

# Fehérjék szerkezeti bioinformatikája és molekuláris dinamikája

Hegedűs Tamás

tamas@hegelab.org

MTA-SE Molekuláris Biofizikai Kutatócsoport  
SE Biofizikai és Sugárbiológiai Intézet



# Fehérjék szerkezetének és dinamikájának jelentősége

A betegség molekuláris szintű oka?  
A gyógyszer-kötő zseb alakja?

37°C-on, oldatban nem egy szerkezet létezik,  
hanem egy konformációs sokaság.

# Számítógépes modellezés jelentősége

**Atomi szintű információt ad mozgásokról.**

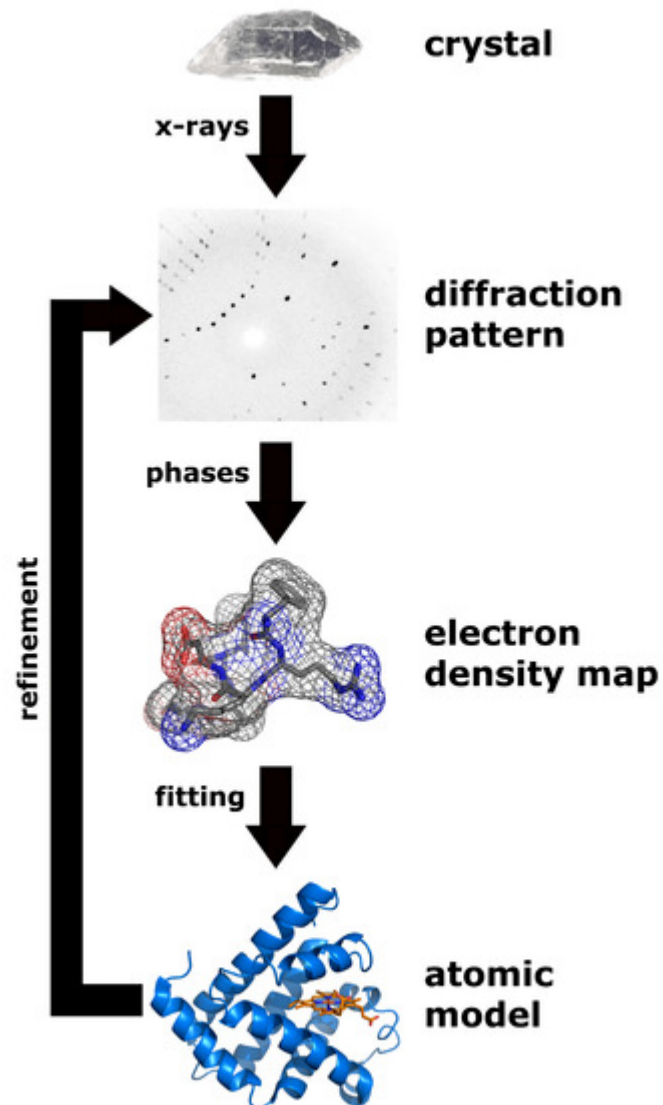
**Kísérletes módszerek  
általában nem szolgáltatnak  
közvetlen információt az atomi szintű történésekről.  
(pl. EPR, X-links: nem, NMR: igen, de korlátoltan)**

# Membránfehérjék szerkezeti modellezése

- Kísérleti adatok alapján
- Elméleti módszerekkel
  - homológia modellezés
  - mélytanulósos módszerek



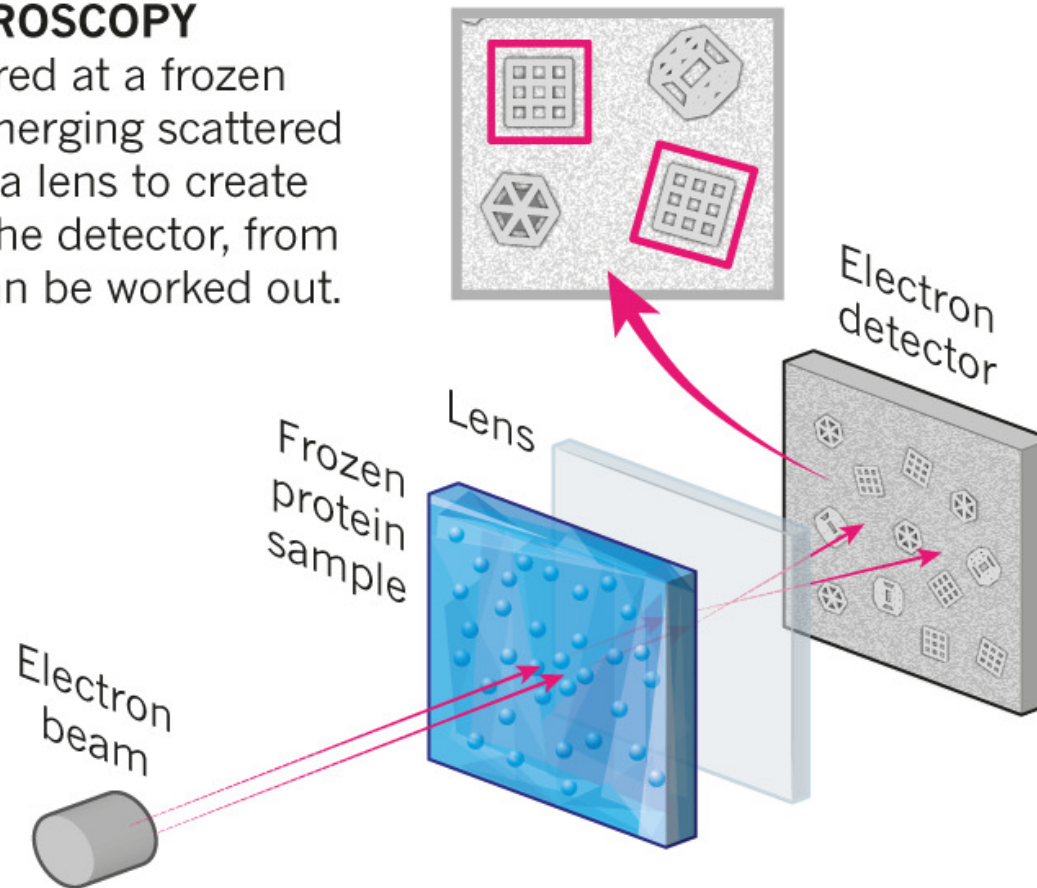
# Szerkezet meghatározás – Röntgen kristallográfia



# Szerkezet meghatározás – krio-EM

## CRYO-ELECTRON MICROSCOPY

A beam of electron is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, from which their structure can be worked out.



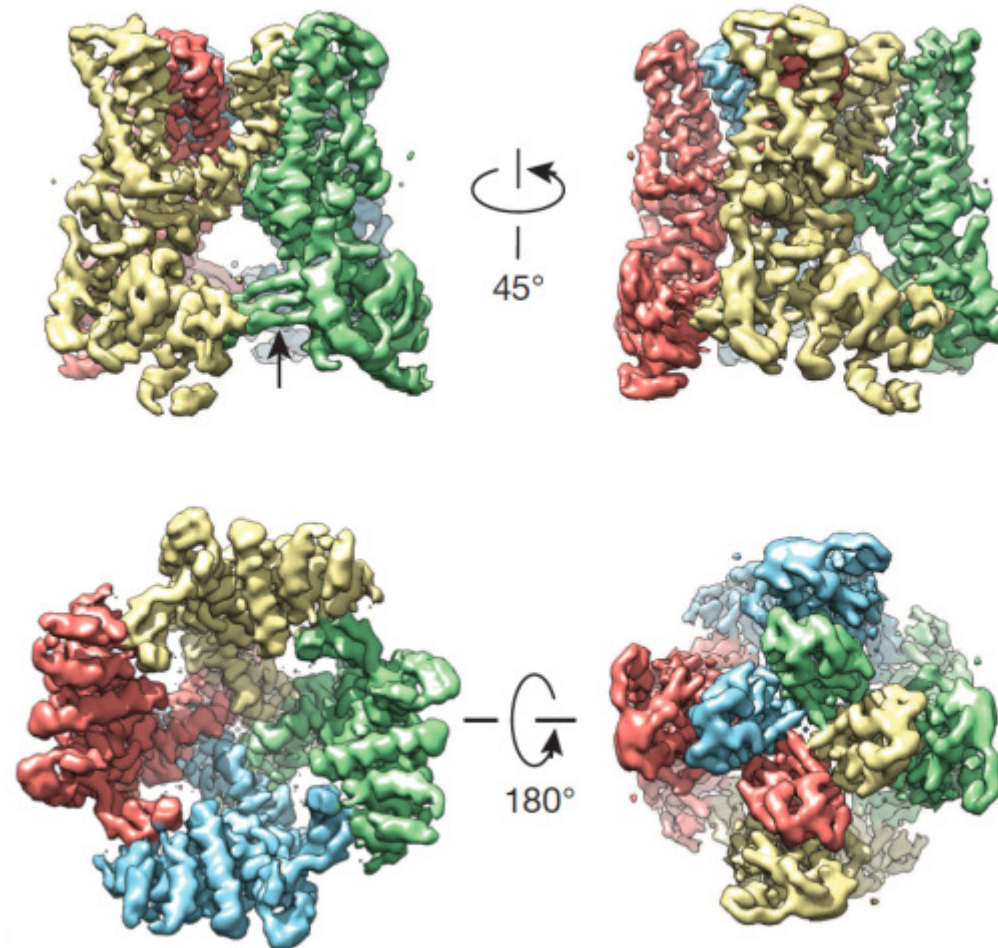
© nature

Ewen Callaway, Nature | News Feature

The revolution will not be crystallized: a new method sweeps through structural biology, 09 September 2015

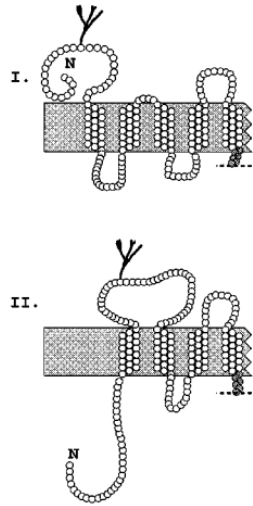
# Szerkezet meghatározás – „single particle”

The TRPV1 channel detects the burn of chilli peppers, and this 3.4-Å structure is considered super-hot in the structural-biology world.

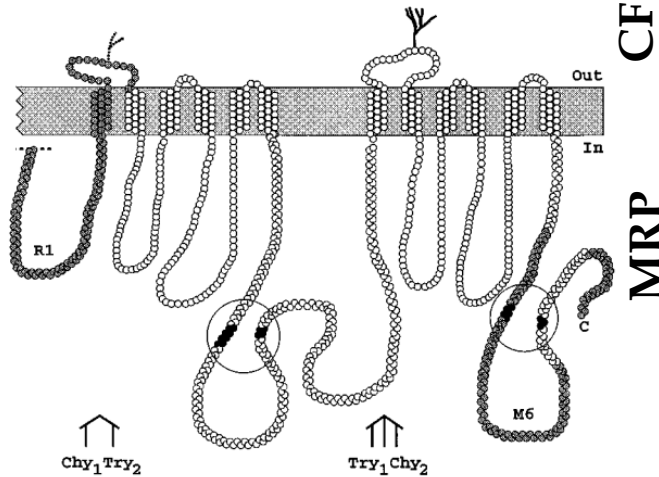


# Membránfehérjék topológiája

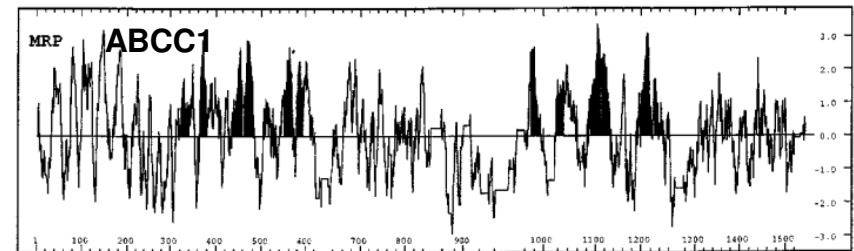
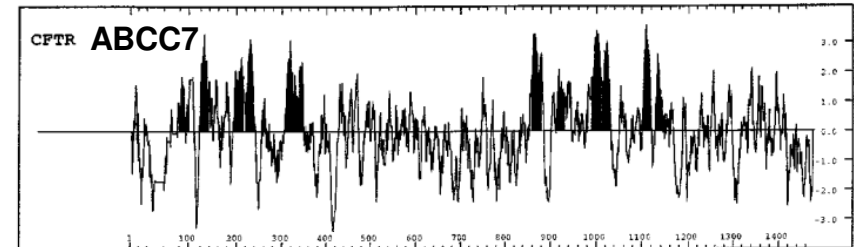
MRP



Bakos *et al.* JBC 1996



Hidrofóbicitási görbék:



CFTR topológiája kísérletekből ismert  
Chang *et al.* J Biol Chem. 1994 269(28):18572-5

Kísérletes topológia meghatározási módszerek:

- Glikozilációs hely és mintázat
- *tag* – felismerés
- Cys hozzáférhetőség

# Membrántopológia és TM hélixek predikciója szekvencia alapján

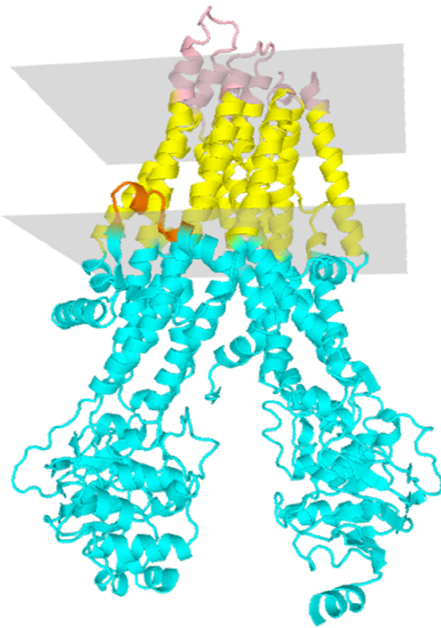
- Szekvencia alapján
  - *positive-inside rule*
  - TM hélixek és nem-TM részek aminosav eloszlása más
- Biológiai tudás figyelembevételének lehetősége
- Több prediktor integrálása  
pl. CCTOP.enzim.hu



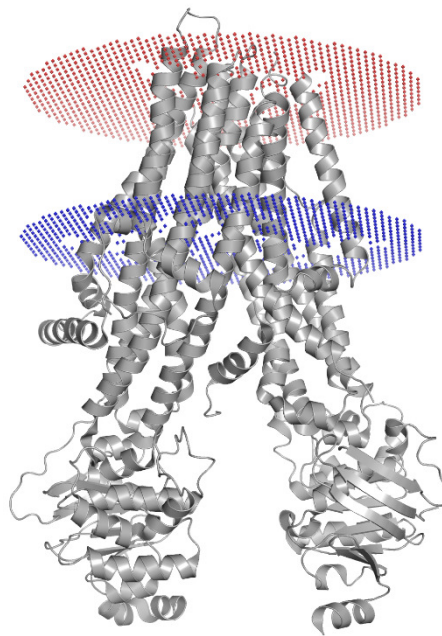
# TM hélixek predikciója szerkezet alapján

1. „water accessible area”, lipid kettősréteg lehetséges helye, pontozás – fitness függvény
2. molekula dinamika szimulációk

