

Fehérjék szerkezete és dinamikája

Hegedűs Tamás

tamas@hegelab.org



**MTA-SE Molekuláris Biofizikai Kutatócsoport
SE Biofizikai és Sugárbiológiai Intézet**



Témák

- **Bevezetés – szimulációk és a fehérje dinamika jelentősége**
- **Fehérjék szerkezetének jellemzése**
- **Fehérjék dinamikájának modellezése**
- **Fehérjék feltekeredésének szimulációja**
- **Dinamika szerepe a ligandfelismerésben**
- **Informatikai eszközök – biológus szempontból**
- **Gén és egyéb hálózatok**

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.

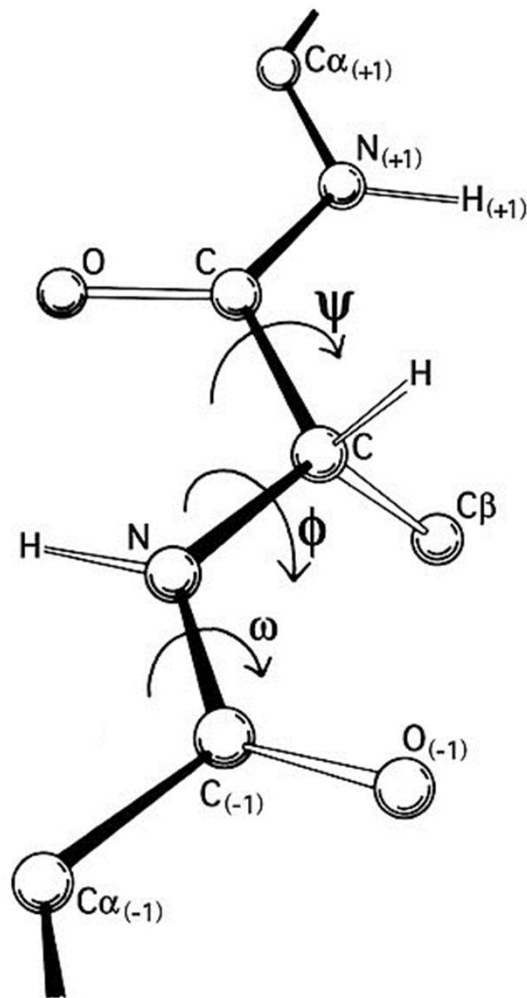
**Kísérletes módszerek általában nem
szolgáltatnak közvetlen információt az
atomi szintű történésekről.**

Pl. NMR és MD - igen

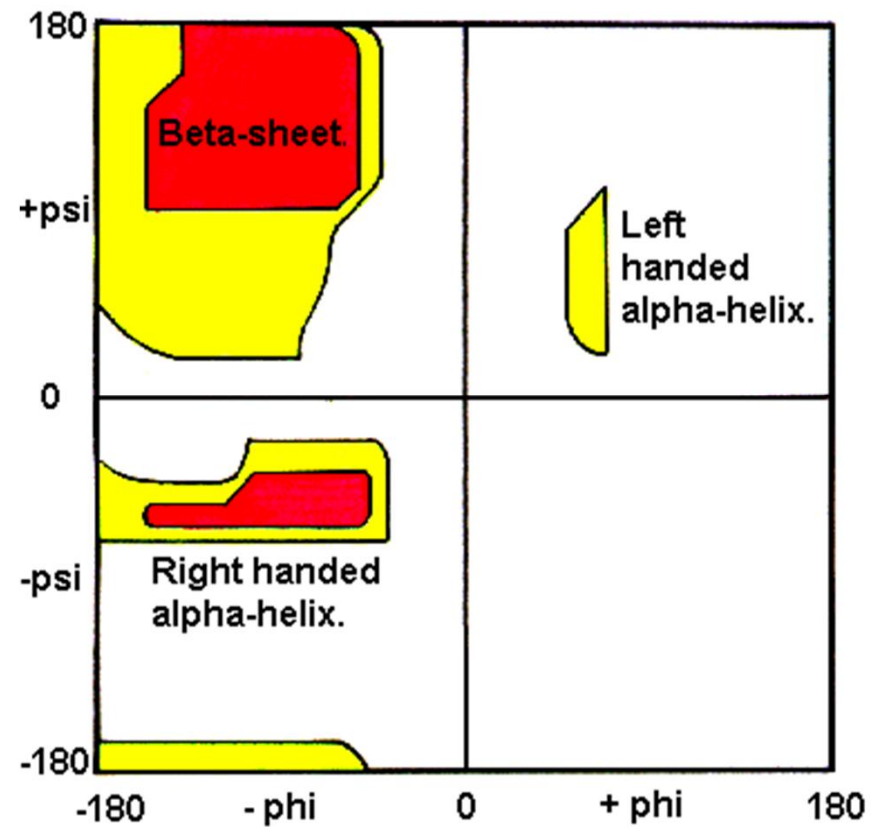
Mai témák

- **Bevezetés – szimulációk és a fehérje dinamika jelentősége**
- **Fehérjék szerkezetének jellemzése**
 - Másodlagos szerkezeti mintázatok jóslása
 - Rendezetlen fehérjék
 - A harmadlagos és negyedleges szerkezet

Másodlagos szerkezeti elemek



The Ramachandran Plot.



Másodlagos szerkezeti elemek predikciója

Megoldott szerkezetekből minden aminosavra meghatározott
helix, β -redő, *coil* formáló hajlamból 60 %

Ezek kombinálása szekvenciák illesztésével 70-80 %

Megvalósítási lehetőségek:

- neurális hálózatok,
- support vector machines,
- rejtett Markov modellek, stb.

Megbízhatósági érték minden pozícióra

GOR4, HNN, Prof, JPred/JNet

Rendezetlen fehérjék I.

Intrinsically Disordered Proteins

Becslések alapján a fehérjéknek akár 25 %-a rendezetlen lehet.

Komplexitással nő a rendezetlen fehérjék aránya

Az emberi fehérjék felében van min. 30 a.a. hosszú rendezetlen szakasz

Nem teljesen random.

Strukturálisan igen flexibilisek.

Nincs kompakt globuláris hajtogatódás, reziduális szerkezet.

**Megdőlt a paradigma,
mely szerint csak jól definiált 3D szerkezethez kapcsolható fehérje funkció.**

Rendezetlen fehérjék II.

Miért jó?

Specifikus és adaptálódó
Rendezetlen/rendezett reverzibilis átmenete
Nagy kötőfelület
Gyors kötés

Mire jó?

Entrópikus lánc:
Effektor:
Scavangers:
Összeszerelődés:
Bemutató felület:

K⁺ csatorna inaktiválása
peptid inhibitorok
kazein
calmodesmon, F-aktin
foszforilációs és proteolitikus helyek

Rendezetlen fehérjék III.

DisProt adatbázis: <http://www.disprot.org>

K. Dunker – Indiana University

Tompa Péter, Kalmár Lajos, Dosztányi Zsuzsa – MTA Enzimológiai Intézet

A rendezetlenség jóslása

↙

Tanuló algoritmusok
PDB-ben előforduló
rendezetlen fehérjék szekvenciája alapján
(nincs bennük sok hidrofób a.a.)

Disopred2

↘

Kölcsönhatási energiák becslése

IUPred.enzim.hu

IUPred

Fizikai alapok!

Ha van szerkezet:

$$E_{\text{calculated}} = \sum_{i,j} M_{ij} C_{ij}$$

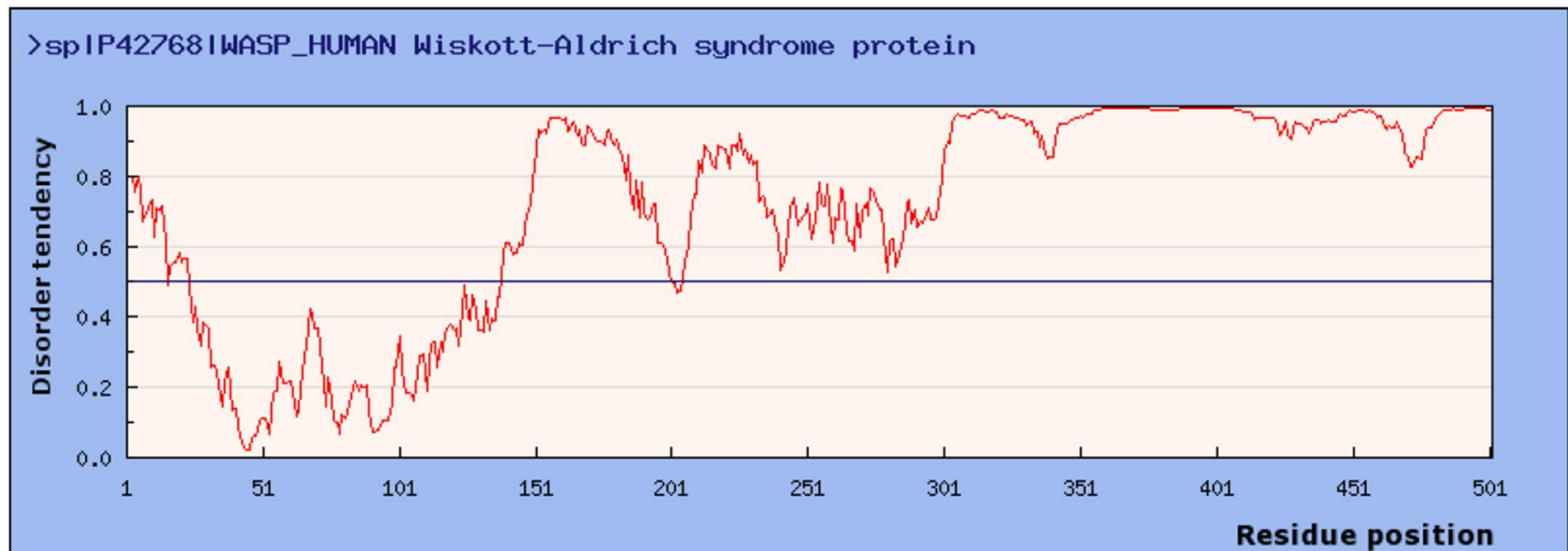
Ha csak szekvencia van:

$$E_{\text{estimated}} = L \sum_{i,j} P_{ij} f_i f_j$$

Egy aminosav rendezetlensége:

$$E_j^k = \sum_{i=1}^{20} P_{ij} f_i^k (w_0)$$

IUPred kimenete



Harmadlagos szerkezet jóslása

Ab initio folding

- **CASP** (Critical Assessment of Techniques for Protein Structure Prediction)
- kényszerfeltételek kísérletekből

Homológia modellezés

- feltételezi: konzervált szekvencia == konzervált struktúra
- > 30% hasonlóság
- a szekvencia-illesztés jósága a legmeghatározóbb

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	-1	0	0	-3	0	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	

Basic Local Alignment Search Tool (BLAST)

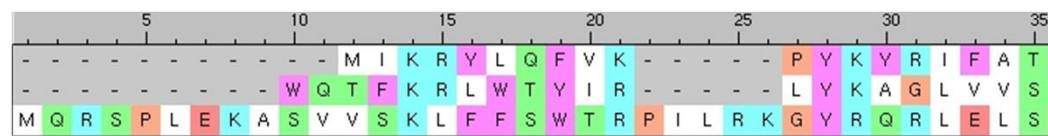
CLUSTAL W (1.83) multiple sequence alignment

Alignement – pl. ClustalW

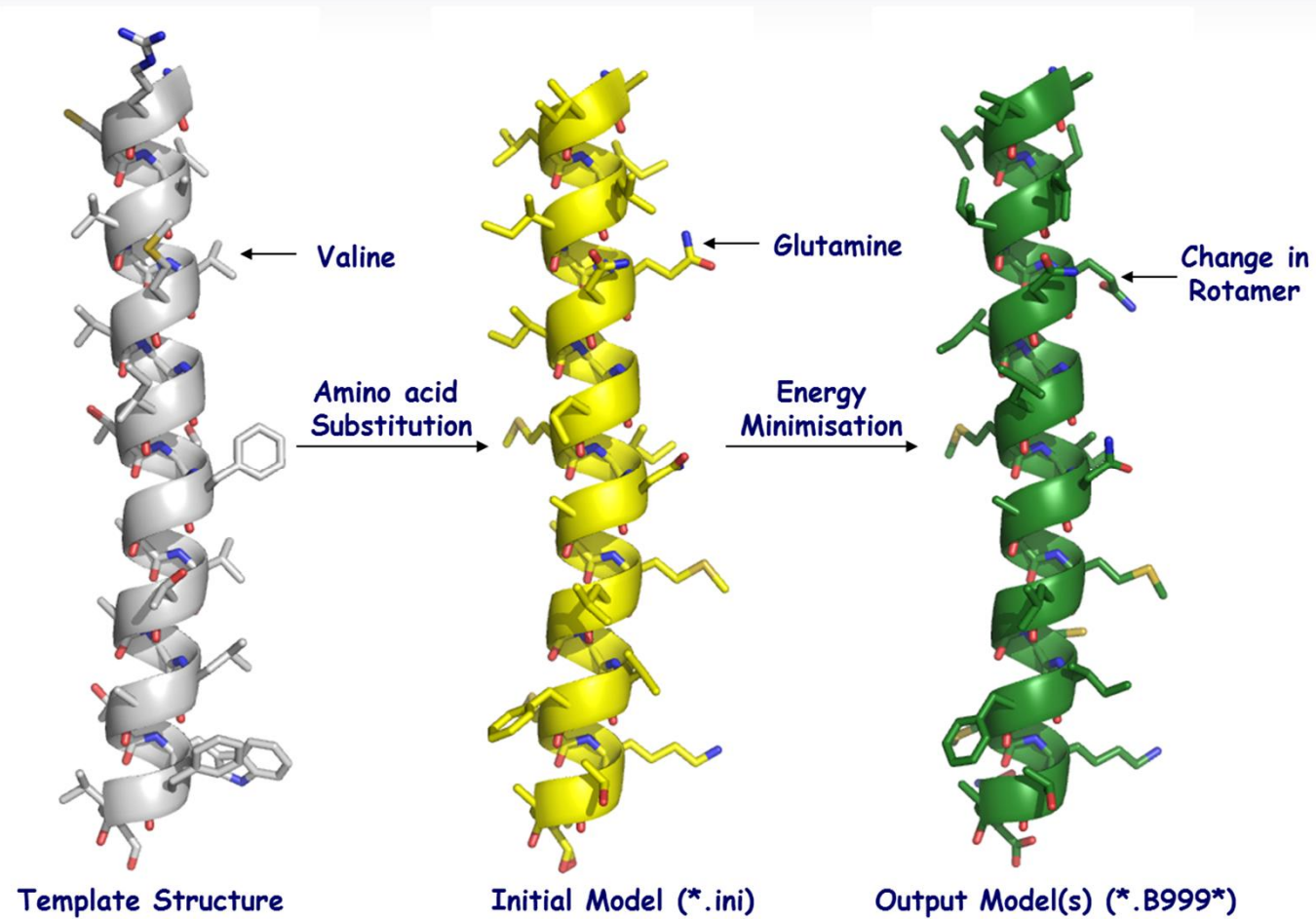
```

2HYD      -----MIKRYLQFVK-----PYKYRIFATIIVGIIKFGIPMLIP
3B5X      -----WQTFKRLWTYIR-----LYKAGLVVSTIALVINAAADTYMI
CFTR_HUMAN MQRSPLEKASVSVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLS
              *      :      :      *      :      :      *      :      :

```



Homológia modellezés II.

