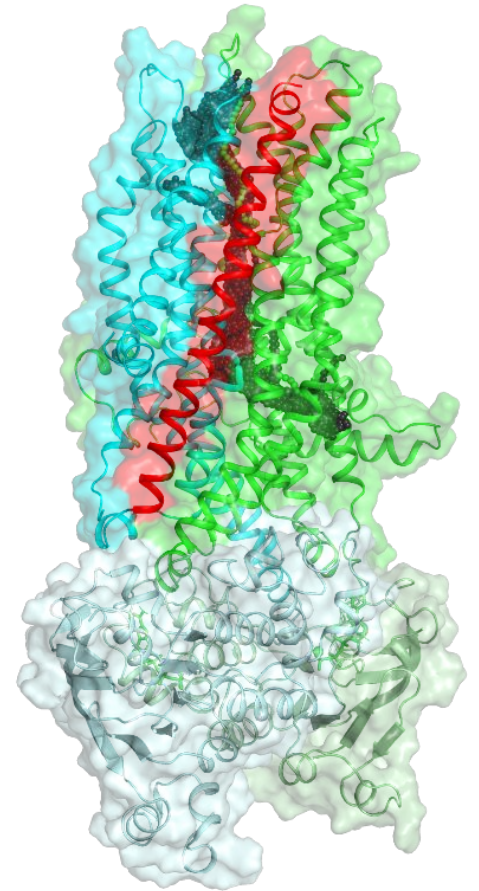
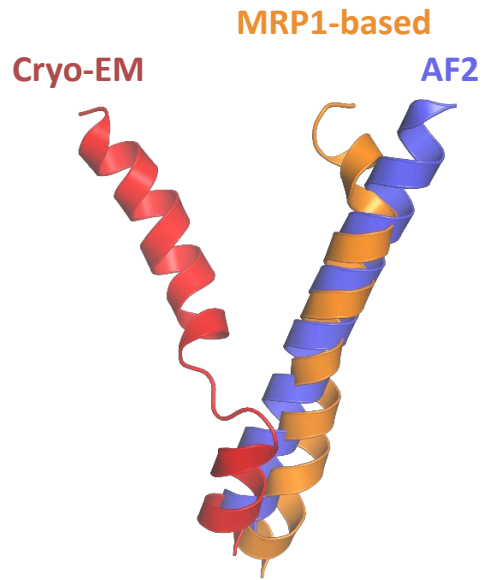
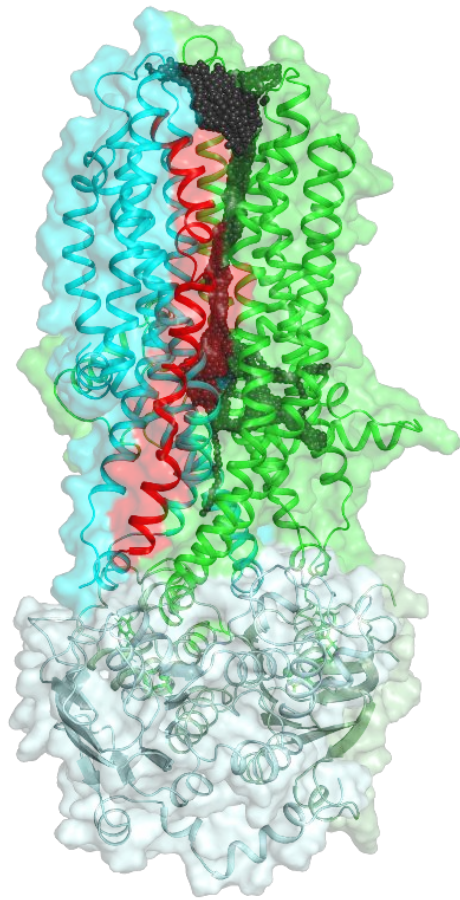
Támogatások:

CFE HEGEDU2010

NKFI-127961

(2018-2022)

# CFTR TM8





# Az ABCG8 regiszter korrekciója

<b>G2</b>	<b>6HCO</b>	AVLSFHNICY	}	<b>X</b>
<b>G8</b>	<b>5DO7</b>	NSLYFTYSGQ		
<b>G2</b>	<b>6HCO</b>	AVLSFHNICY	}	✓
<b>G8</b>	<b>seq</b>	NTLEVRDLNY		
<b>G2</b>	<b>6HCO</b>	AVLSFHNICY	}	✓
<b>G8</b>	<b>AF</b>	NTLEVRDLNY		