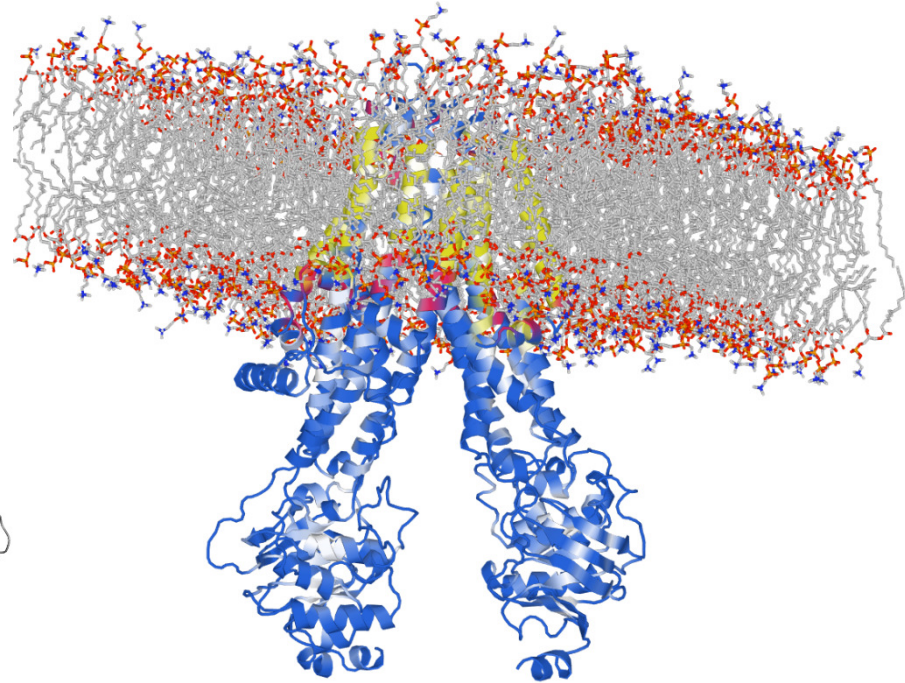
TMDet  
PDBTM



PPM  
OPM

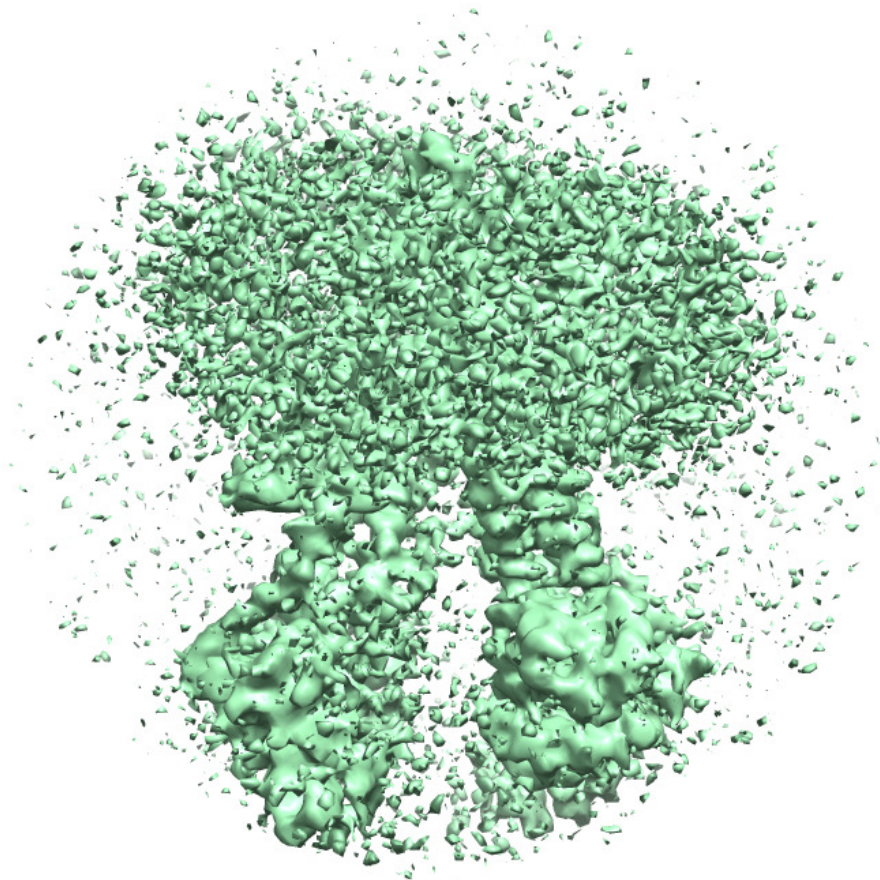


MemProtMD

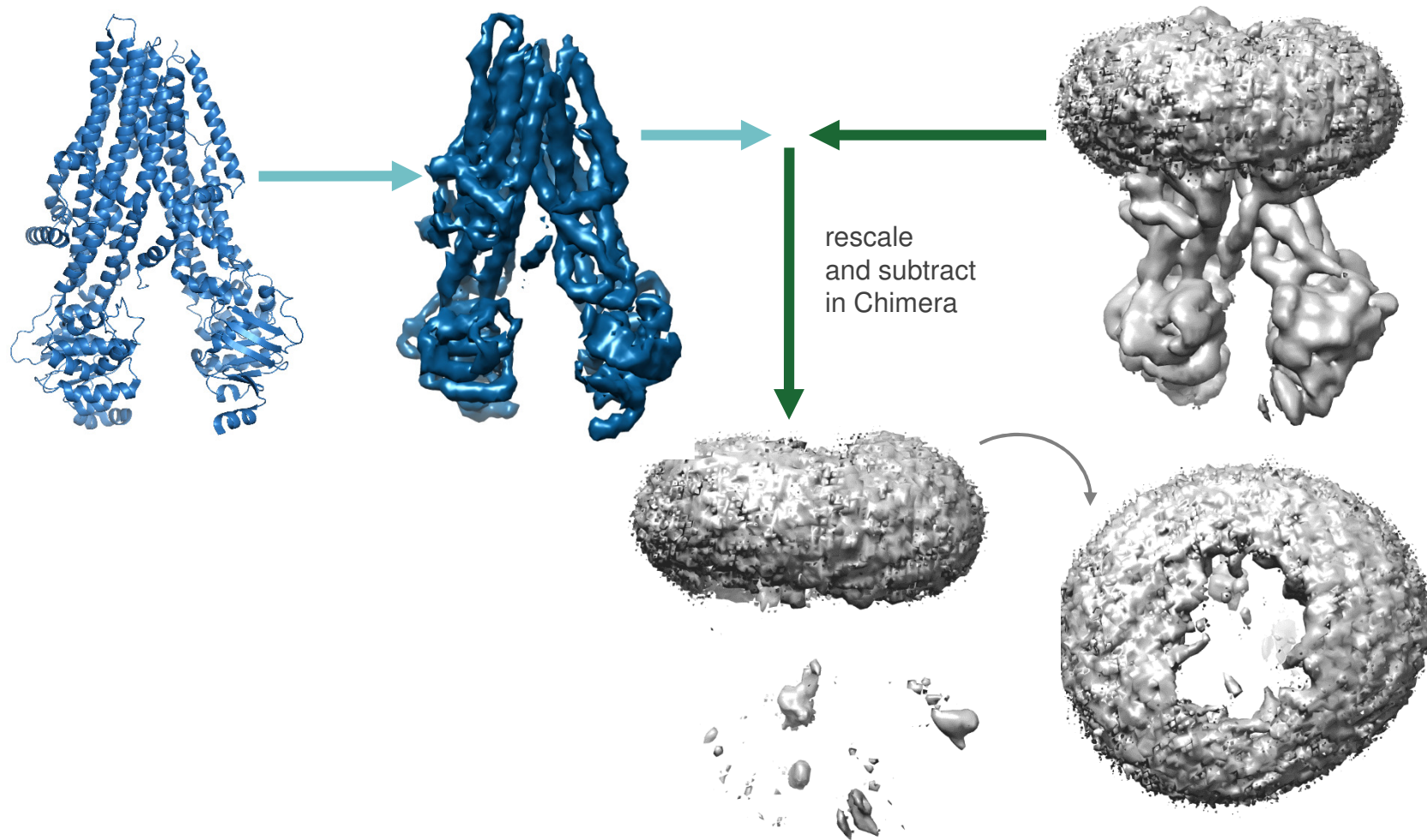


Membrane embedding data is in the  
electron density maps

**CFTR (PDBID: 5UAK) EMD**

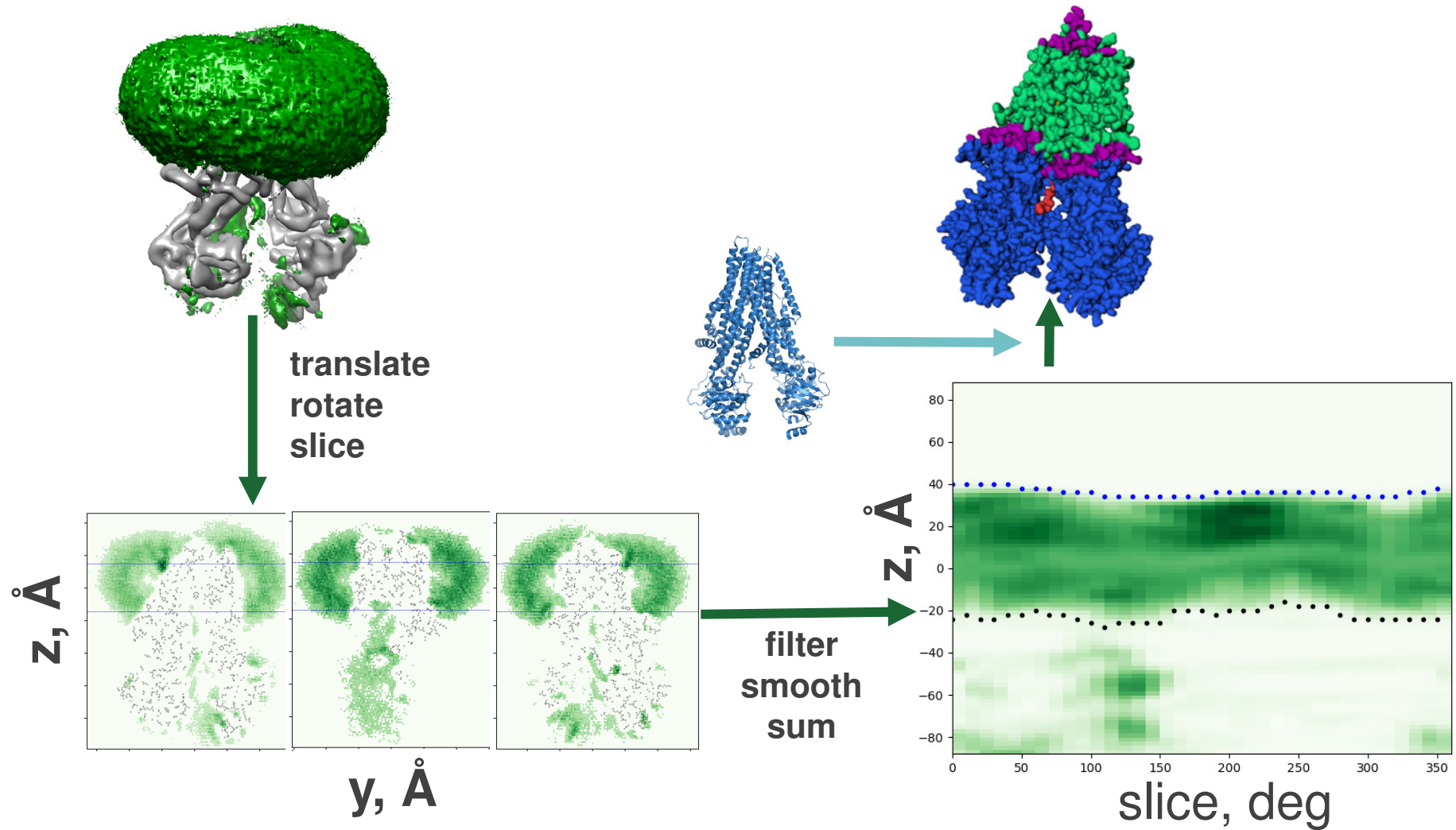


# A membrane blob can be extracted from the electron density map





# The MemBlob can be converted to membrane boundaries



# Homológia modellezés I.

- **Templát keresés**
- **Szekvencia illesztés**
- **Modellezés**
- **Energia-minimalizálás**

**BLOSUM (BLOcks of Amino Acid SUBstitution Matrix) matrix**  
is a substitution matrix

**BLOSUM (BLOCKs of Amino Acid Substitution)**  
is a substitution matrix

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-3	-2	-3	-1	1	-4	-3	-2	11			
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	2	7		
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	0	-3	-1	4	

CLUSTAL W (1.83) multiple sequence alignment

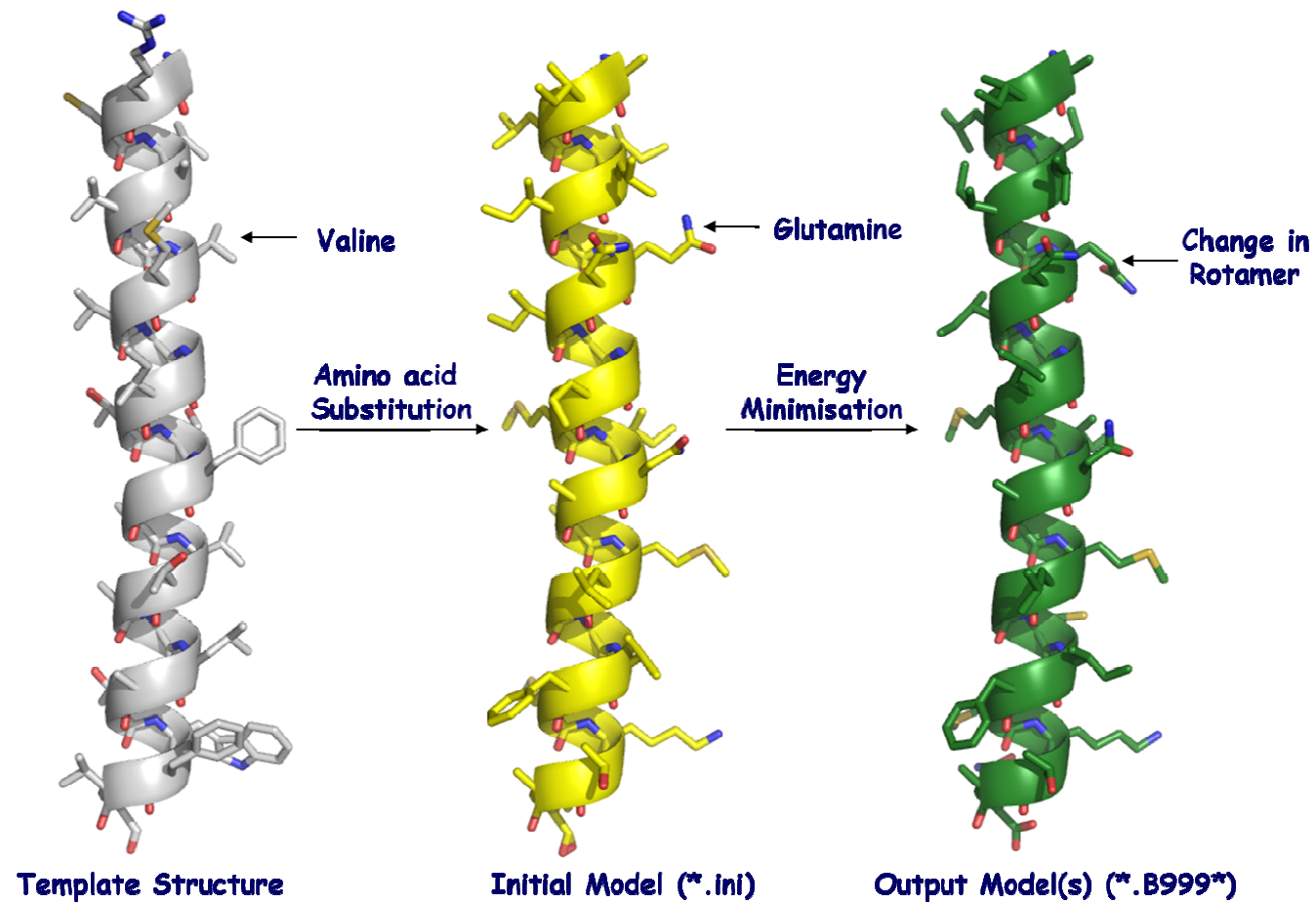
## Alignement – pl. ClustalW

```

2HYD      -----MIKRYLQFVK-----PYKYRIFATIIVGIIKFQIPMLIP
3B5X      -----WQTFKRLWTYIR-----LYKAGLVVSTIALVINAAADTYMI
CFTR_HUMAN MQRSPLEKASVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLS
              *      :      :      *      :      :      *      :      :

```

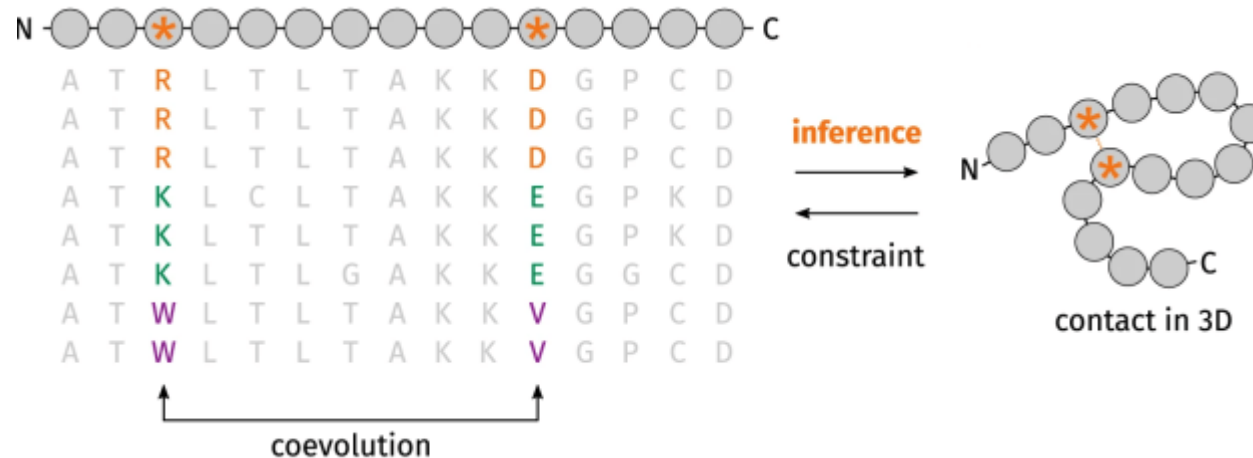
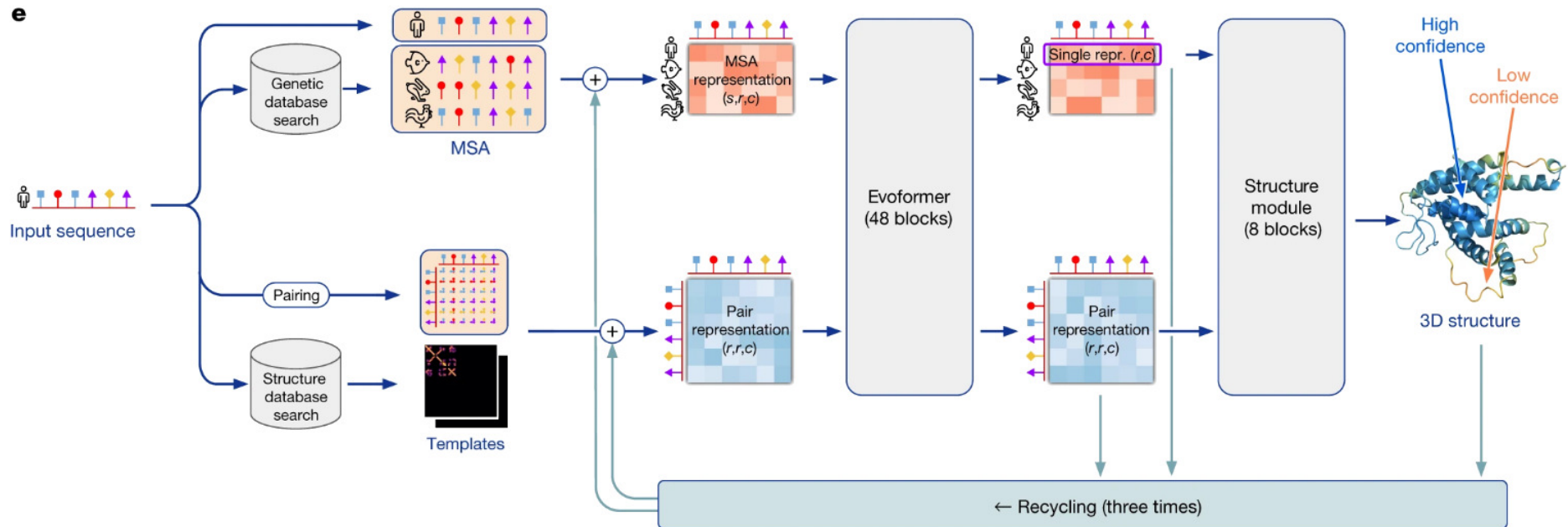
# Homológia modellezés II.



