Protein structure and dynamics

Tamás Hegedűs

tamas.hegedus@hegelab.org

Dept. of Biophysics and Radiation Biology

Thanks for your attention!

hegedus.tamas@hegelab.org

- I did not include all slides, which contained important unpublished data
- I did include some slides, which we did not discussed
- If you would like to read more about any of the topics, write me to ask references

Importance of protein dynamics

The atomic level basis of a disease…
The shape of a drug binding site…

There is no single structure
but a conformational ensemble at 37°C

Importance of computational modelling

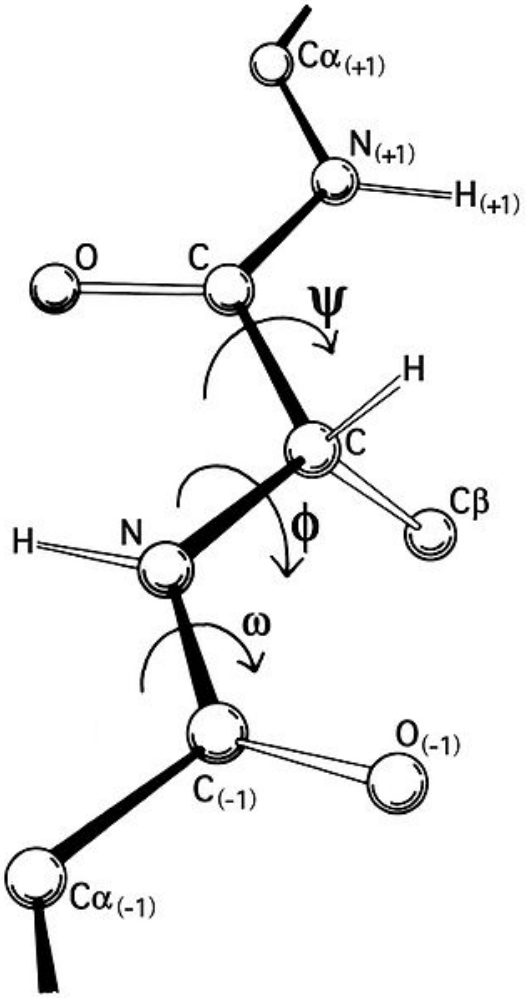
Atomic level information on motions

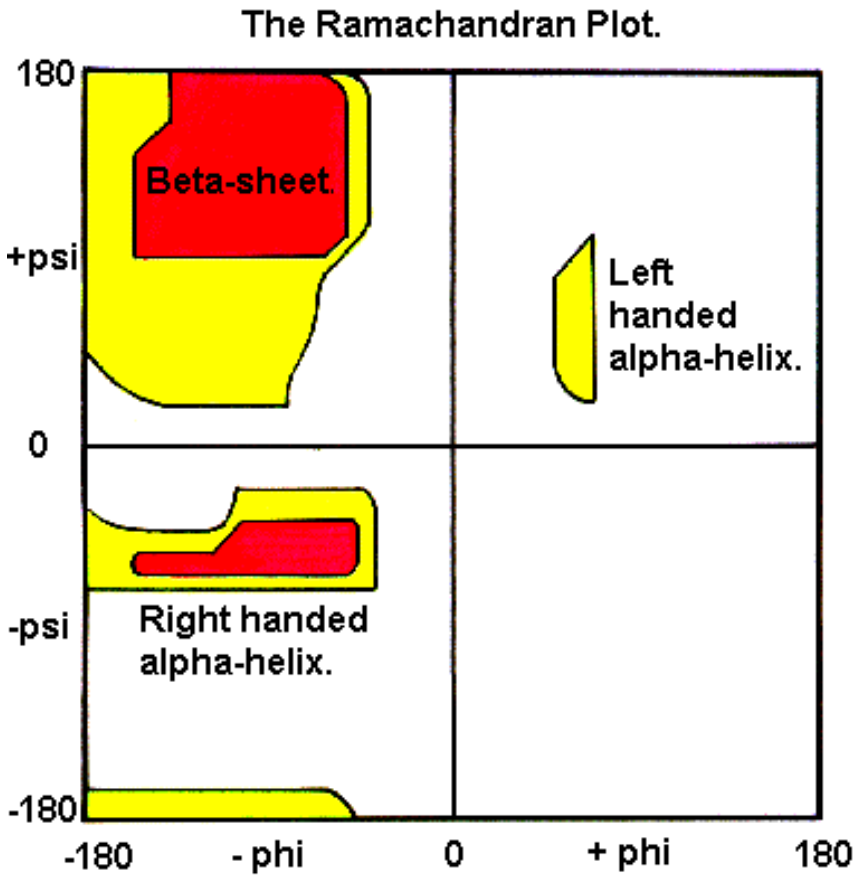
Experiments usually do not provide atomic level information (there are exceptions, such as NMR)

Topics

- **Introduction to protein structure**
- **Characterization of protein structure**
  - **Prediction of secondary structure**
  - **Intrinsically disordered proteins**
  - **Tertiary structure**

Secondary structure





wikipedia

Prediction of secondary structure

Using only available structures **60 %**

Combining with sequence aligments **70-80 %**

Implementations:

- **neural networks,**
- **support vector machines,**
- **hidden Markov models, etc.**

Scoring each positions
GOR4, HNN, Prof, JPred/JNet

Intrinsically Disordered Proteins

- **25 % of proteins are predicted disordered**
- **Increased disorder with increased complexity**

  50 % of human proteins contains a disordered region with 30 a.a. or longer

- **Not fully random**
- **Structural flexibility**
- **No compact globular folding and no residual structure**

The paradigm
protein function needs a well-defined 3D structure
has changed.

Intrinsically Disordered Proteins

Benefits

- Specificity and adaptation
- Reversible transition between ordered/disordered states
- Large binding surface
- Fast binding

Roles

- Entropic chain: inactivation of K^+ channels
- Effectors: peptide inhibitors
- Scavengers: casein
- Assembly: calmodesmon, F-actin
- Presentation: phosphorylation and cleavage sites

Intrinsically Disordered Proteins

K. Dunker, Péter Tompa, Zsuzsa Dosztányi, M. Fuxreiter

DisProt database: http://www.disprot.org

Disorder prediction

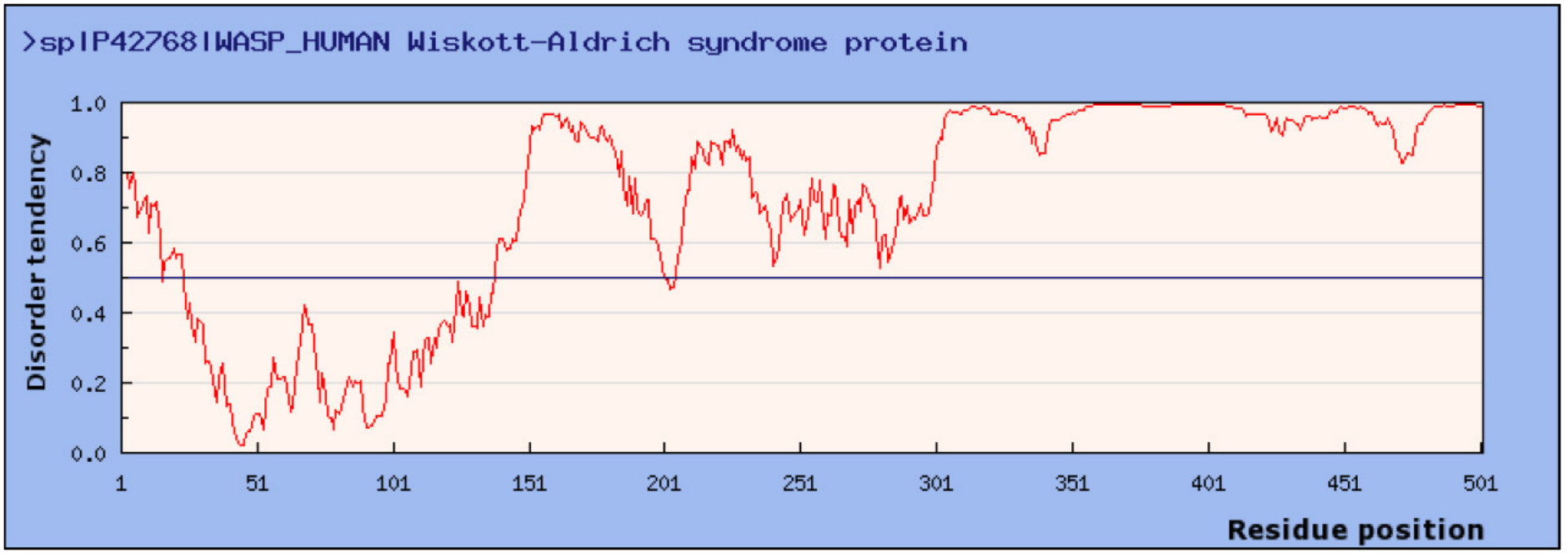
Learning algorithms
Based on disordered sequences
in the PDB database

Disopred2
AlphaFold2

Predicting interaction energies

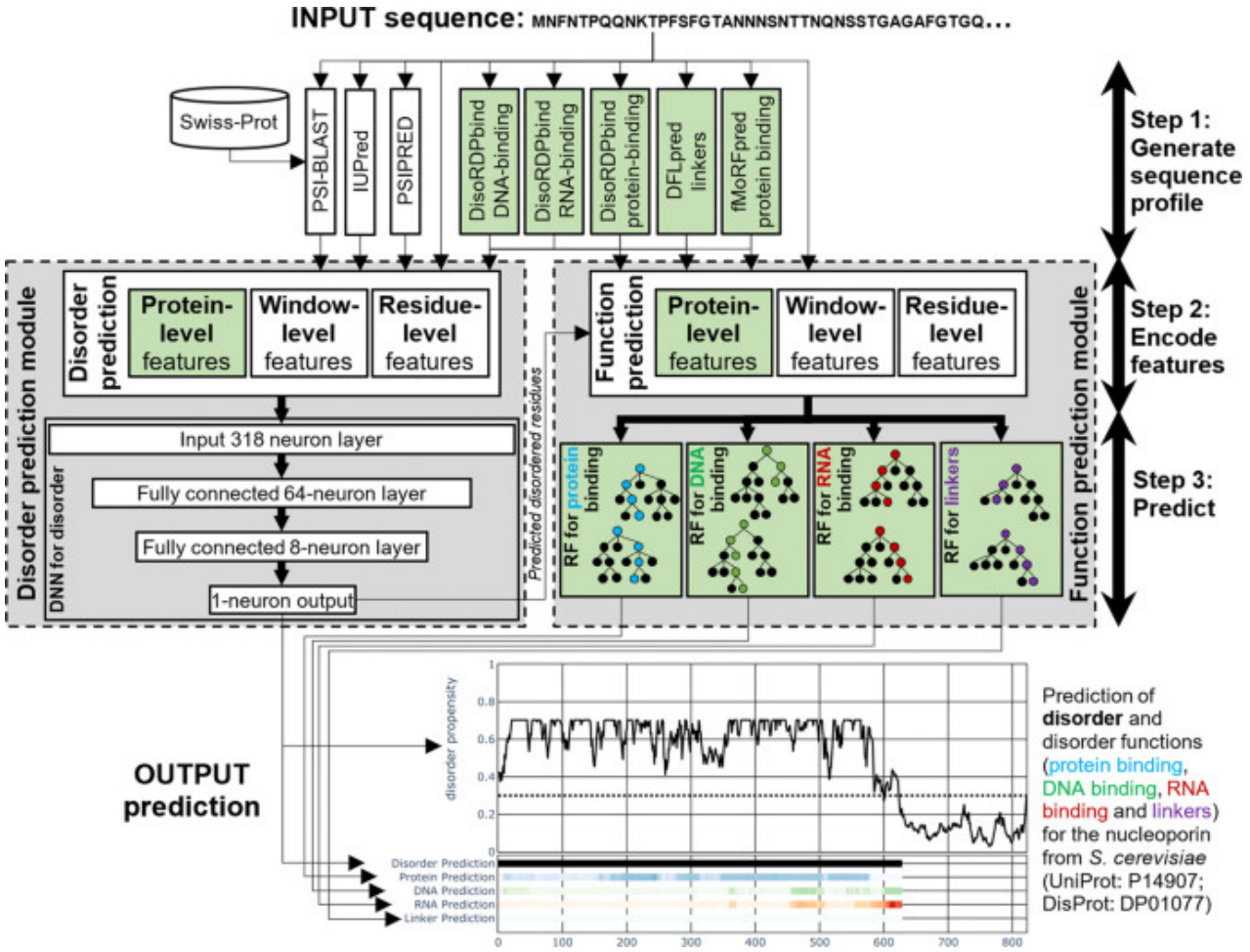
IUPred.enzim.hu

An IUPred output



CAID winner flDPnn

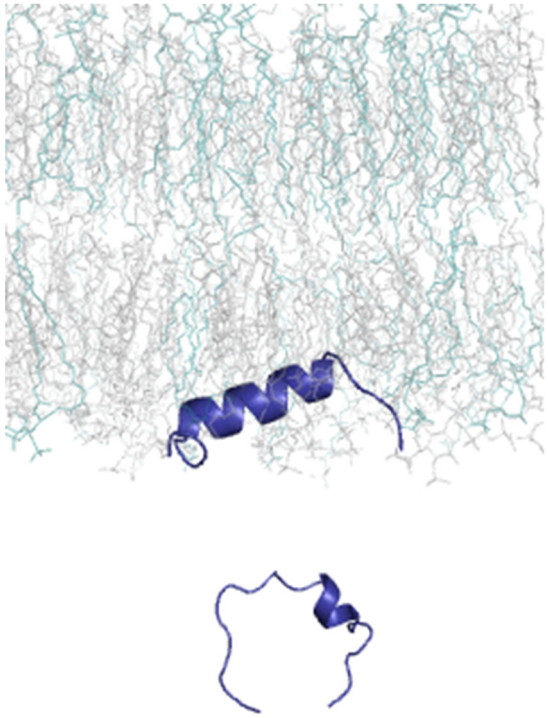
Critical Assessment of protein Intrinsic Disorder



Hu *et al.* Nature Communications volume 12, Article number: 4438 (2021)

MemMoRF

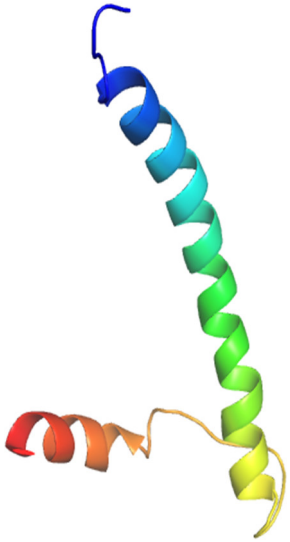
Membrane Molecular Recognition Feature, https://memmorf.hegelab.org



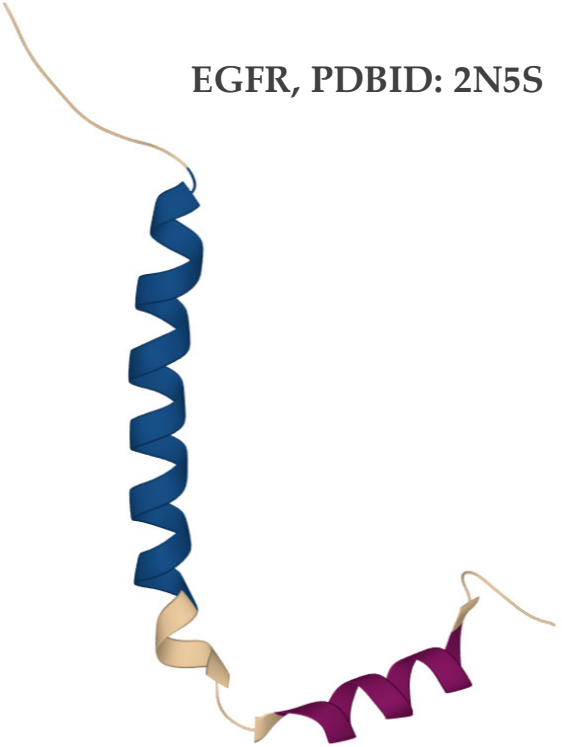
MemMoRF

Membrane Molecular Recognition Feature, https://memmorf.hegelab.org

E protein



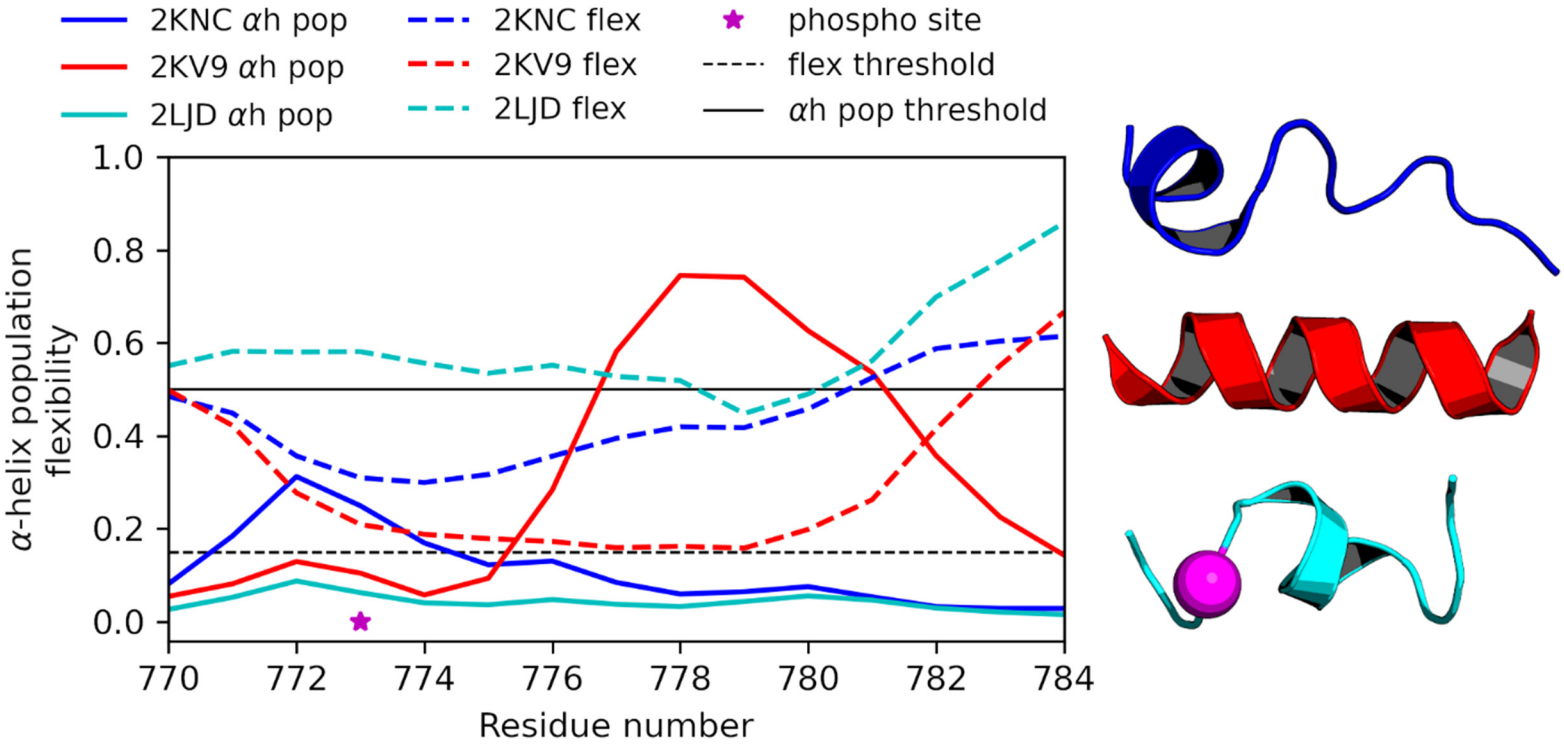
EGFR, PDBID: 2N5S



MemMoRF

Membrane Molecular Recognition Feature

Integrin beta-3



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.

http://memmorf.hegelab.org

3D structure prediction

„*Ab initio*" folding

- CASP (Critical Assessment of Techniques for Protein Structure Prediction)
- constraints from experiments
- deep learning (e.g. AlphaFold2, RoseTTAFold)

Homology modelling

- conserved sequence == conserved structure
- > 30% similarity
- most important: the sequence alinement

Homology modelling

- Searching a template
- Sequence alignment
- Modelling
- Energy minimization

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

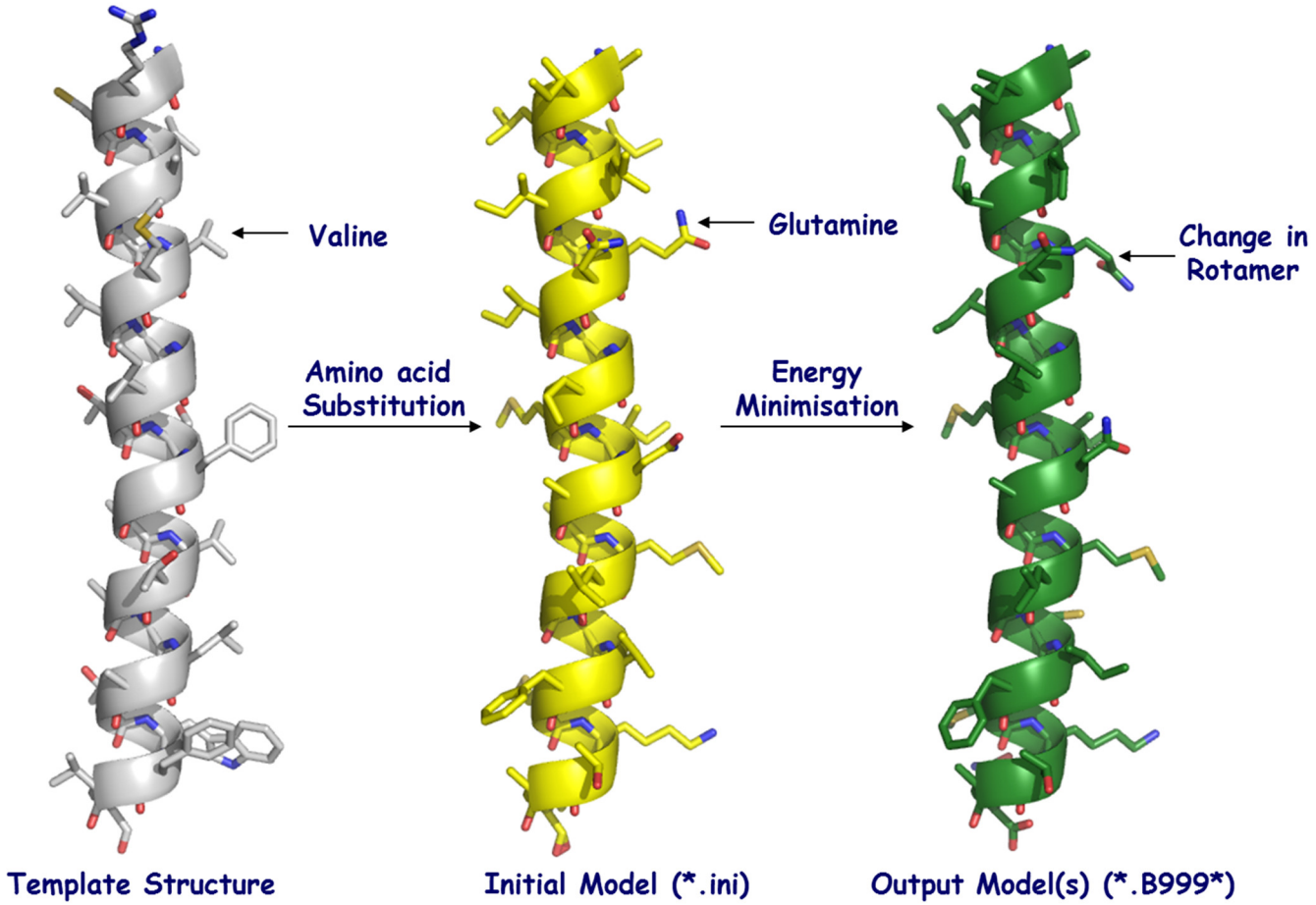
	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1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