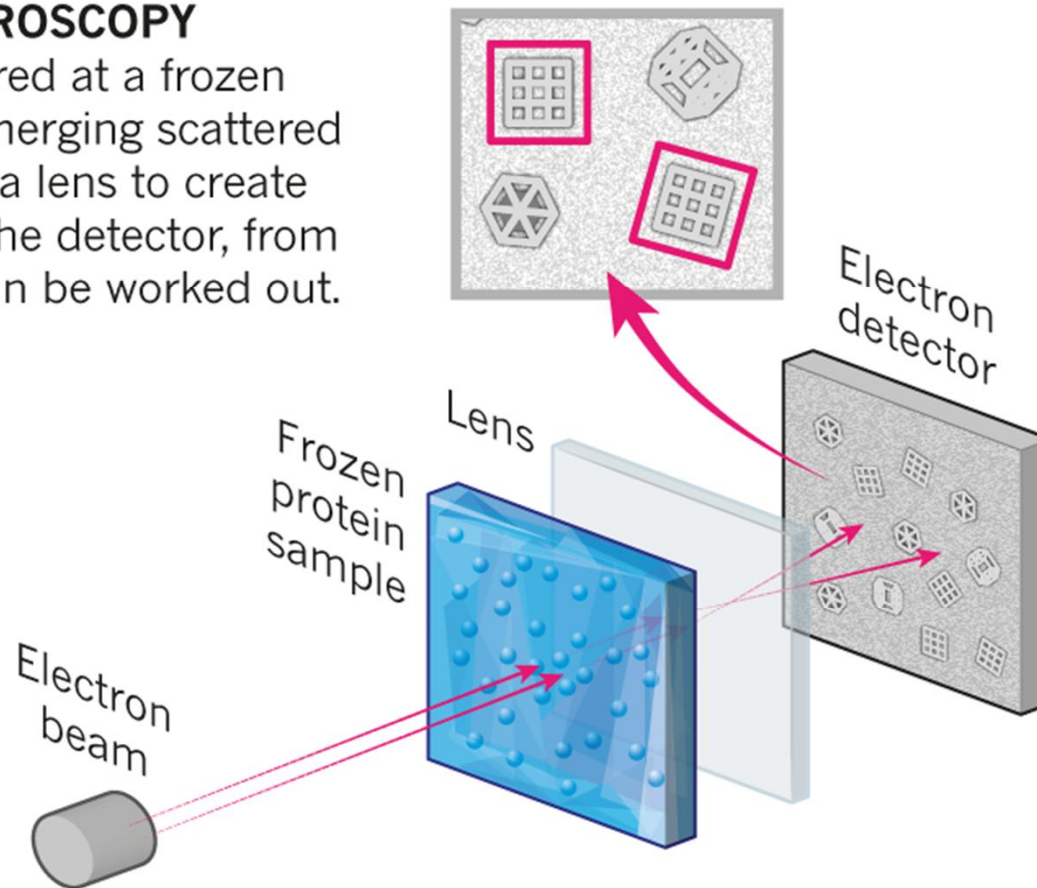


Szerkezet meghatározás – „single particle”

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.



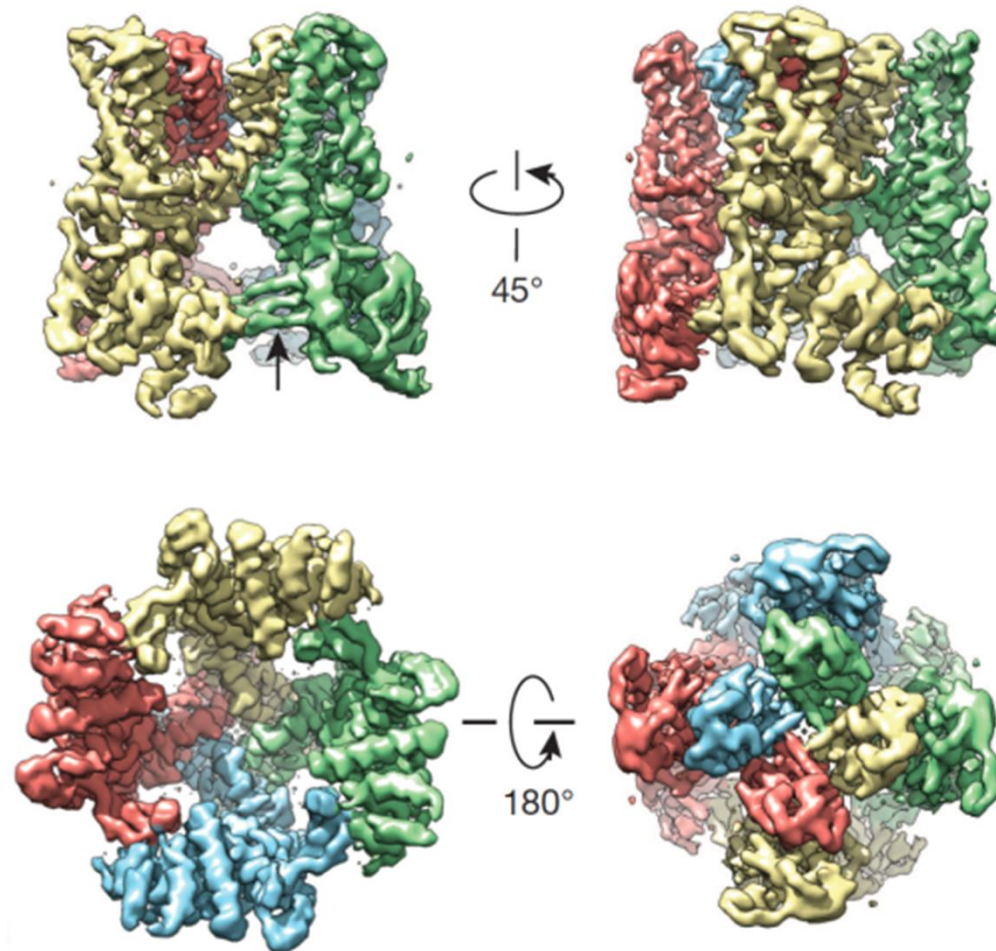
© nature

Ewen Callaway, Nature | News Feature

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

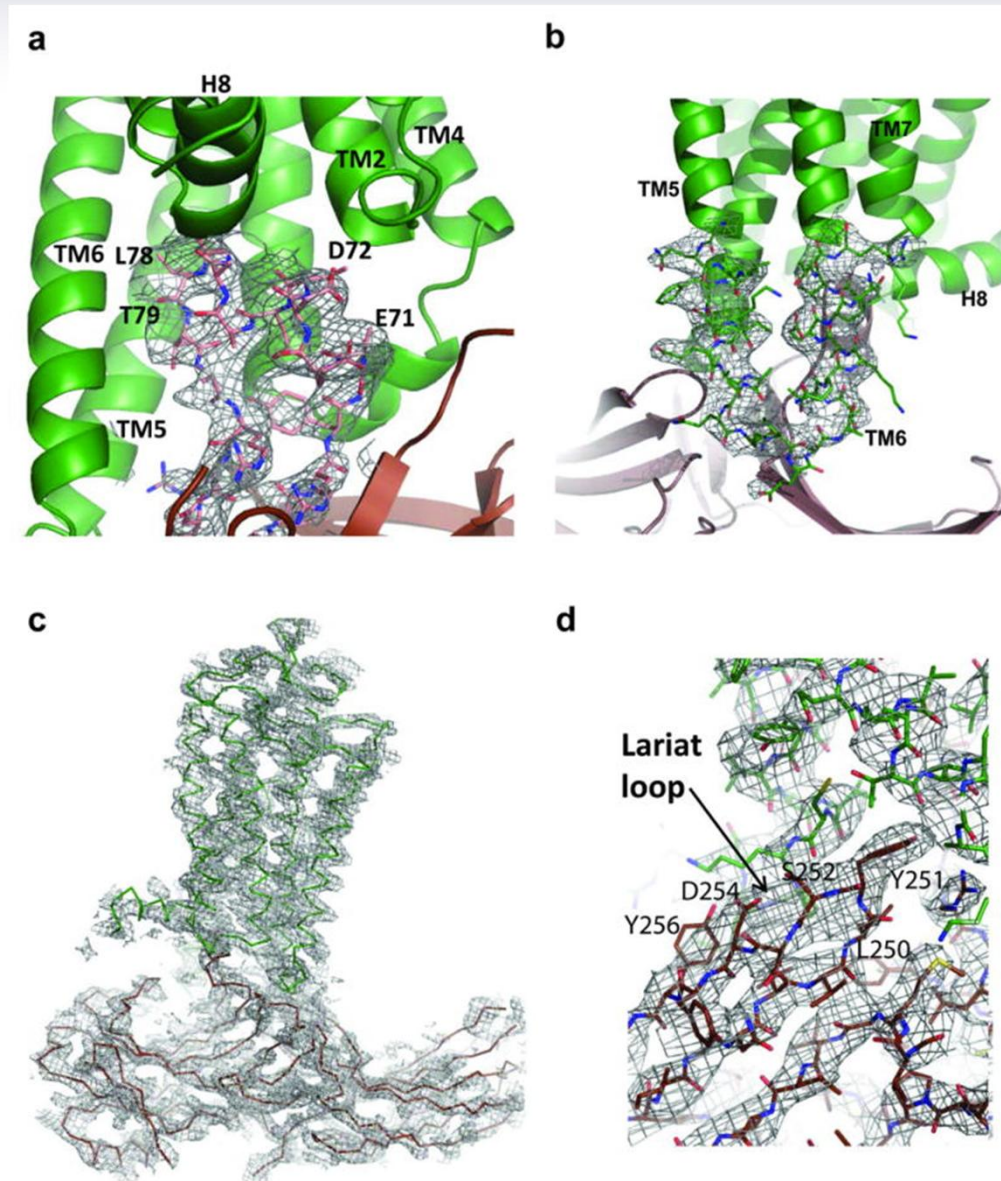
Szerkezet meghatározás – „single particle”

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



Szerkezet meghatározás – „single particle”

Free Electron Laser (FEL)



Nature. 2015 Jul
30;523(7562):561-7.

Crystal structure of
rhodopsin bound to
arrestin by
femtosecond X-ray
laser.