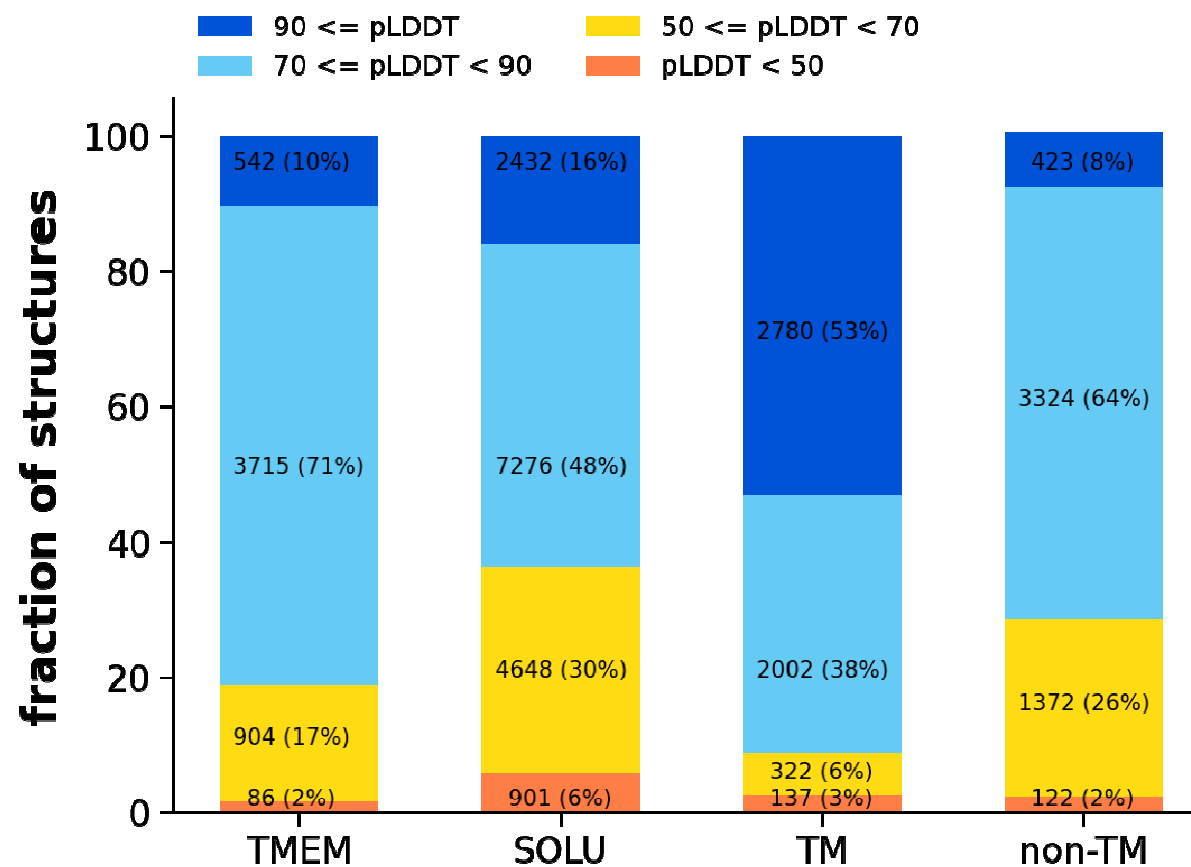
# AlphaFold2

machine learning, deep learning, AI

Jumper et al. Nat 2021

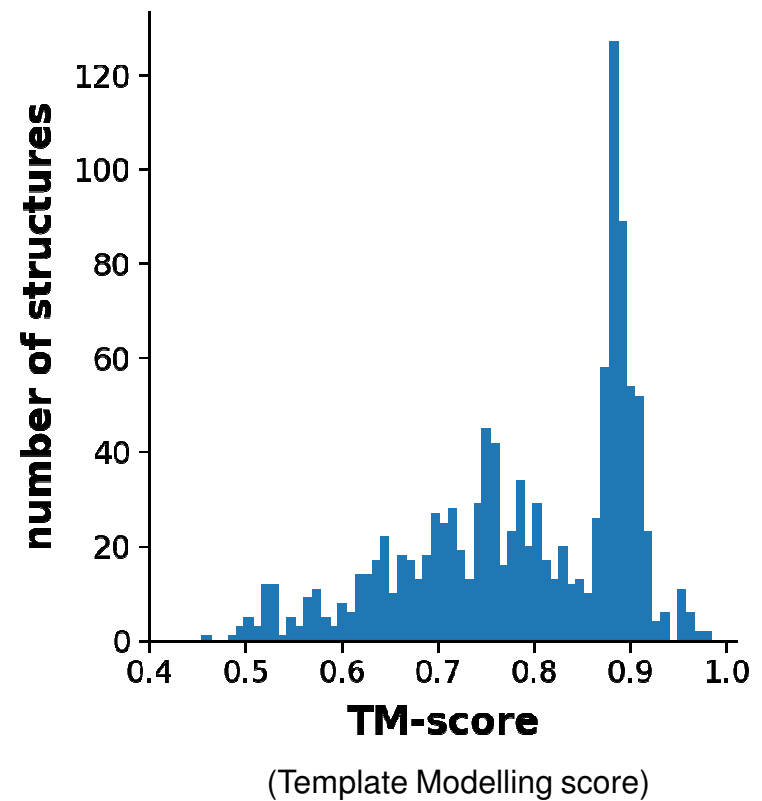
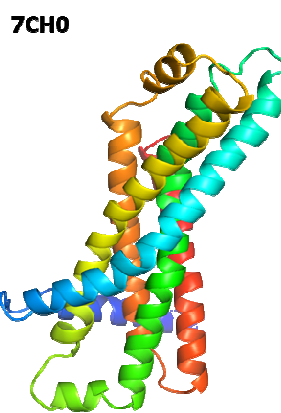
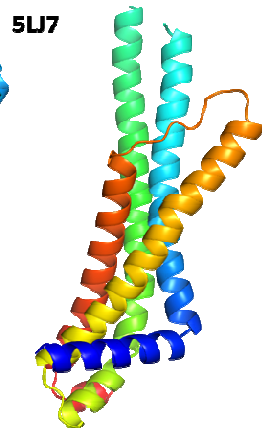
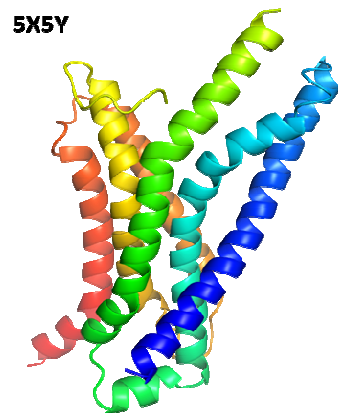
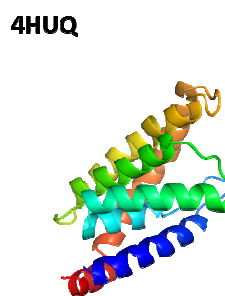
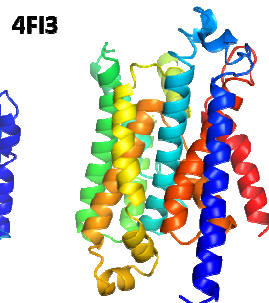
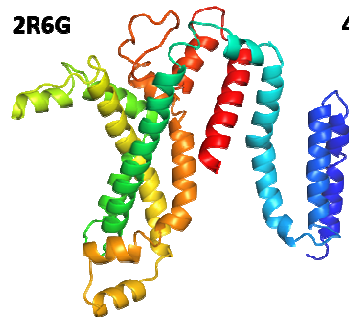
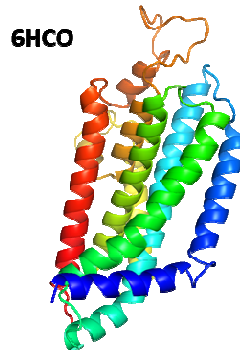
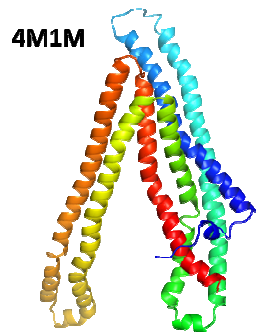


# TM protein structure prediction by AF2

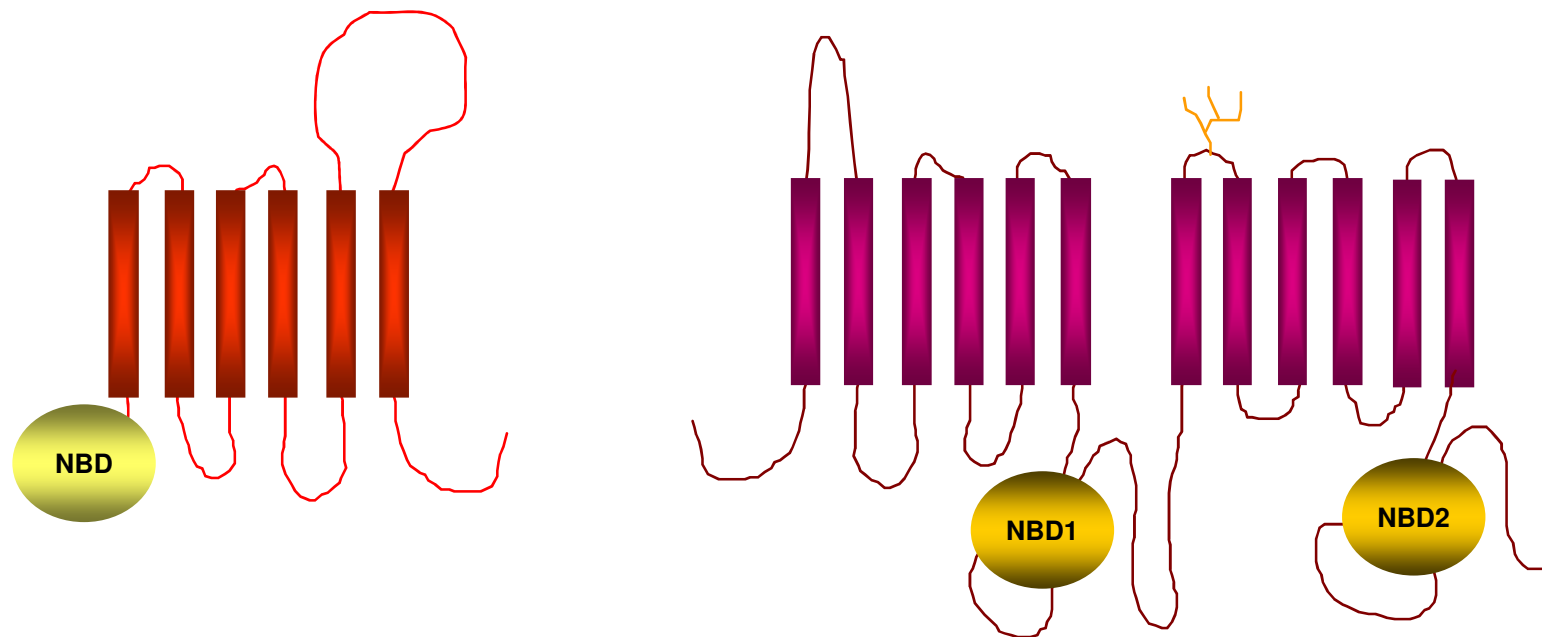


# ABC protein folds

fold class	reference PDB
Pgp-like	4M1M
ABCG2-like	6HCO
MalFG-like	2R6G
BtuC-like	4FI3
EcT-like	4HUQ
LptFG-like	5X5Y
MacB-like	5LJ7
MlaE-like	7CH0



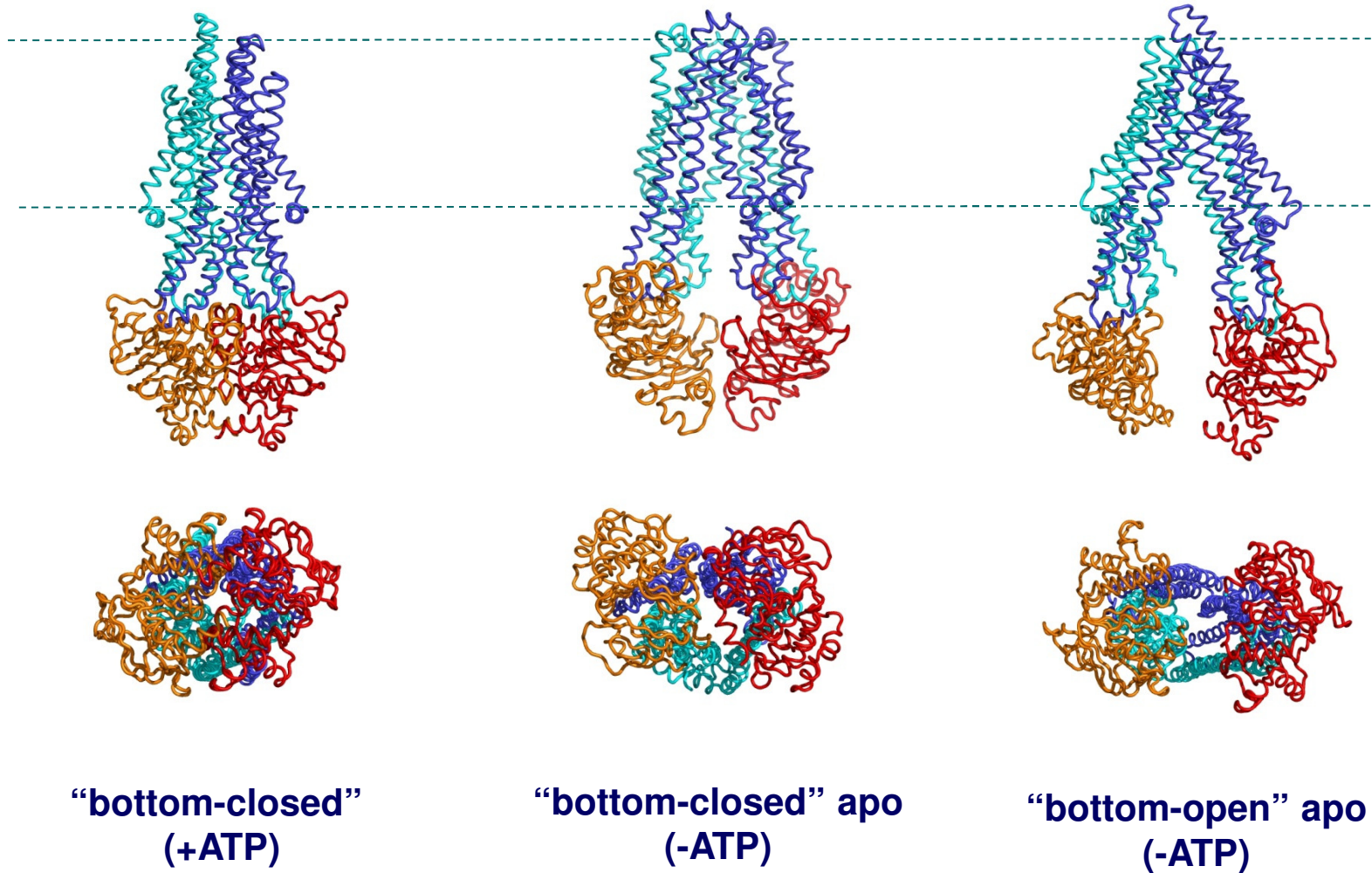
# ATP Binding Cassette (ABC) proteins





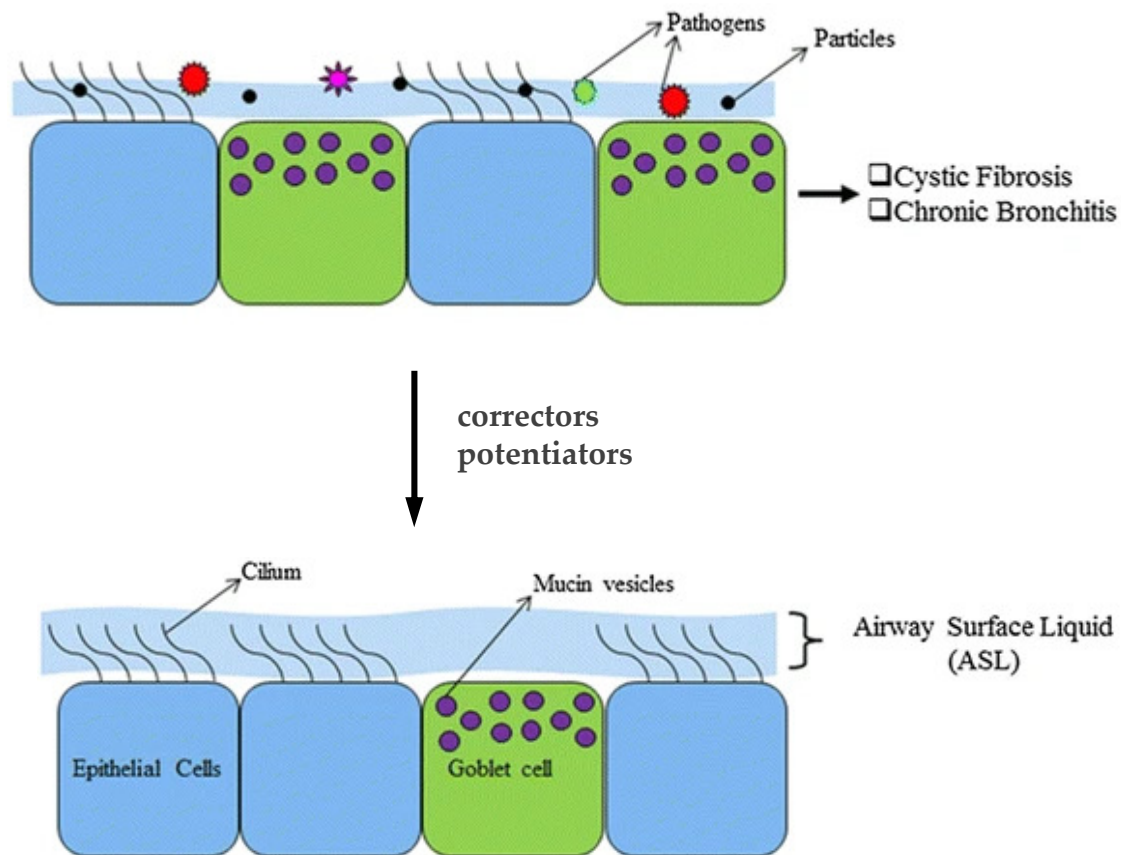
# Conformation of ABC proteins

(Pgp-like)