Basic Local Alignment Search Tool (BLAST)

Alignement – pl. ClustalW

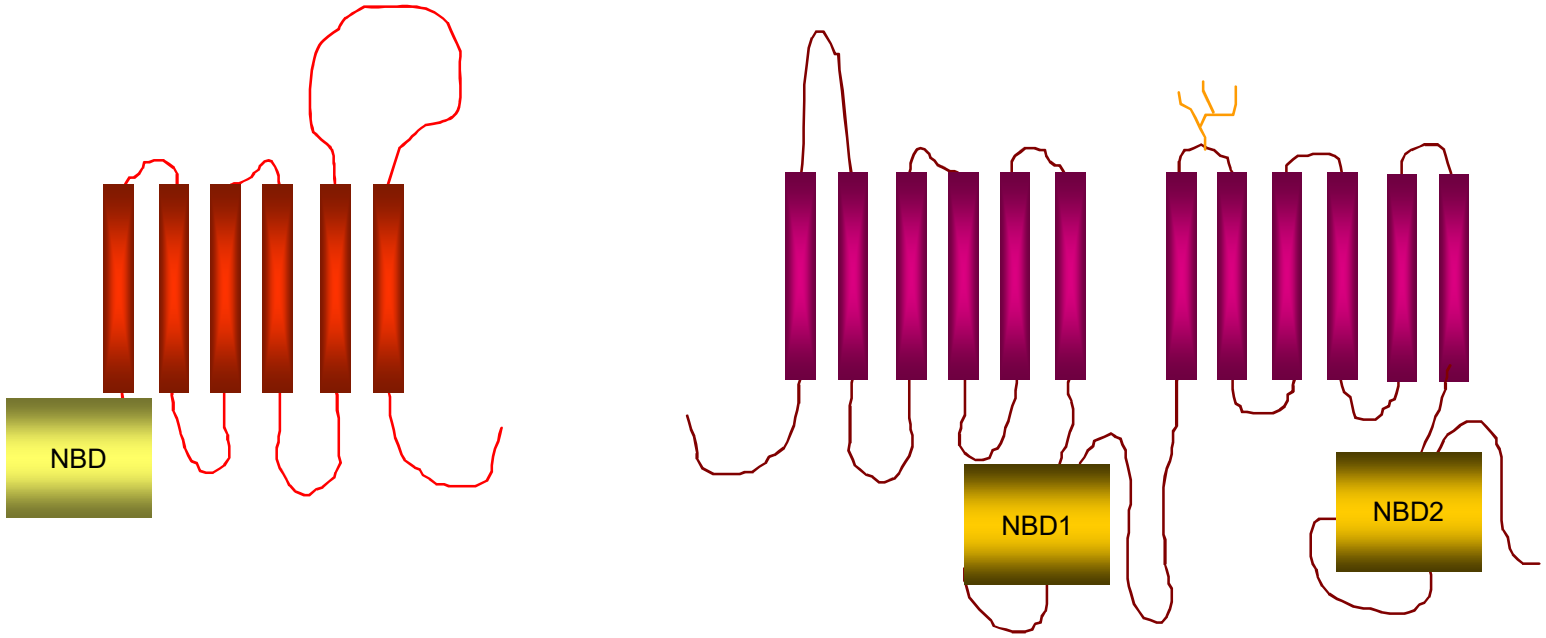
```
CLUSTAL W (1.83) multiple sequence alignment


2HYD        -----------MIKRYLQFVK-----PYKYRIFATIIVGIIKFGIPMLIP
3B5X        ---------WQTFKRLWTYIR-----LYKAGLVVSTIALVINAAADTYMI
CFTR_HUMAN  MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLS
                              *    : :      *:  :  : *  : .      :
```

Homology modelling



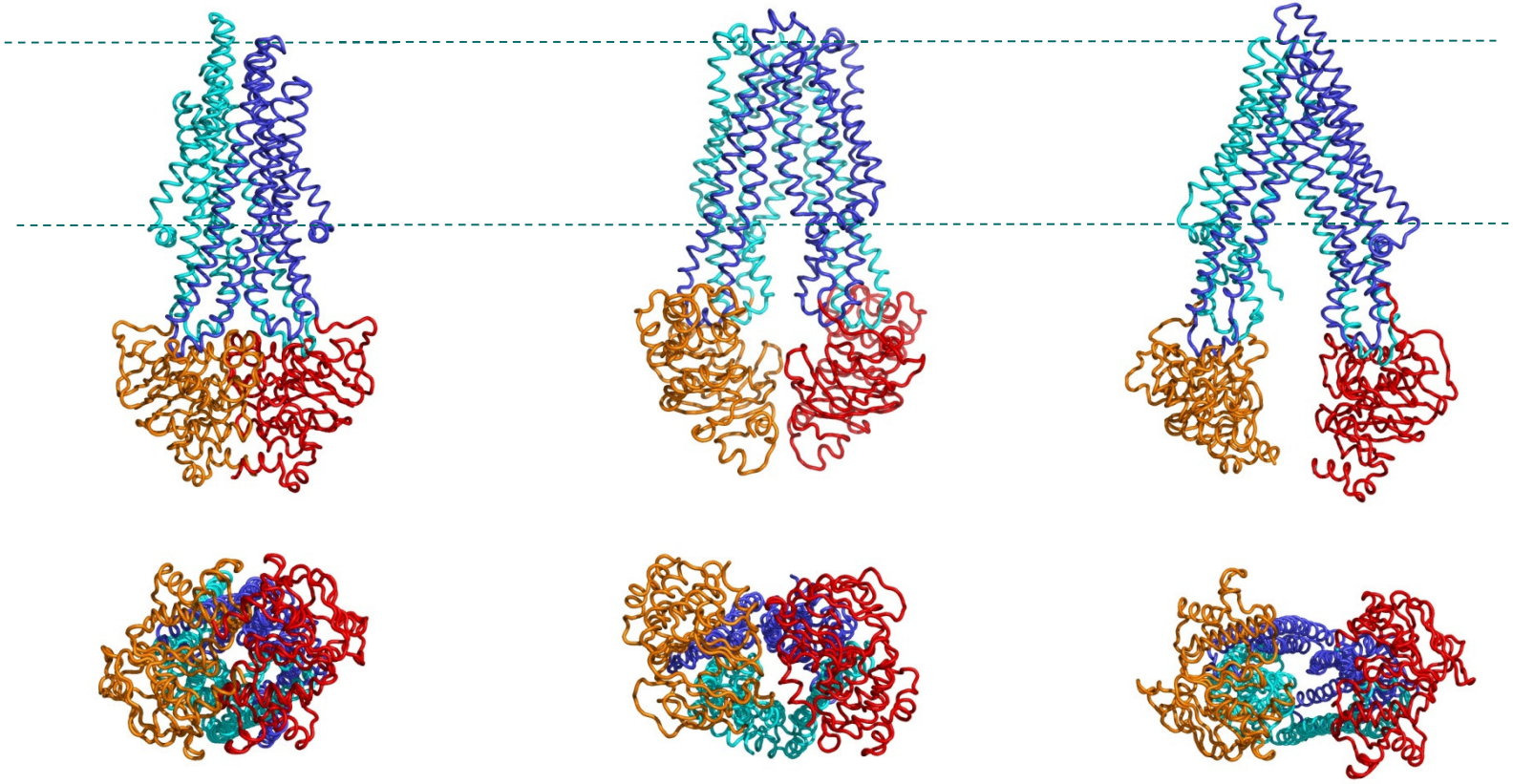
source: SBCB, Oxford, UK

ATP Binding Cassette (ABC) proteins



Conformation of ABC proteins

(Pgp-like)

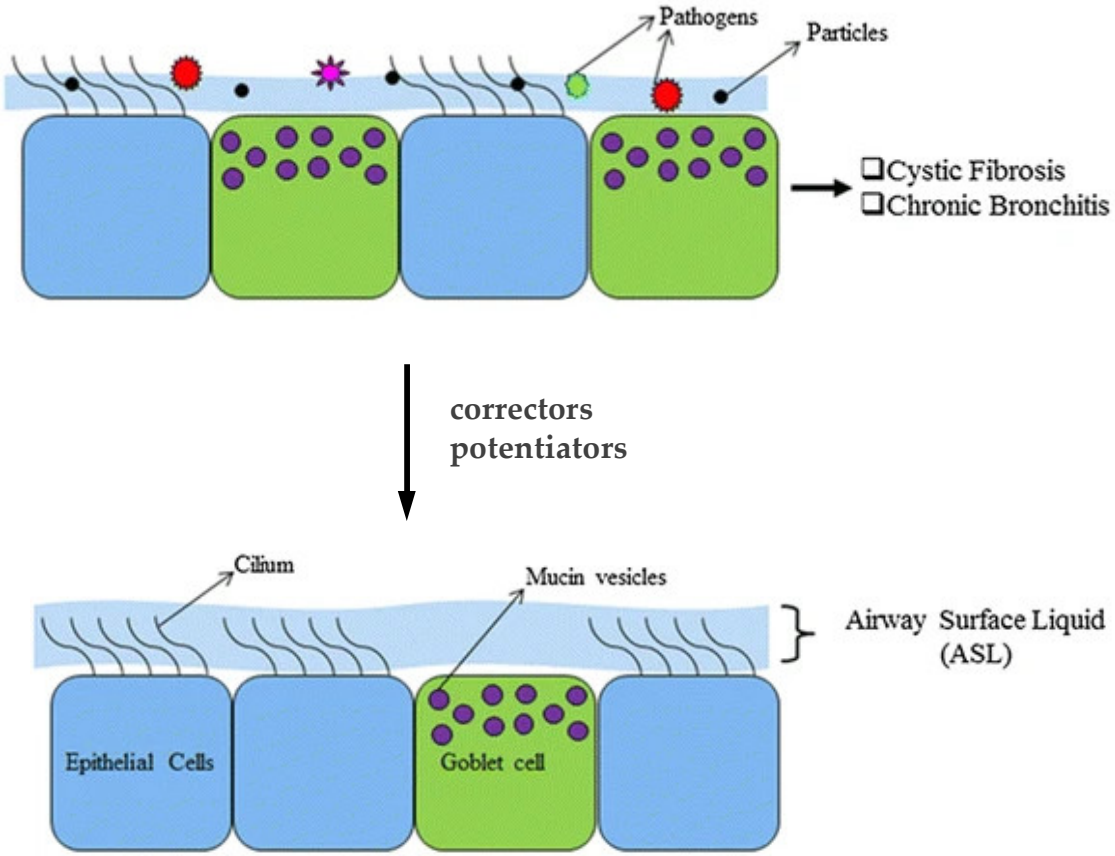


“bottom-closed” (+ATP)

“bottom-closed” apo (-ATP)

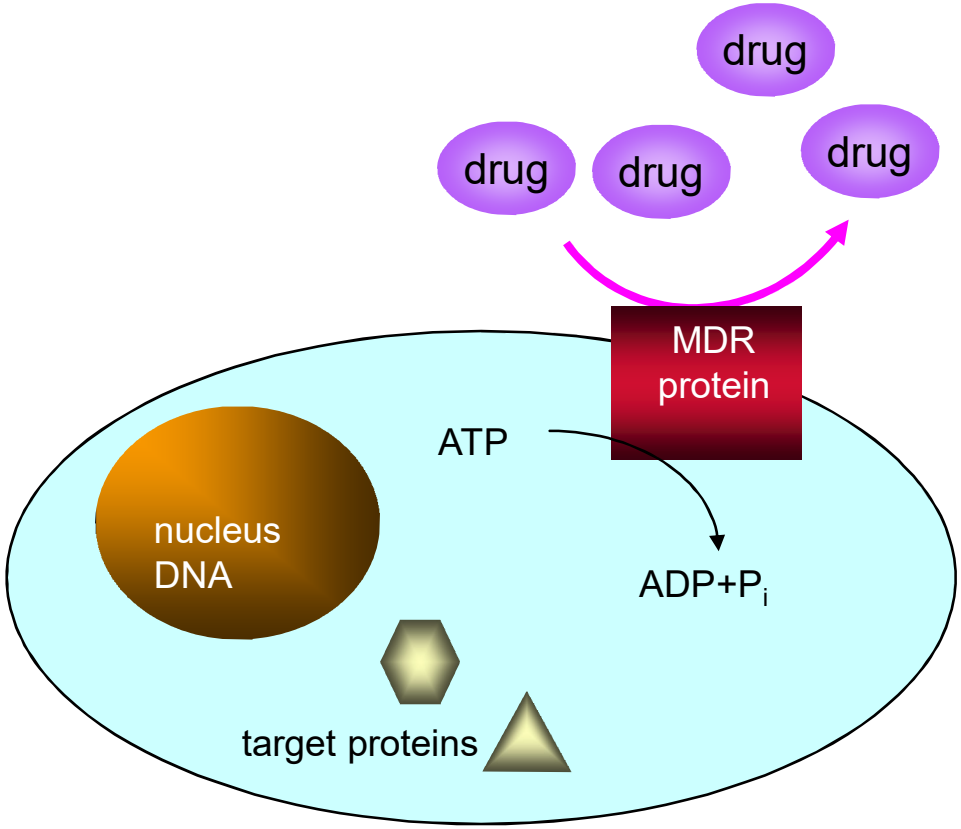
“bottom-open” apo (-ATP)

Cystic fibrosis (CF)