Negyedleges szerkezet

**Fehérje-fehérje dokkolás – rendkívül nehéz feladat
(felületek leírása, dinamika)**

**PISA - Protein Interfaces, Surfaces and Assemblies
Molecular Dynamics**

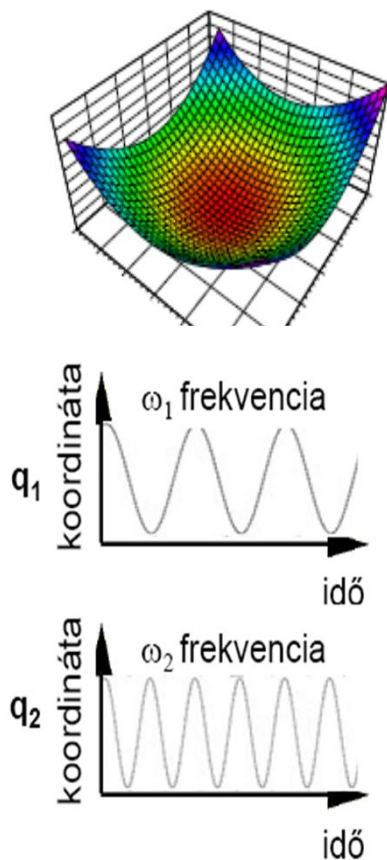
Mai témák

- Bevezetés – a fehérje dinamika és a szimulációk jelentősége
- Fehérjék szerkezetének jellemzése
 - Másodlagos szerkezeti mintázatok jóslása
 - Rendezetlen fehérjék
 - A harmadlagos és negyedleges szerkezet
- Fehérjék dinamikájának modellezése
- Fehérjék feltekeredésének szimulációja

Fehérje dinamika vizsgálata

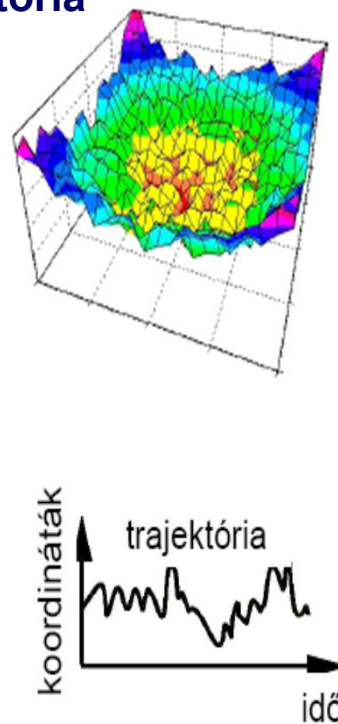
Normál-módus elemzés

- harmonikus potenciál
- analitikus mozgásegyenletek
- normál módusok



Molekuláris dinamika (MD)

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



A „force field“

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

$$E_{\text{solv}} = - \sum_i^{\text{natom}} \sum_{j>i}^{\text{natom}} \left\{ \frac{2\Delta G_i^{\text{free}}}{4\pi\sqrt{\pi}\lambda_i r_{ij}^2} \exp(-d_{ij}^2) V_j + \frac{2\Delta G_j^{\text{free}}}{4\pi\sqrt{\pi}\lambda_j r_{ij}^2} \exp(-d_{ji}^2) V_i \right\} \quad \text{Lazaridis (2003)}$$

TABLE I. Solvation Parameters[†]

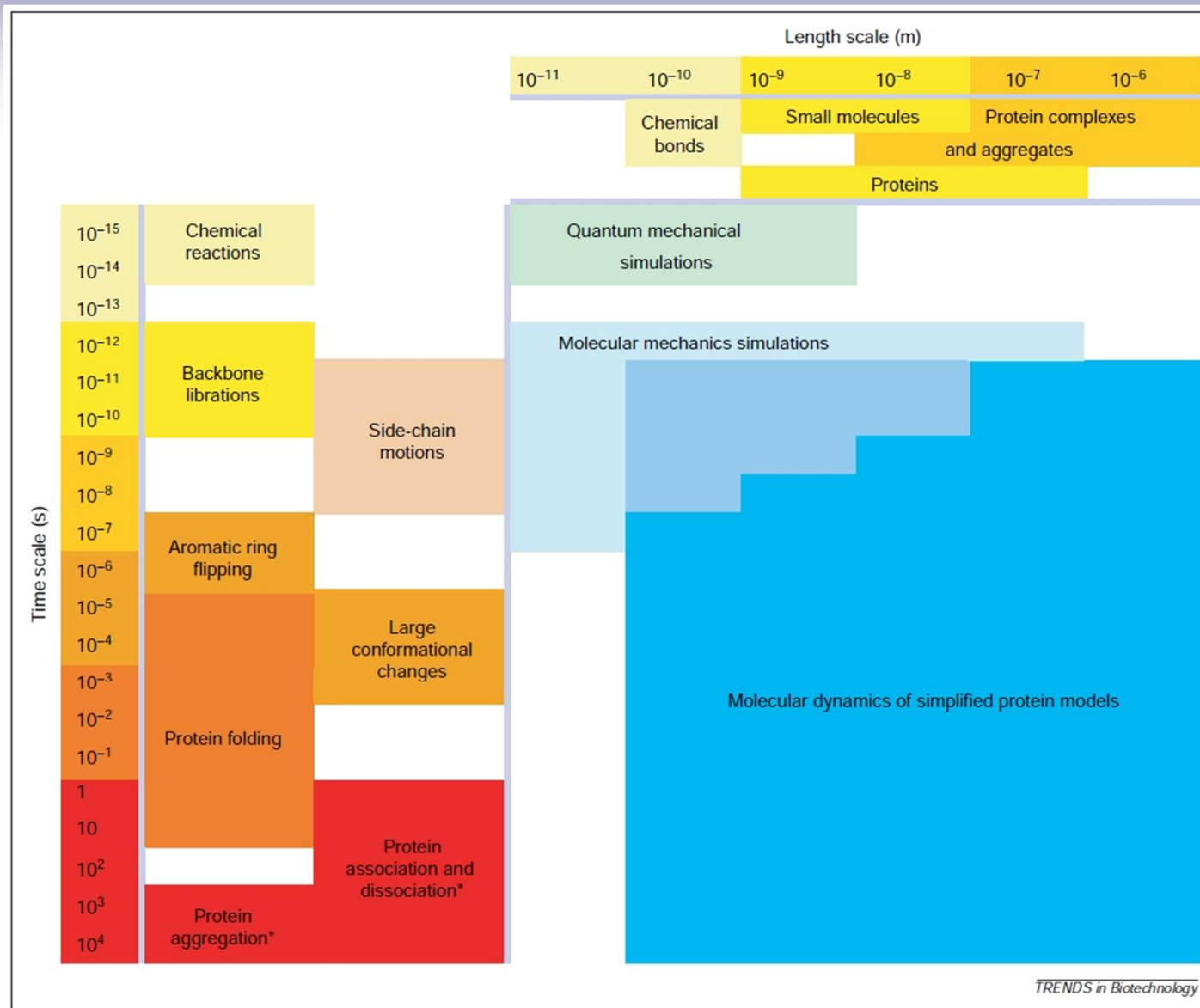
Atom types ^a	Volume	$\Delta G_1^{\text{ref b}}$	$\Delta G_1^{\text{free c}}$	$\Delta H_1^{\text{ref b}}$	$\Delta C p_1^{\text{ref d}}$
C	14.7	0.000	0.00	0.000	0.00
CR	8.3	-0.890	-1.40	2.220	6.90
CH1E	23.7	-0.187	-0.25	0.876	0.00
CH2E	22.4	0.372	0.52	-0.610	18.60
CH3E	30.0	1.089	1.50	-1.779	35.60
CR1E	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)

Az MD korlátjai

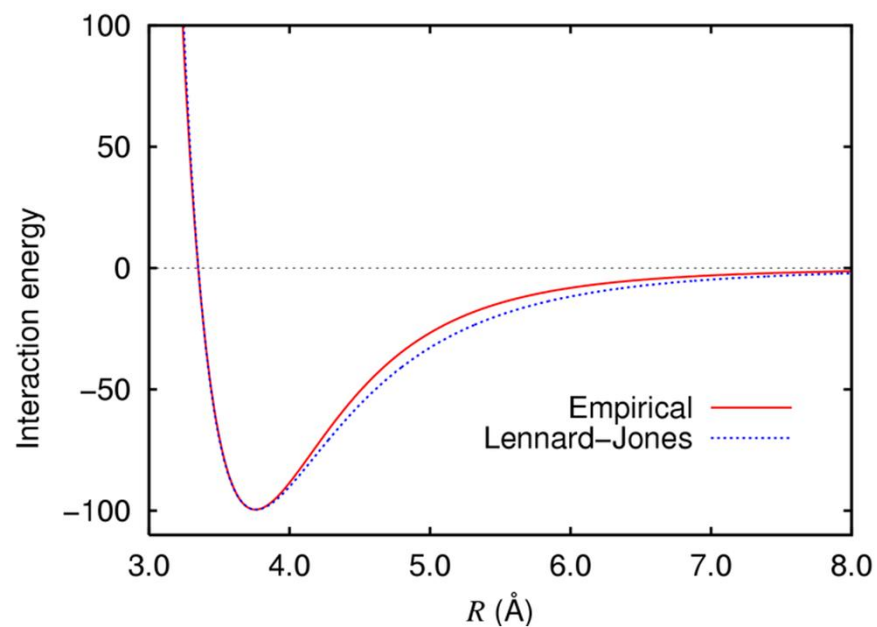
- idő (CPU, valós)
- potenciál kiszámolása a szűk keresztmetszet
- numerikus integrálás hibája
- fs-os integrációs lépések
- oldószer (explicit/implicit)
- „boundary condition”

„Események” időskálája

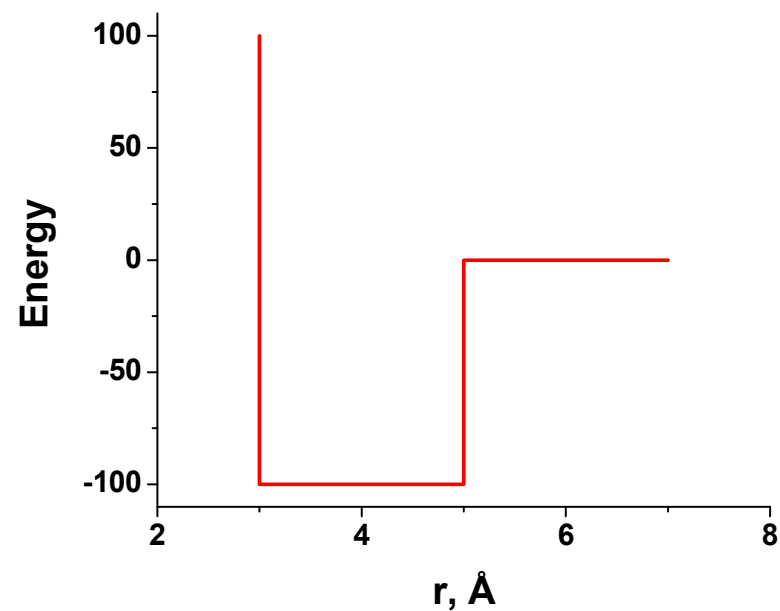


F. Ding and N.V. Dokholyan, TRENDS in Biotechnology, **23**:450 (2005)

Diszkrét Molekuláris Dinamika (DMD)



wikipedia



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

Egyszerűsített (Coarse Grain) modellek

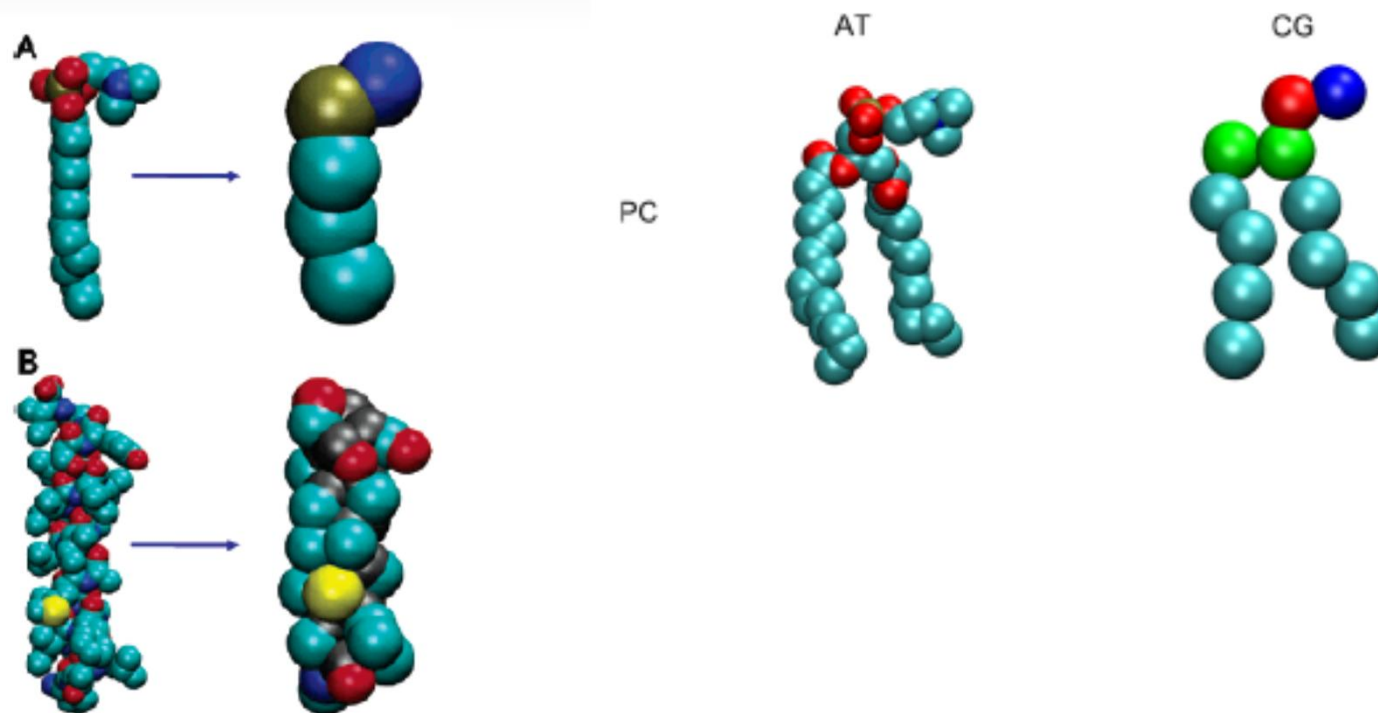


Figure 1. Atomistic (left-hand) and coarse-grained (right-hand) models compared for (A) a DPC molecule and (B) a GpA helix. Colors for atoms:

Fehérjére pl. 2 bead vagy 4+ bead modellek

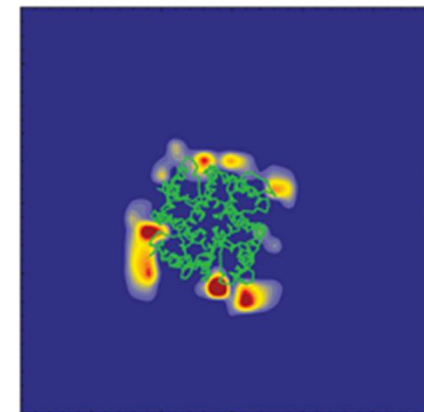
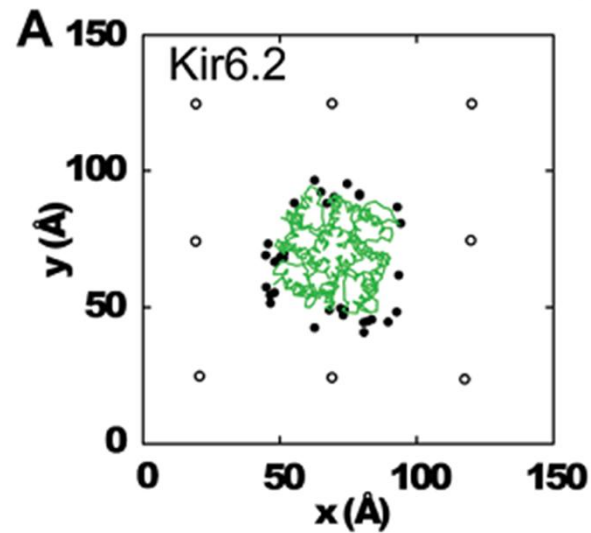
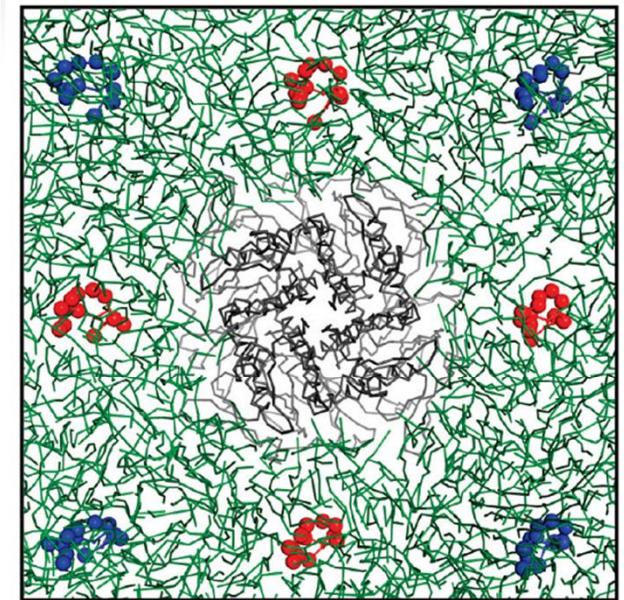
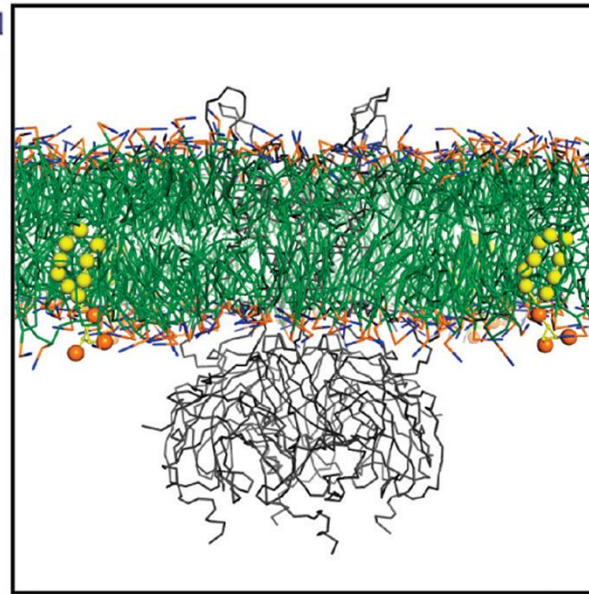
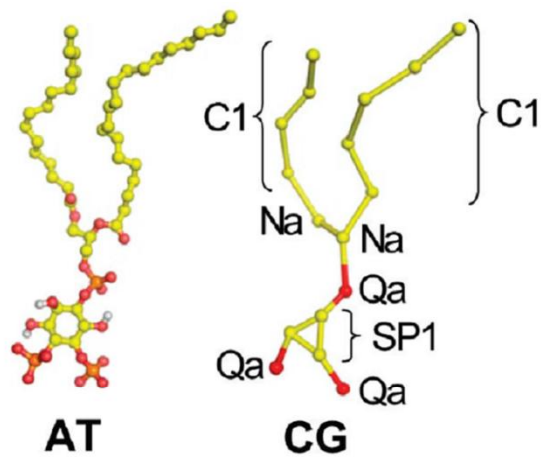
MARTINI CG force field

Kettősréteg felépülése a fehérje köré

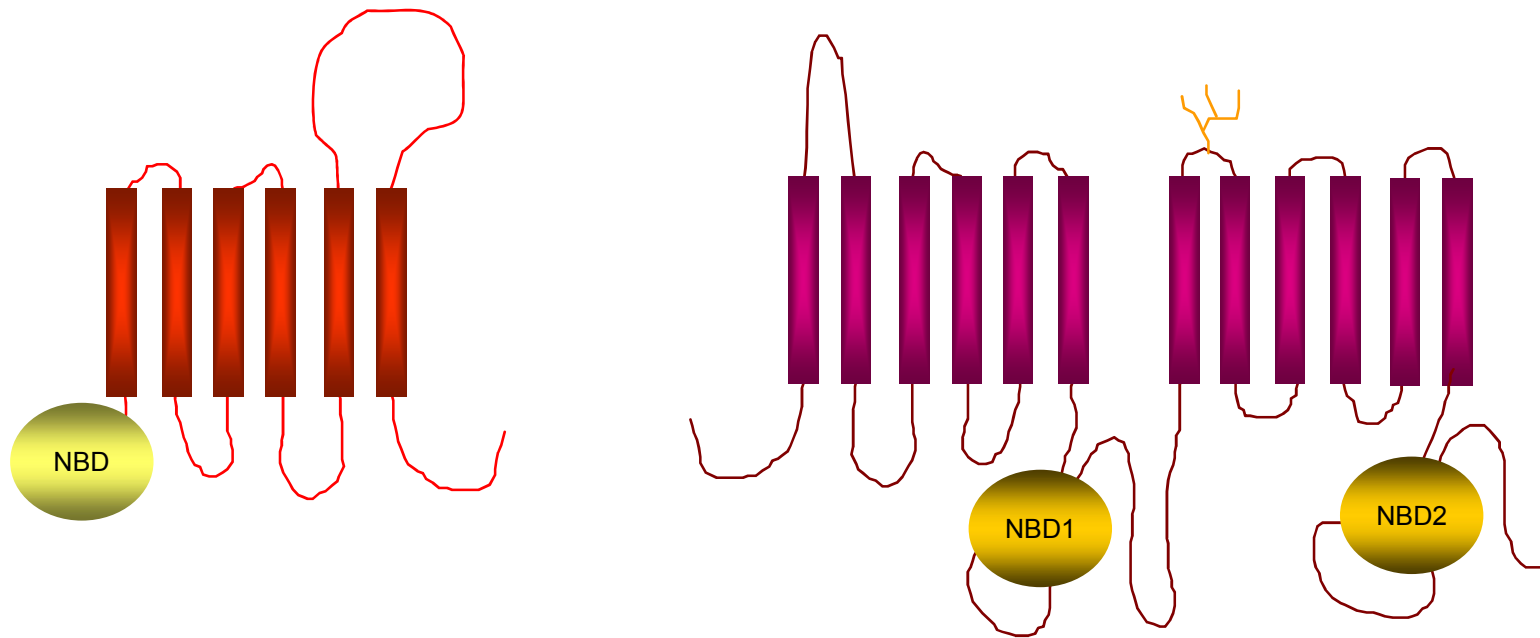
mozi

PIP2 kötődése Kir kálium csatornához

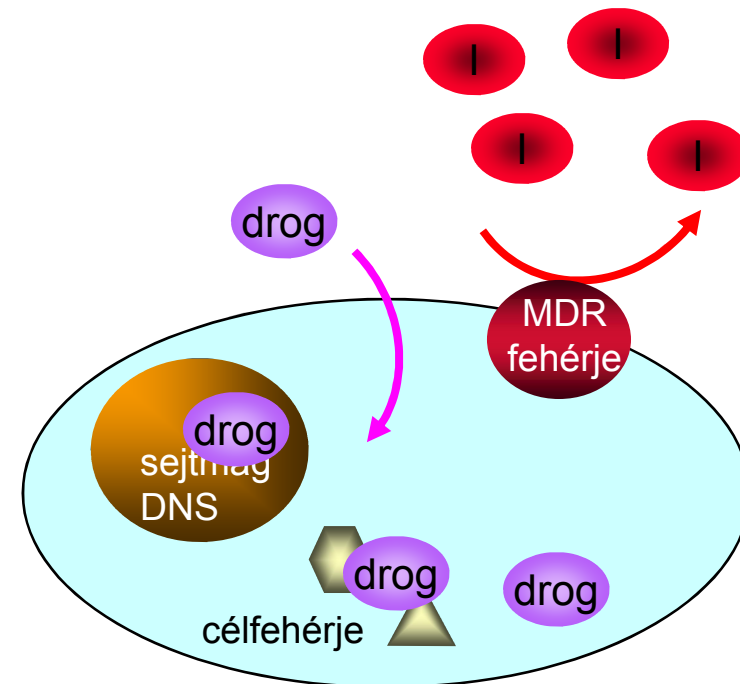
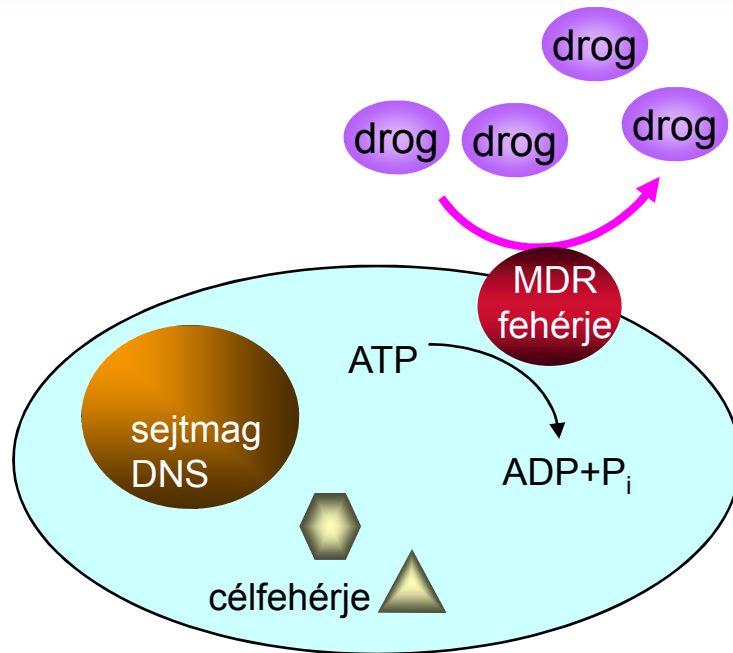
Biochemistry, Vol. 48, No. 46, 2009 1