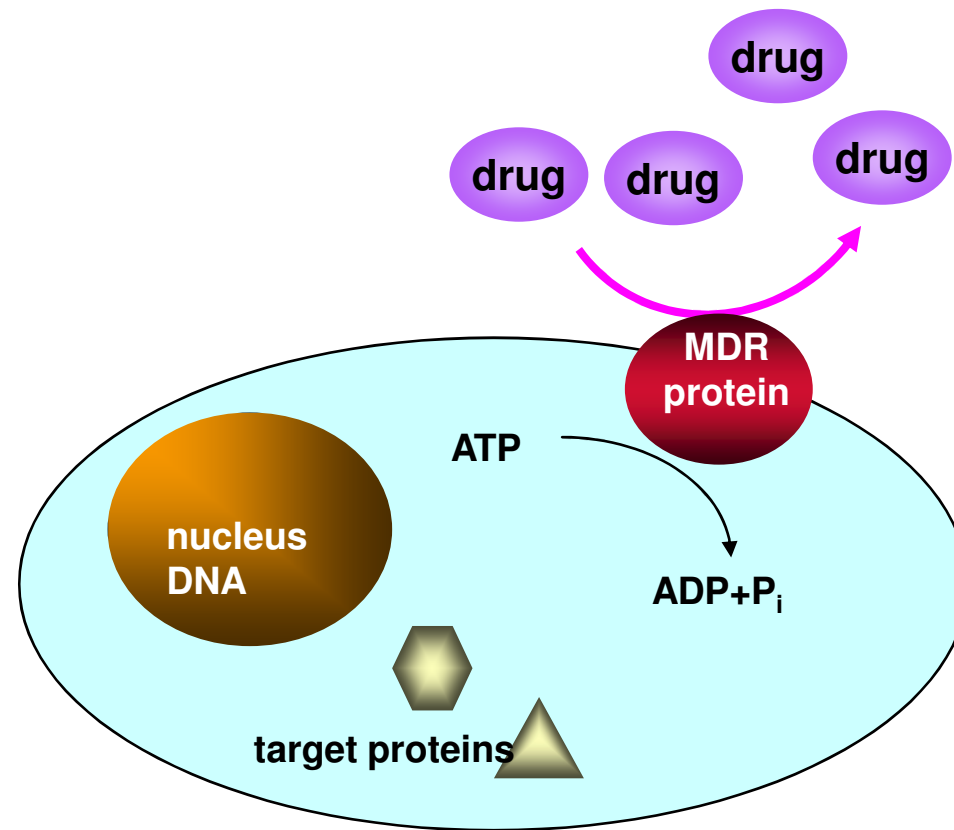


# Cystic fibrosis (CF)

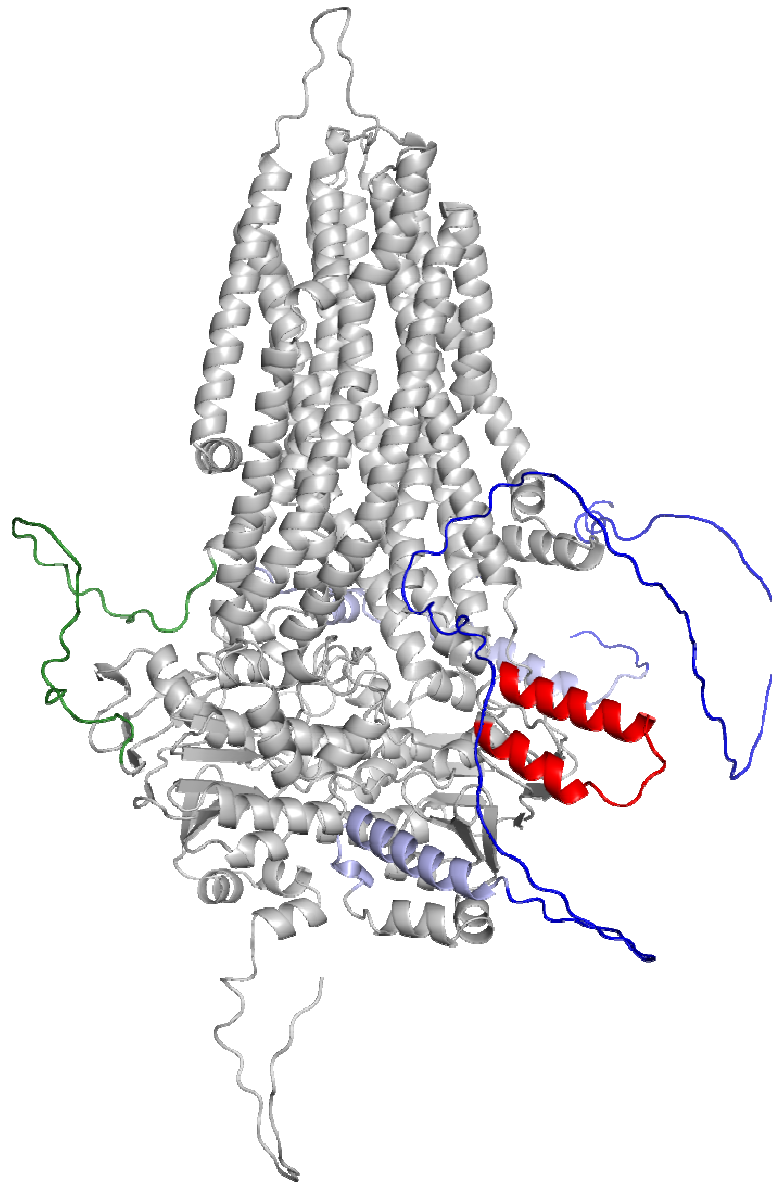


Ghosh, Boucher, Tarran,  
CMLS 2015

# Multidrug transport – ABC proteins

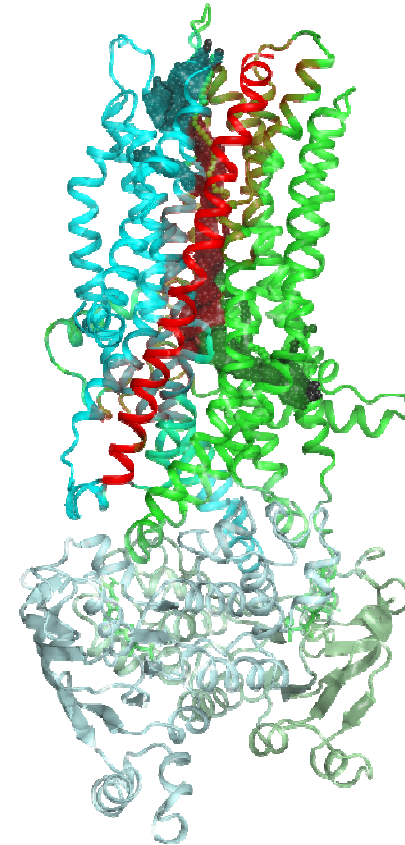
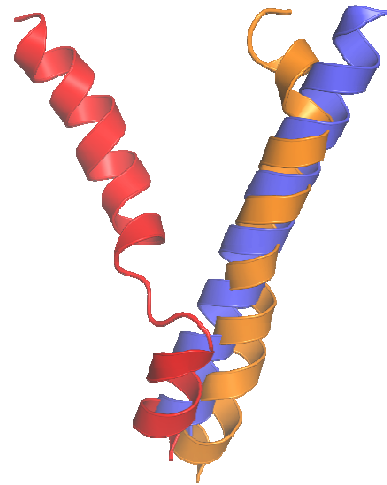
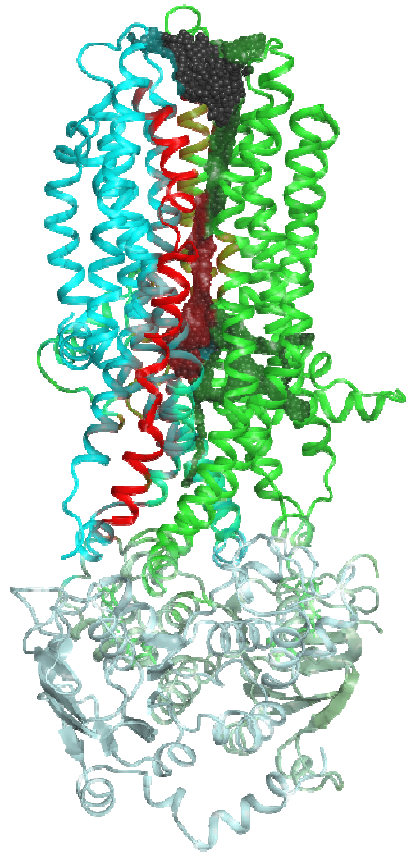


# CFTR – disordered regions

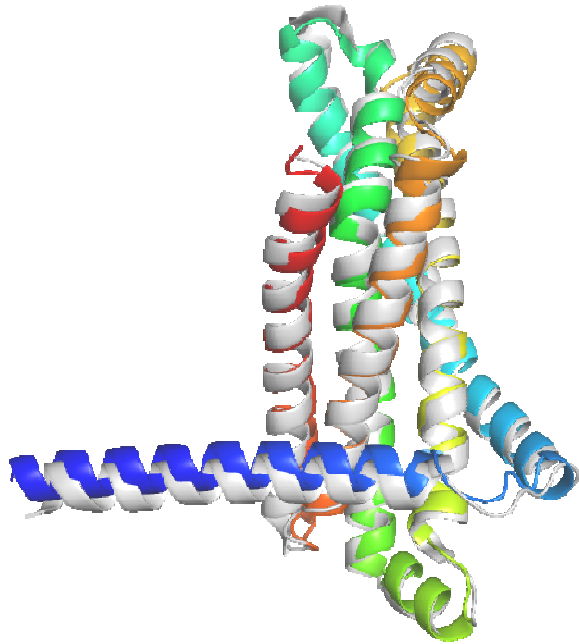


pLDDT score - IDR prediction

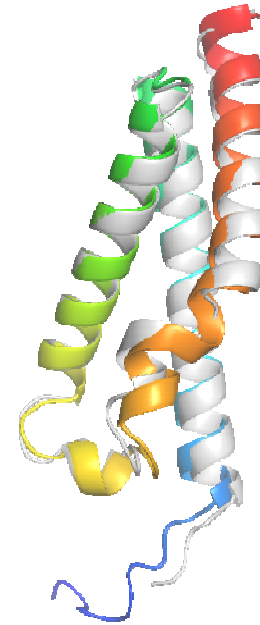
# CFTR TM8



# Prediction of new TM folds



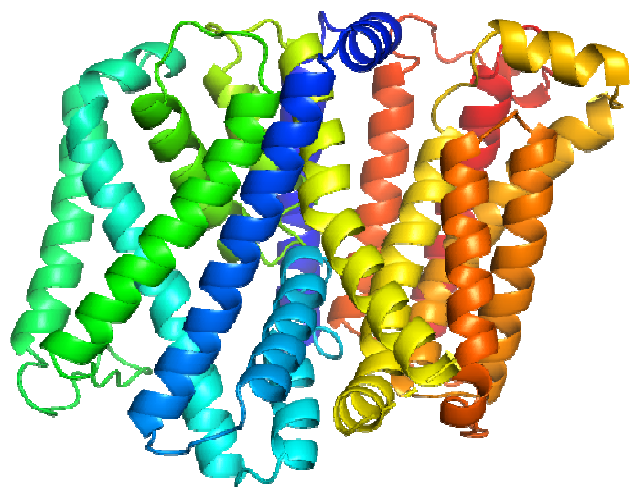
MlaE-like fold  
PDBID: 7ch0  
RMSD of 1.28 Å



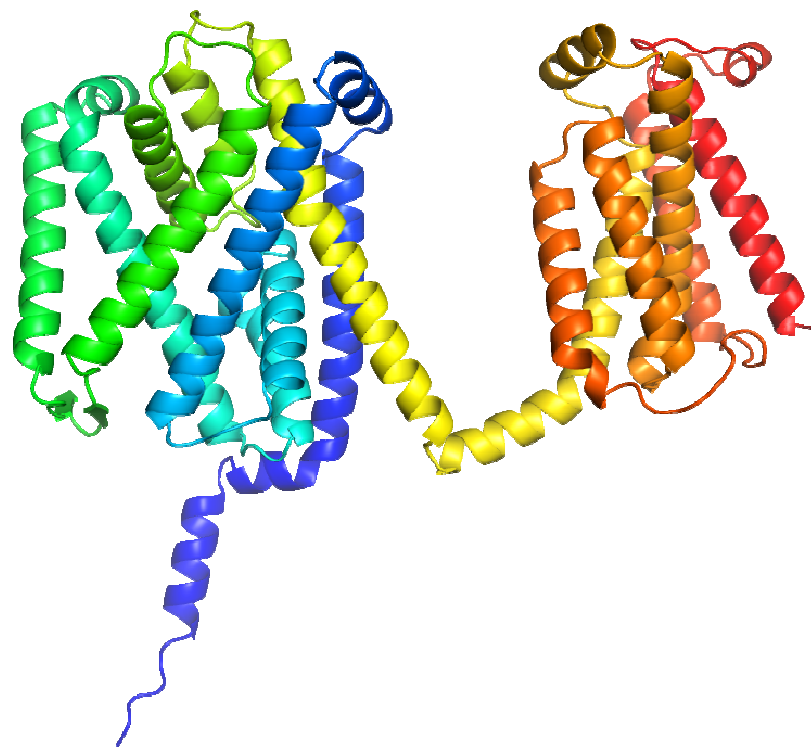
ER membrane protein complex subunit 6  
PDBID: 6ww7  
RMSD of 0.96 Å

# Prediction of new TM folds

MprF (PDBID: 7DUW)



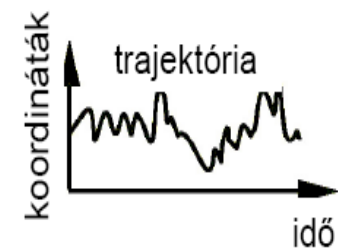
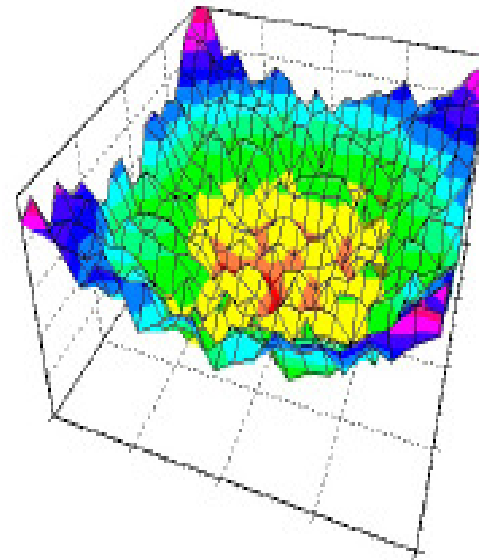
AF2