Ghosh, Boucher, Tarran,
CMLS 2015

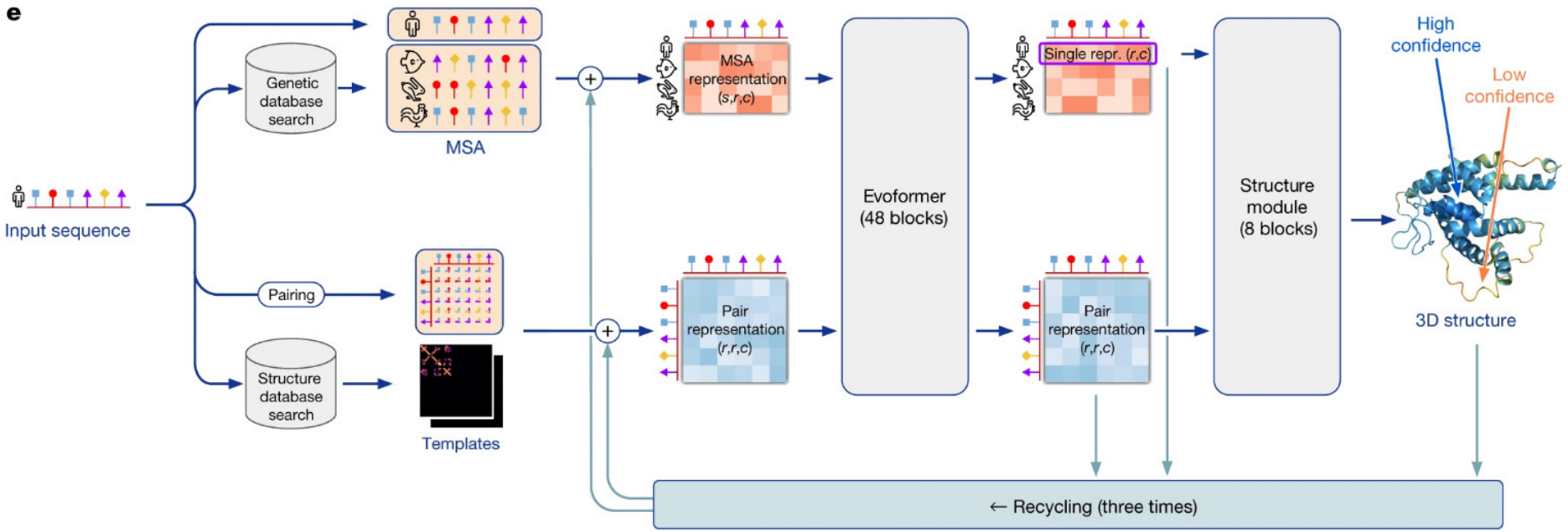
Multidrug transport – ABC proteins

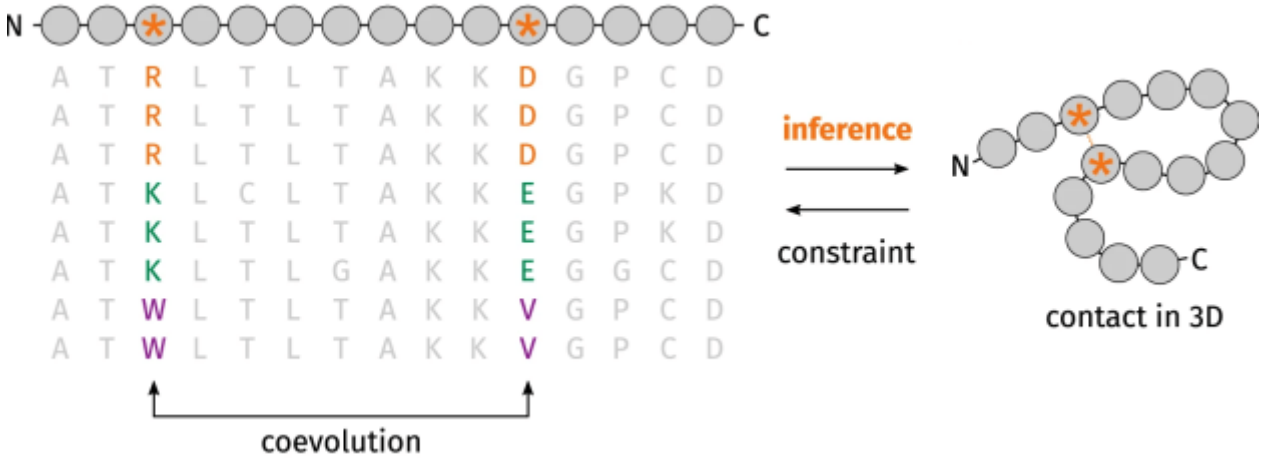


AlphaFold2

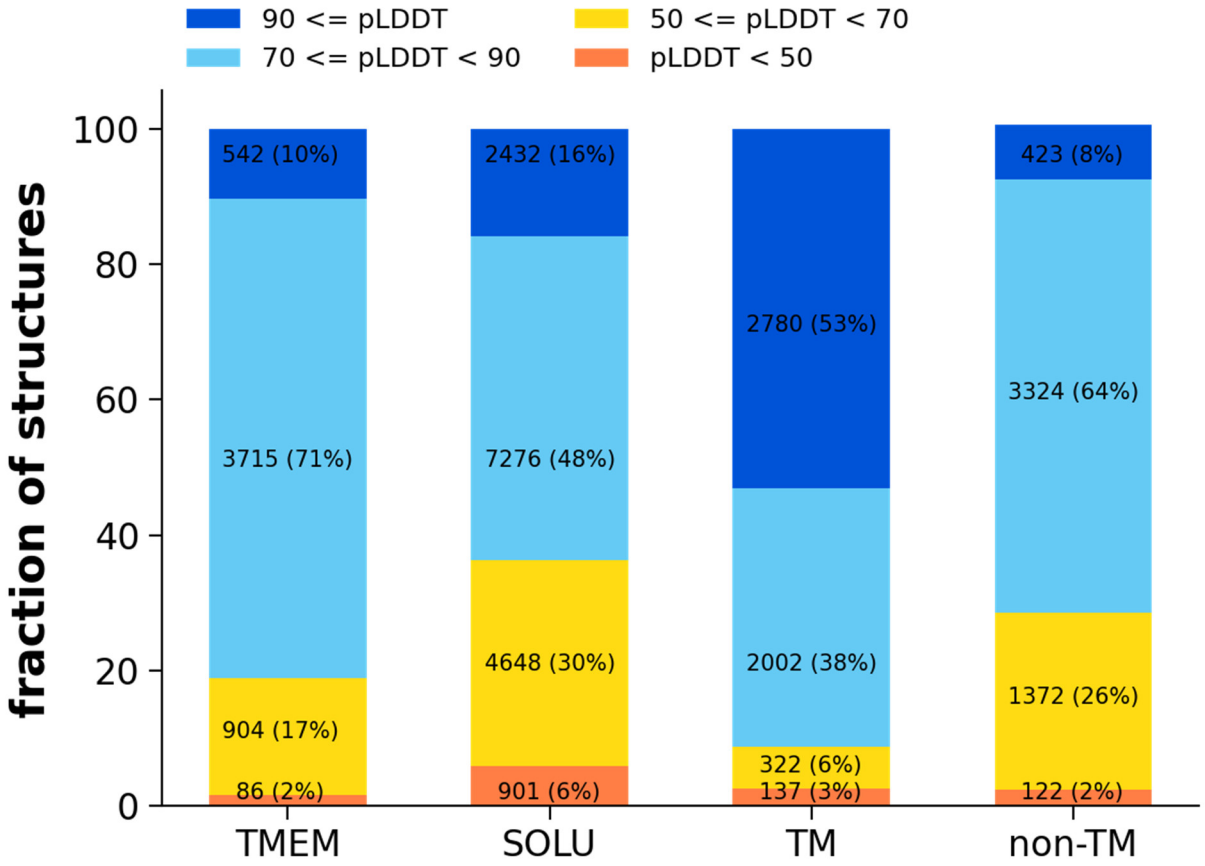
machine learning, deep learning, AI

Jumper et al. Nat 2021

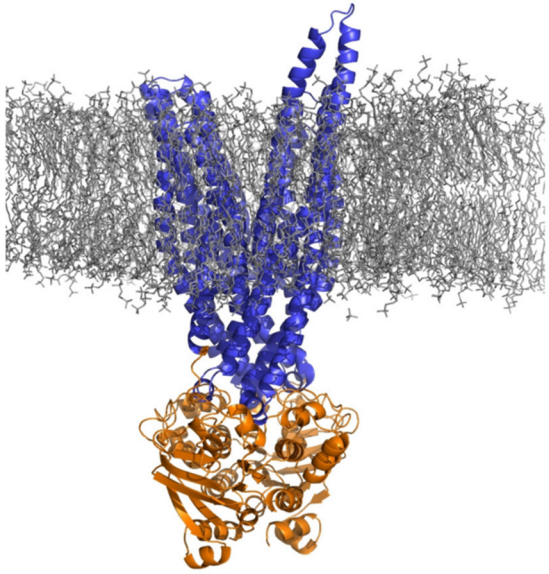


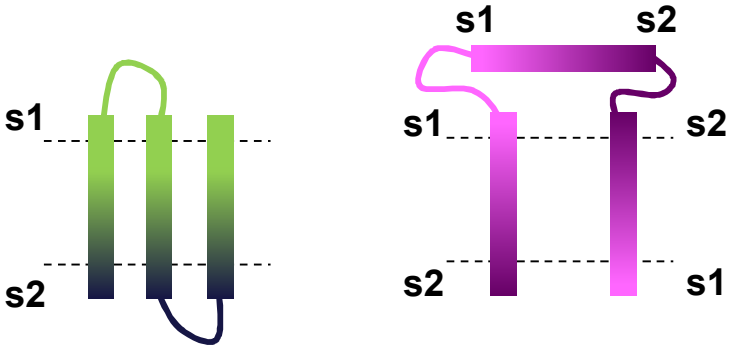


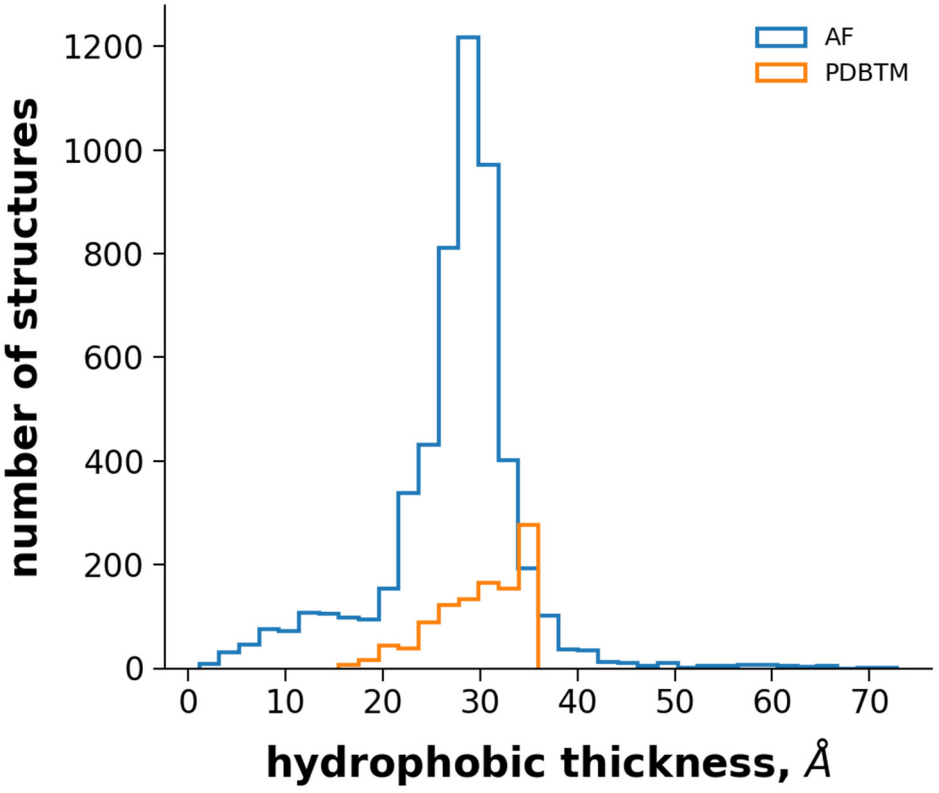
TM protein structure prediction by AF2



TM protein structure prediction by AF2

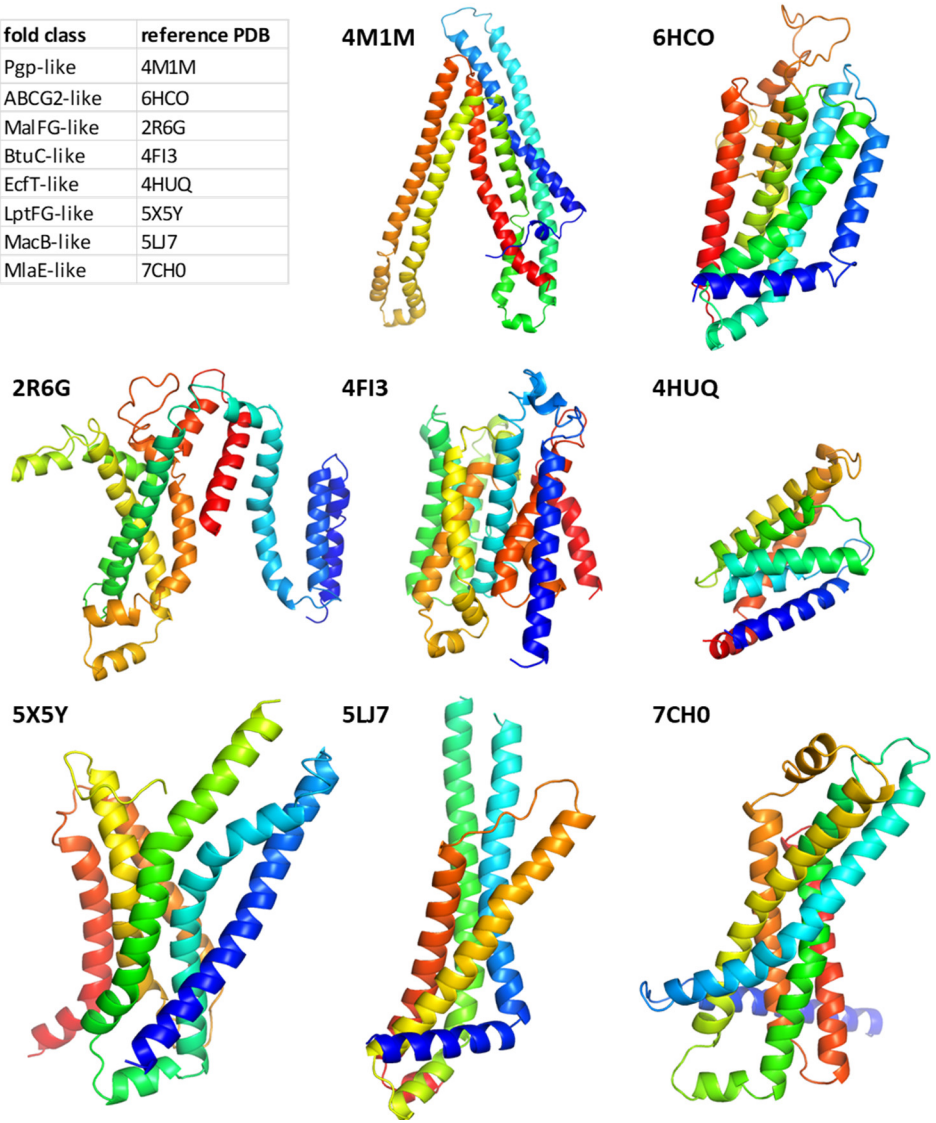


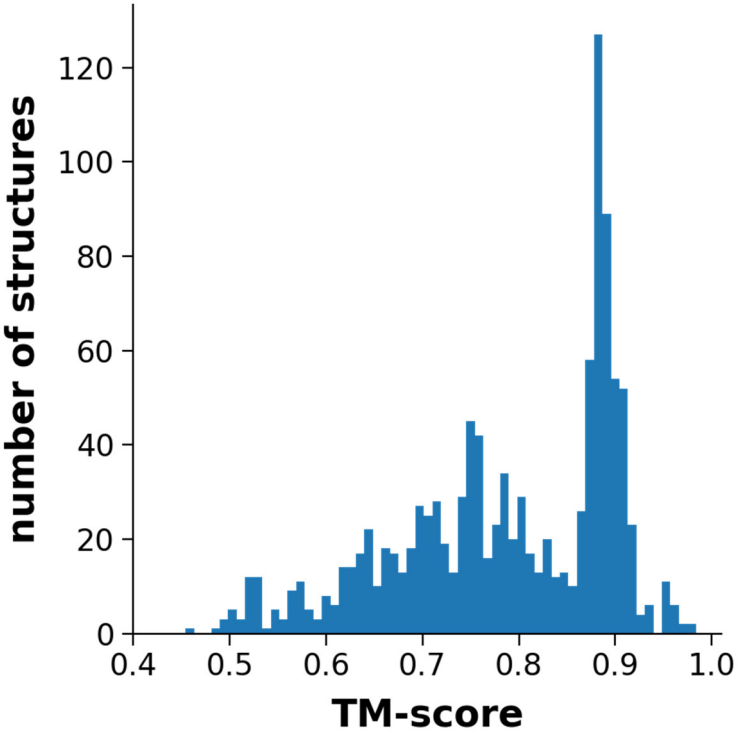




ABC protein folds

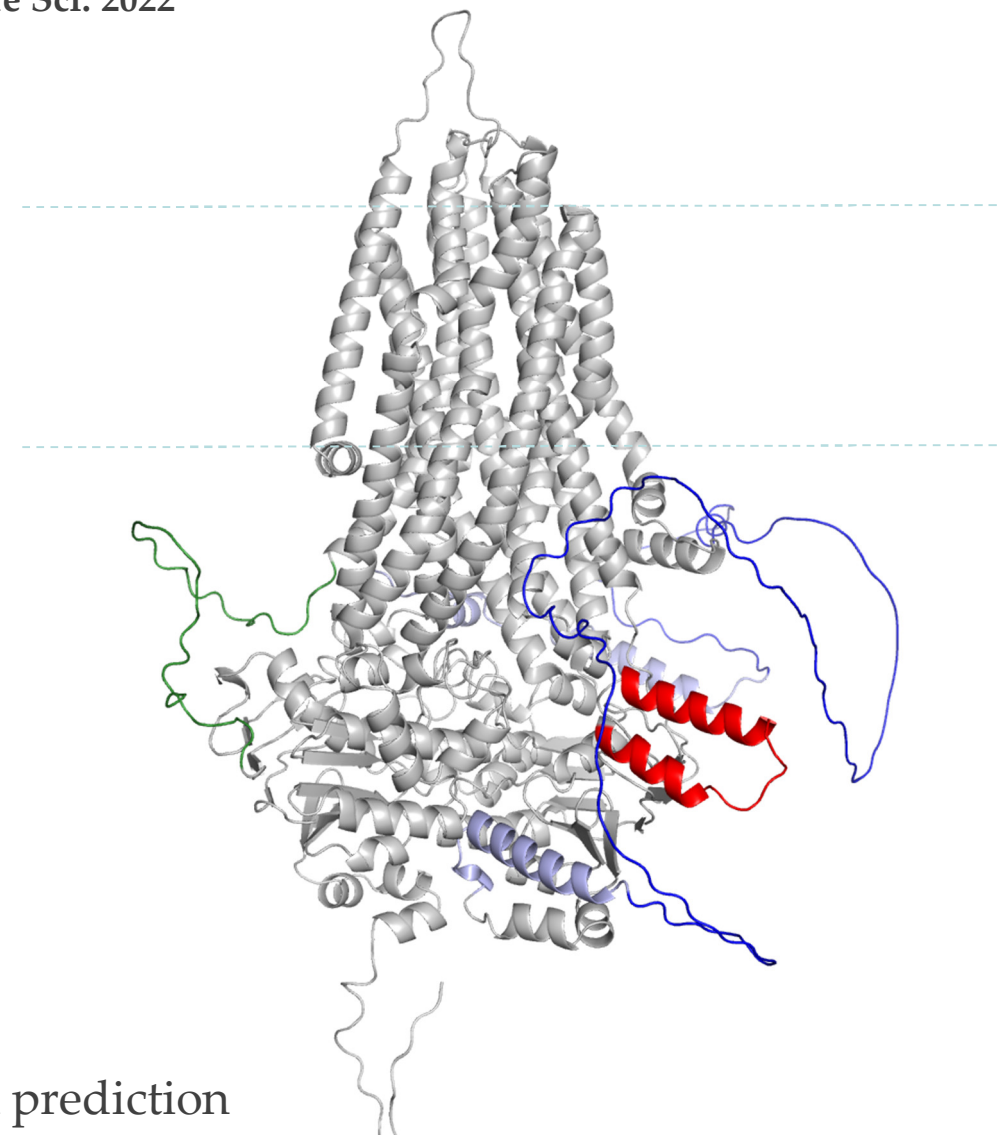
fold class	reference PDB
Pgp-like	4M1M
ABCG2-like	6HCO
MalFG-like	2R6G
BtuC-like	4FI3
EcfT-like	4HUQ
LptFG-like	5X5Y
MacB-like	5LJ7
MlaE-like	7CH0





AlphaFold – TM – ABC – CFTR

Hegedus *et al.* Cell Mol Life Sci. 2022

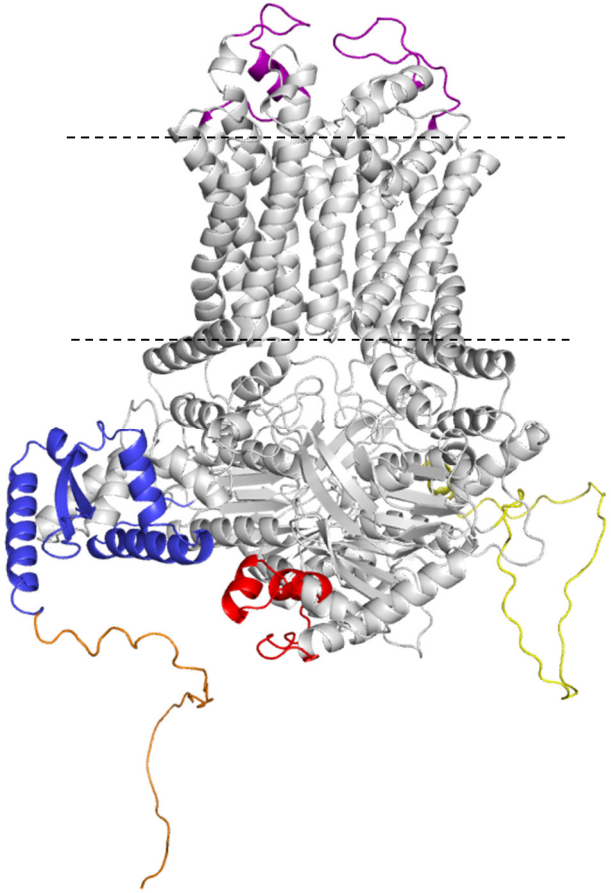


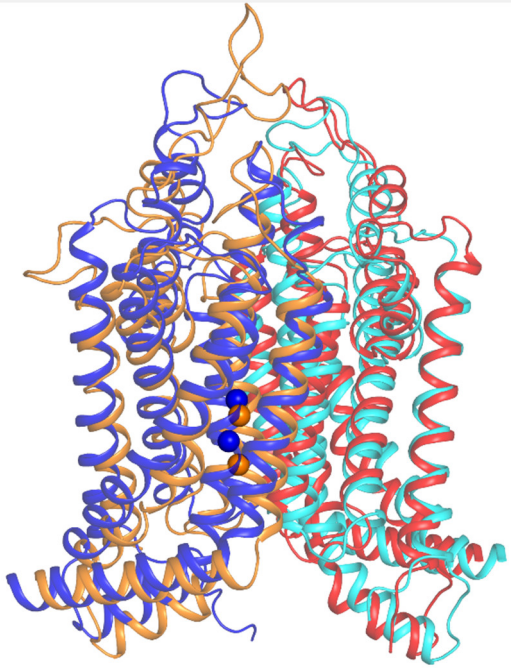
pLDDT score - IDR prediction

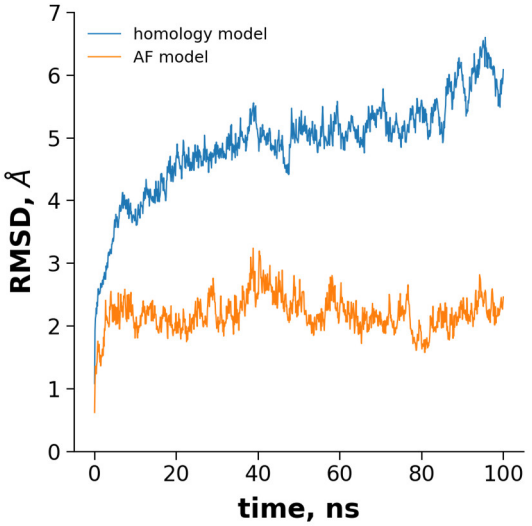
Tusnády GE *et al.*, https://tmalphafold.ttk.hu, NAR 2022

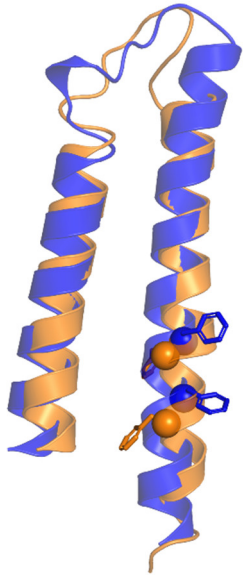
A plant transporter, AtABCG36

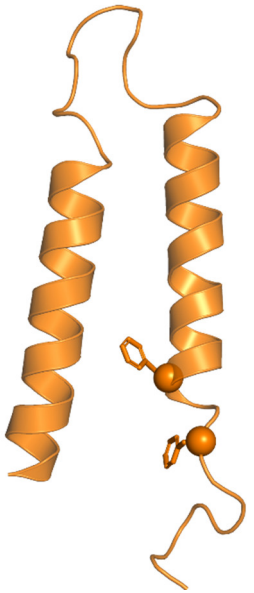
hormones, Umbrella Sampling



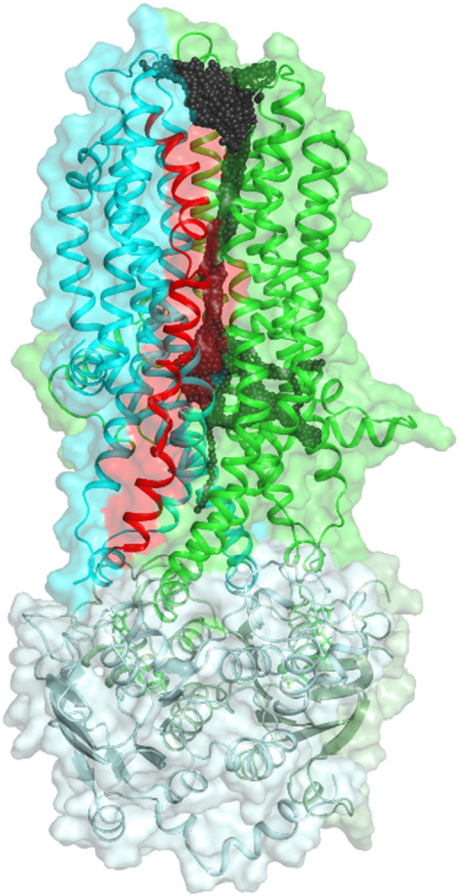




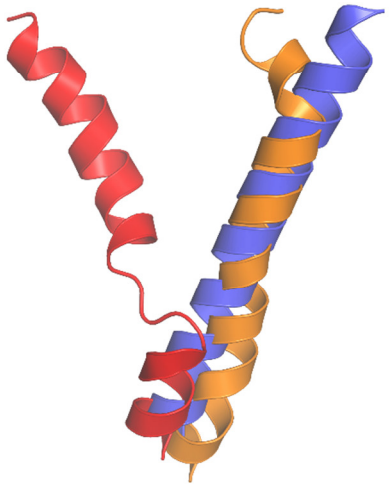




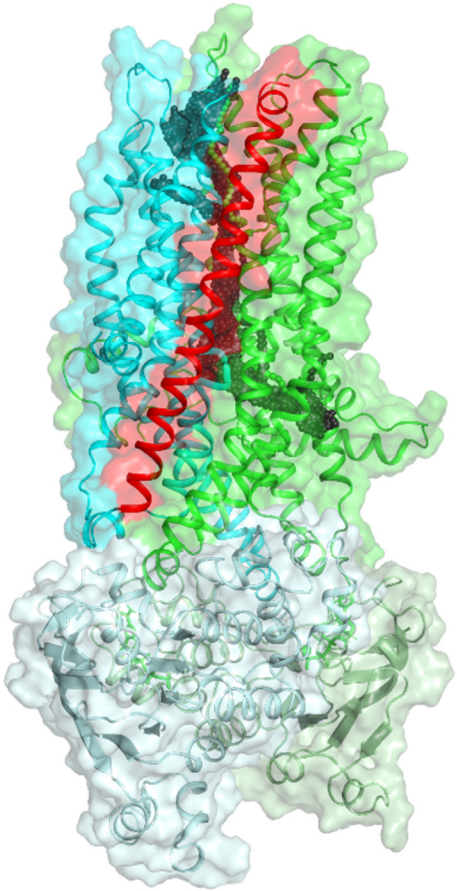
CFTR TM8



cryo-EM AF2 homo



cryo-EM	
	1/22 simulations
	0.36%
homo	
	5/6 simulations
	3.74%

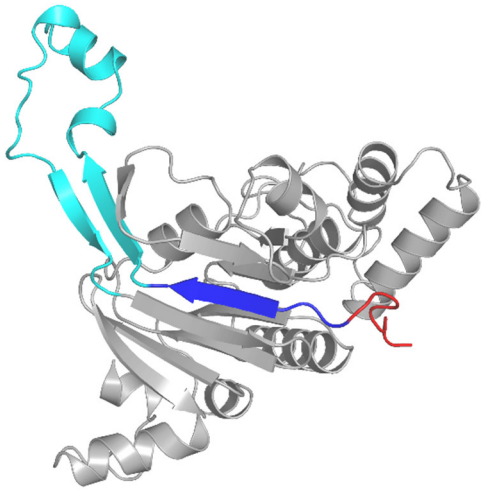


AF2 corrects an experimental structure

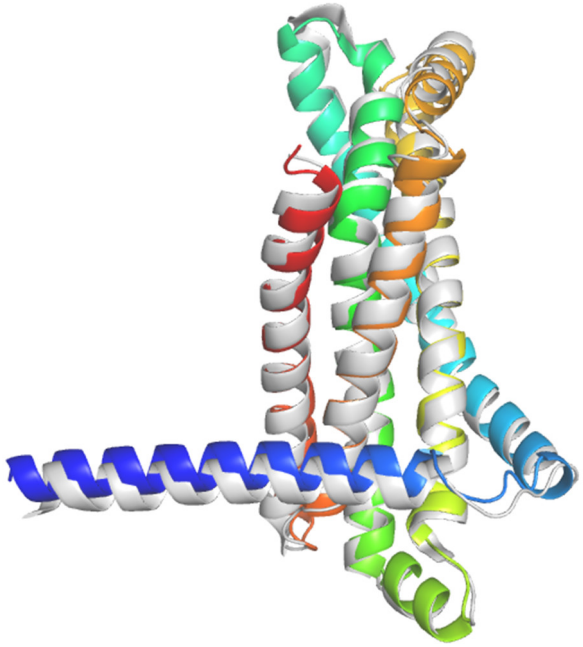
G2 6HCO AVLSFHNICY
G8 5DO7 NSLYFTYSGQ ✗

G2 6HCO AVLSFHNICY
G8 seq NTLEVRDLNY ✓

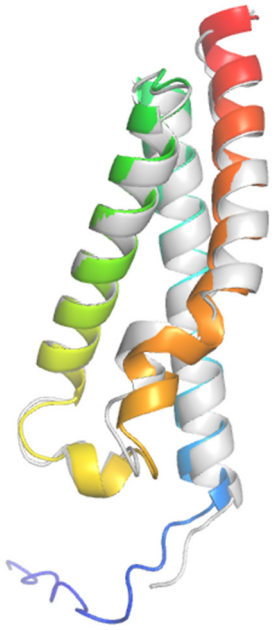
G2 6HCO AVLSFHNICY
G8 AF NTLEVRDLNY ✓



Prediction of new TM folds



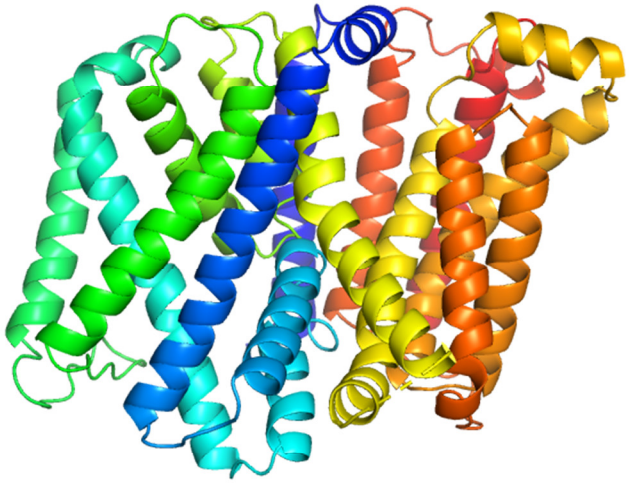
MlaE-like fold
PDBID: 7ch0
RMSD of 1.28 Å