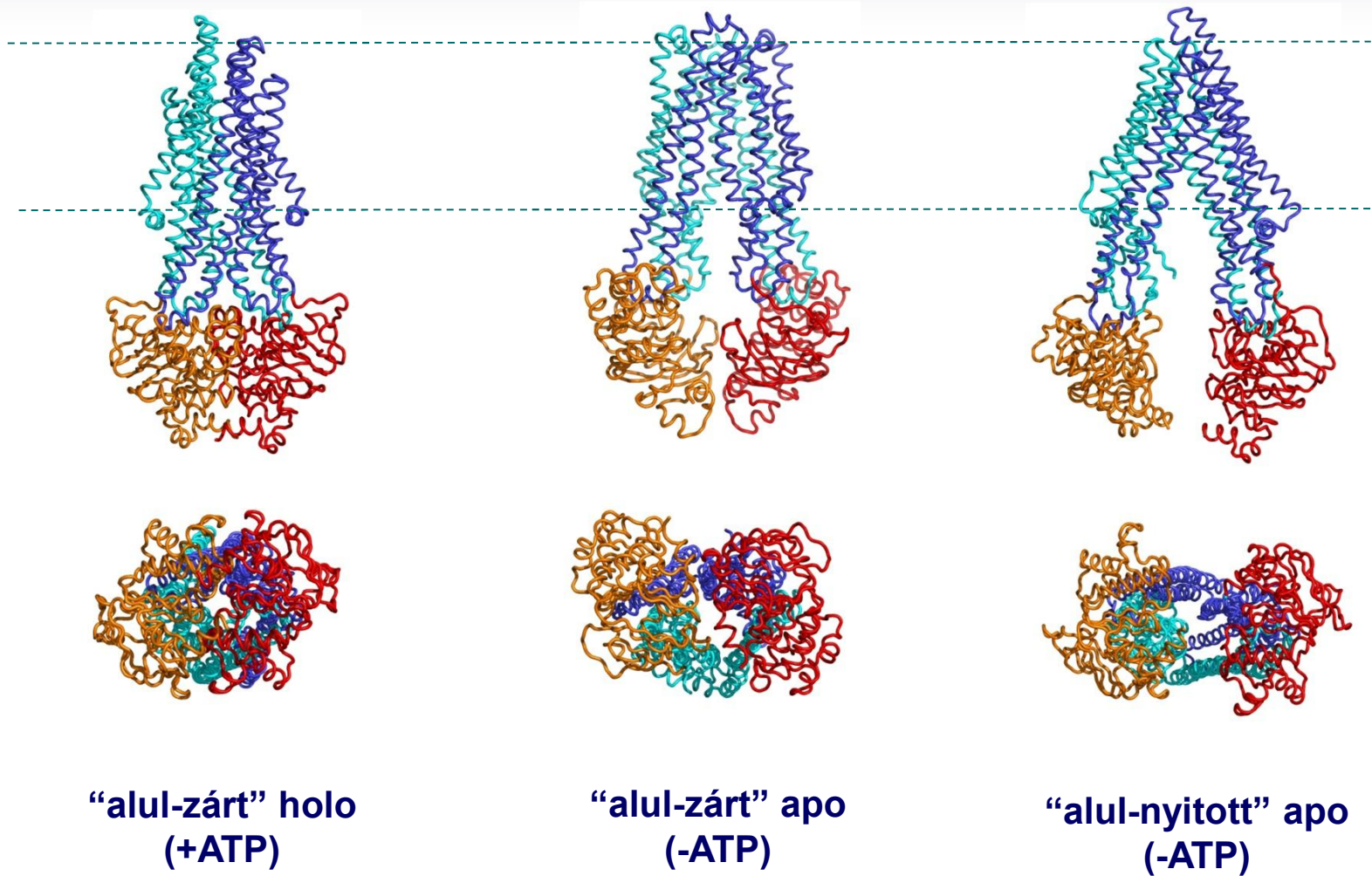
ATP Binding Cassette (ABC) fehérjék



A multidrog-rezisztencia és felfüggesztése



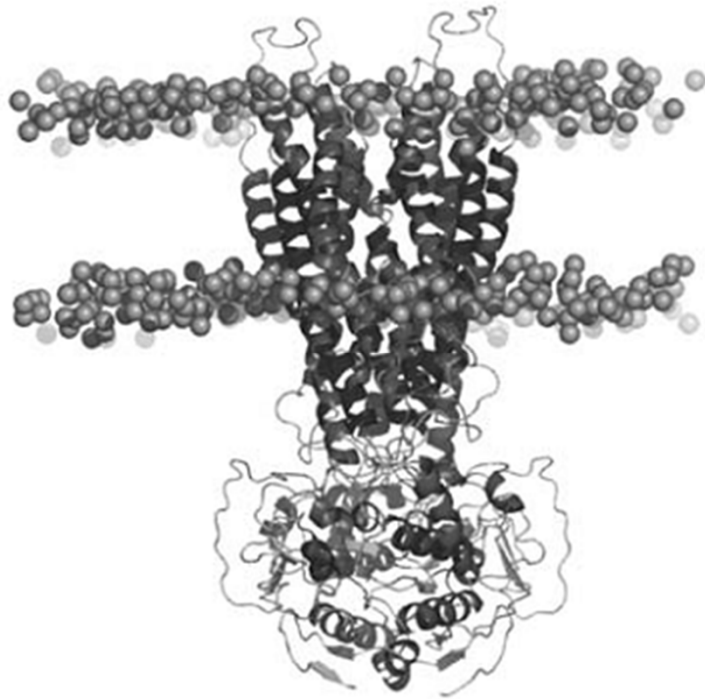
ABC fehérjék konformációi



Fehérjék konformációinak stabilitása

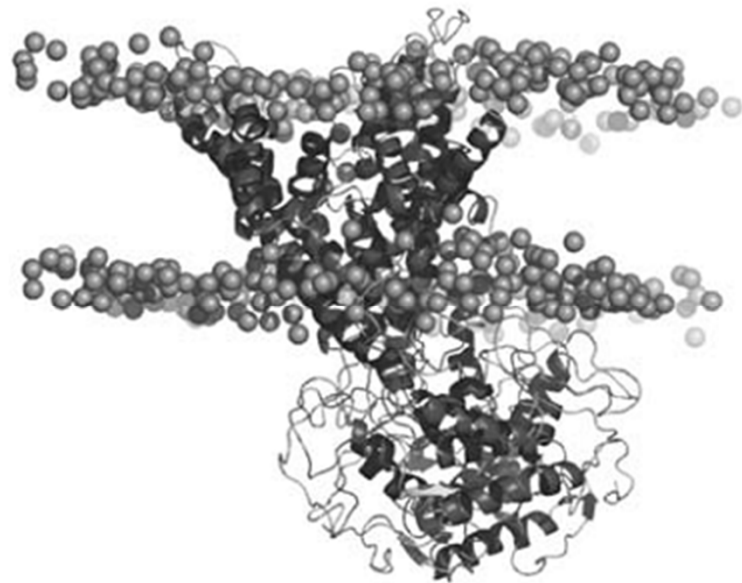
Eur Biophys J (2008) 37:403–409

B



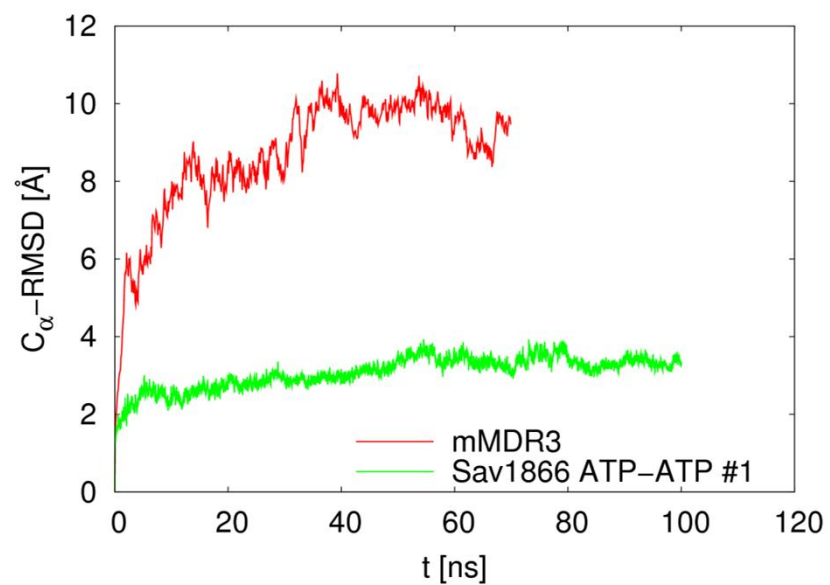
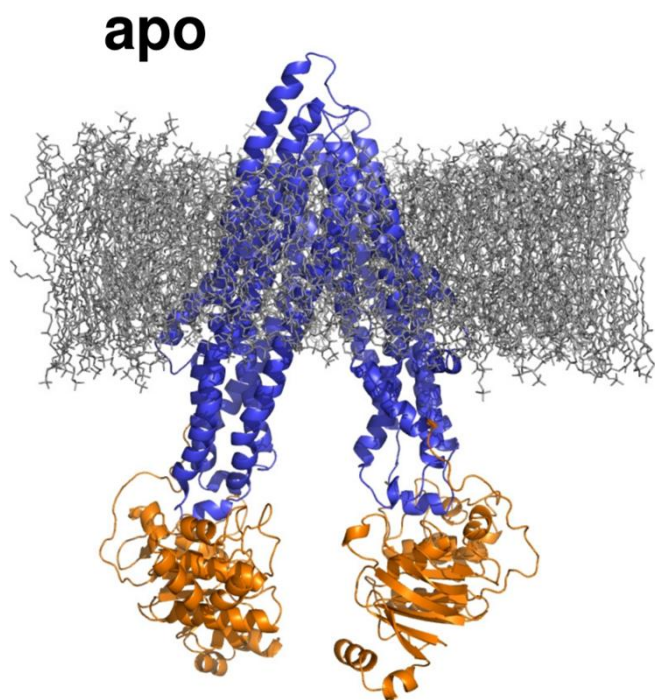
0 ns

C



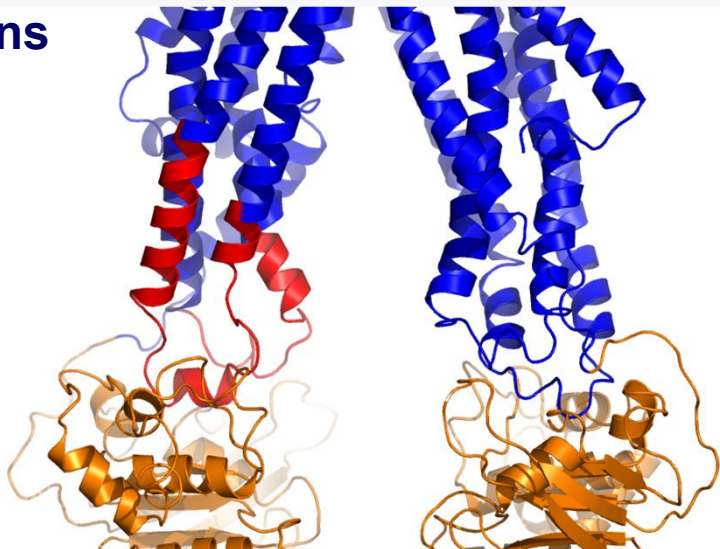
20 ns

Az alul nyitott apo szerkezet nem stabil

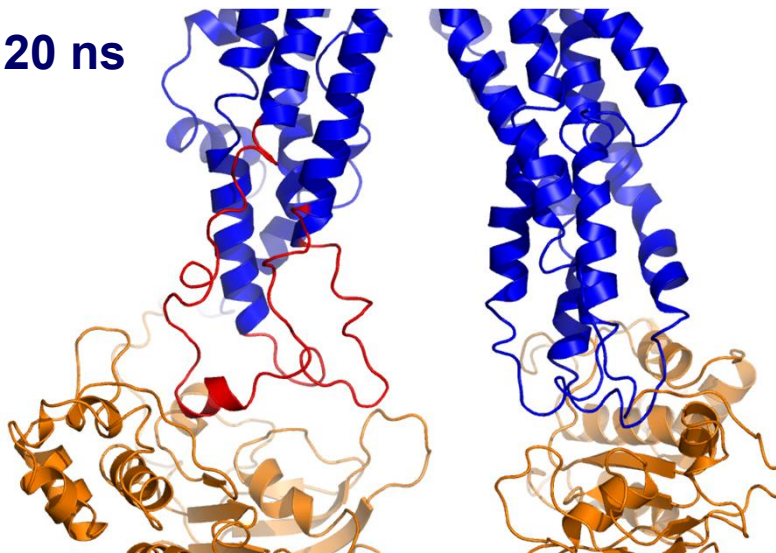


Az alul nyitott apo szerkezet nem stabil

t = 0 ns

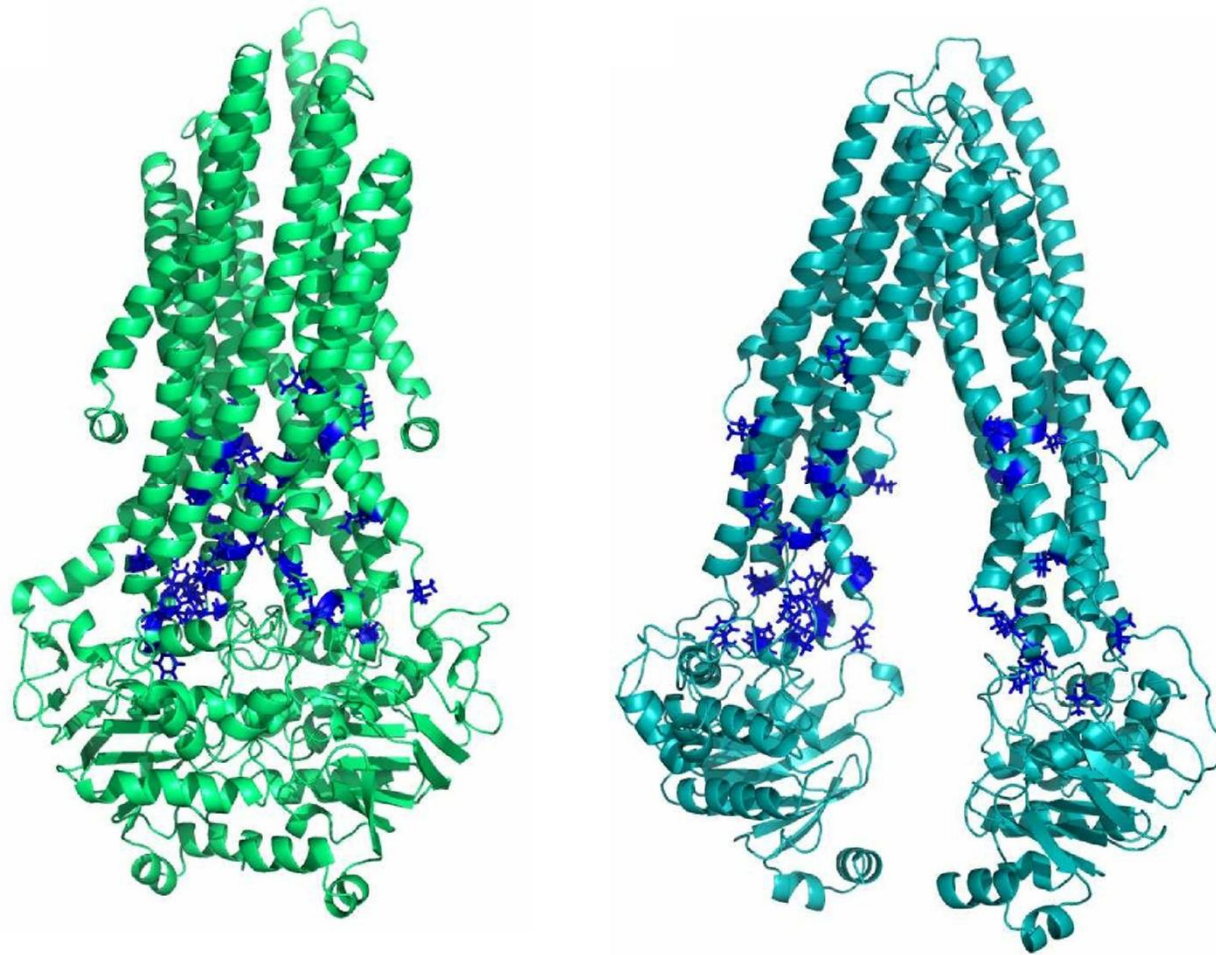


t = 20 ns



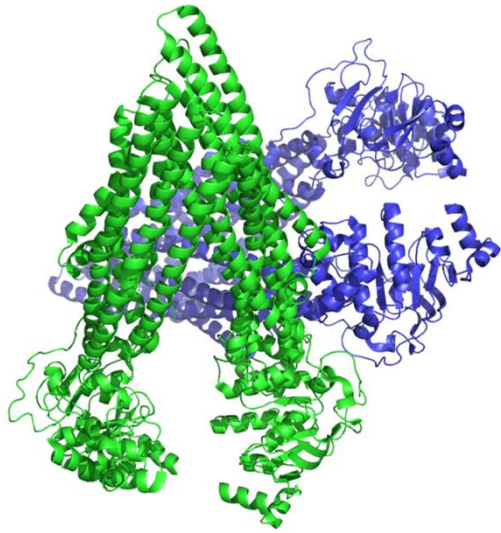
rendszer	megtartott hélixtartalom
Sav1866 ATP/ATP #1	90.04%
hMDR1 holo	91.84%
hMDR1 apo	64.30%
mMDR3	63.13%