**Van szerkezet, mozgassuk meg!**

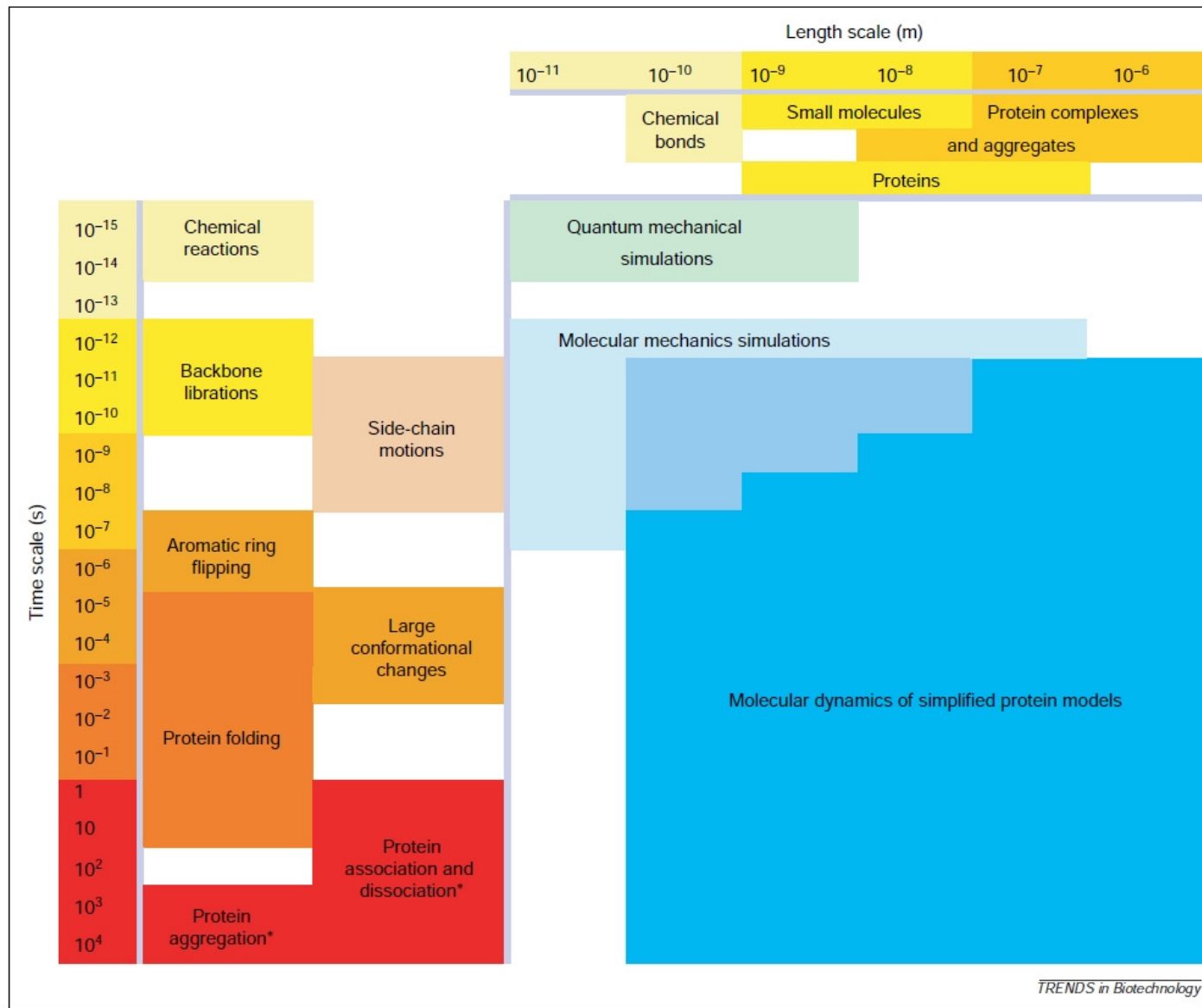
# Molekula dinamika/Molecular dynamics

- valós potenciálfelület
- mozgásegyenletek  
idő-lépésenkénti numerikus megoldása
- trajektória





# Események időskálája



F. Ding and N.V. Dokholyan, TRENDS in Biotechnology, 23:450 (2005)

# Egyszerűsített (Coarse Grain) modellek

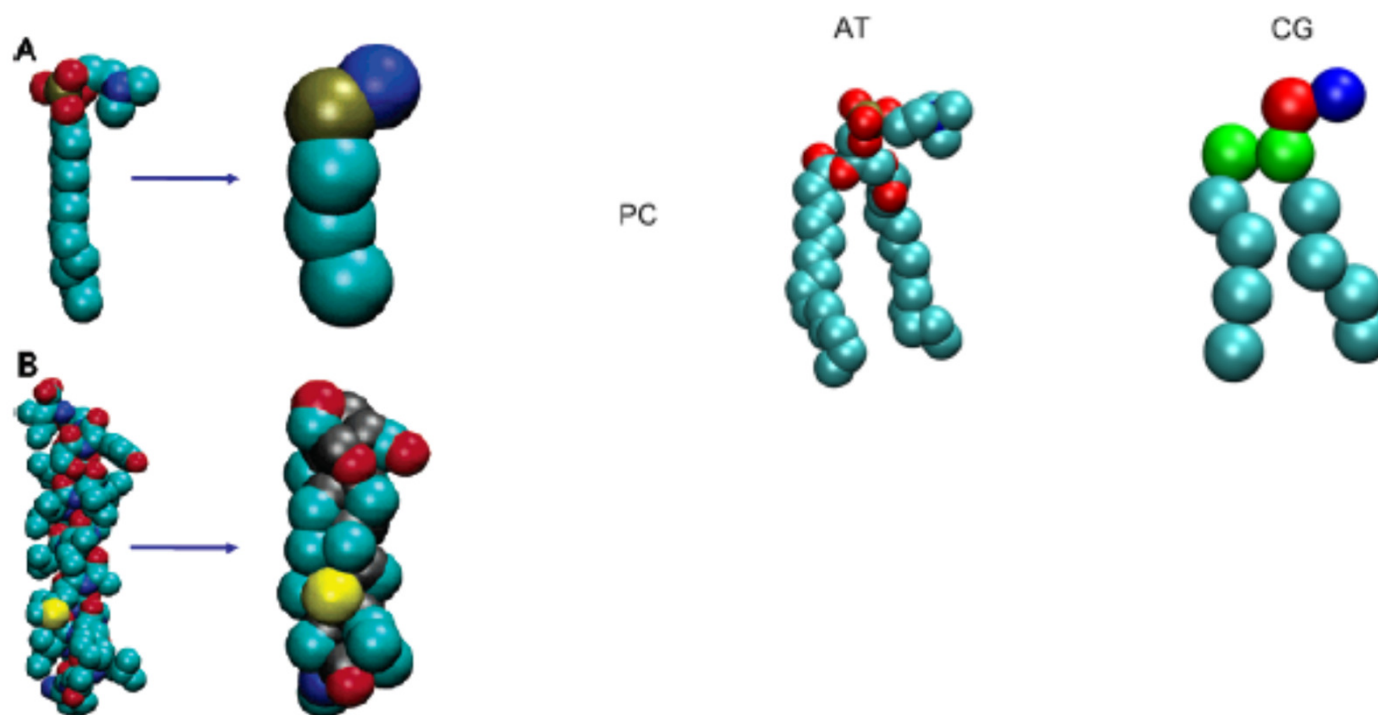
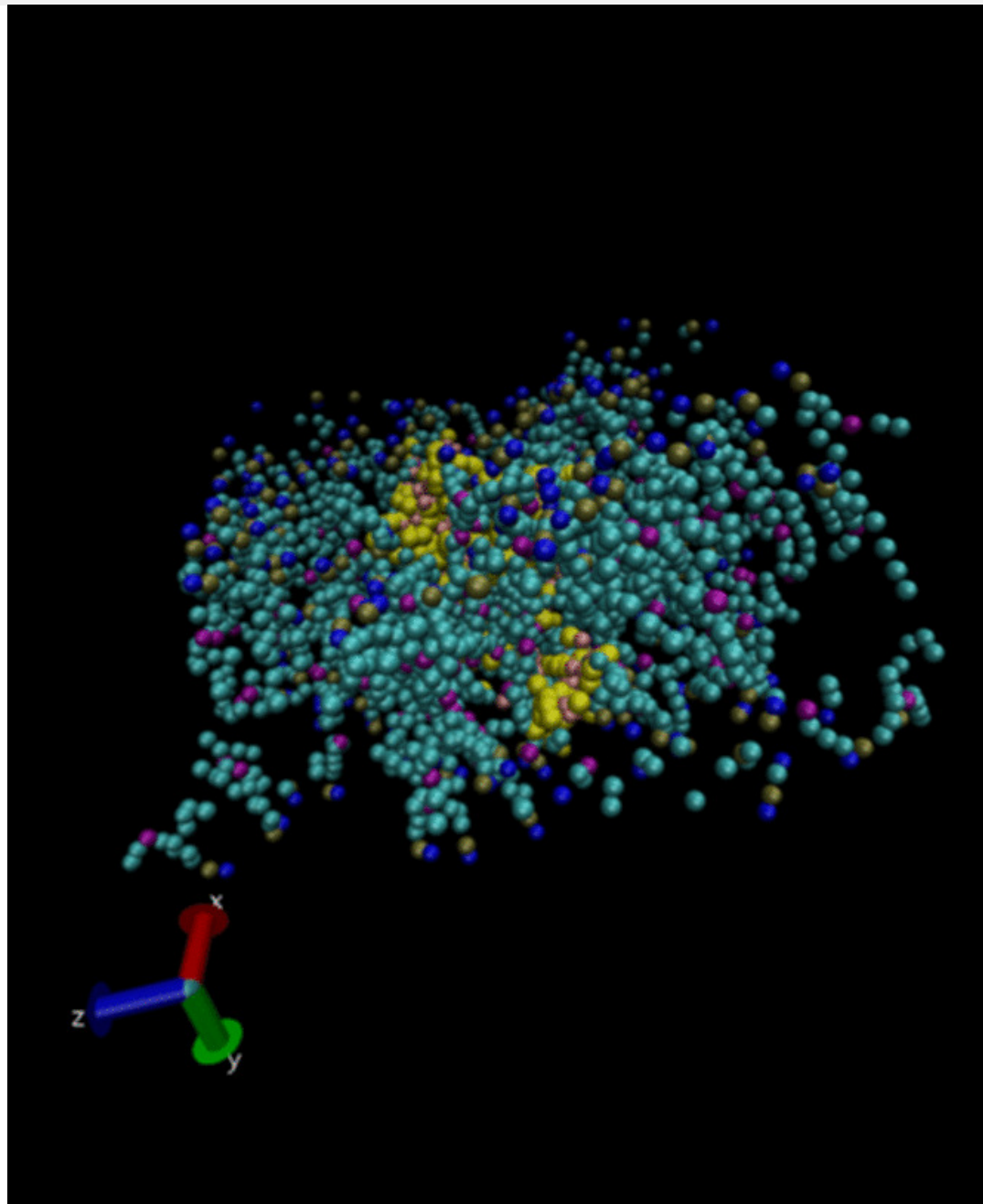


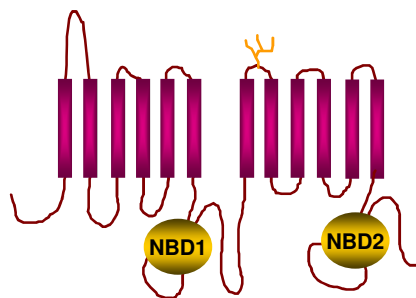
Figure 1. Atomistic (left-hand) and coarse-grained (right-hand) models compared for (A) a DPC molecule and (B) a GpA helix. Colors for atoms:

Sansom, Oxford University  
Marrink, Groningen, MARTINI force field

# Kettősréteg felépülése a fehérje köré

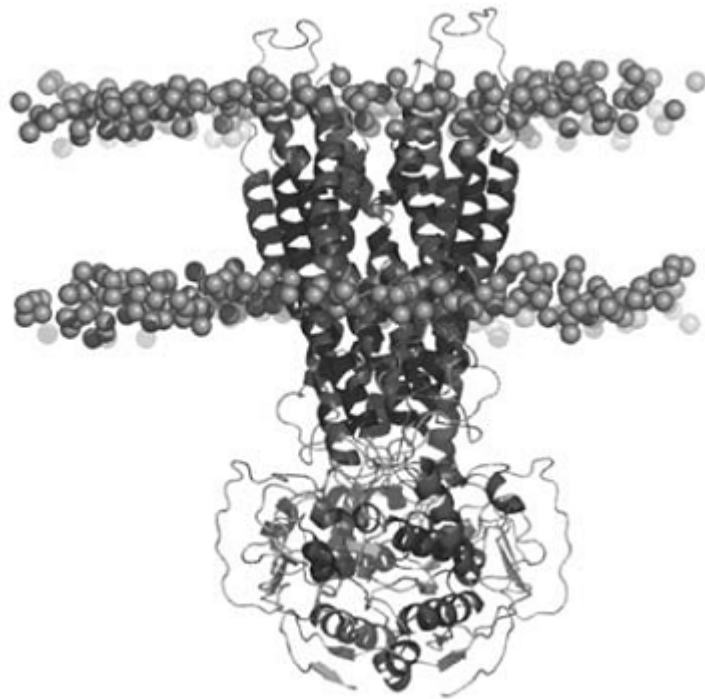


# Fehérjék konformációinak stabilitása



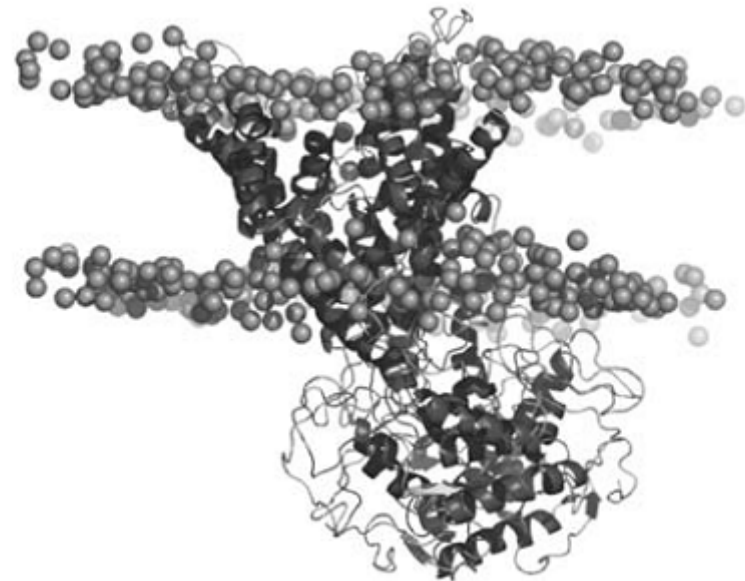
Eur Biophys J (2008) 37:403–409

**B**



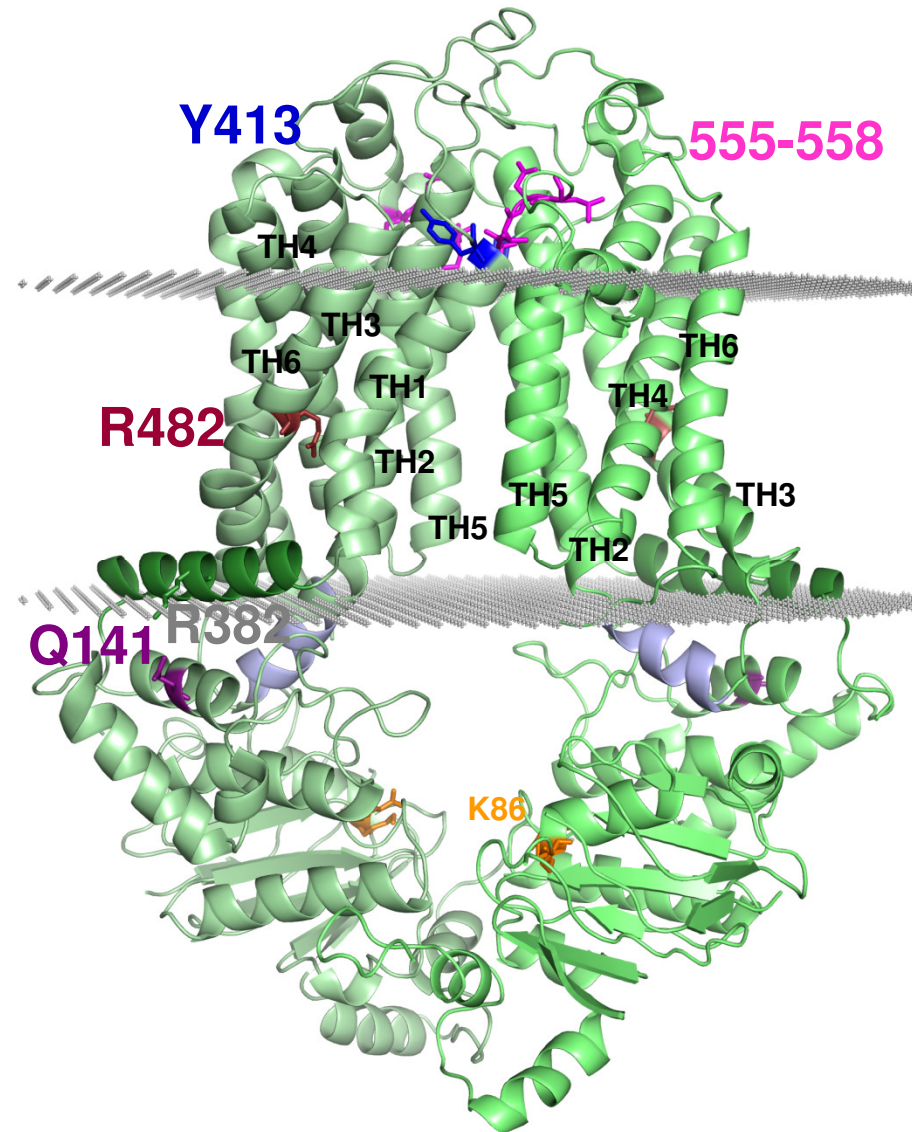
**0 ns**

**C**



**20 ns**

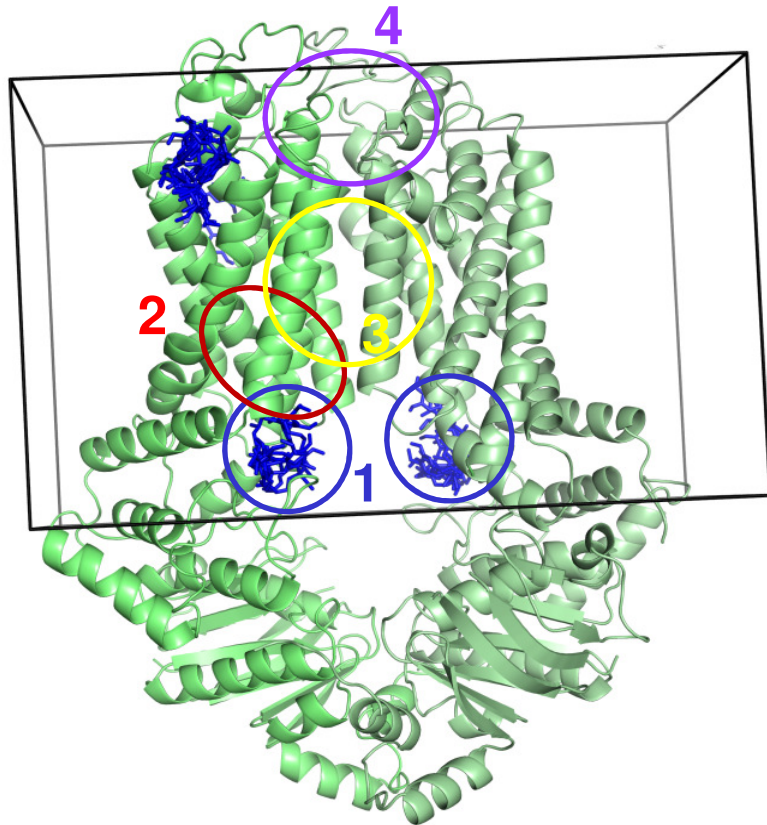
# Az ABCG2 modell



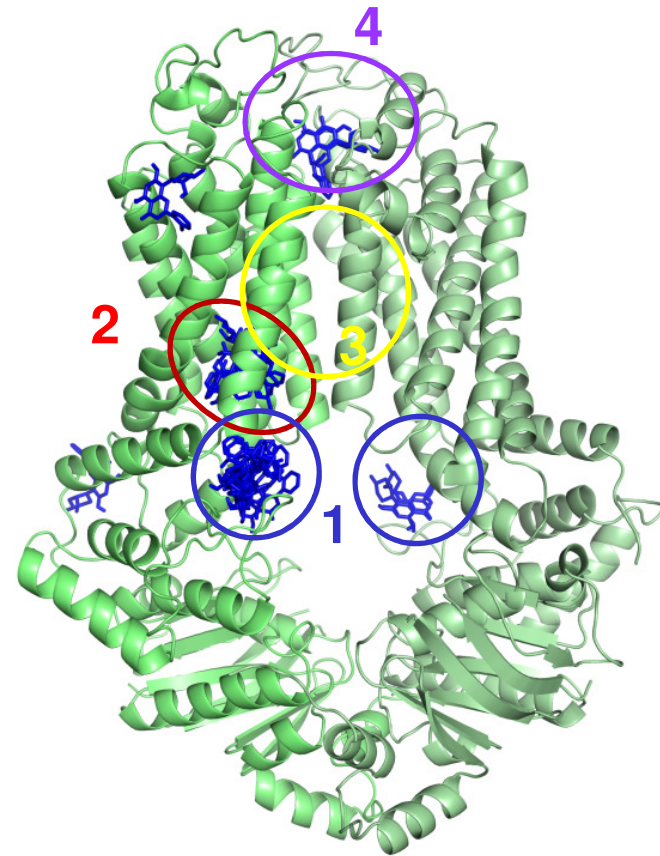


# Drogkötő helyek vizsgálata *in silico* dokkolással

**verapamil**

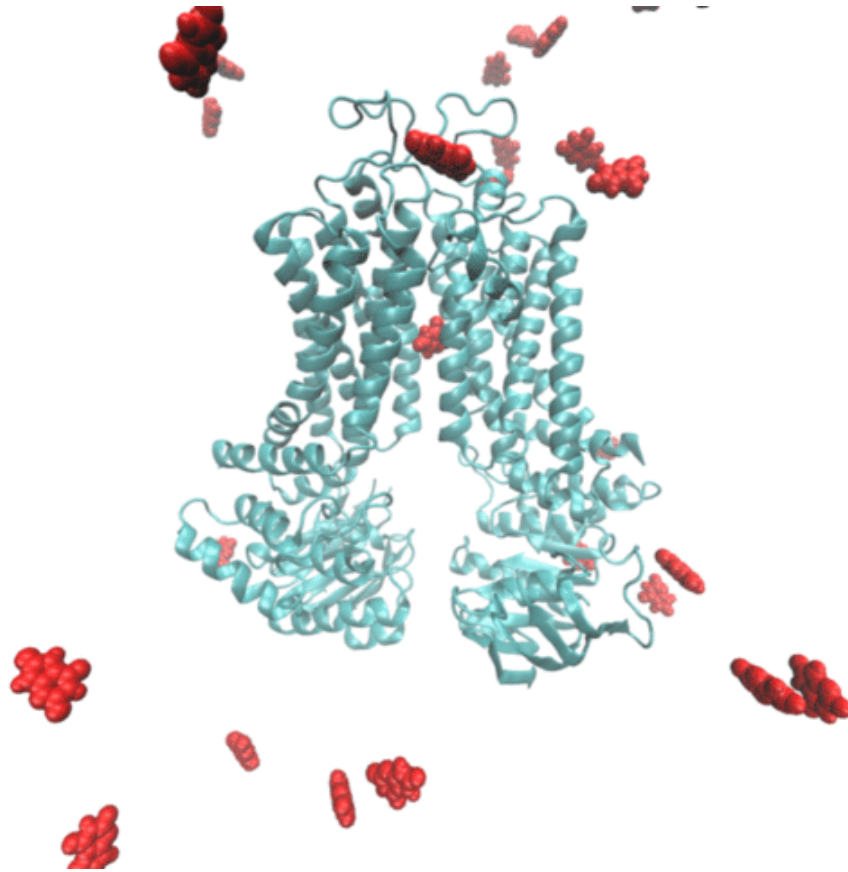


**flavopiridol**



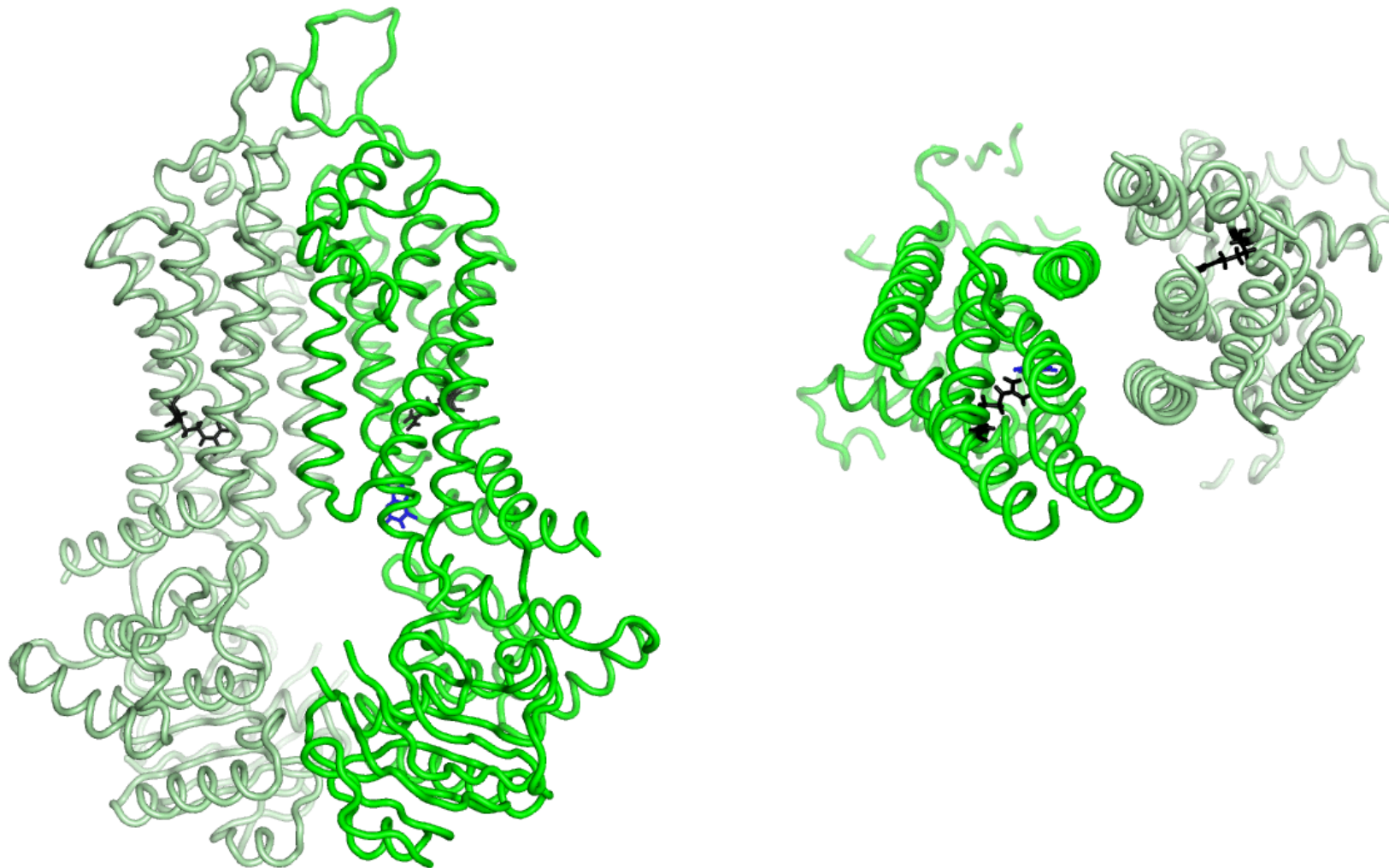
# Describing the transport using MD

equilibrium simulations, uric acid molecules



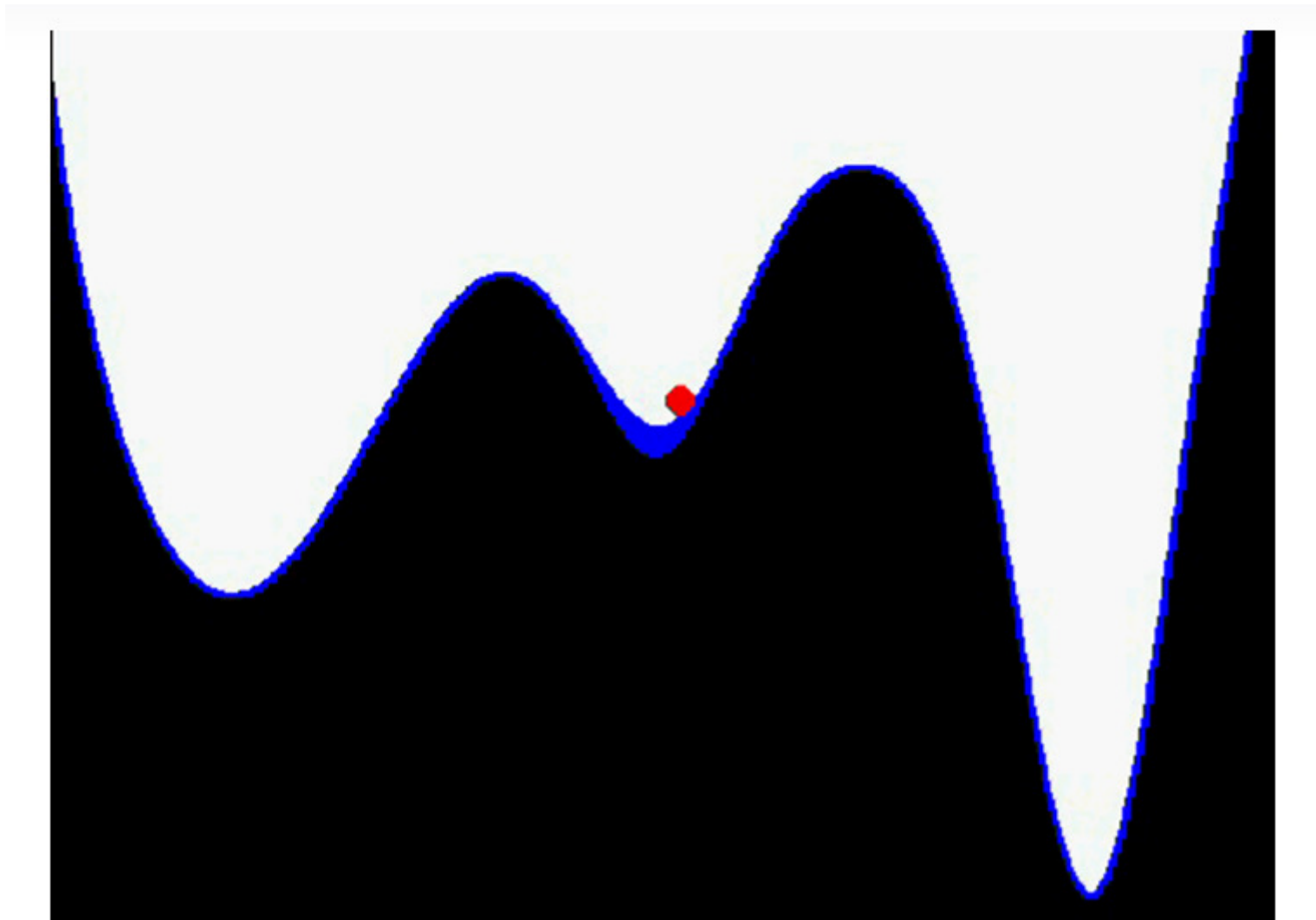
# Describing the transport using MD

equilibrium simulations, uric acid molecules

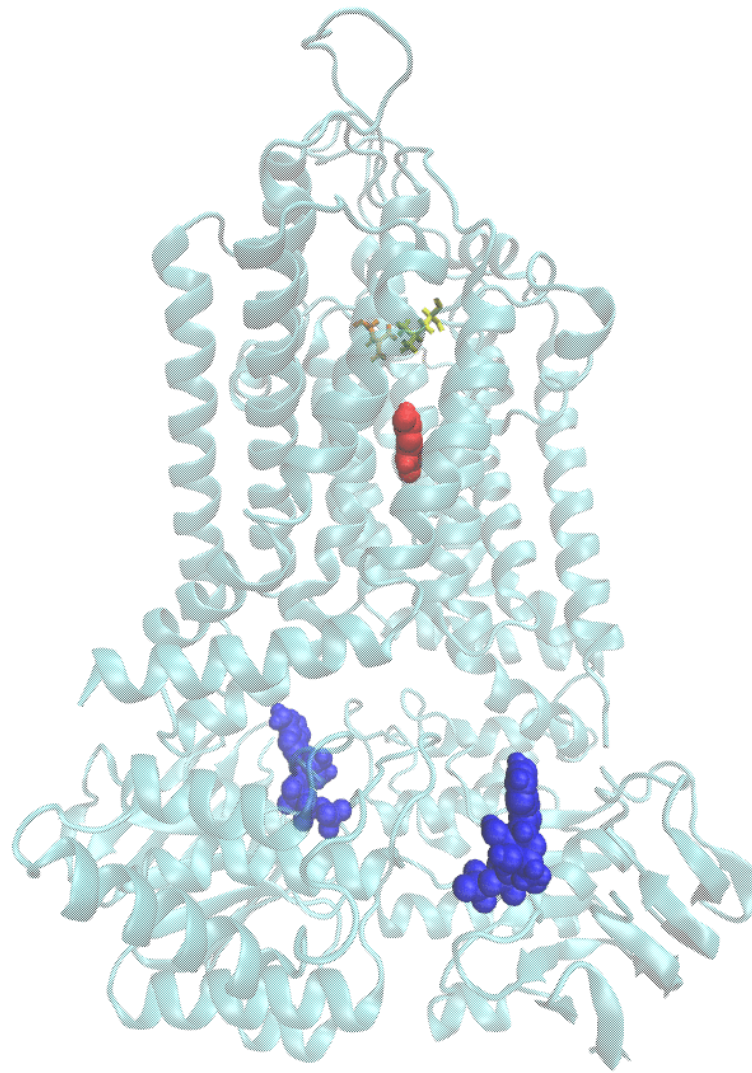




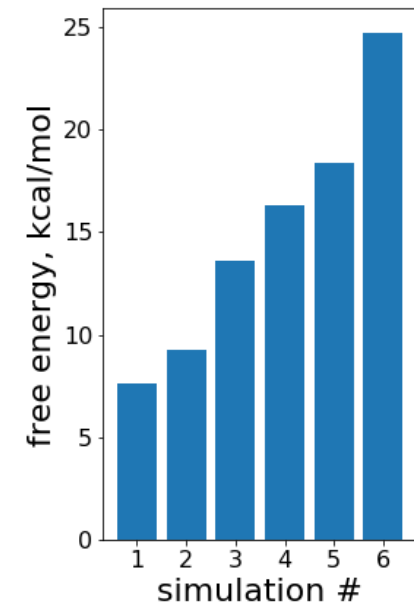
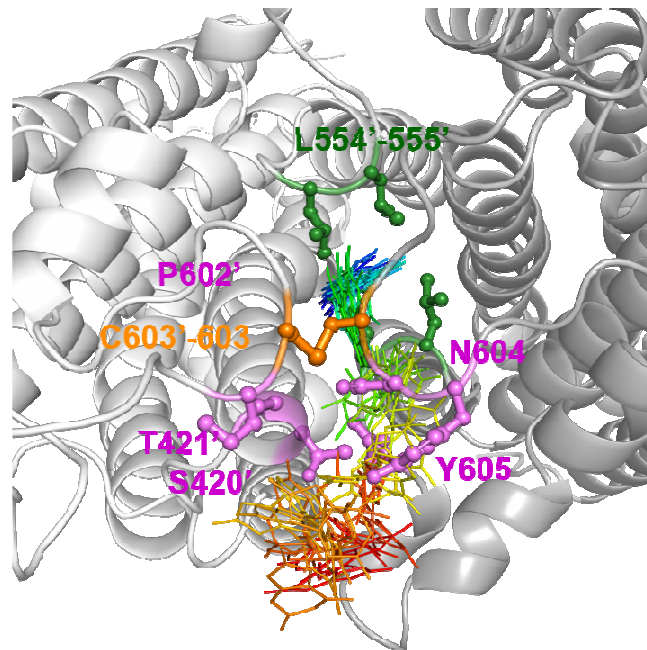
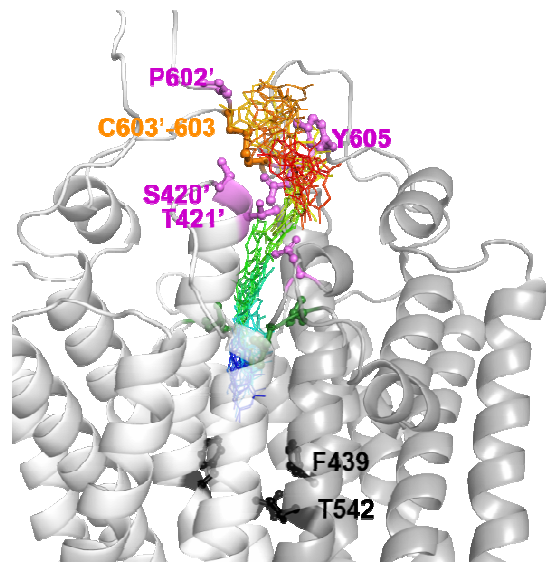
## A „exit” jellemzése metadinamikával