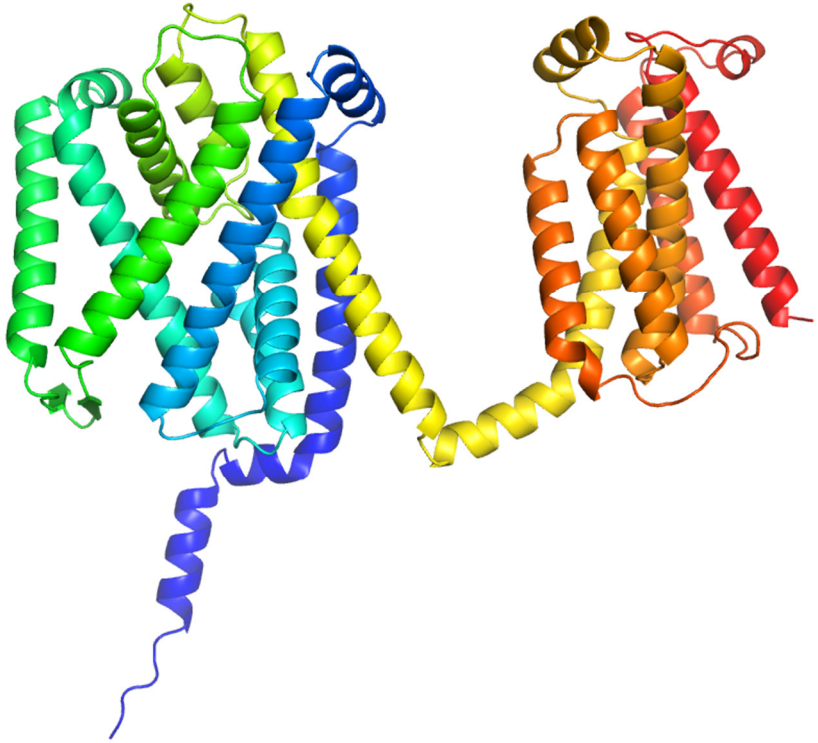
ER membrane protein complex subunit 6
PDBID: 6ww7
RMSD of 0.96 Å

Prediction of new TM folds

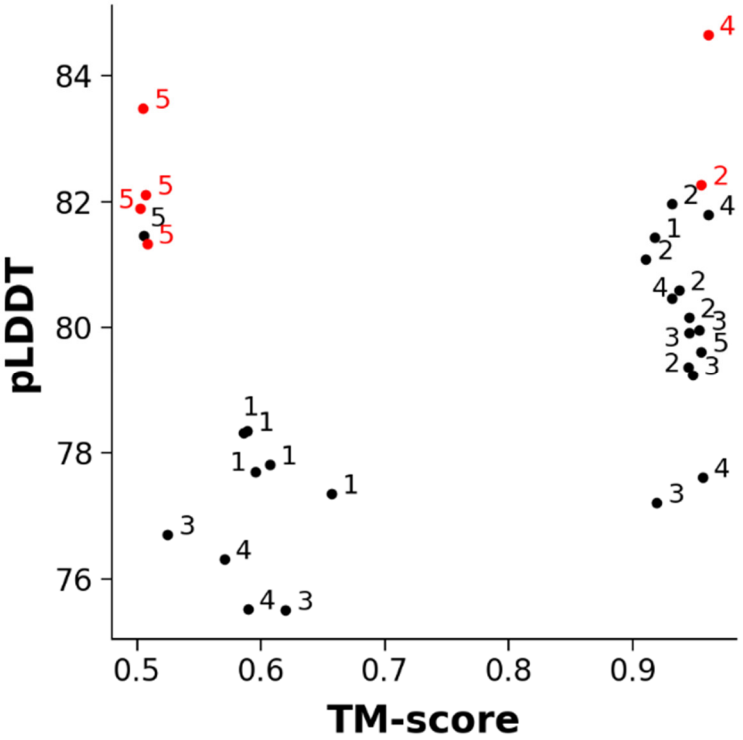
MprF (PDBID: 7DUW)

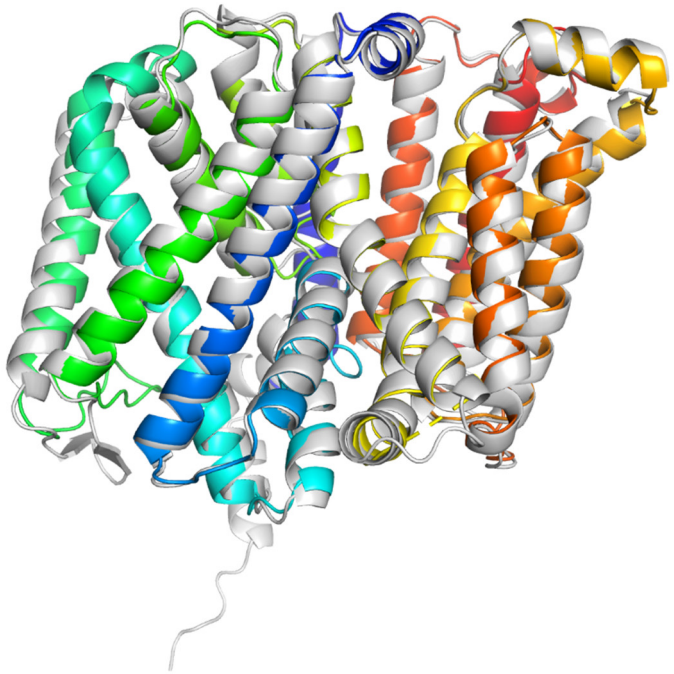


AF2



Prediction of new TM folds

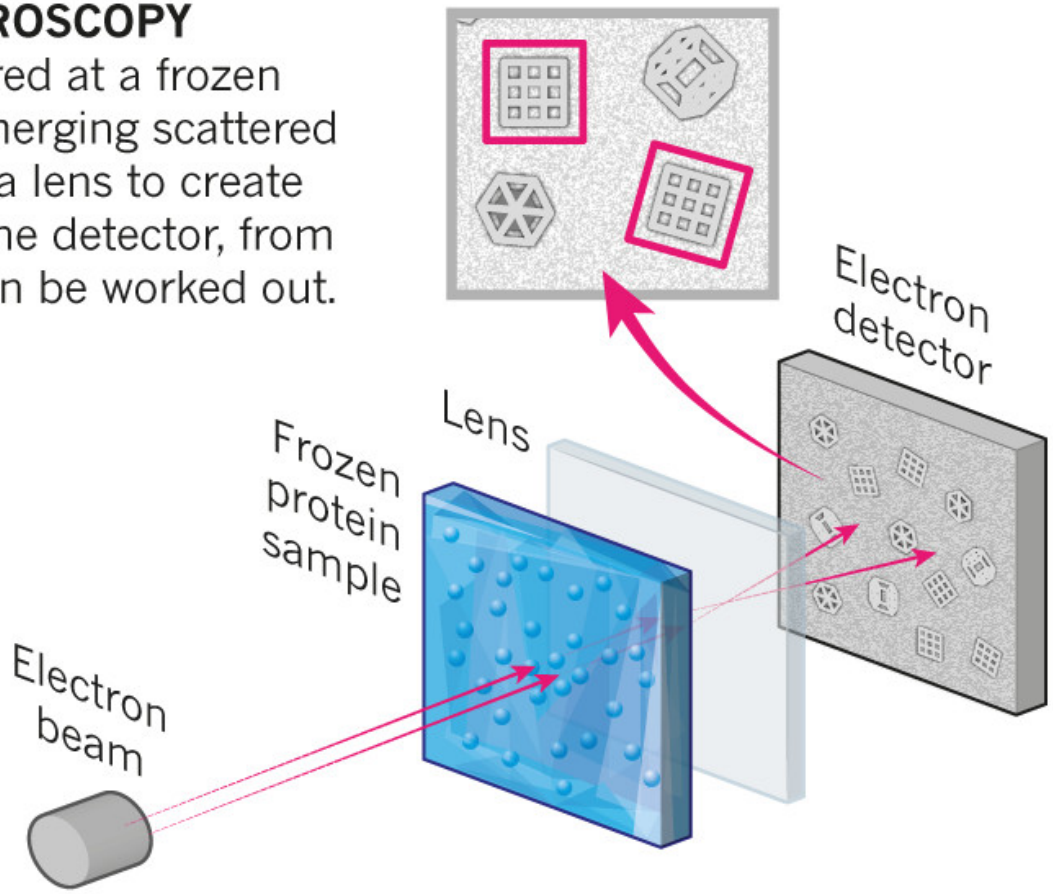




Structure determination – „single particle”

Cryo-electron microscopy

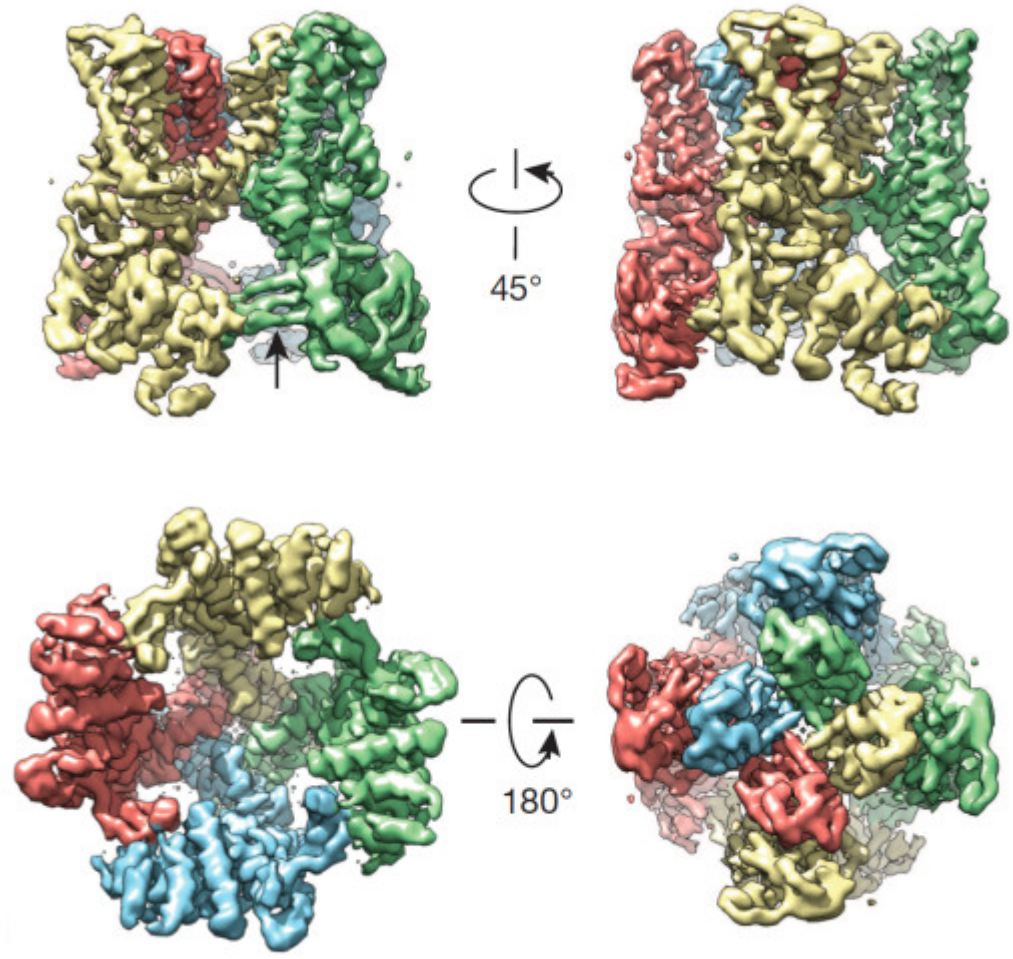
CRYO-ELECTRON MICROSCOPY
A beam of electron is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, from which their structure can be worked out.



Ewen Callaway, Nature | News Feature
The revolution will not be crystallized: a new method sweeps through structural biology, 09 September 2015

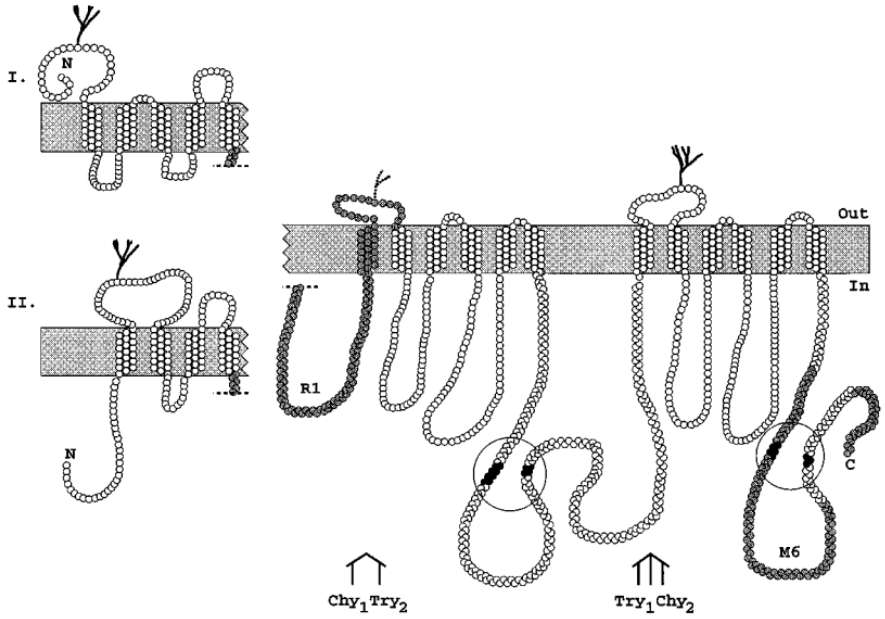
Structure determination – „single particle”

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



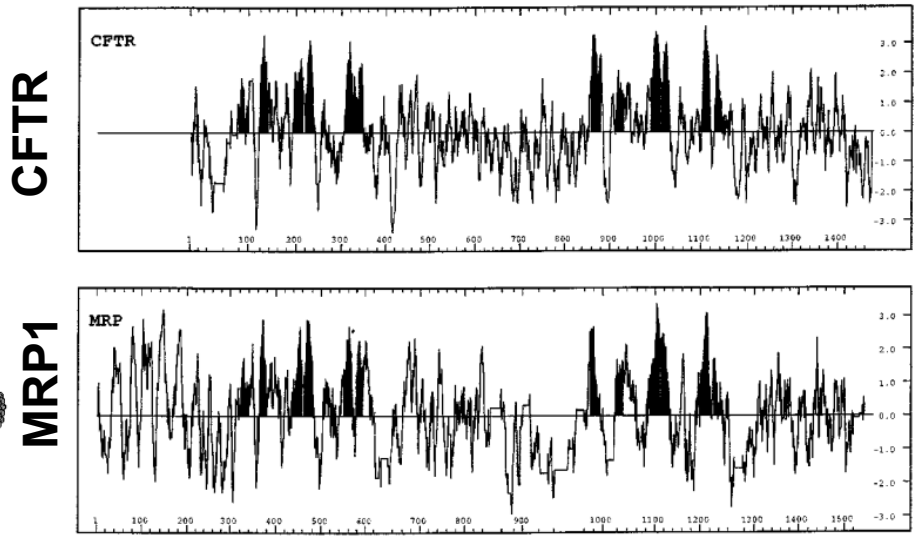
Membrane protein topology

MRP1



Bakos *et al.* JBC 1996

Hydrophobicity profiles:



CFTR topology is known from experiments

Chang et al. J Biol Chem. 1994 Jul 15;269(28):18572-5

Experimental methods for topology determination:

- *tag* recognition
- Cys accessibilitly

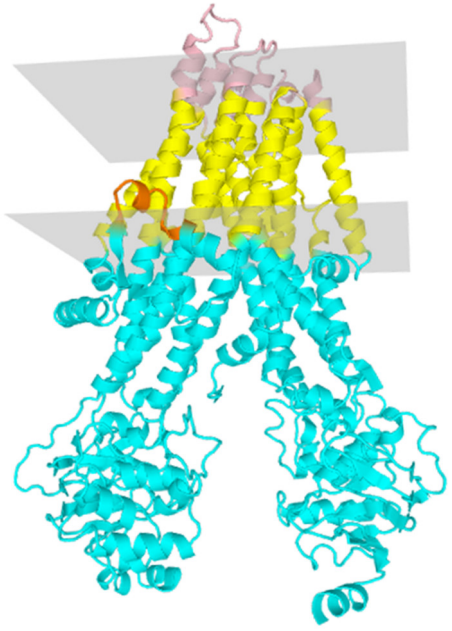
Prediction of membrane topology

- **Based on chemical properties of amino acids**
- **a.a. distribution in TM and soluble regions (statistics)**
- **Incorporation of experimental knowledge**
- **Integration of several predictors**

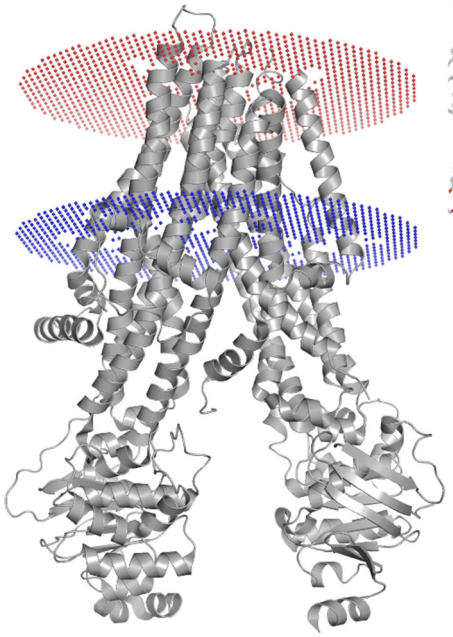
UNIfied database of TransMembrane Proteins
https://www.unitmp.org