Hidrofób aminosavak kerülnek felszínre

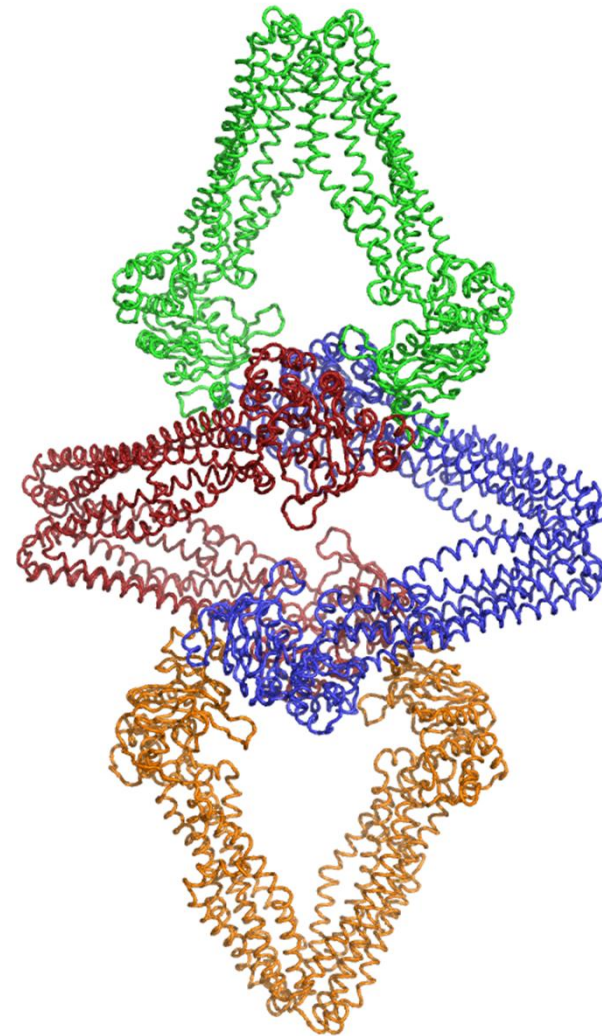


Alul nyitott apo szerkezet elemi cellája

mMDR3, PDBID:3G5U



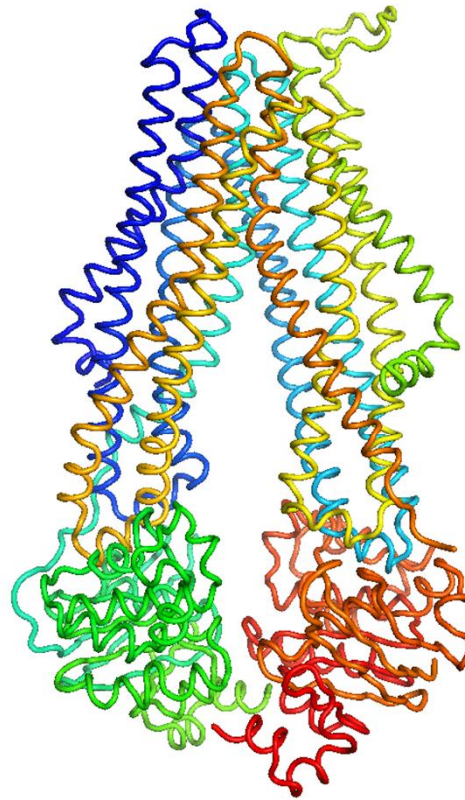
MsbA, PDBID:3B5W



Események modellezése

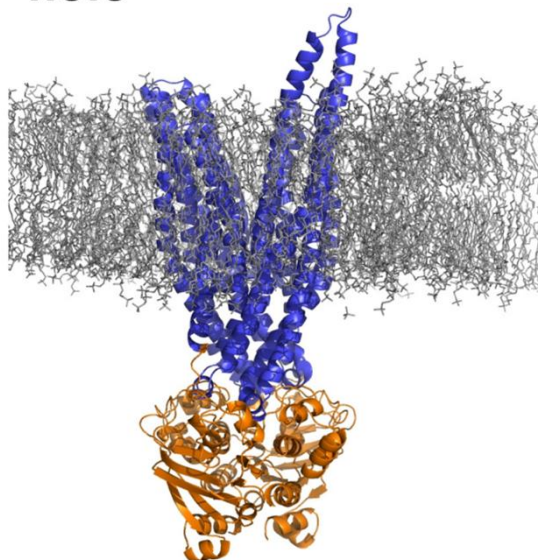
- Hogyan befolyásolja az ATP hidrolízise a fehérje dinamikáját?
Pl. steered MD
- Hogyan történik meg az átmenet az „alul-zárt” konformációból az „alul-nyitott” konformációba?
Pl. targeted MD, METADINAMIKA, stb.

Targeted MD



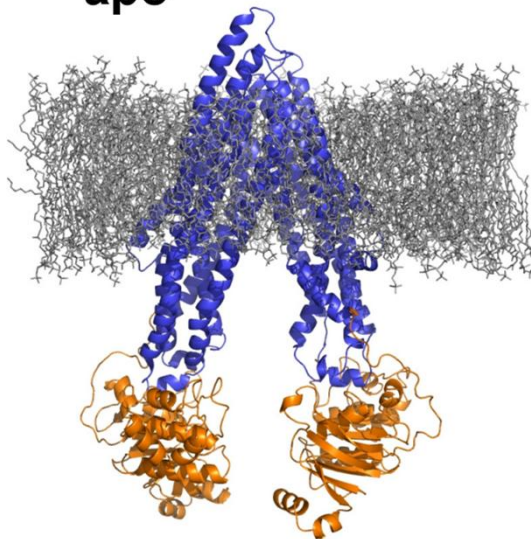
Zárt-nyitott átmenet jellemzése molekuláris dinamikával

holo



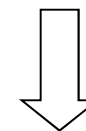
hMDR1 homologia modell
(3x100 ns)

apo

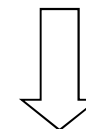


hMDR1 homologia modell

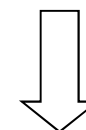
**molekularis dinamika
trajektóriák**



**esszenciális
dinamika**

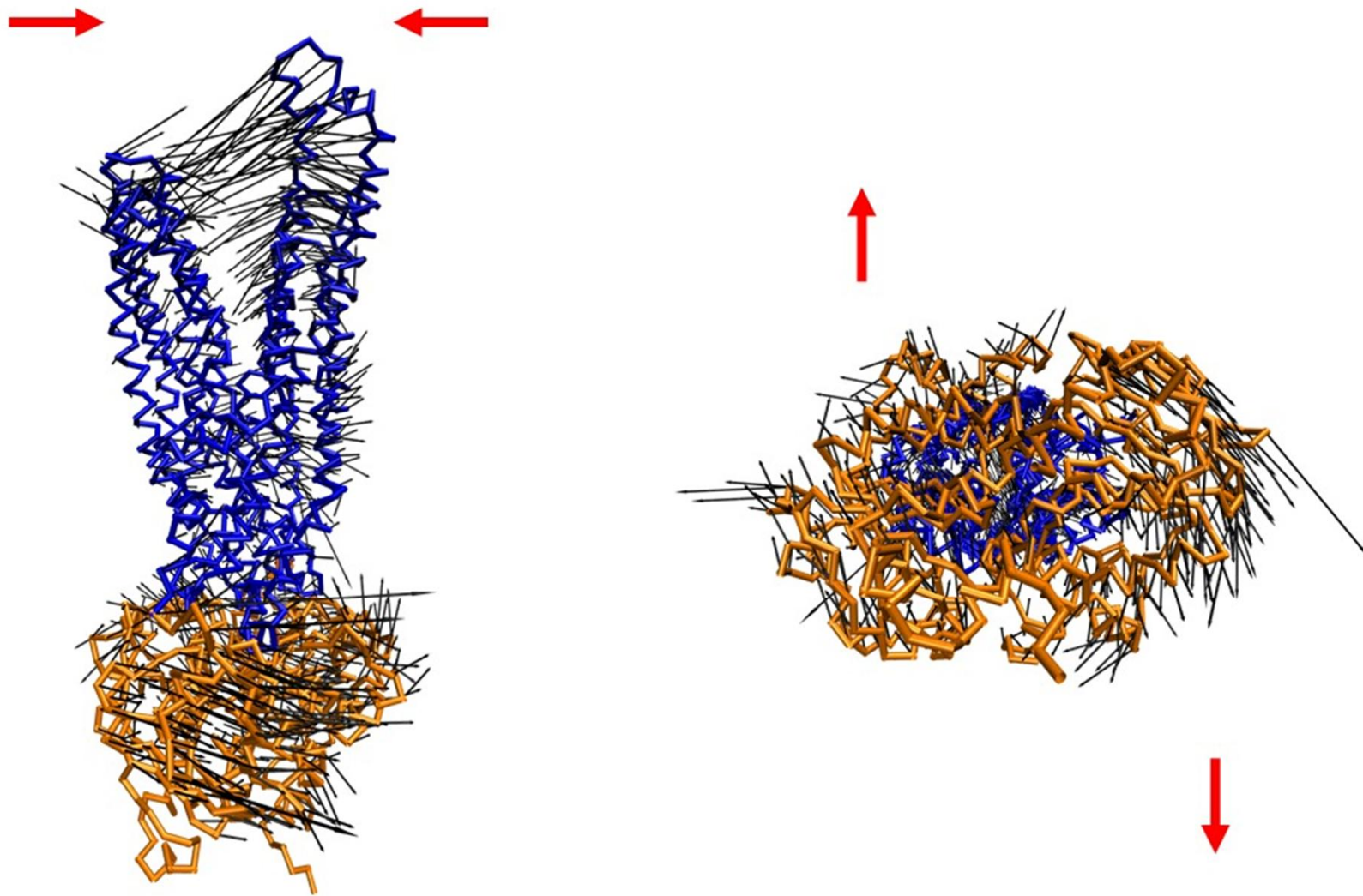


módus kiválasztás



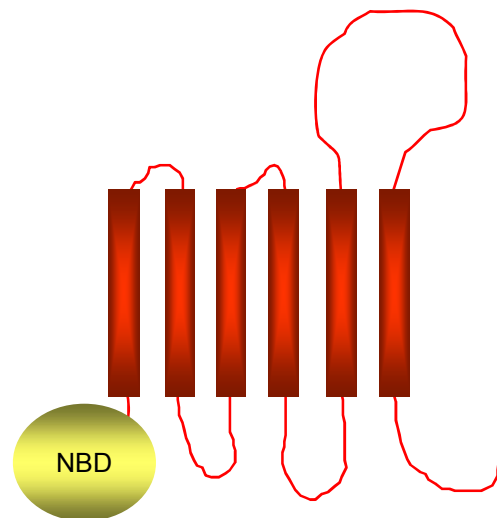
kollektív mozgások

Zárt-nyitott átmenet



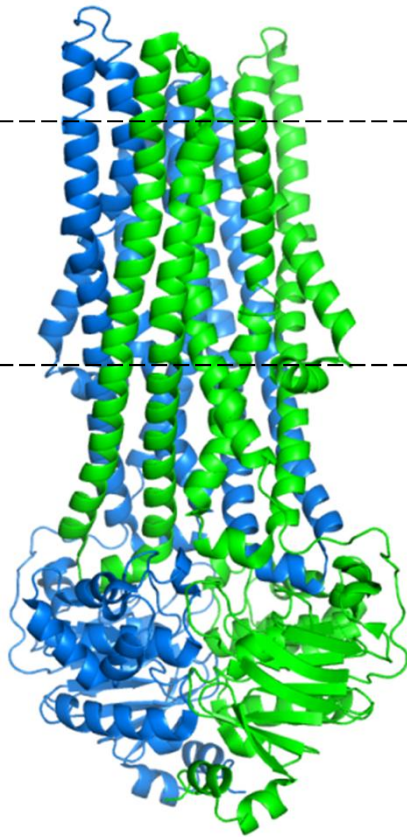
1. komplex példa – ABCG2

- **Szerkezet modellezése**
 - **Mutációk hatásának elemzése molekuláris dinamikával**
 - **Kis molekulák dokkolása a fehérjéhez**
-
- **Őssejtekben, szöveti barrierekben, rákos sejtekben fejeződik ki**
 - **Multidrog transzporter (xenobiotikumok és endobiotikumok):**
 - rákellenes gyógyszerek
 - húgysav
 - **Leggyakoribb variáns (Q141K) csökkent expresszió és funkció**