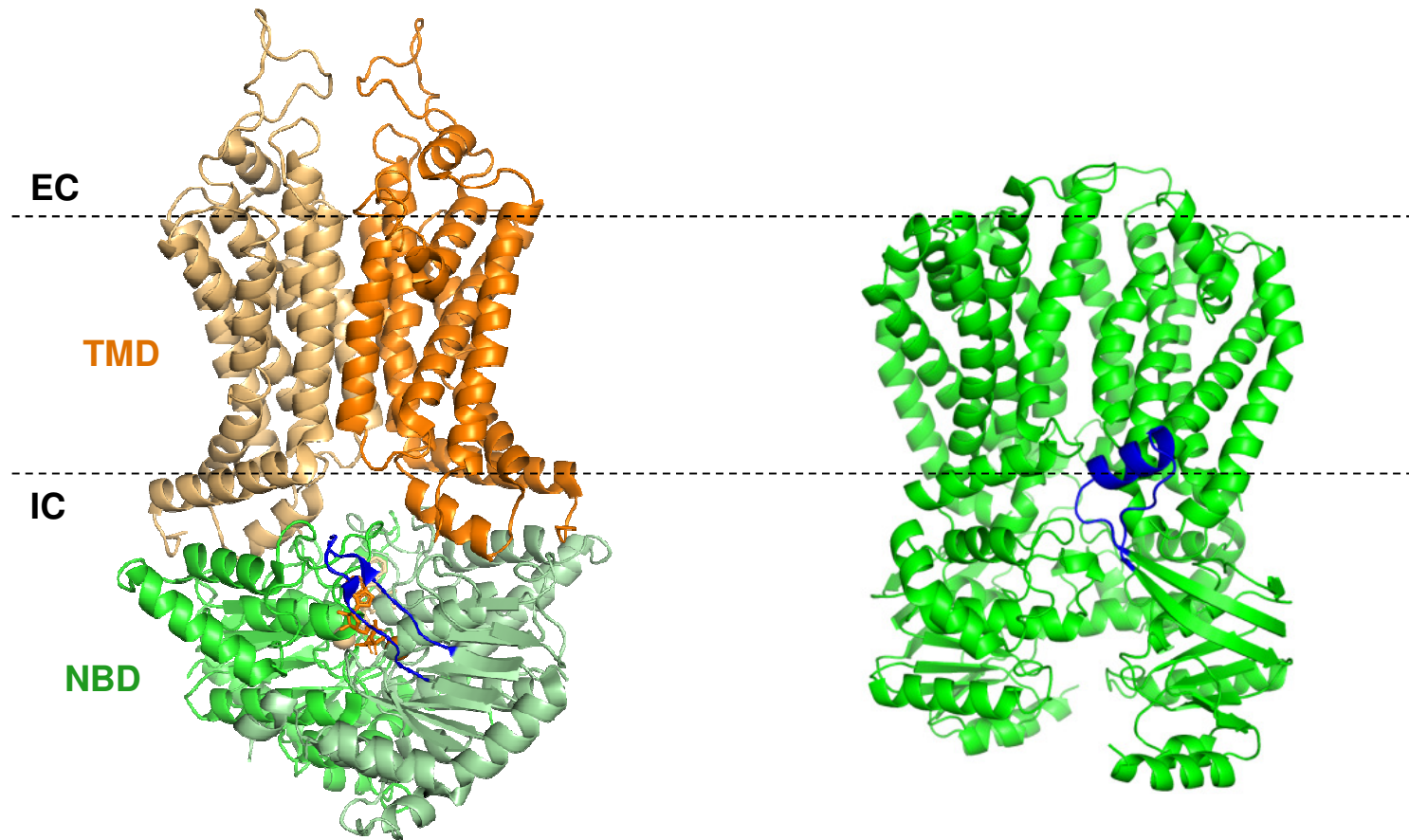
## A „exit” jellemzése metadinamikával



# A „exit” jellemzése metadinamikával



# Membránkölcsönható fehérjerégiók

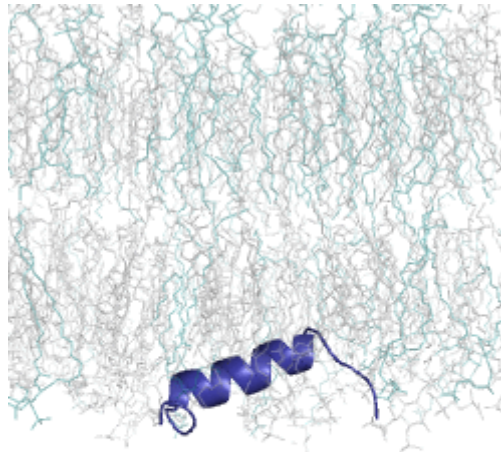


ABCG2  
6hbu

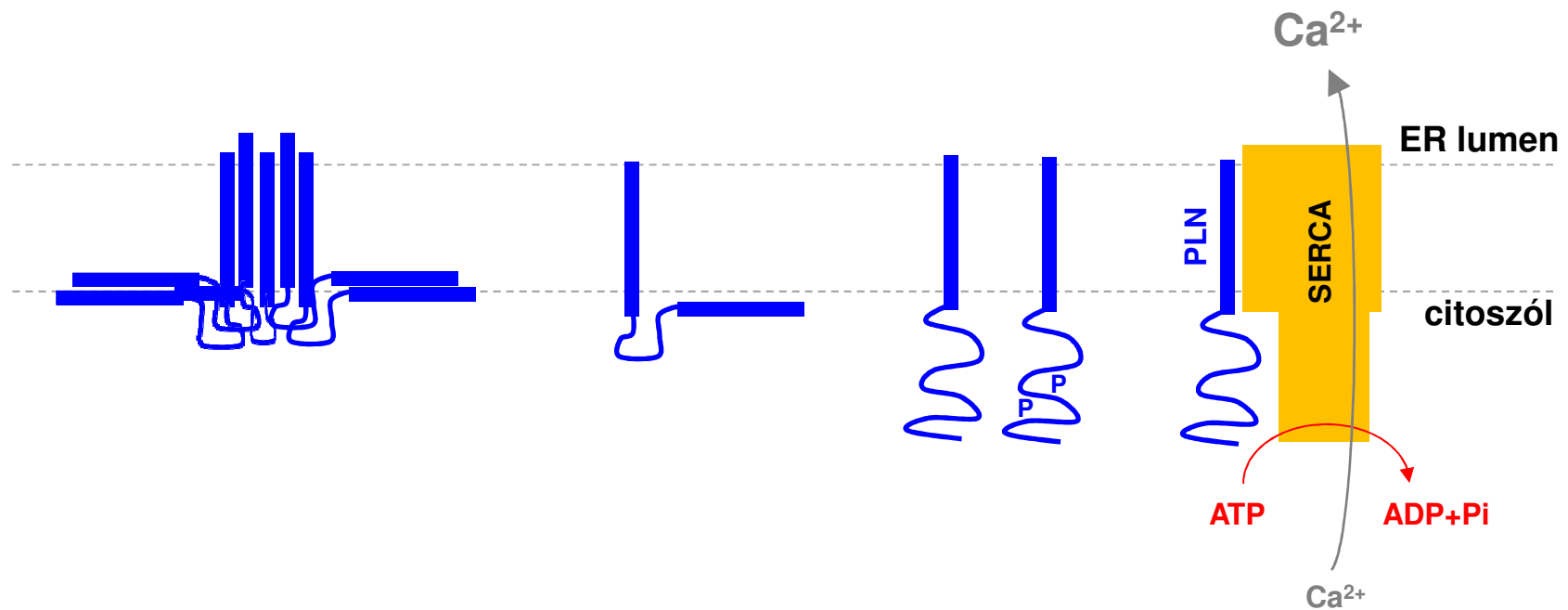
O-antigen polysaccharide ABC-transporter  
6oih

# MemMoRFs

Membrane Molecular Recognition Features, <https://memmorf.hegelab.org>



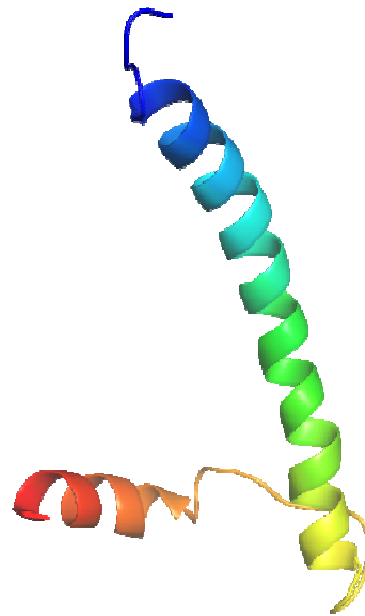
# Kalcium ATP-ázok (SERCA, PMCA)



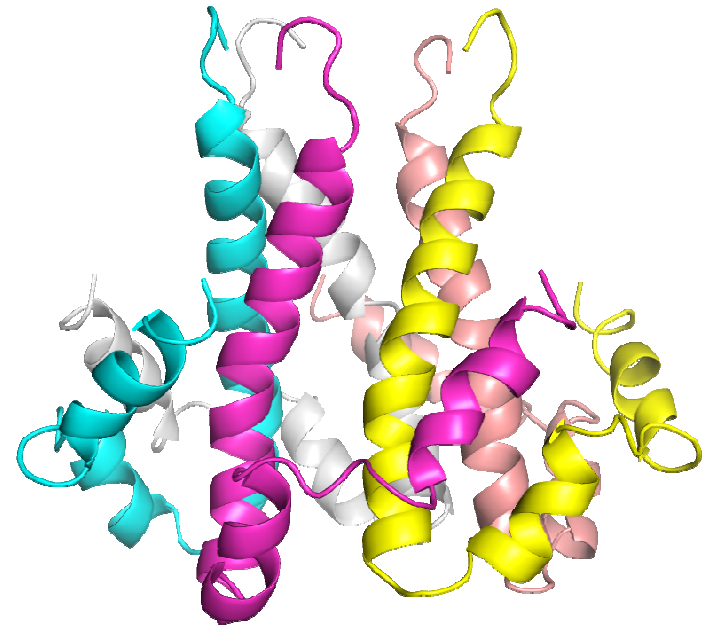
# Regulinok és exoregulinok

SLN  
ELN  
ALN  
Dworf  
???

SARS Cov-2  
Envelope (E) protein



- ERGIC
- viroporin
- PDZ-kötés
- egyéb PPI?



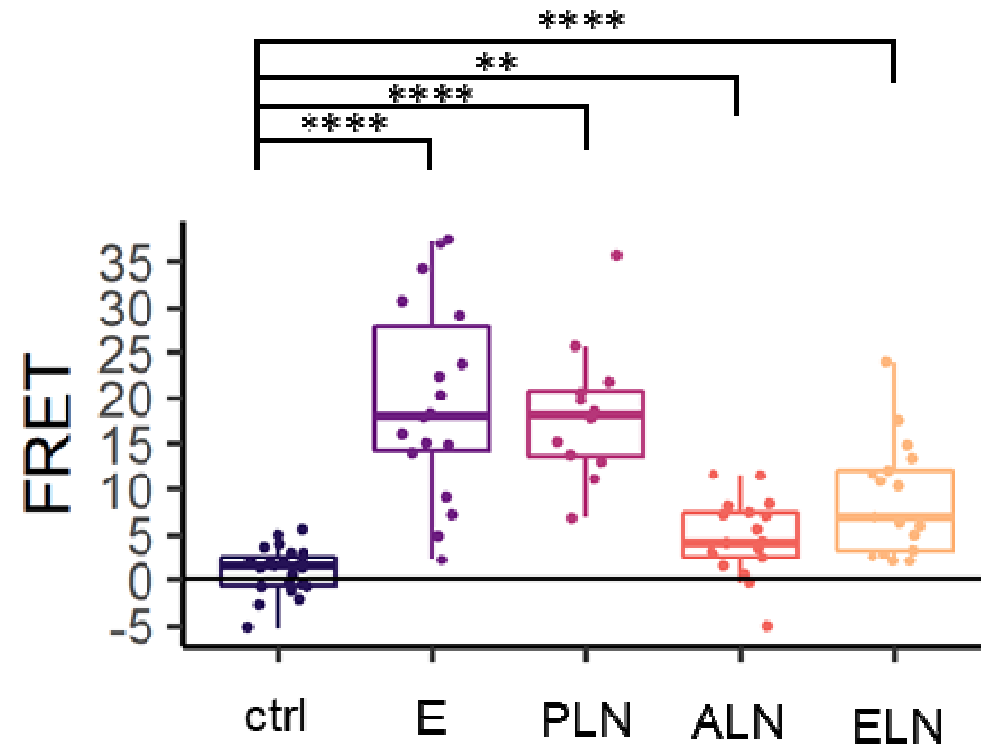
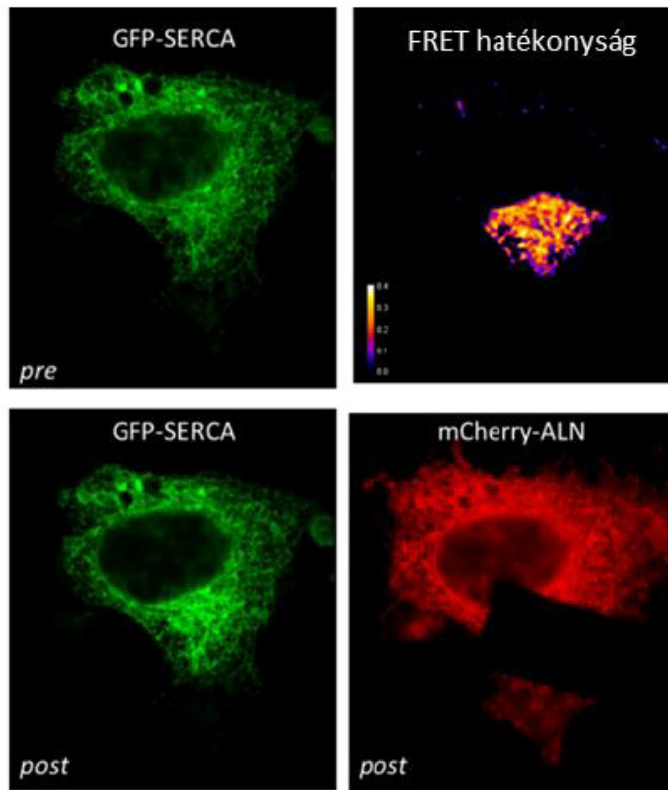


# Célkitűzéseink

- Fizikai kölcsönhatás bemutatása
- Funkcionális relevancia demonstrálása
- Atomi szintű modell építése