Prediction of TM helices based on structure

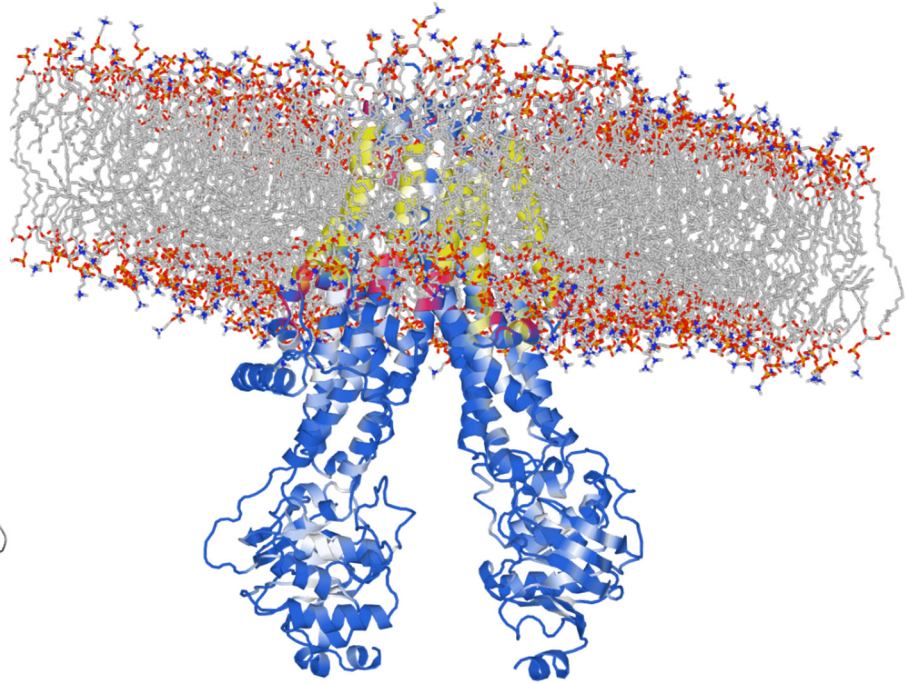
TMDET
PDBTM



PPM
OPM

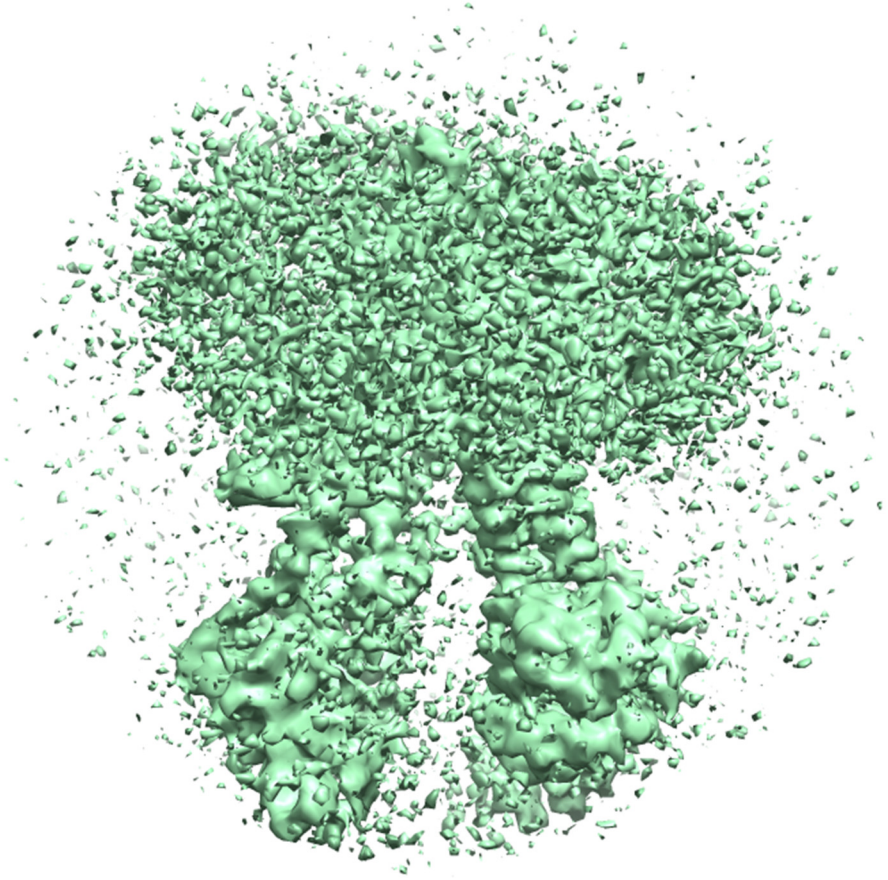


MemProtMD



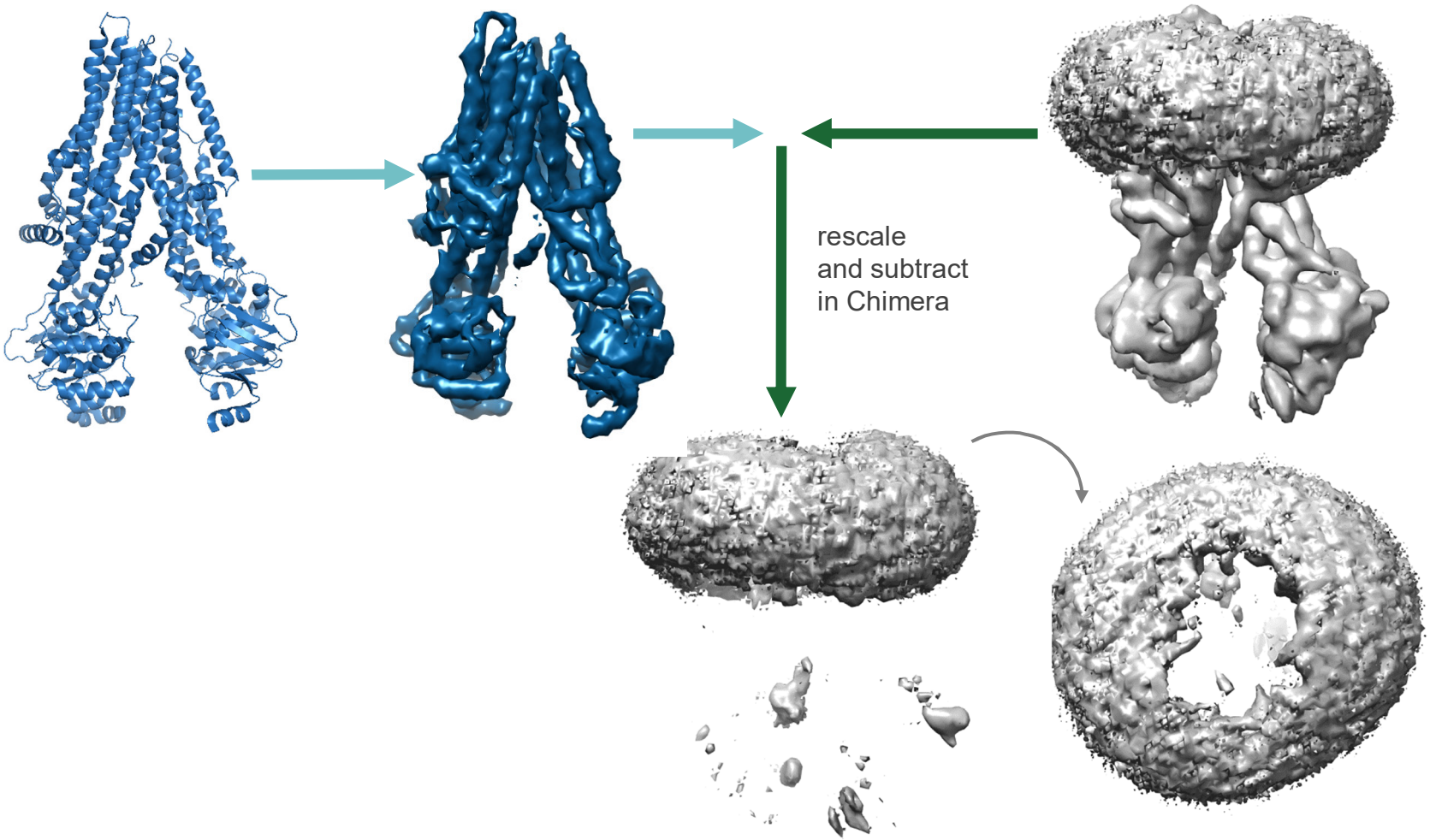
Membrane embedment in cryo-EM

CFTR (PDBID: 5UAK) EMD



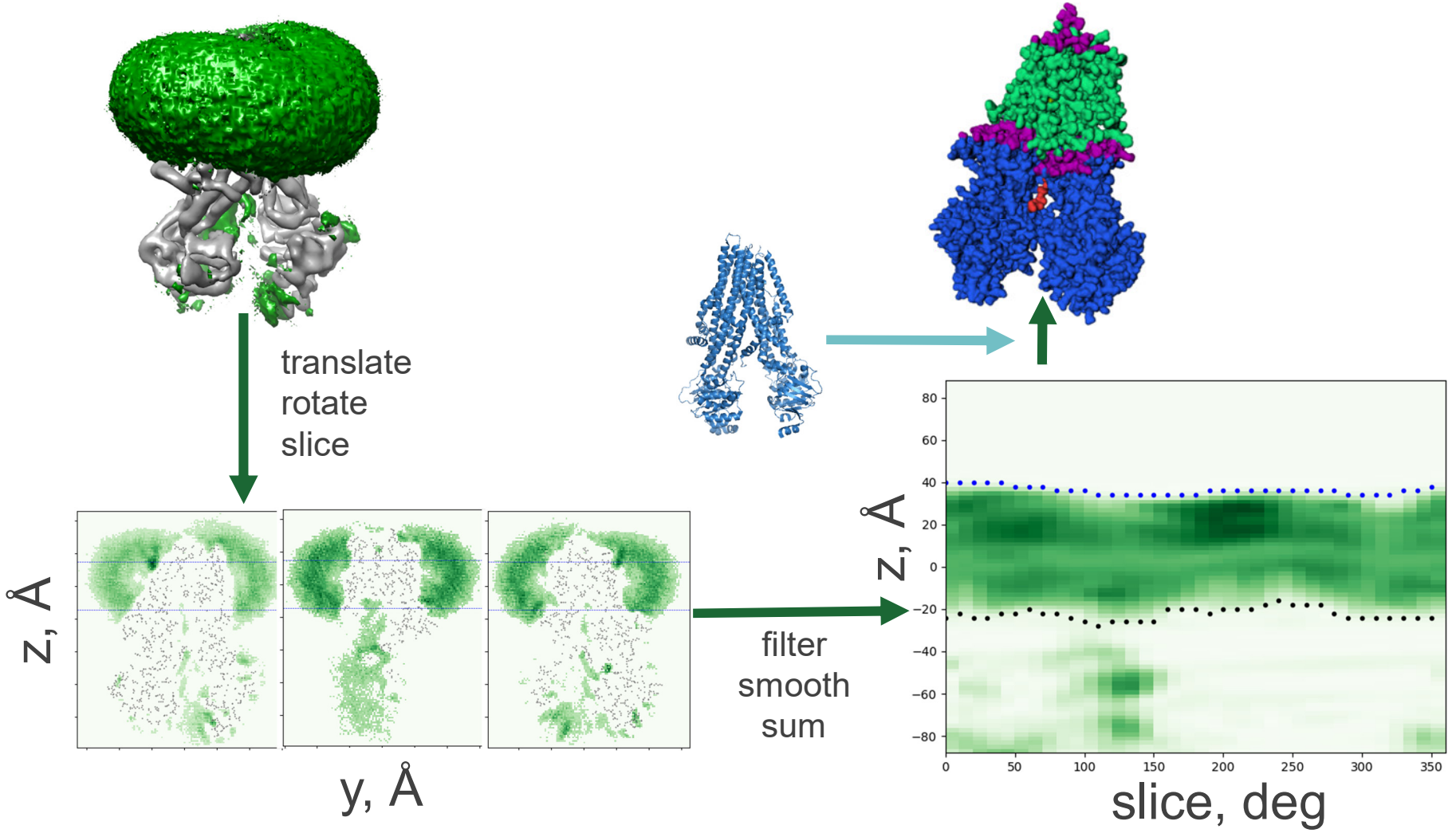
http://memblob.hegelab.org

Data extraction from cryo-EM



http://memblob.hegelab.org

Data extraction from cryo-EM



http://memblob.hegelab.org

Protein-protein interactions

Docking of proteins – challenging (surface shape, dynamics)

PISA - Protein Interfaces, Surfaces and Assemblies

Molecular Dinamics

AlphaFold2-Multimer

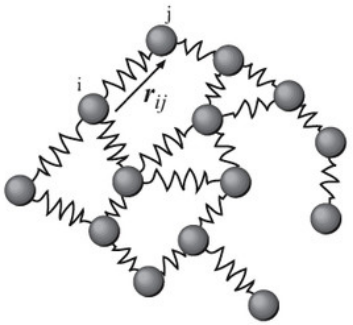
Topics

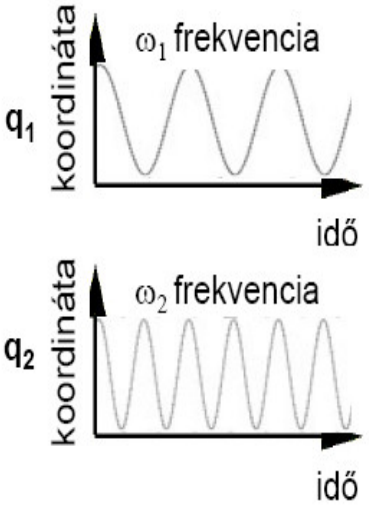
- **Protein structure**
- **Protein dynamics**

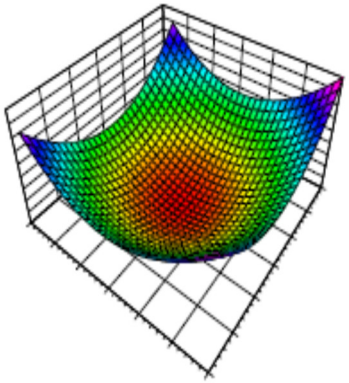
Methods for studying protein dynamics

Normal mode analysis

- harmonic potential
- analytic equation of motions
- normal modes







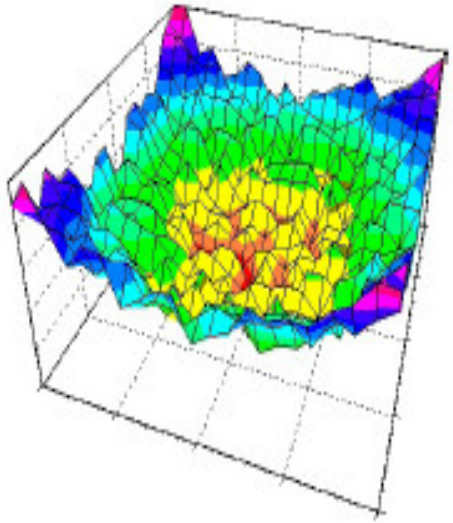
- Gaussian network model (GNM)
  - mean squared displacements
  - cross-correlations between fluctuations
- Anisotropic network model (ANM)
  - directionality by projection of motions to a mode space of N dimensions

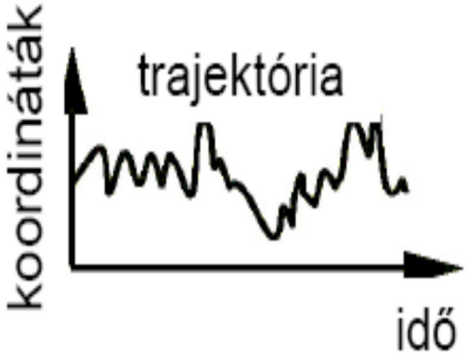
Tools: http://prody.csb.pitt.edu

Methods for studying protein dynamics

Molecular dynamics

- realistic potential surface
- numerical integration of Newton's equations
- a system of interacting particles
- forces between the particles and their potential energies are calculated by using interatomic potentials (molecular mechanics force fields)
- output: trajectory





The force field

$$E_{\text{prot}} = W_{\text{rot}} E_{\text{rot}} + W_{\text{atr}} E_{\text{atr}} + W_{\text{rep}} E_{\text{rep}} + W_{\text{solv}} E_{\text{solv}} + W_{\text{pair}} E_{\text{pair}}$$
$$+ W_{\text{mbenv}} E_{\text{mbenv}} + W_{\text{hbond}} E_{\text{hbond}} - E_{\text{ref}}$$

$$E_{\text{solv}} = -\sum_{i}^{\text{natom}} \sum_{j>i}^{\text{natom}} \left\{ \frac{2\Delta G_i^{\text{free}}}{4\pi\sqrt{\pi}\lambda_i r_{ij}^2} \exp\left(-d_{ij}^2\right) V_j + \frac{2\Delta G_j^{\text{free}}}{4\pi\sqrt{\pi}\lambda_j r_{ij}^2} \exp\left(-d_{ji}^2\right) V_i \right\}$$

Lazaridis (2003)

TABLE I. Solvation Parameters[†]

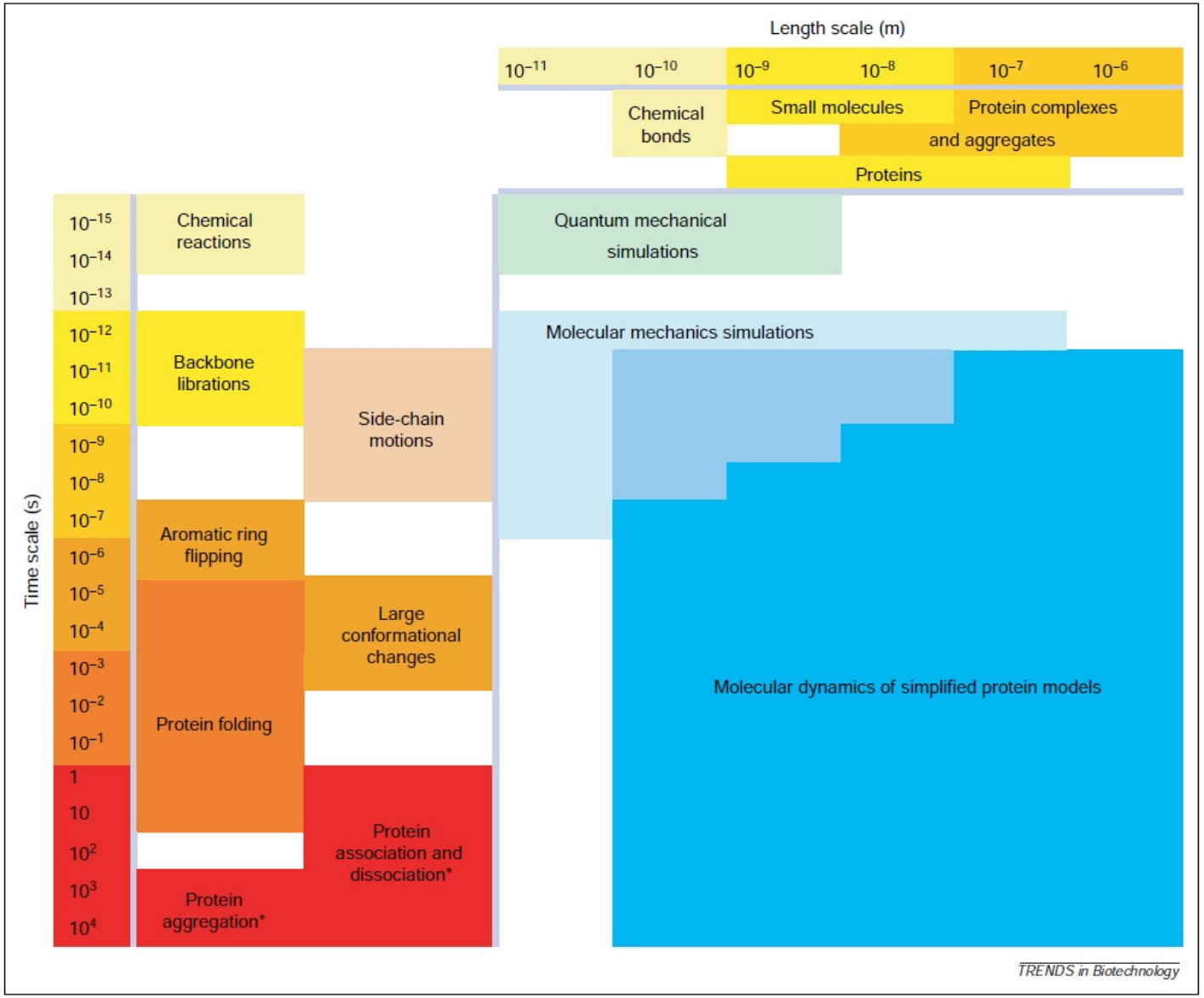
Atom types[a]	Volume	ΔG_i^{ref} [b]	ΔG_i^{free} [c]	ΔH_i^{ref} [b]	ΔCp_i^{ref} [d]
C	14.7	0.000	0.00	0.000	0.00
CR	8.3	−0.890	−1.40	2.220	6.90
CH1E	23.7	−0.187	−0.25	0.876	0.00
CH2E	22.4	0.372	0.52	−0.610	18.60
CH3E	30.0	1.089	1.50	−1.779	35.60
CR1E	18.4	0.057	0.08	−0.973	6.90
NH1	4.4	−5.950	−8.90	−9.059	−8.80
NR	4.4	−3.820	−4.00	−4.654	−8.80
NH2	11.2	−5.450	−7.80	−9.028	−7.00
NH3	11.2	−20.000	−20.00	−25.000	−18.00
NC2	11.2	−10.000	−10.00	−12.000	−7.00
N	0.0	−1.000	−1.55	−1.250	8.80
OH1	10.8	−5.920	−6.70	−9.264	−11.20
O	10.8	−5.330	−5.85	−5.787	−8.80
OC	10.8	−10.000	−10.00	−12.000	−9.40
S	14.7	−3.240	−4.10	−4.475	−39.90
SH1E	21.4	−2.050	−2.70	−4.475	−39.90

Lazaridis (1999)

The limitations of MD

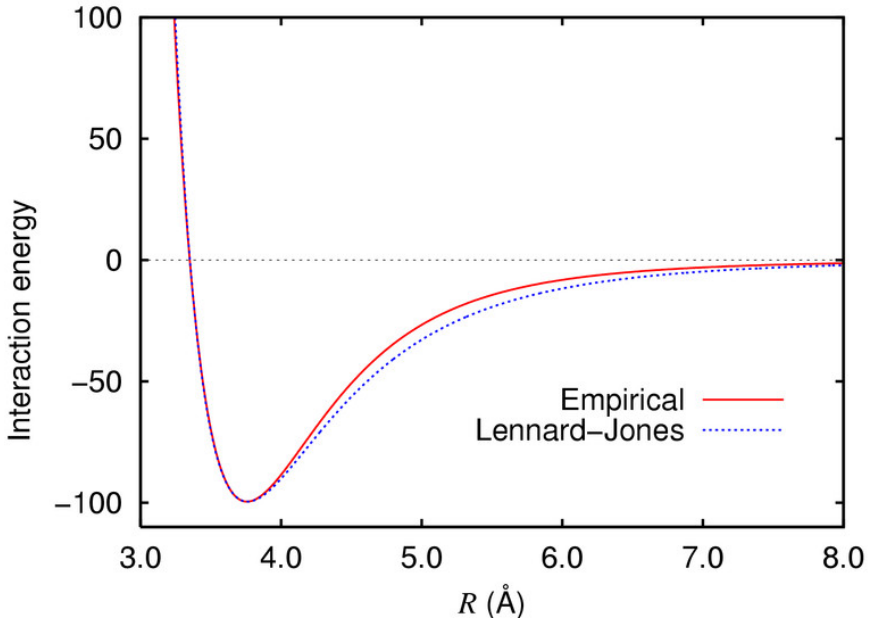
- time (computation time versus real time)
- calculation of the potential is the bottle-neck
- fs long integration steps
- „periodic boundary condition”
- solvent (explicit/implicit)

The time scale of various molecular events

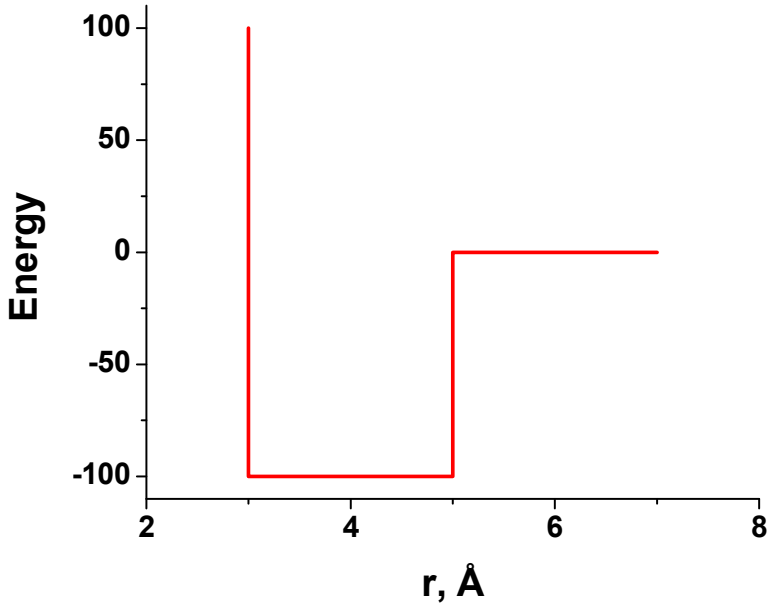


F. Ding and N.V. Dokholyan, TRENDS in Biotechnology, **23:**450 (2005)

Discrete Molecular Dynamics (DMD)



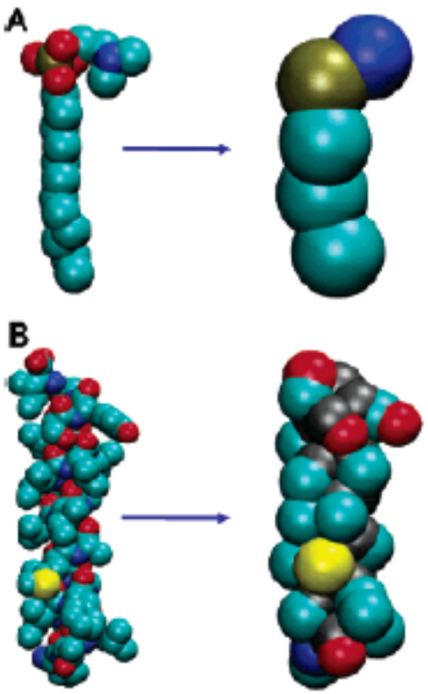
wikipedia

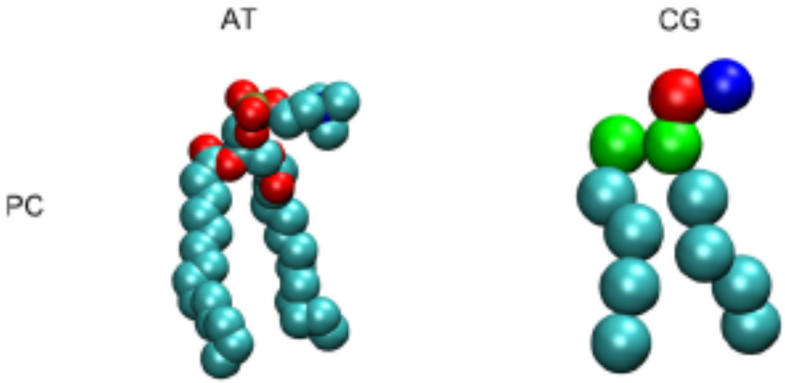


Ding, F., Dokholyan, N. V. PLoS Comput Biol 2:e85

$$\mathcal{V}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] = \varepsilon\left[\left(\frac{R_{min}}{r}\right)^{12} - 2\left(\frac{R_{min}}{r}\right)^{6}\right]$$