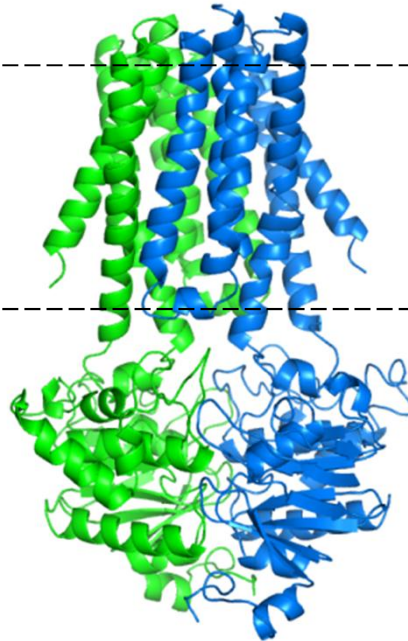


ABCG alcsalád - Type II

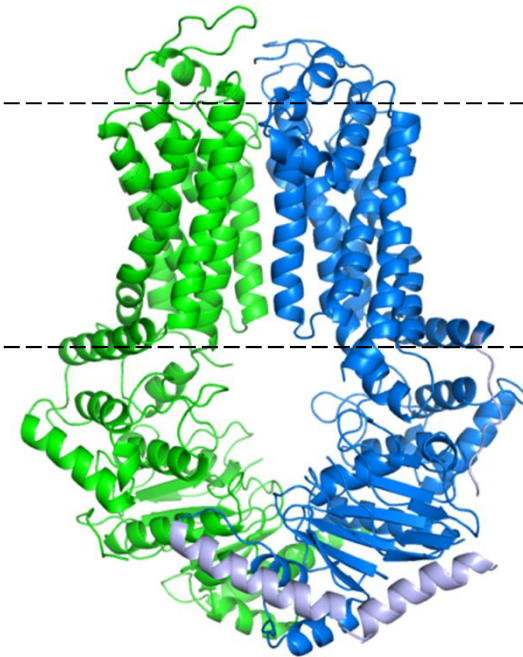
Type I +ATP



ABCG2 from 2010

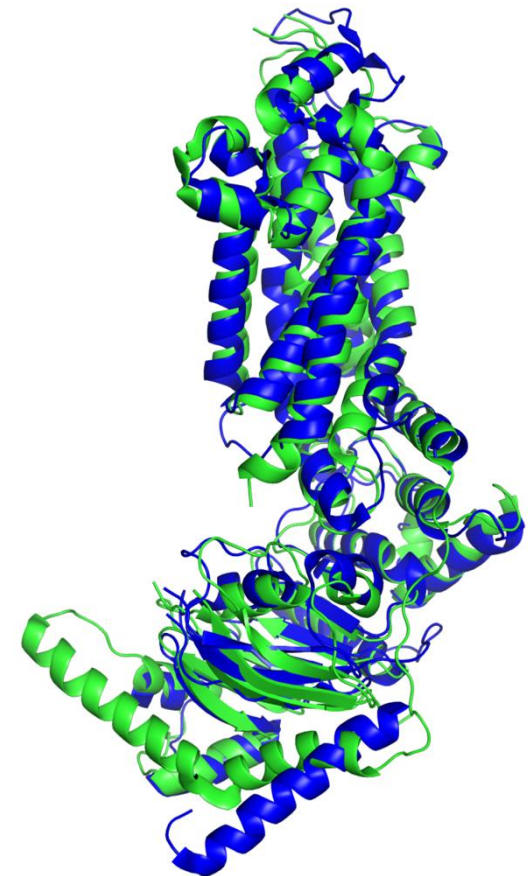


ABCG5-ABCG8



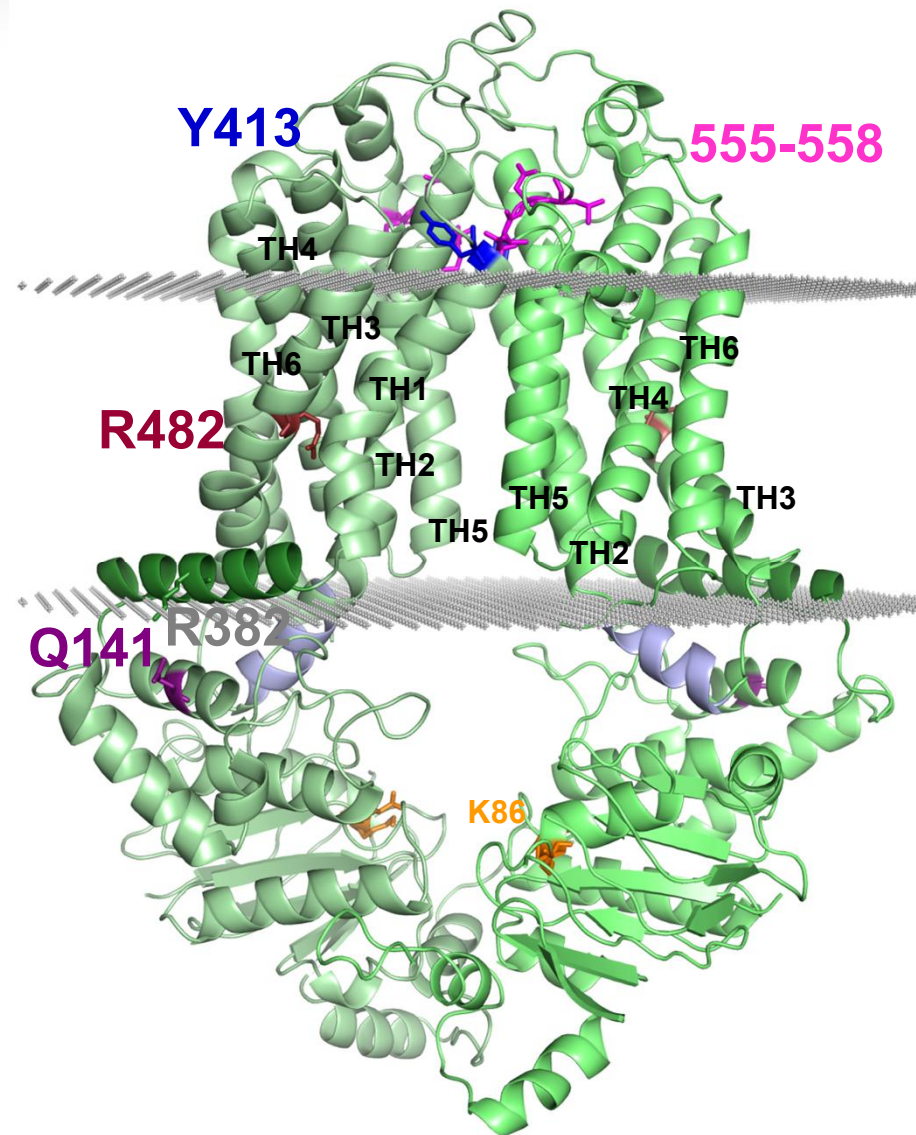
Az ABCG2 homodimer homológia modellezése

- Kb. 25% megegyező és 45% hasonló a.a.
- A szekvenciaillesztés nagy része triviális volt
- 100 modellt építettünk Modeller-vel
- A legjobb modellt (DOPE score) kiválasztottuk és használjuk

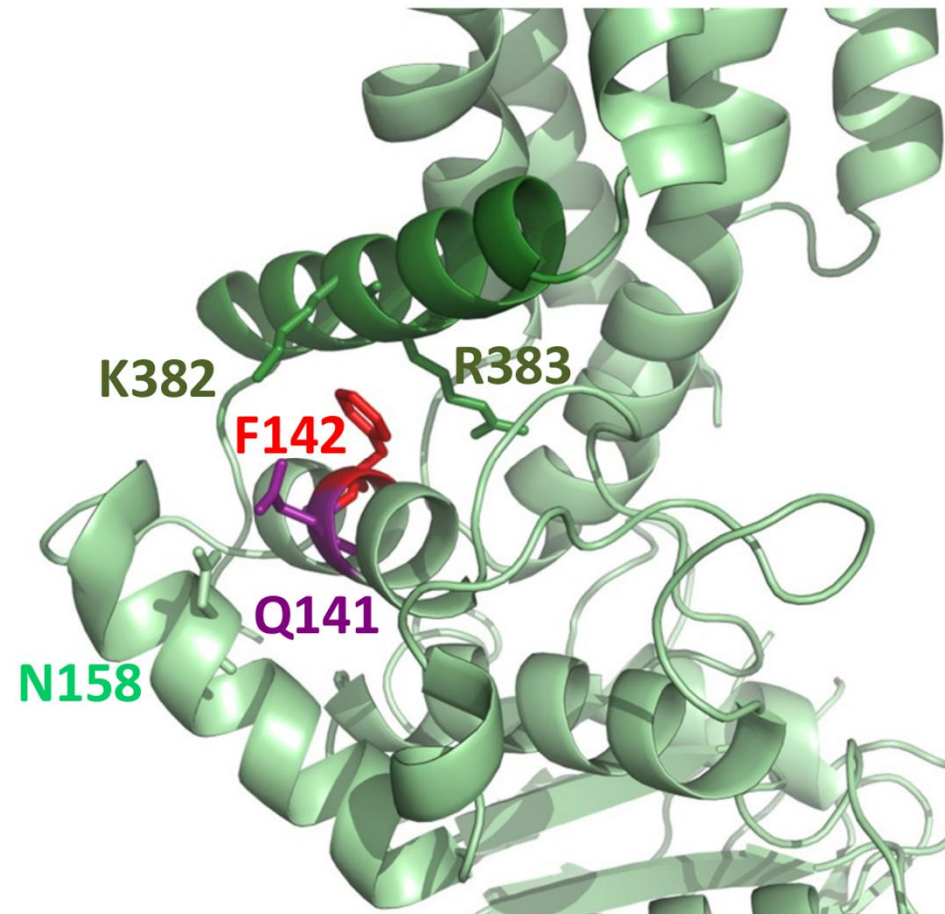


ABCG5-ABCG8
PDBID:5DO7

Az ABCG2 modell



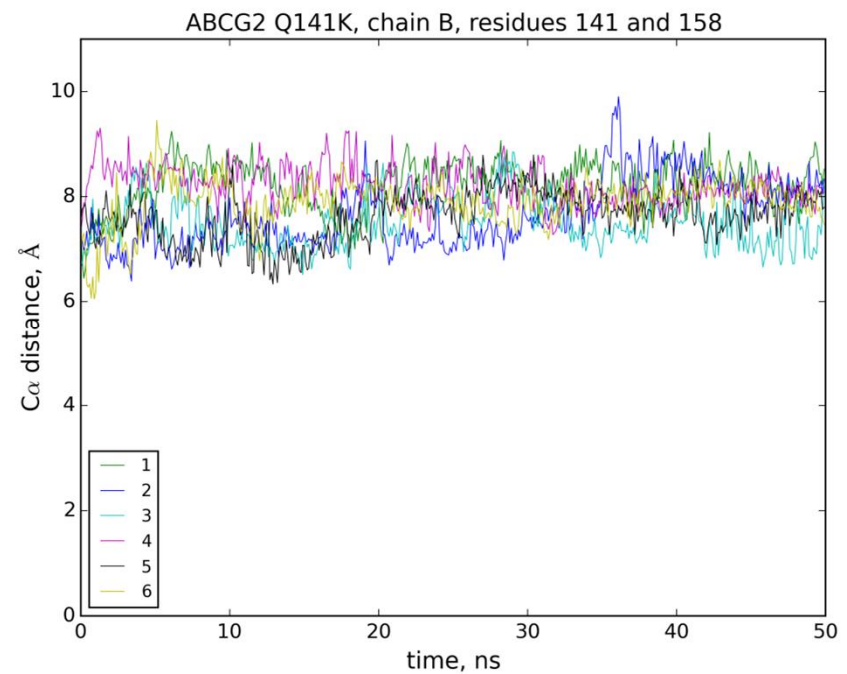
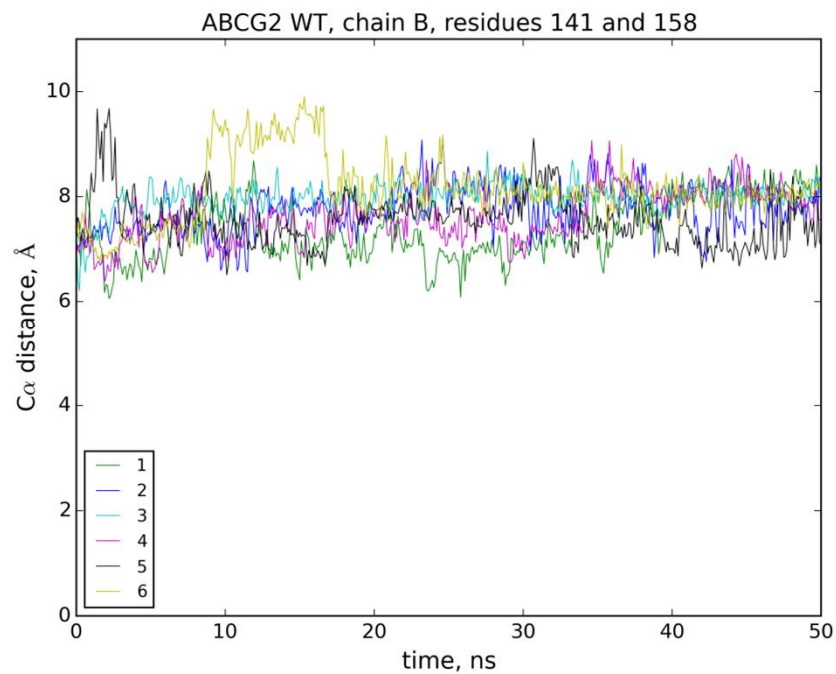
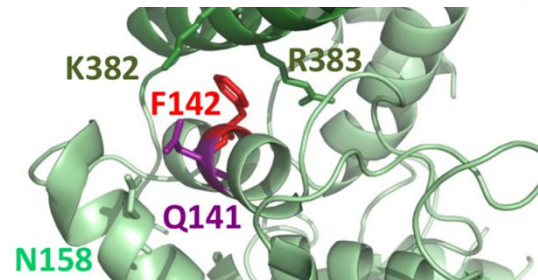
A Q141 helyzete