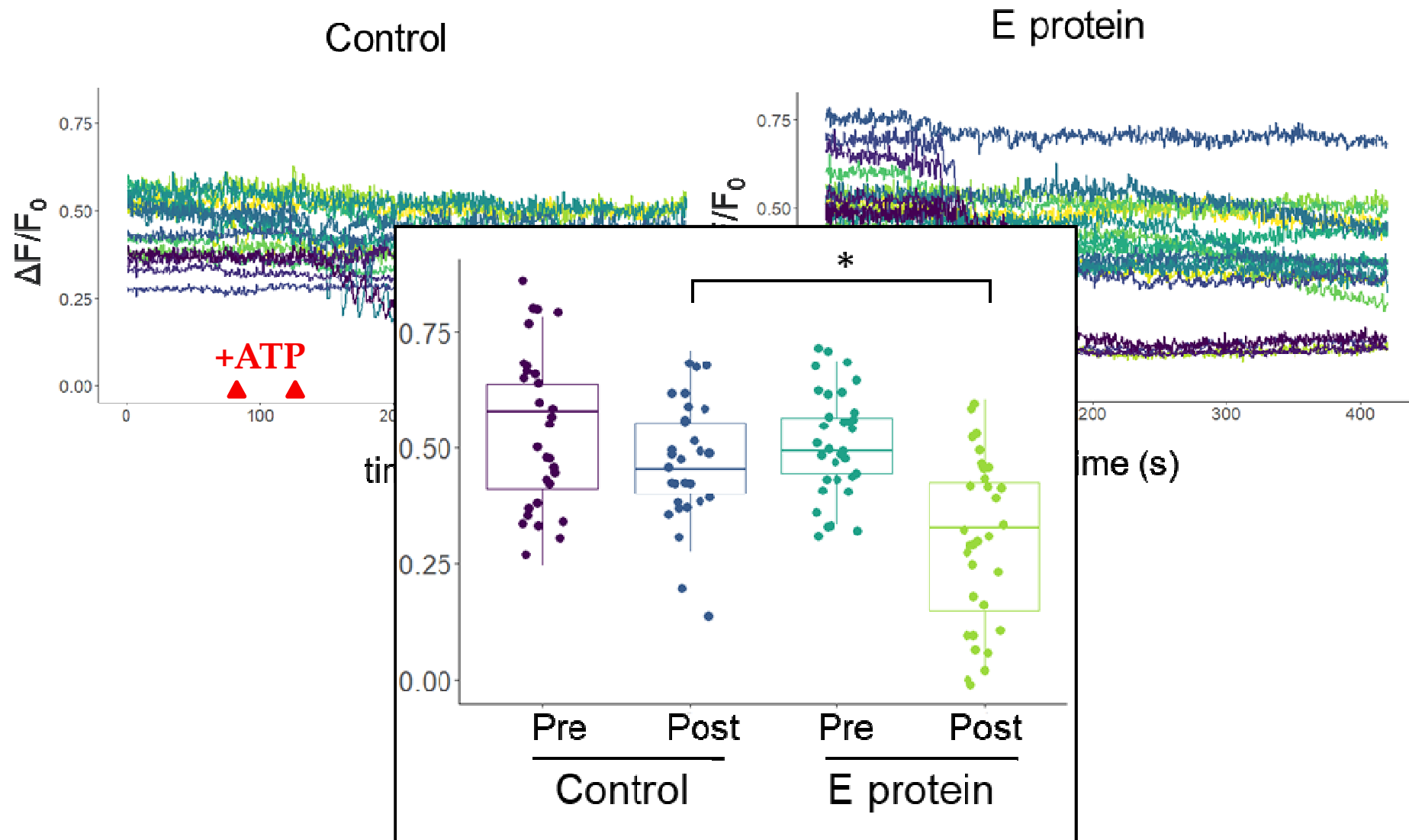
# Az E fehérje kölcsönhat a SERCA pumpával

- eGFP és mCherry jelölt fehérjék
- koexpresszió HeLa sejtekben
- Lokalizáció átfedés
- AP-FRET



# Az E fehérje gátolja a SERCA működését

ER-GCaMP6-150  $\text{Ca}^{2+}$  sensor



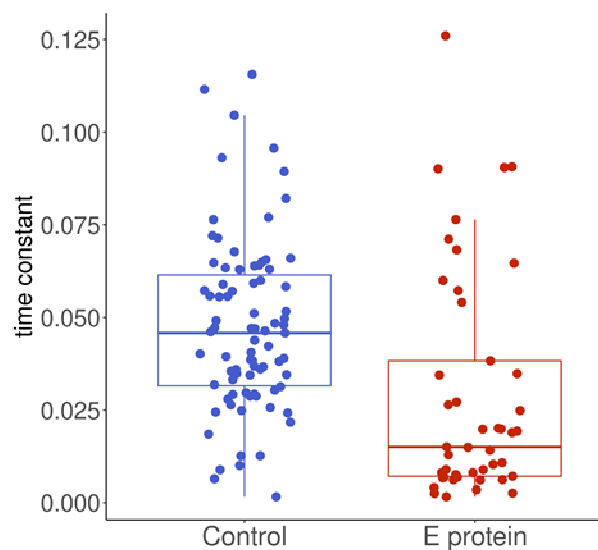
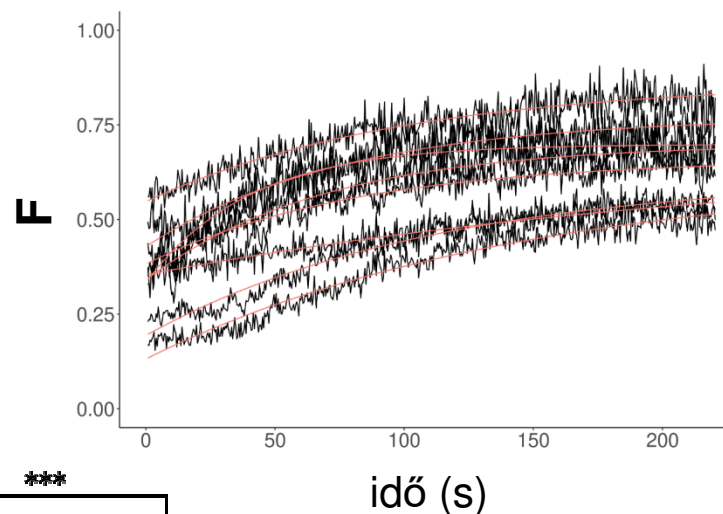
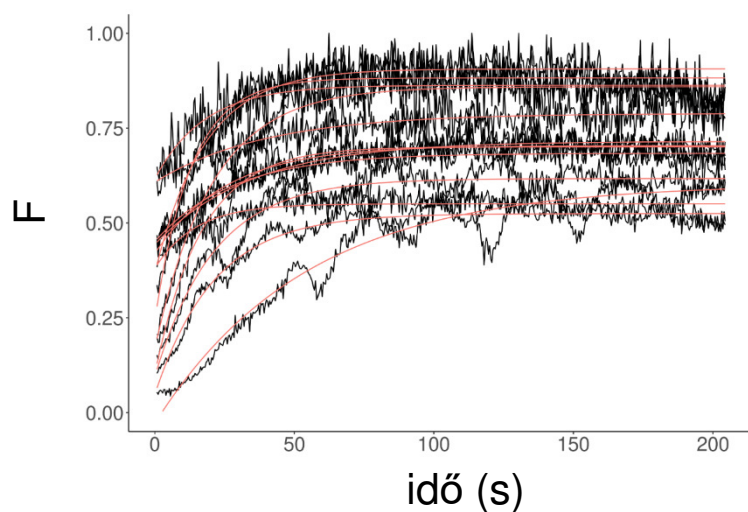
# Az E fehérje gátolja a SERCA működését

nincs EC  $\text{Ca}^{2+}$

ER ürítése

+  $\text{Ca}^{2+}$

+ E protein

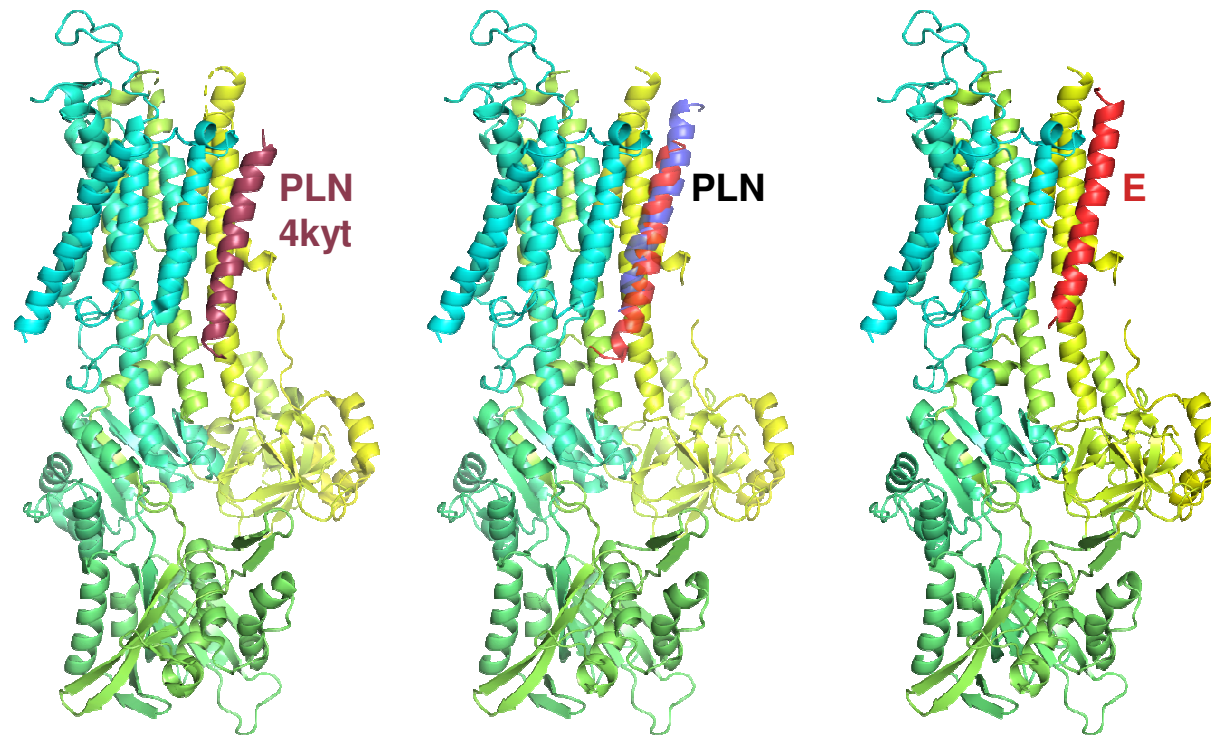


# Kísérletes SERCA/E fehérje eredmények

- Közvetlen fehérje-fehérje kölcsönhatás
- E fehérje gátolja a SERCA pumpát
- Rendszerünkben az E fehérje nem viroporin

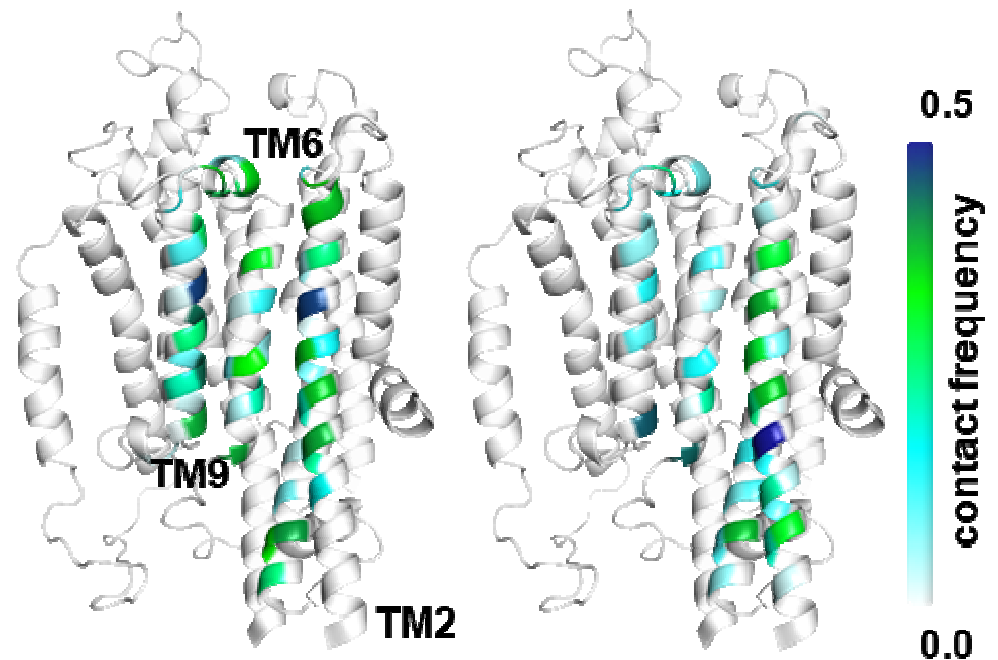
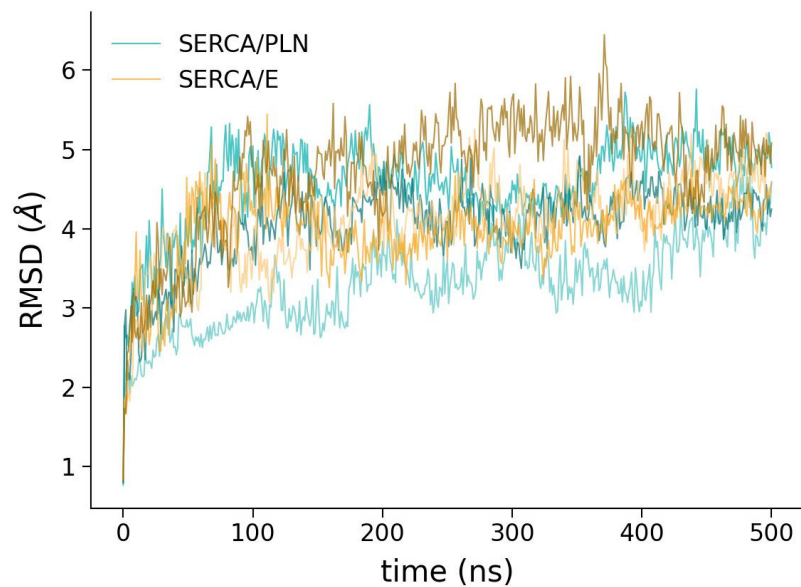
# A SERCA / E protein komplex szerkezete

AlphaFold-Multimer  
PIPER/ClusPro



# SERCA/regulin komplexek stabilitása

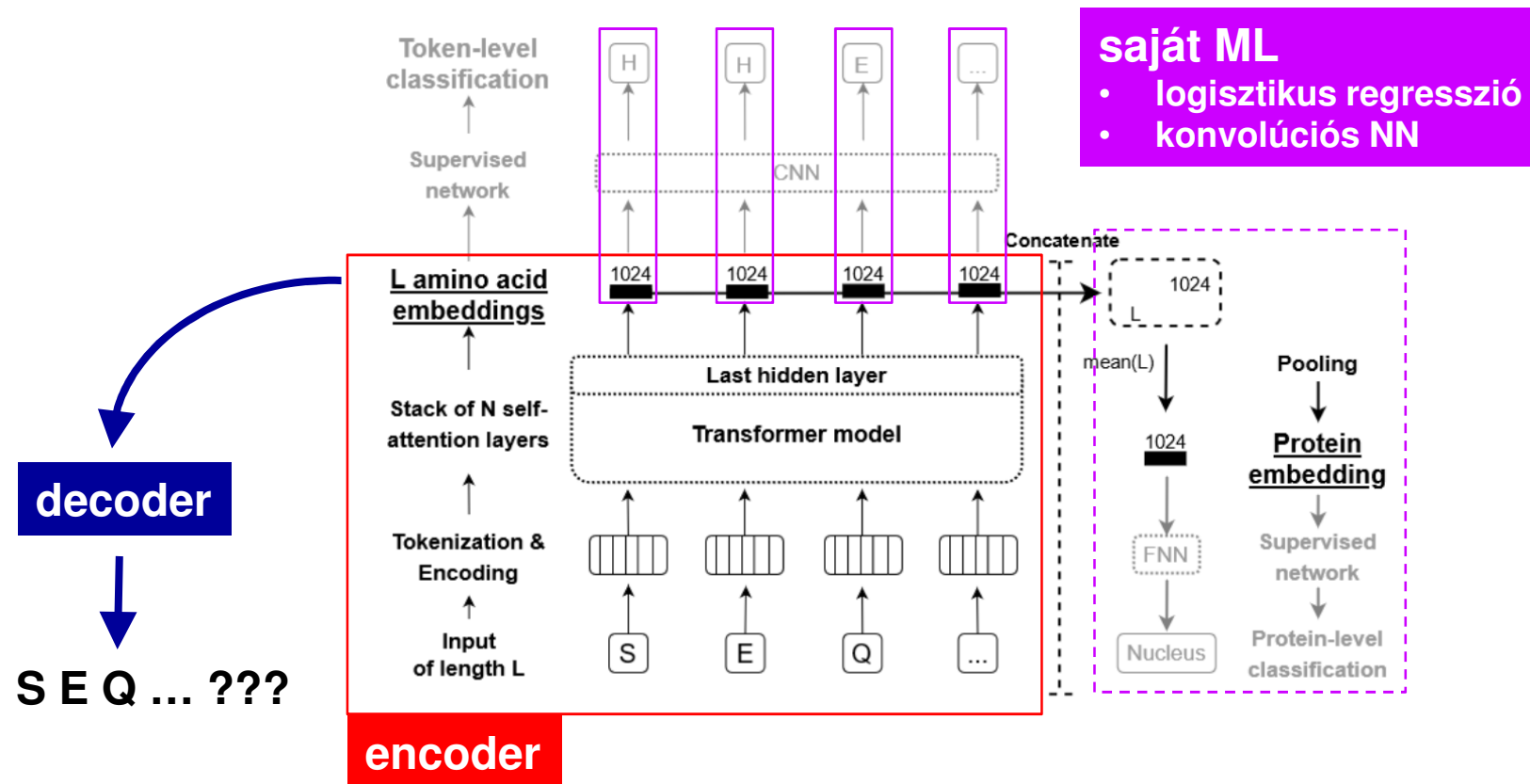
## MD szimulációk





# MemMoRF prediktor

## protein Language Models



# MemMoRF prediktor

pLM + CNN

	Training	Testing
Accuracy	0.936	0.881
AUC	0.979	0.928
MCC	0.873	0.767

# Összefoglalás

- Fehérje dinamika
  1. Szerkezet
  2. Szolvatálás
  3. Energiaminimalizálás
  4. Egyensúlyba hozás
  5. *Production run*
  
- Számításos módszerek igen fontosak
  
- Fehérjék működésének és mutációk hatásának megismerése
  
- Gyógyszertervezés

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