Simplified coarse-grained models





AT: atomistic, all-atom
CG: coarse grained

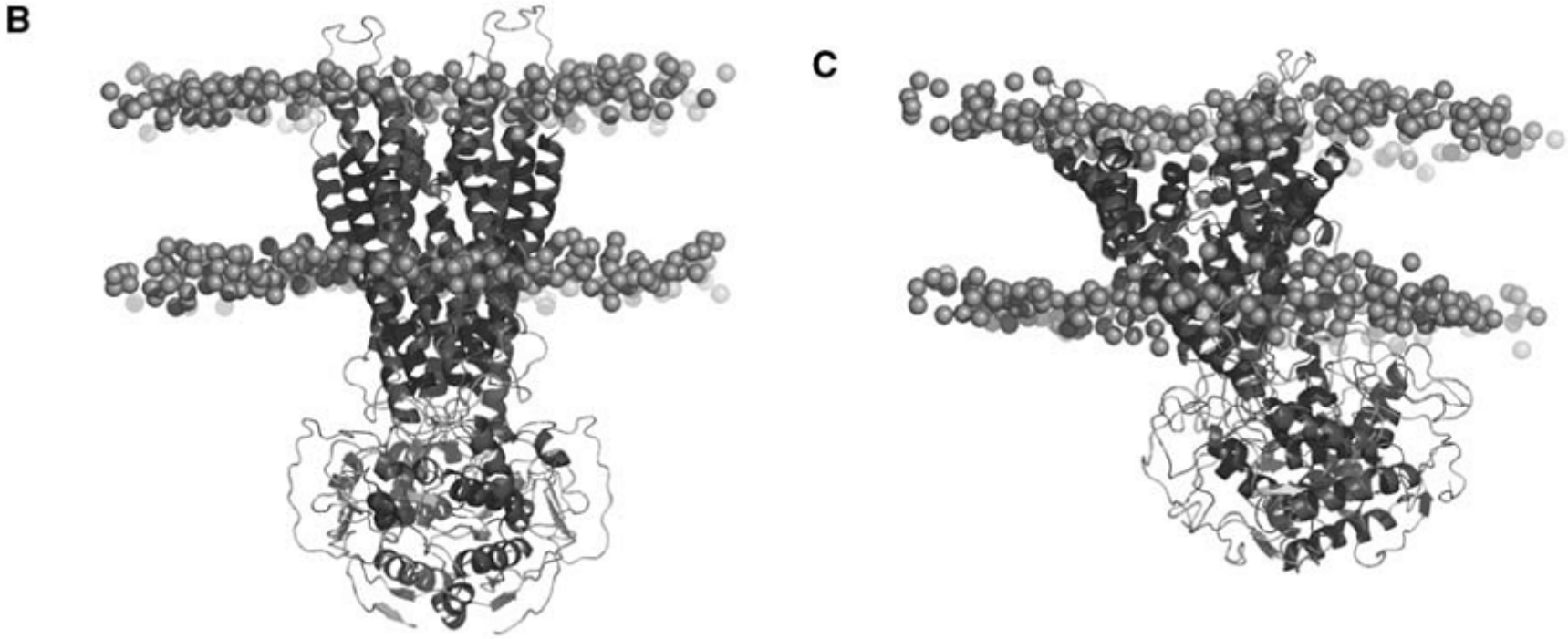
e.g. 2 bead or 4+ bead models for proteins
e.g. MARTINI CG force field

Membrane bilayer formation

CG simulation movie

Stability of simulations

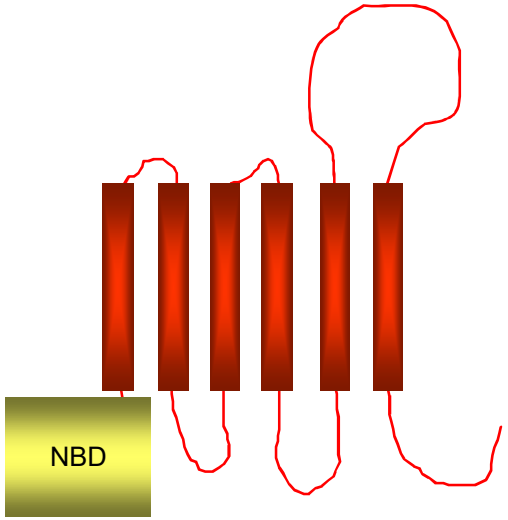
Eur Biophys J (2008) 37:403–409



0 ns **20 ns**

The complex example of ABCG2

- Homology modelling of the structure (2016)
- Investigating the effect of mutations using MD
- Effect of cholesterol on function
- Identification of drug binding sites
- Describing the transport process by MD and METAD

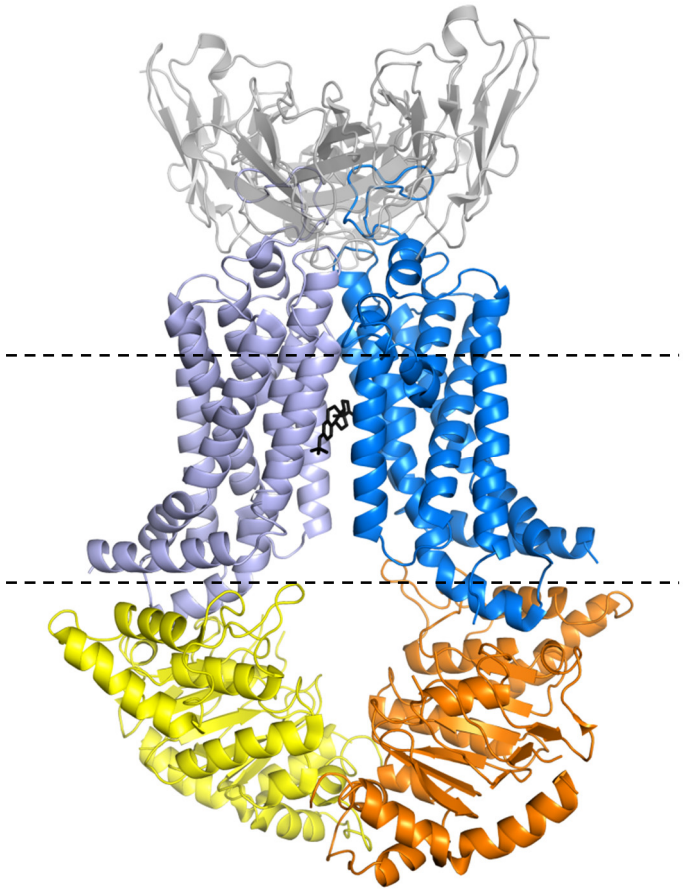


Importance of ABCG2

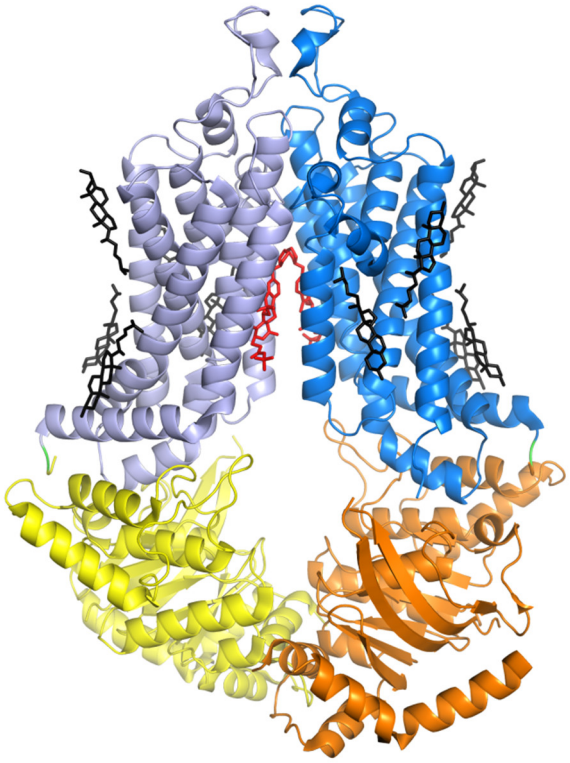
- Expressed in stem cells, tissue barriers, cancer cells
- Multidrug transporter of xenobiotics and endobiotics
  - antitumor agents
  - uric acid
- The Q141K variant exhibit decreased function and expression

ABCG2 structures

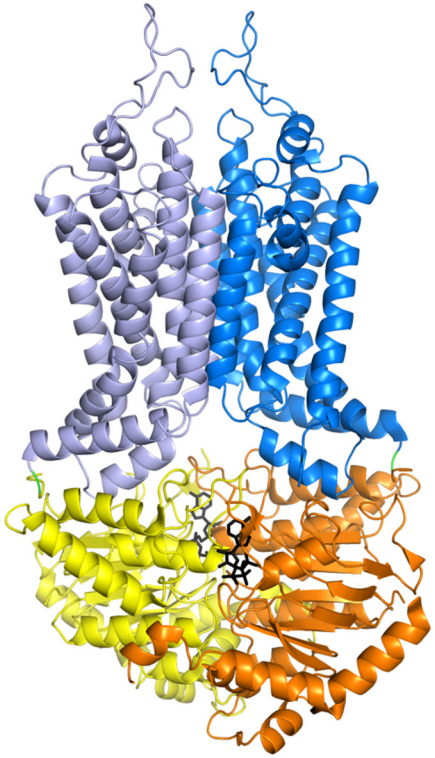
K. Locher, ETH, Zurich



6HCO

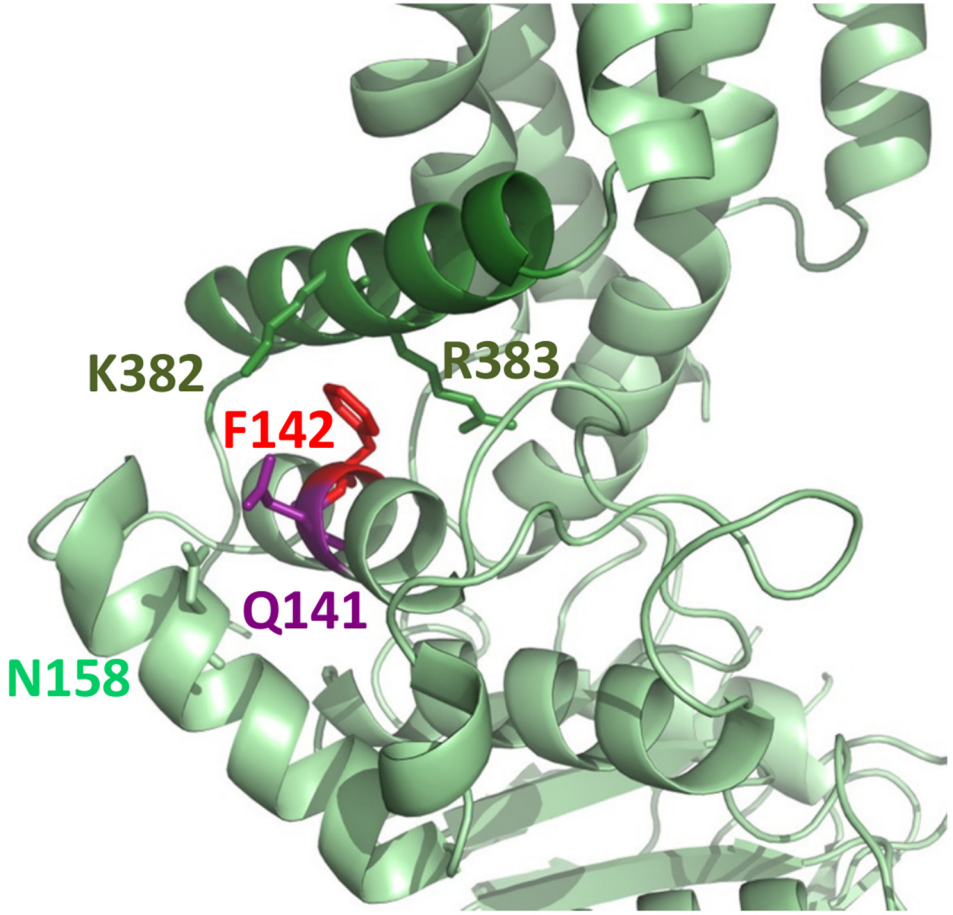


6HIJ



6HZM

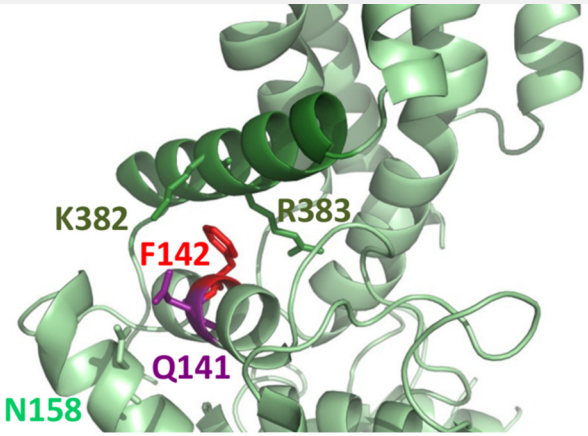
The Q141 position

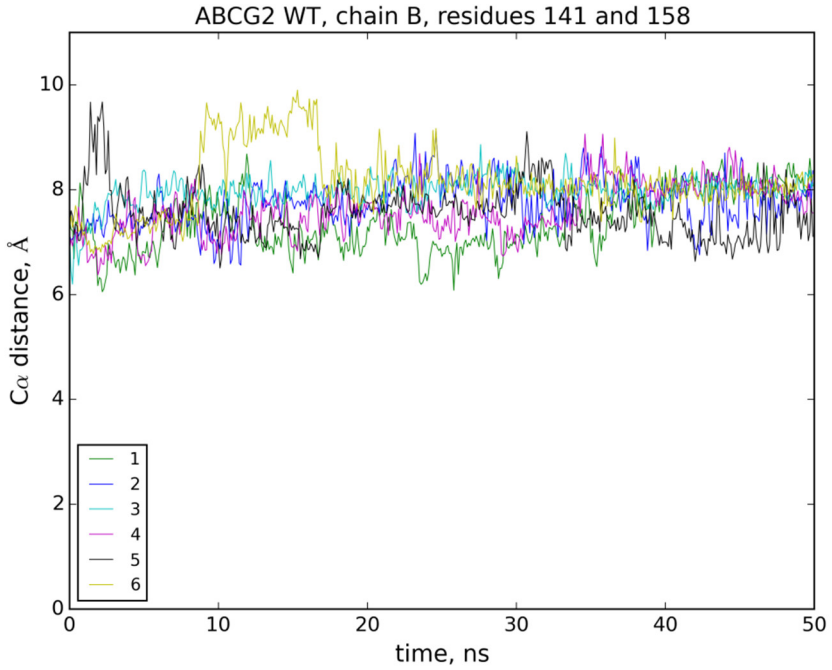


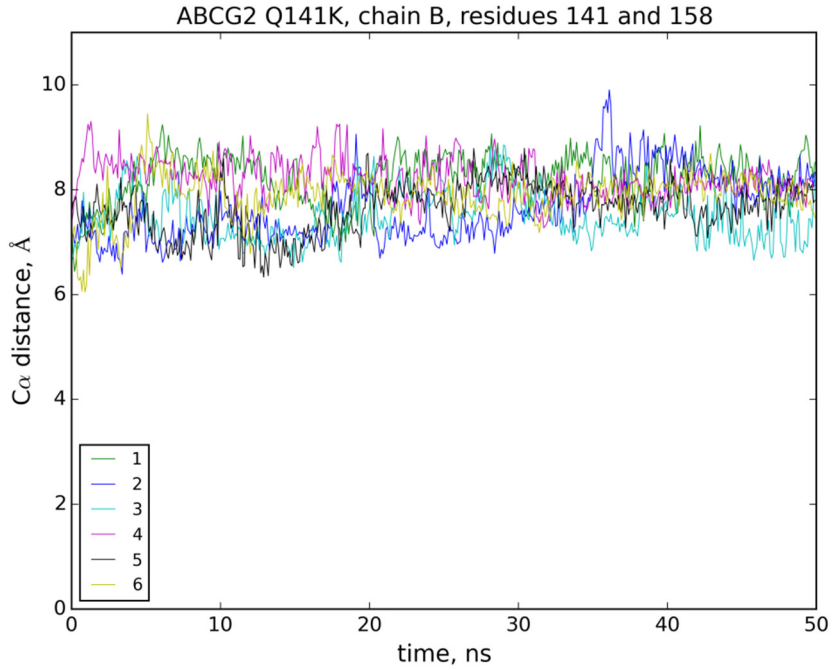
MD simulations

- The protein was embedded in POPC bilayer
- Optimizing the orientation of water, lipids, amino acid side chains:
  - energy minimization
  - equilibration
  - minimal backbone motions (position constrains)
- Production run
  - no constraints
  - 50 ns x 6 = 300 ns
- Comparing WT és mutants (e.g. Q141K, R482G)

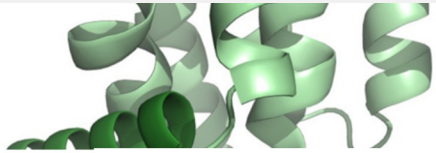
The effect of Q141K on protein dynamics

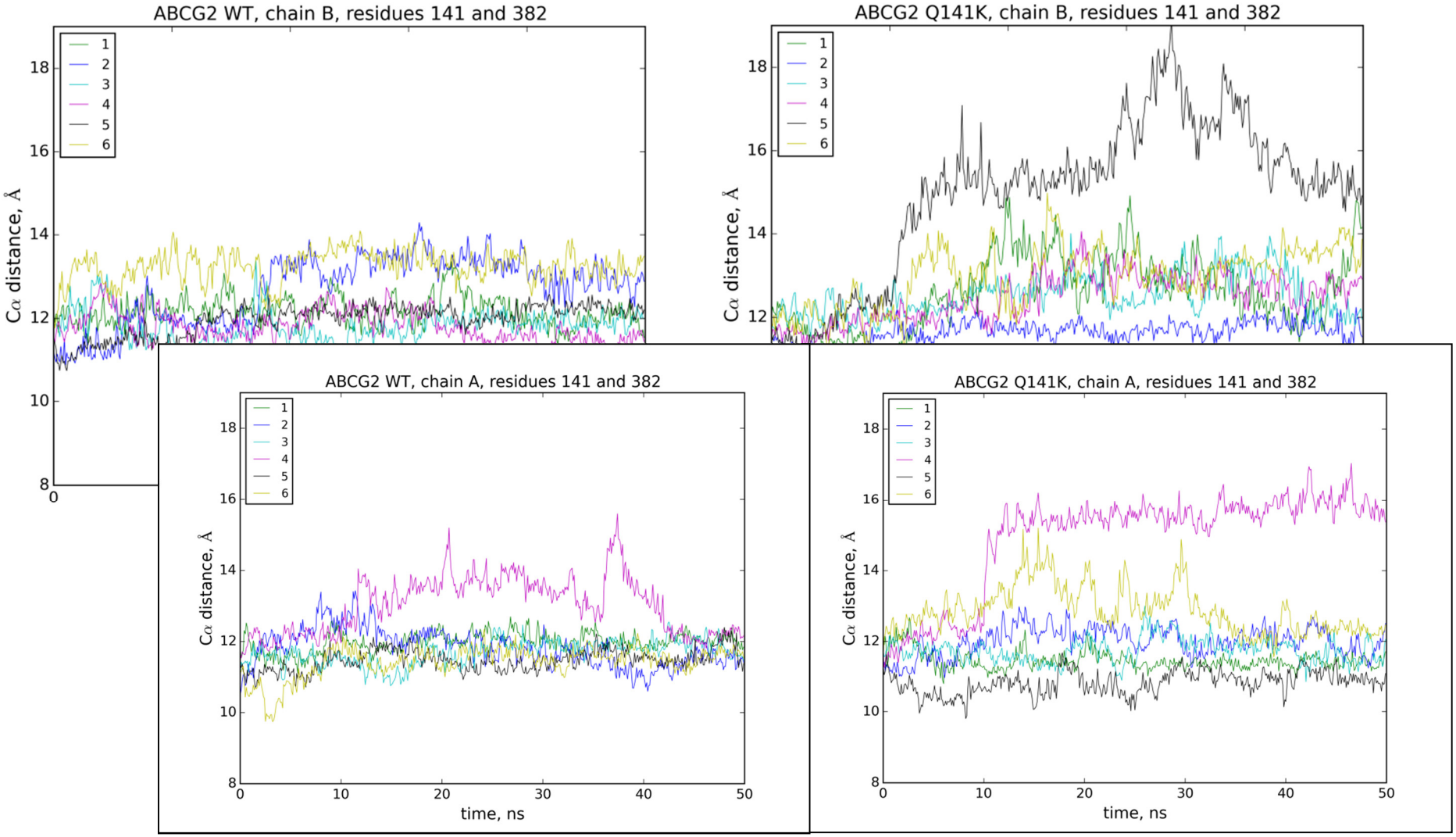




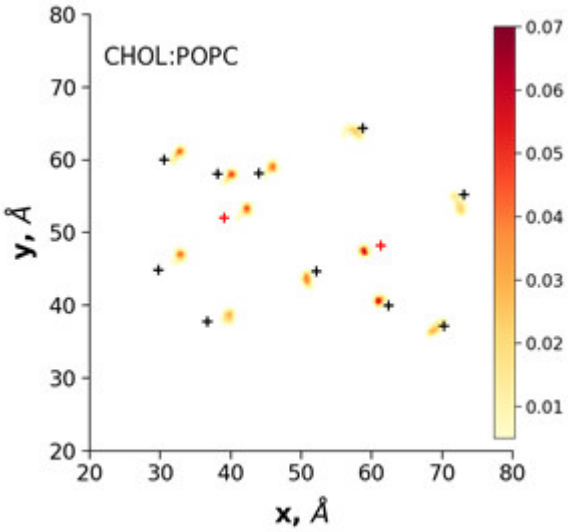


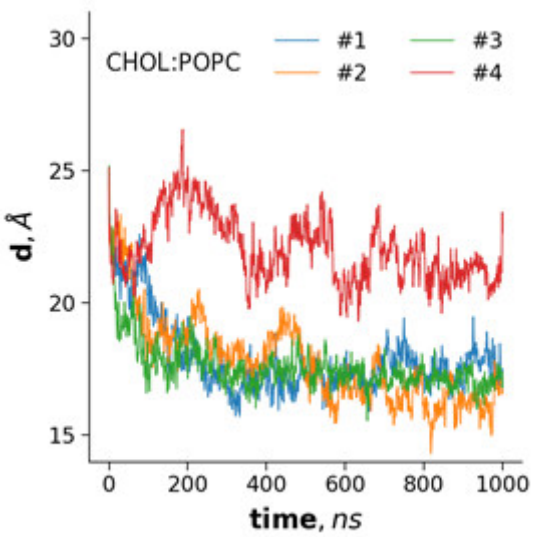
The effect of Q141K on protein dynamics

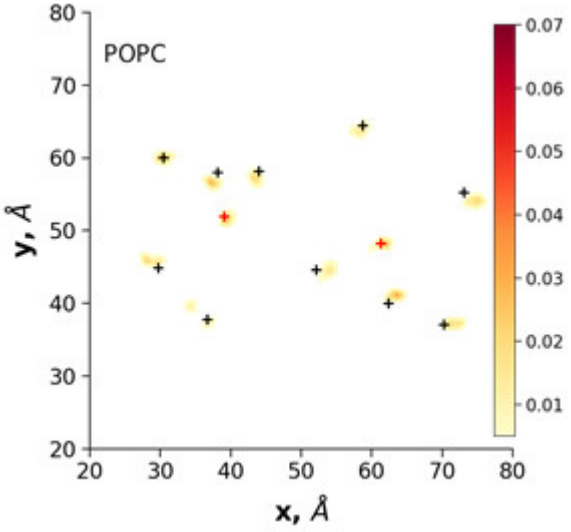


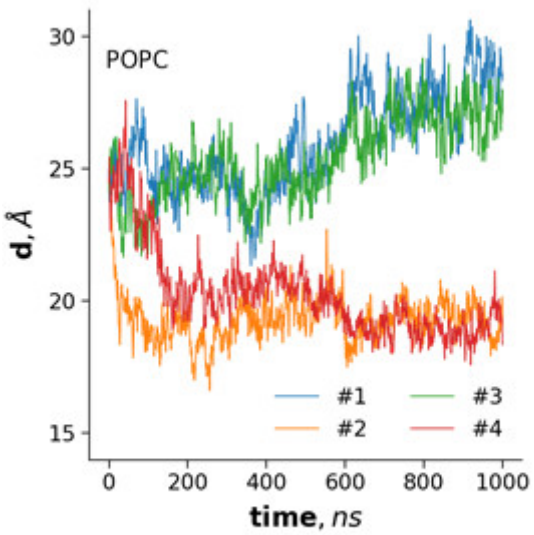


The effect of cholesterol on ABCG2





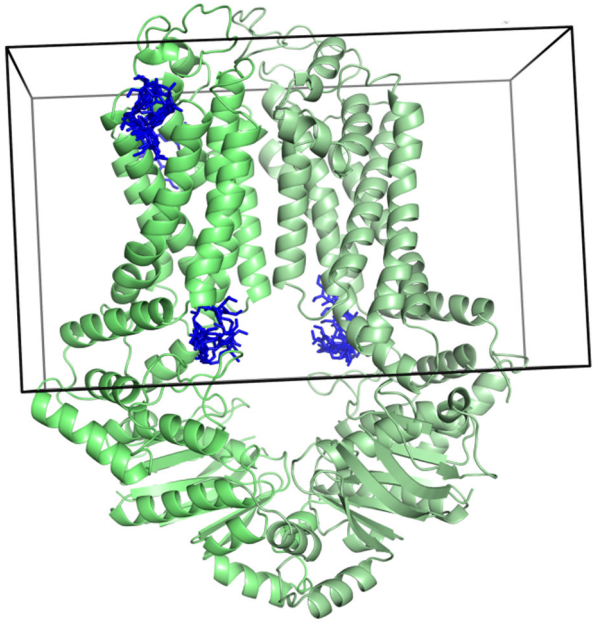




Identification of drug binding sites

in silico docking, AutoDock Vina

- Flexible ligand, non-flexible protein
- Several conformations from simulations
- Search space defined by a box

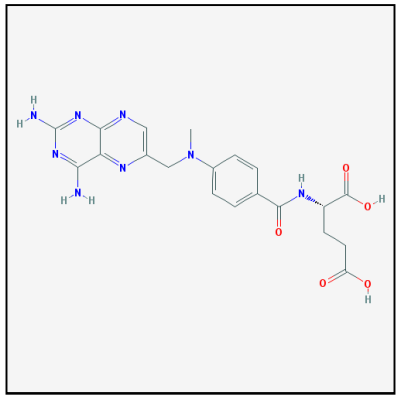


Identification of drug binding sites

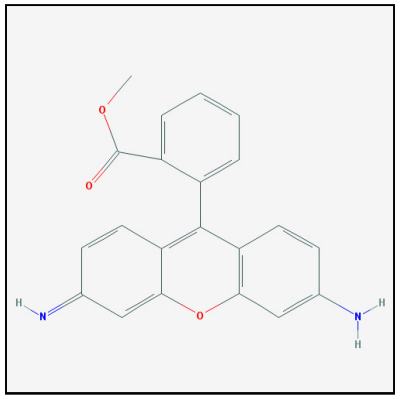
in silico docking, AutoDock Vina

- Flexible ligand, non-flexible protein
- Several conformations from simulations
- Search space defined by a box
- Everything gets docked; docking substrates and non-substrates

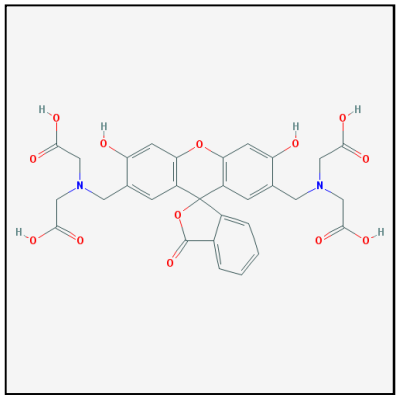
methotrexate



rhodamine123



calcein



Identification of drug binding sites

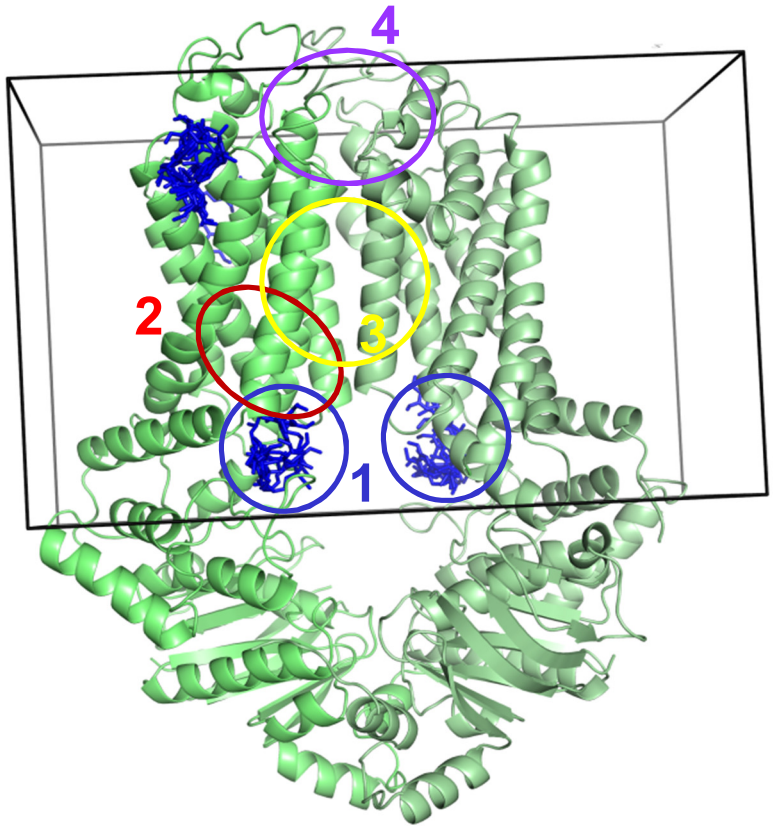
in silico docking, AutoDock Vina

- Flexible ligand, non-flexible protein
- Several conformations from simulations
- Search space defined by a box
- Everything gets docked; docking substrates and non-substrates
- (6 ABCG2 conformations) * (3 parallel dockings) * (20 poses) *
  (25 substrates + 14 non-substrates)
- Clustering poses

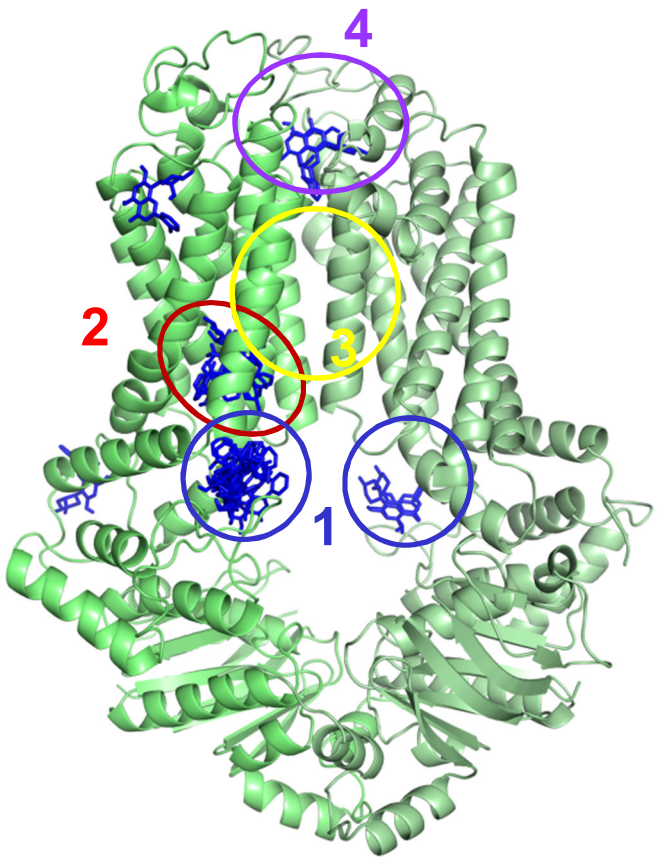
Identification of drug binding sites

in silico docking, AutoDock Vina

verapamil

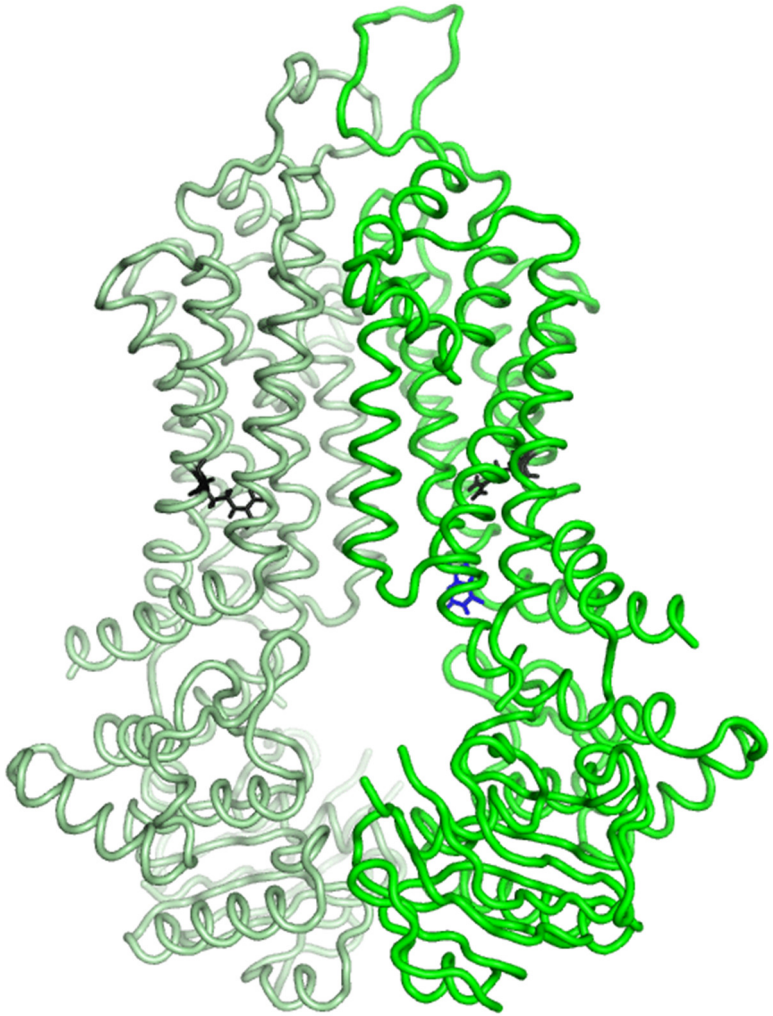


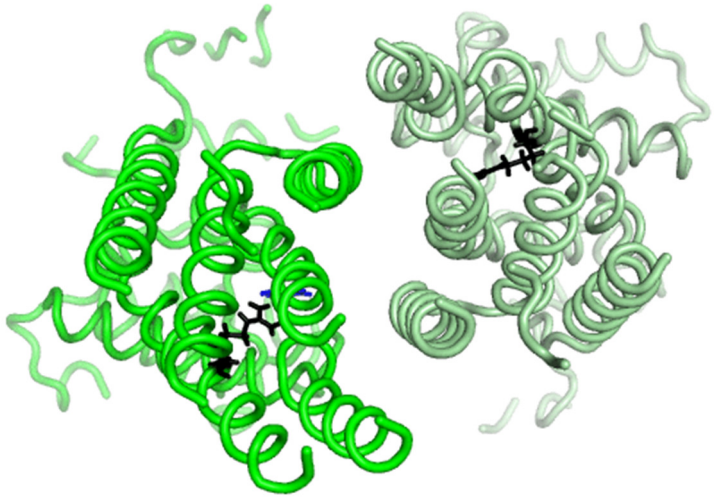
flavopiridol



Describing the transport using MD

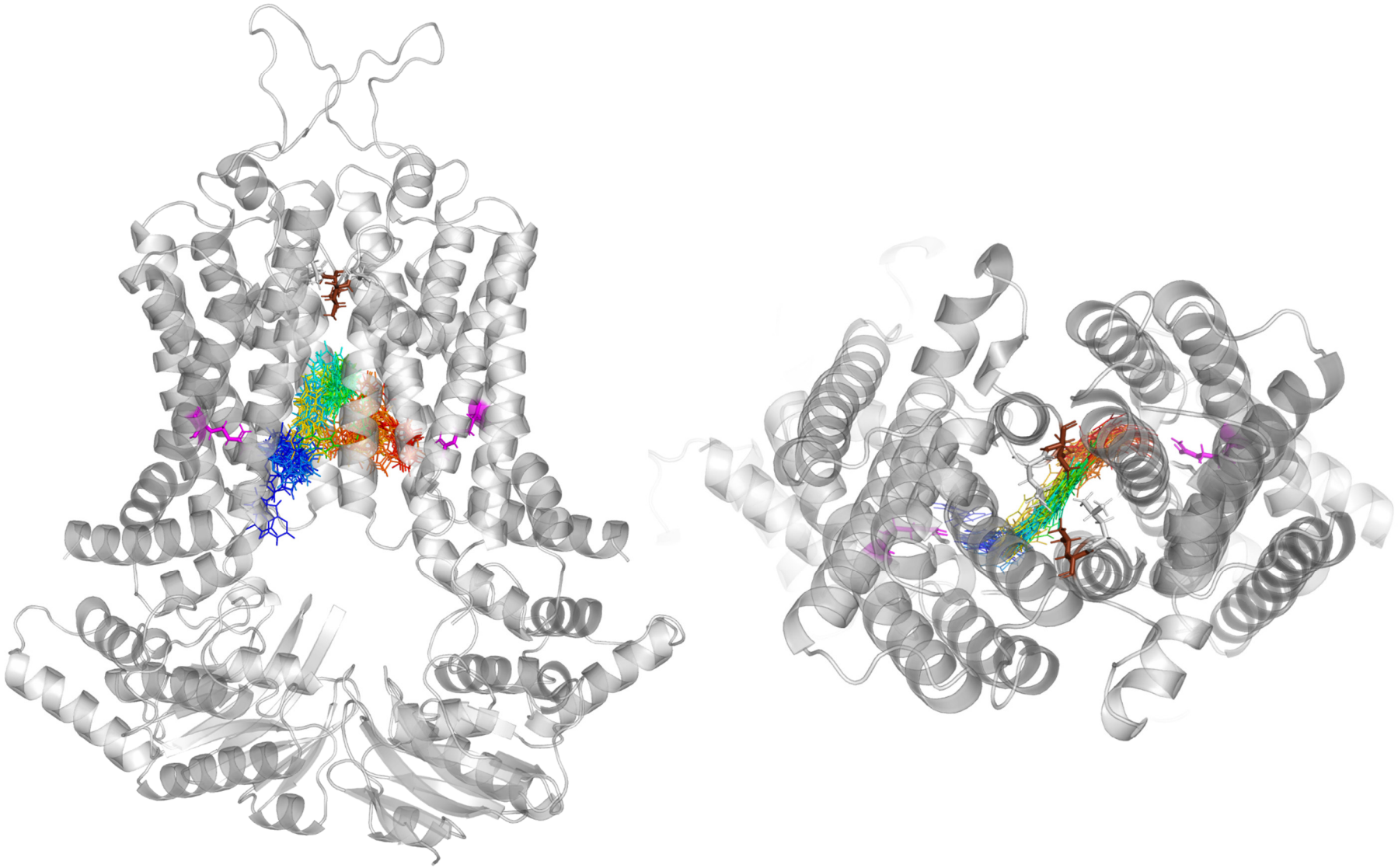
equilibrium simulations, uric acid molecules



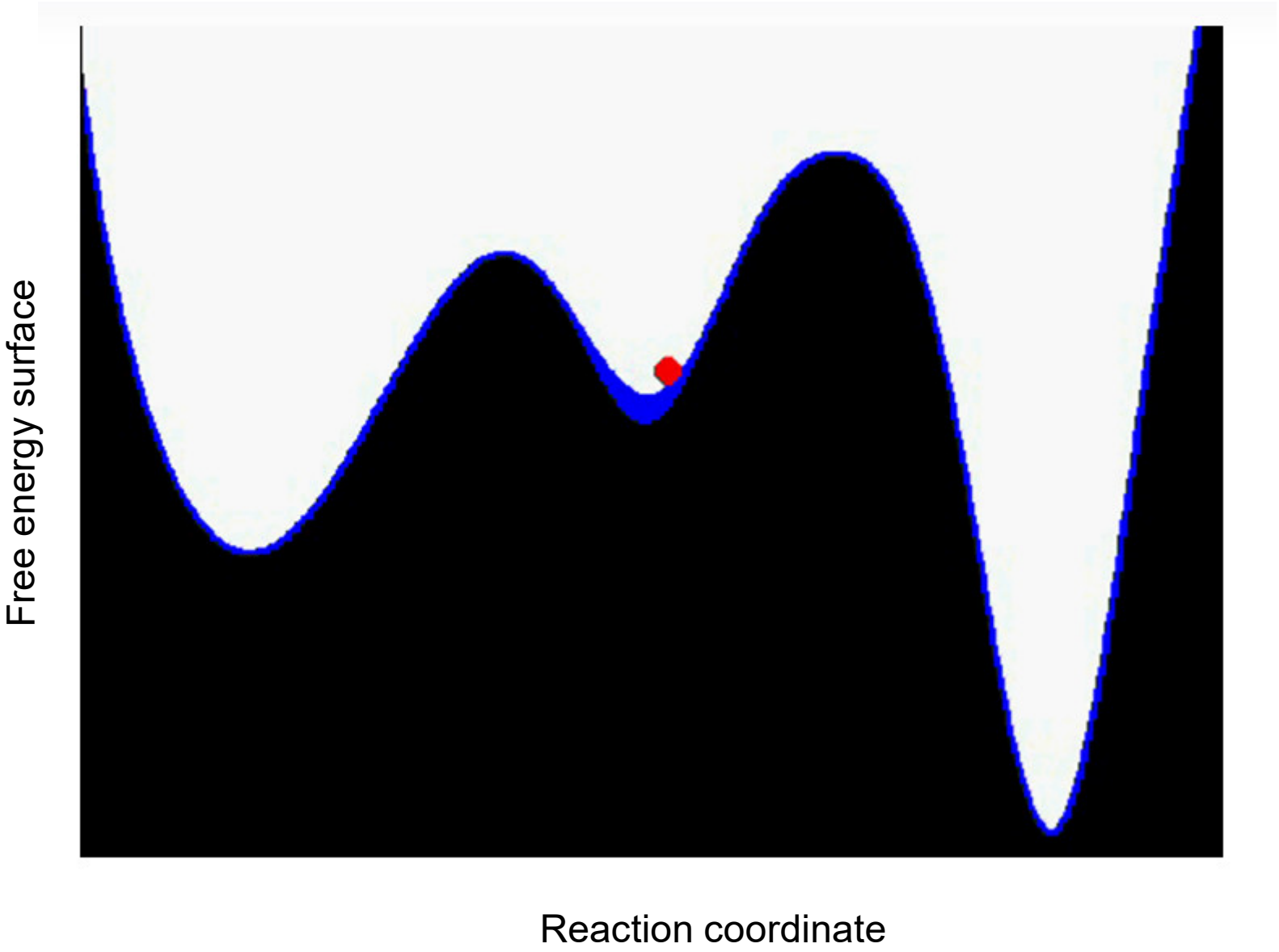


Describing the transport using MD

equilibrium simulations, uric acid molecules

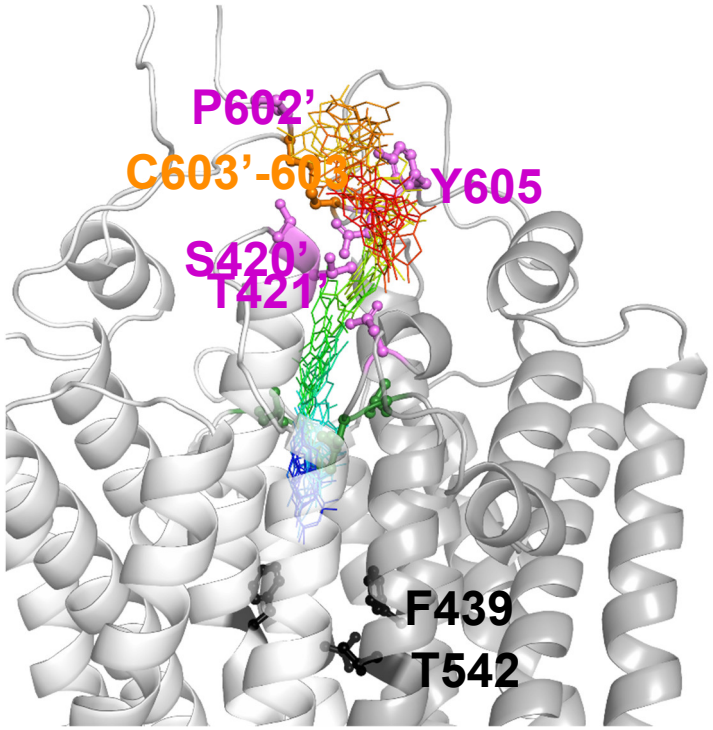


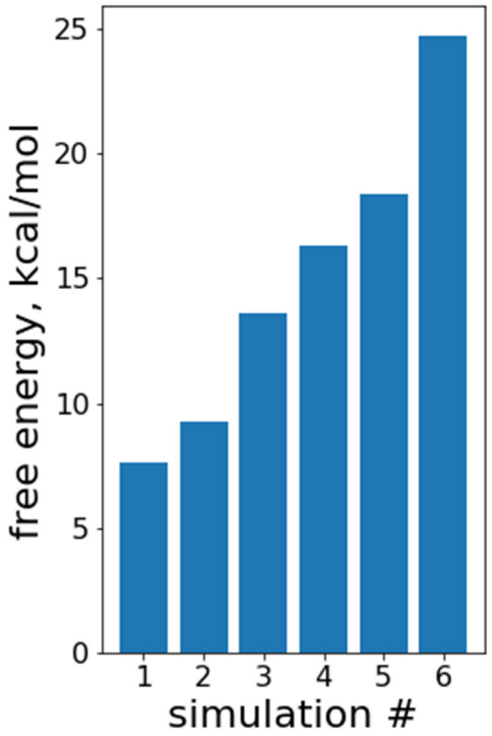
Exploring substrate transport by biased MD simulations



Describing the transport using MD

metadynamics simulations, uric acid molecule

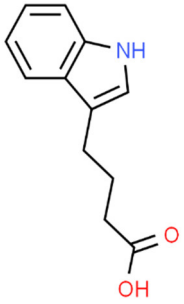




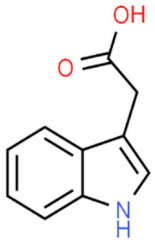
Plant AtABCG36

metadynamics simulations

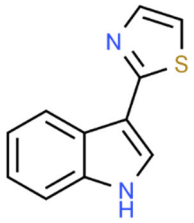
IBA (indole-3-butyric acid)

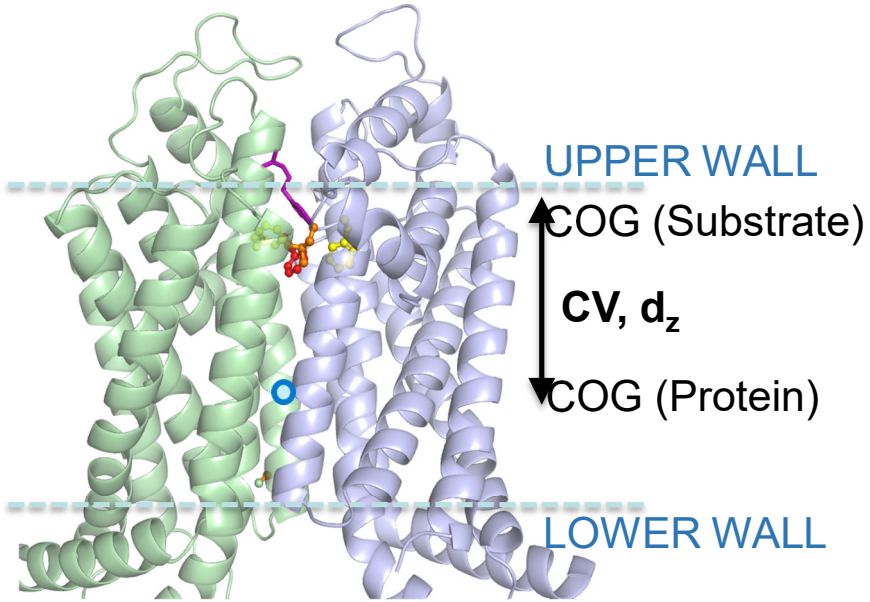


IAA (indole-3-acetic acid)



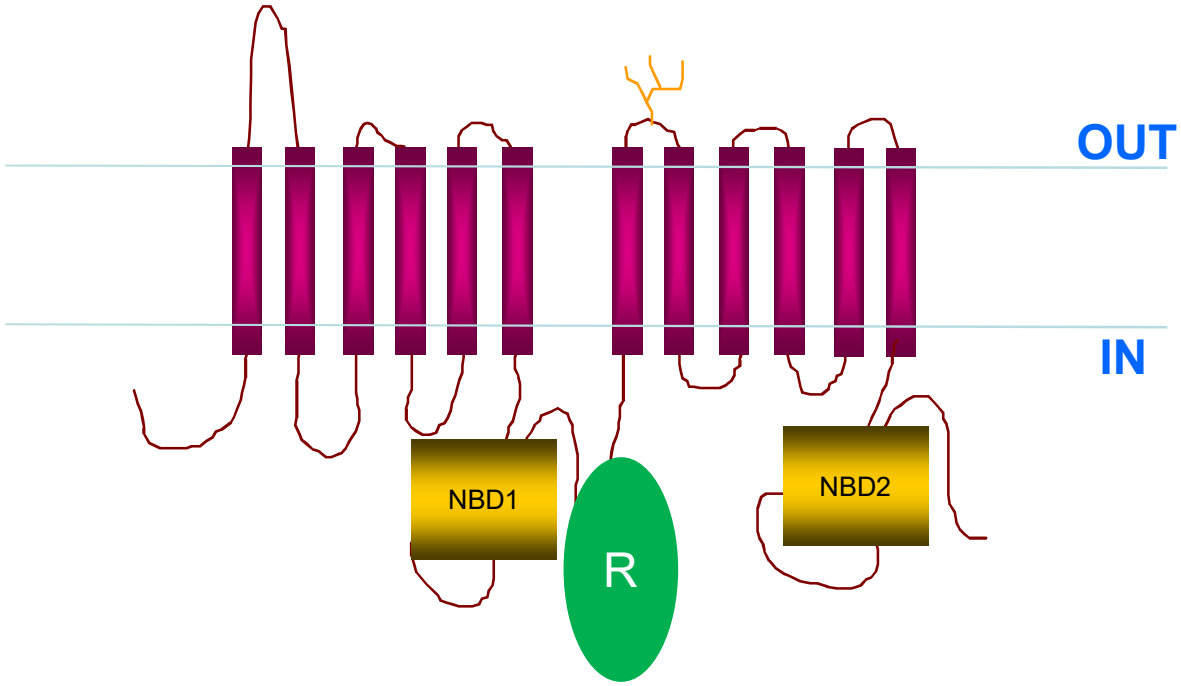
CAM (camalexin)





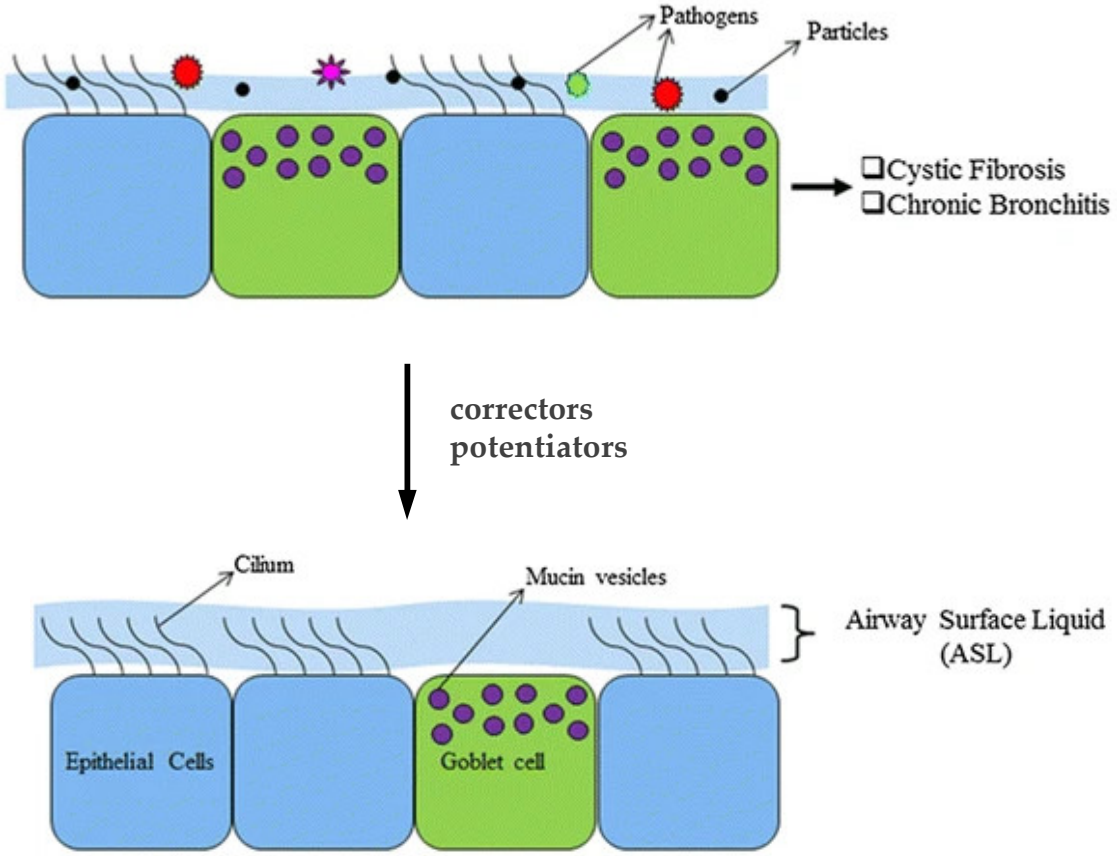
ABCC7/CFTR

Cystic Fibrosis Transmembrane Conductance Regulator



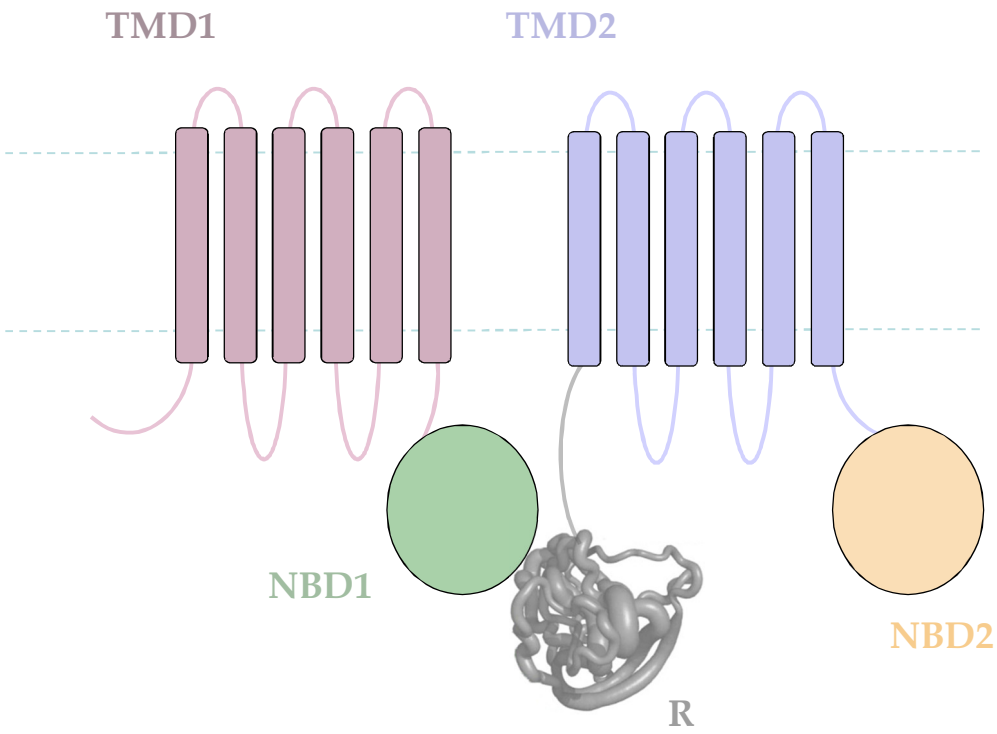
Cystic fibrosis (CF)

Ghosh, Boucher, Tarran, CMLS 2015



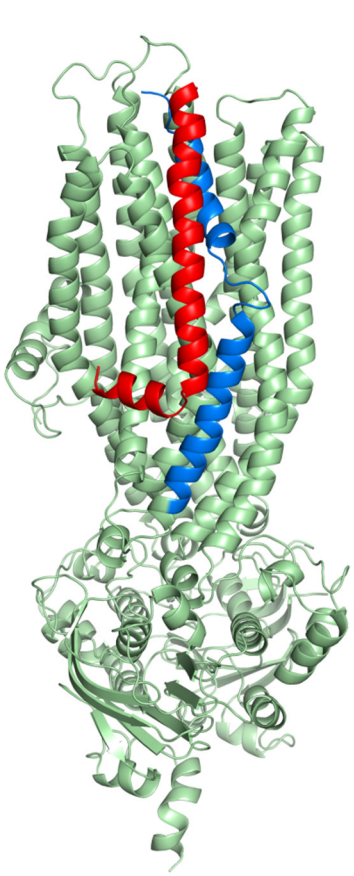
ABCC7/CFTR

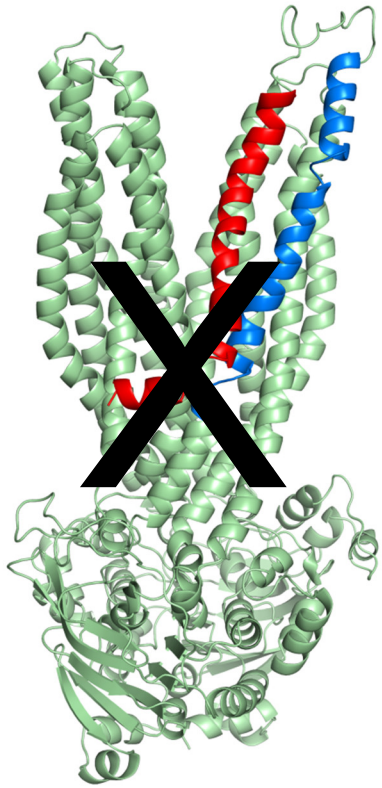
Cystic Fibrosis Transmembrane Conductance Regulator

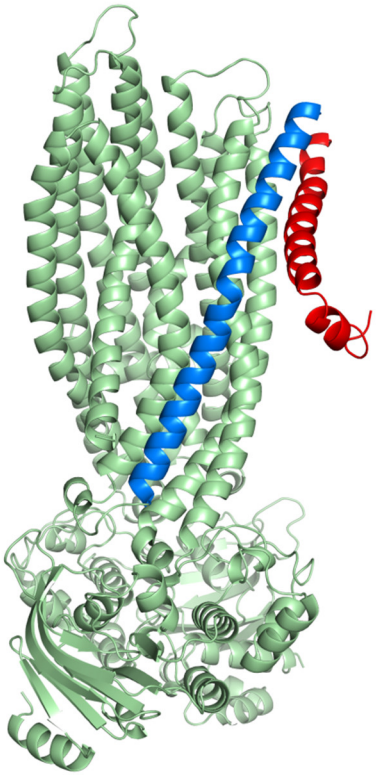


Full-length, ATP-bound CFTR structures

Cryo-EM revolution





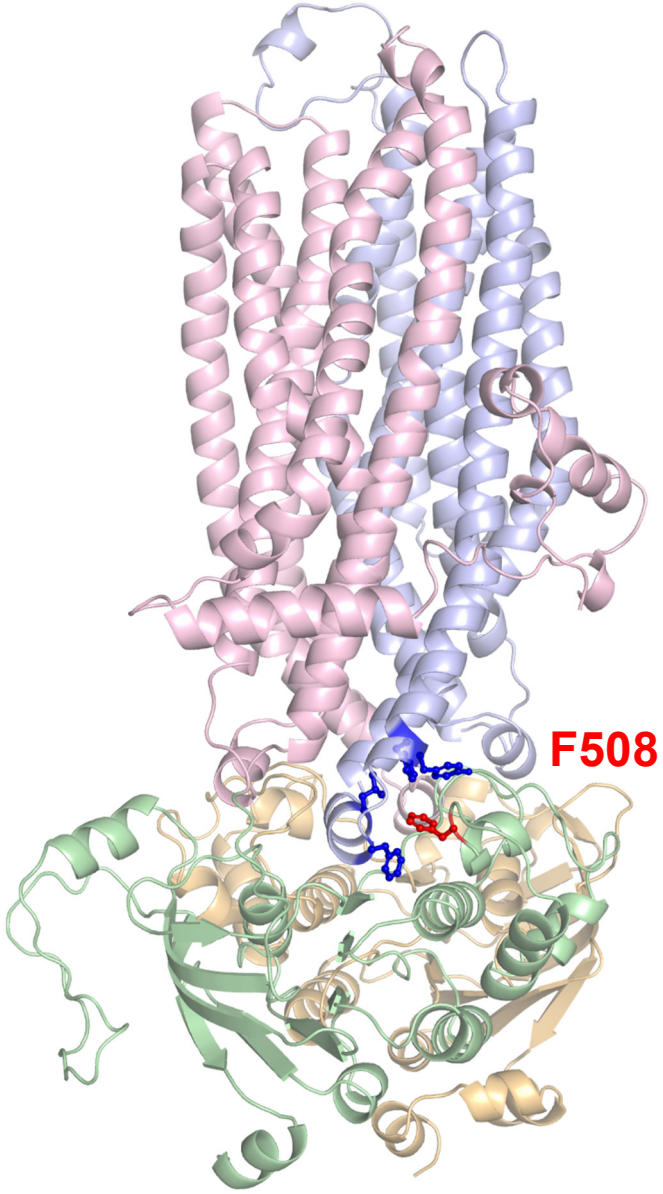


Zhang et al. (2017) Cell 170: 483-491.e8
PDBID:5W81

Bob Ford
University of Manchaster, UK

J. Fay, Jack Riordan
UNC, Chapel Hill, USA

ΔF508 mutation



Many experimental and computational studies

Domain folding
Domain stability
Domain-domain assembly

Transmission of the consequence of a mutation; allosteric propagation of alterations in dynamics

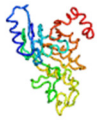
NBD1 folding

Padanyi *et al.* Cell Mol Life Sci. 2022

To learn folding

computationally

experimentally highly challenging

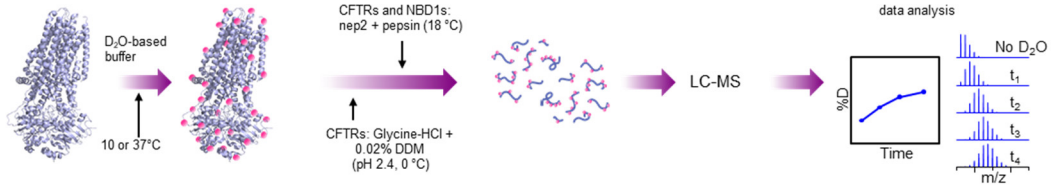


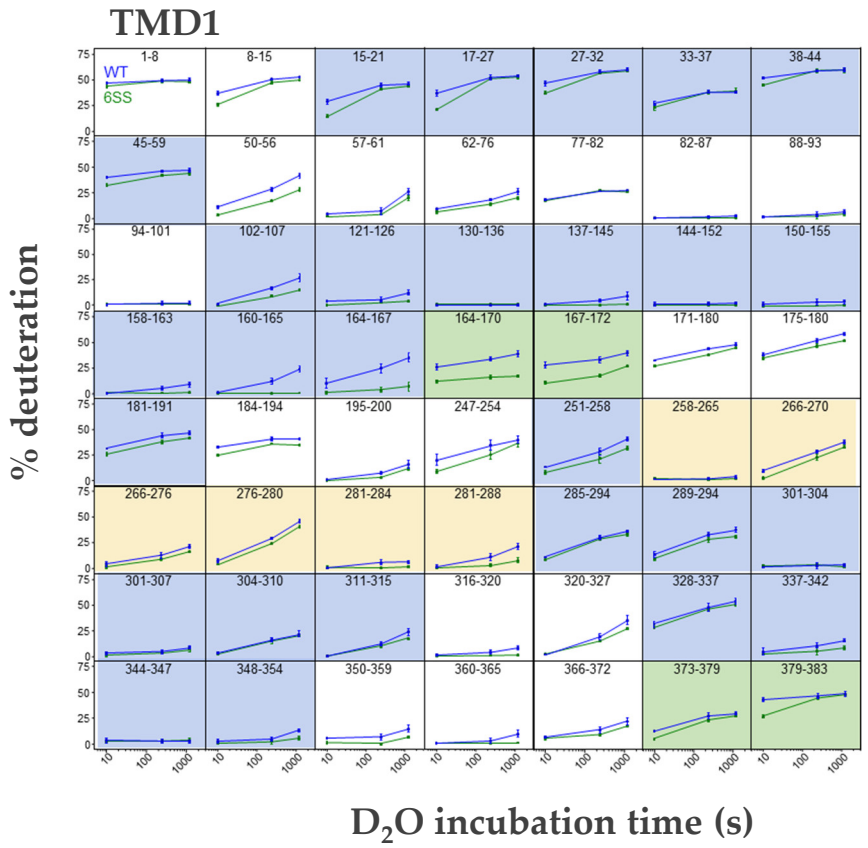
Unfolding

pulling molecular dynamics (MD) simulations

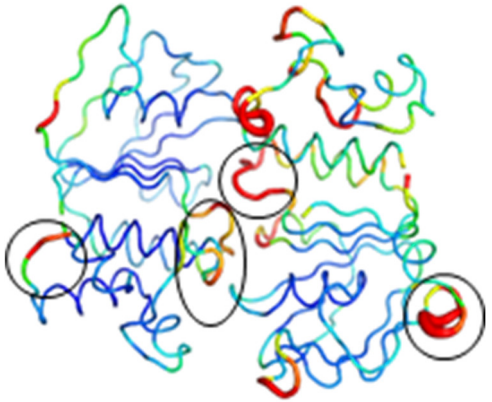
atomic force microscopy (AFM) experiments

Folding of the full length CFTR

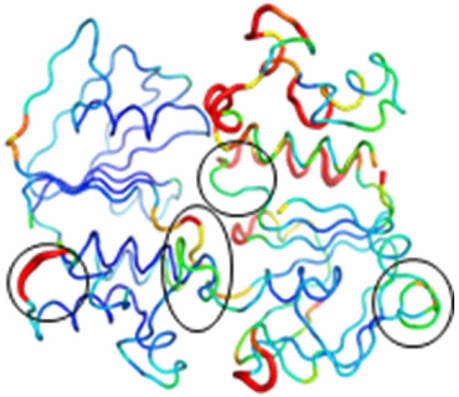




F508G

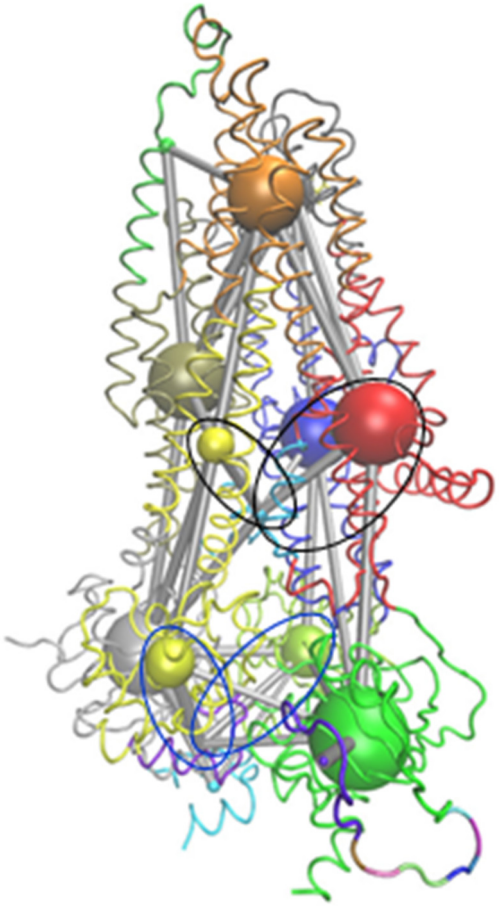


F508G-6SS

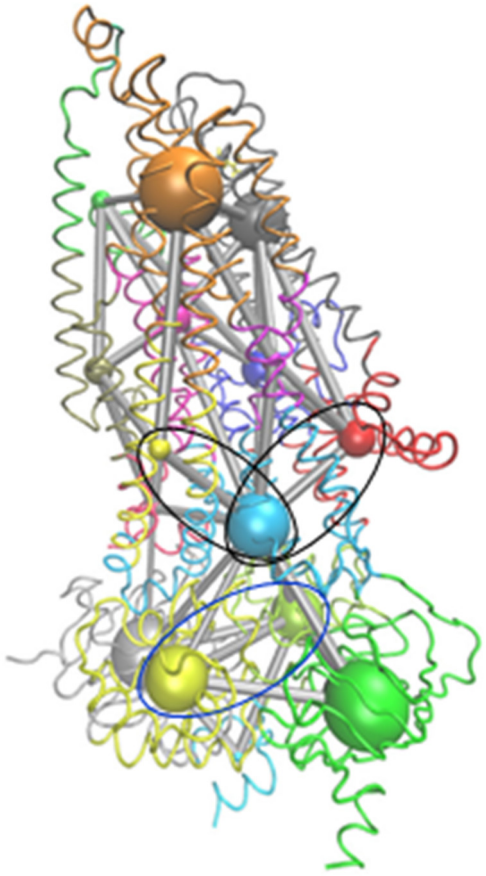


Allosteric stabilization of TM_{IC}

F508G



F508G-6SS



Identification of the chloride permeation pathway

Farkas *et al.* Cell Mol Life Sci. 2019