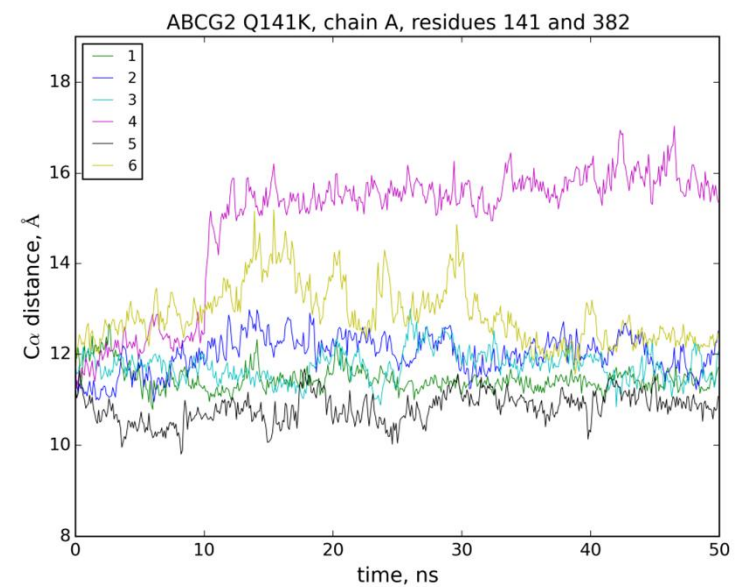
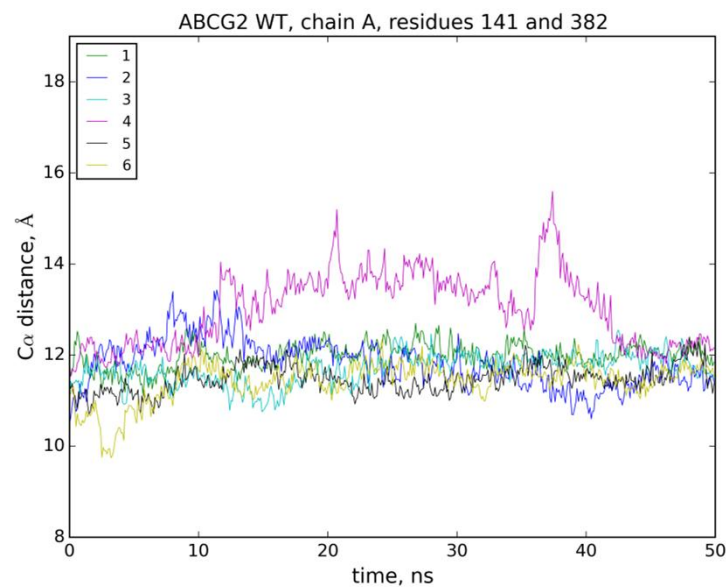
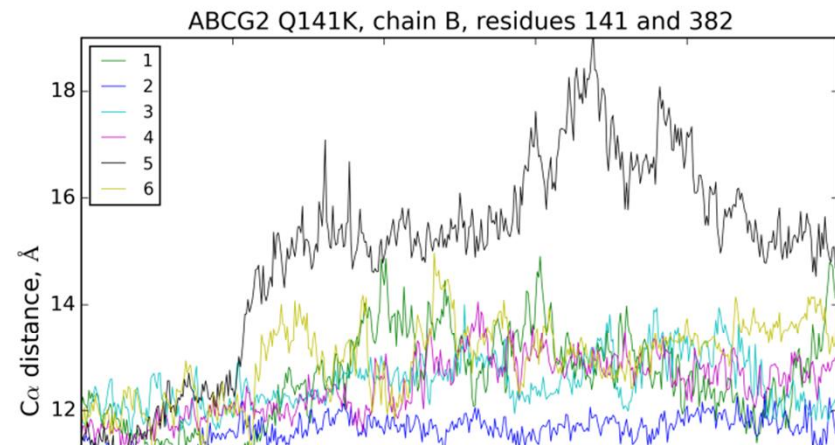
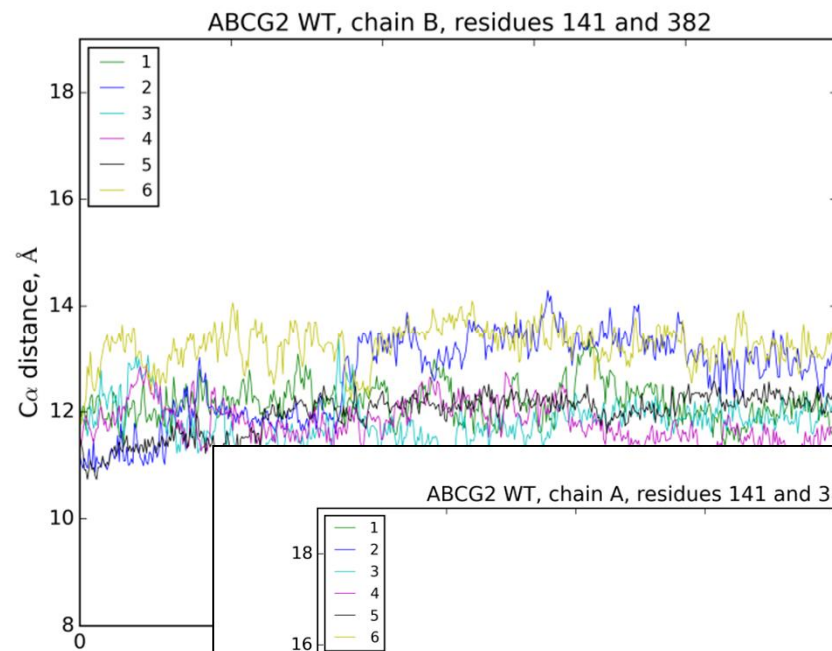
Molekuláris Dinamikai (MD) szimulációk

- **Az ABCG2 fehérjét membrán kettősrétegbe helyeztük (POPC)**
- **Az a.a. oldalláncok, lipidek és vízmolekulák elhelyezkedésének optimalizálása:**
 - energia minimalizáció
 - egyensúlyba hozás (6 db párhuzamos)
 - minimális gerinc-mozgás (position constrains)
- **Production run**
 - no constraints
 - 50 ns x 6 = 300 ns;
- **WT és variánsok összehasonlítása**
(pl. Q141K, R482G)

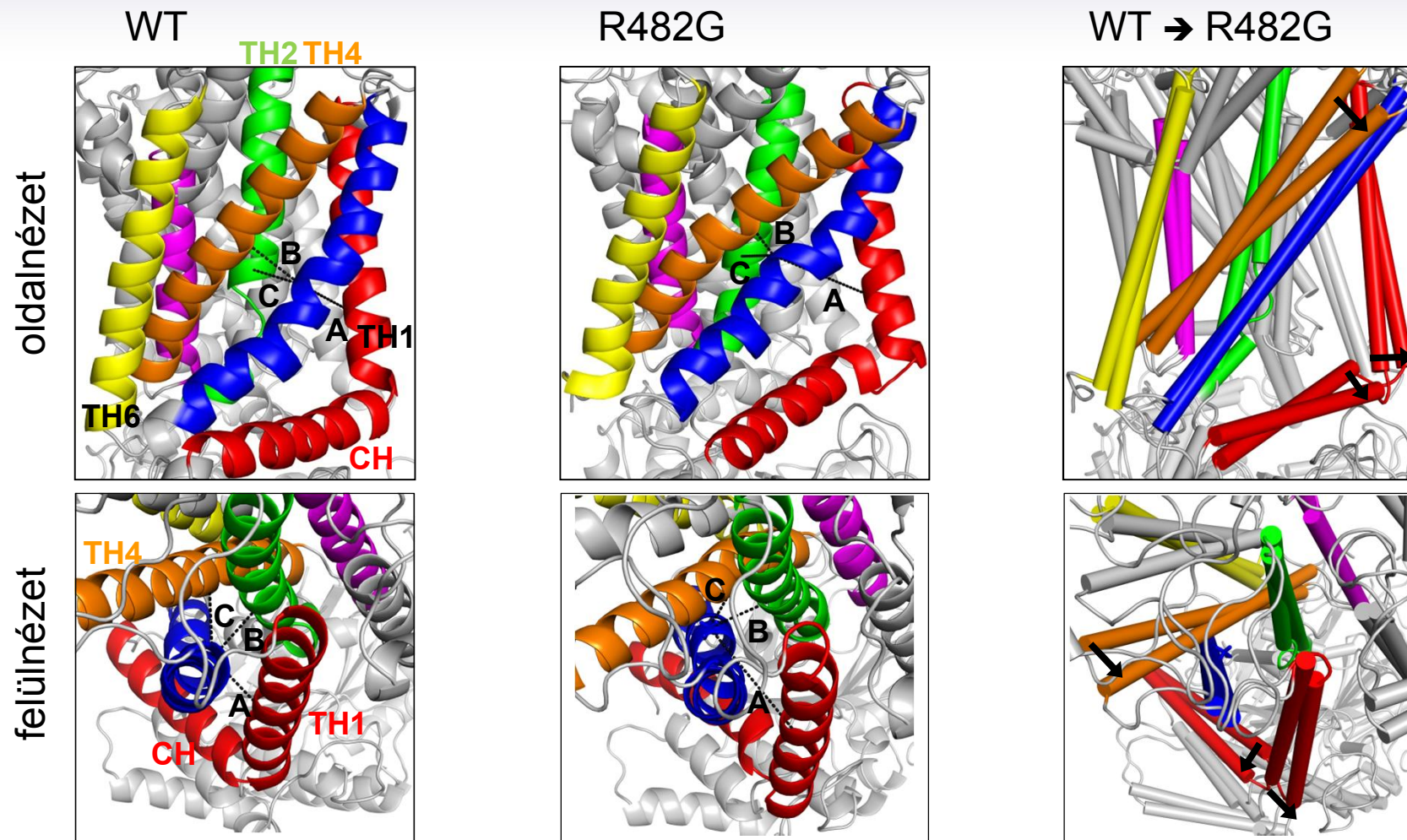
Q141K csere hatása a 141/158 kölcsönhatásra



A Q141K hatása a 141/382 (NBD/TMD) kapcsolatra



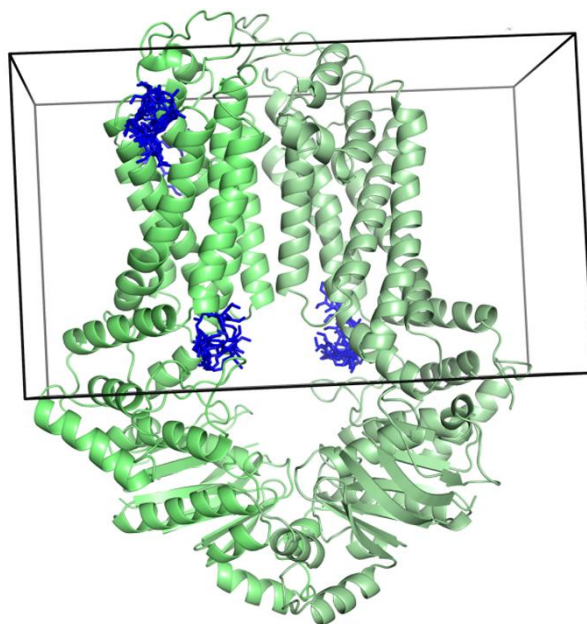
Az R482G által okozott változások



Drogkötő helyek vizsgálata *in silico* dokkolással

AutoDock Vina

- Fix fehérje, flexibilis ligand
- Több konformáció (6 db, egyensúlyba hozott)
- Doboz; ligand odapróbálása nagyon sokszor (ebből 20/futás)

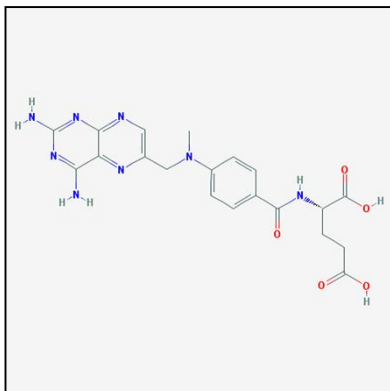


Drogkötő helyek vizsgálata *in silico* dokkolással

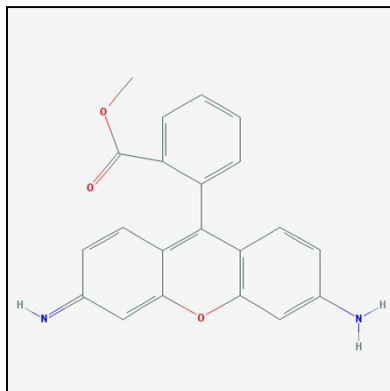
AutoDock Vina

- Fix fehérje, flexibilis ligand
- Több konformáció (6 db, egyensúlyba hozott)
- Doboz; ligand odapróbálása nagyon sokszor (ebből 20/futás)
- Mindent oda lehet dokkolni... ezért szubsztrát + nemszubsztrát

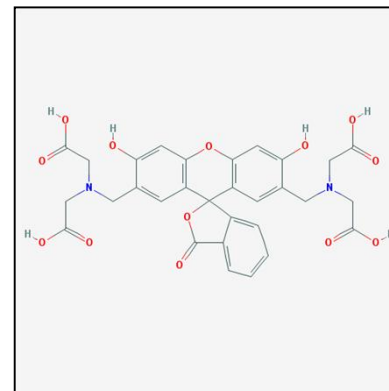
methotrexate



rhodamine123



calcein



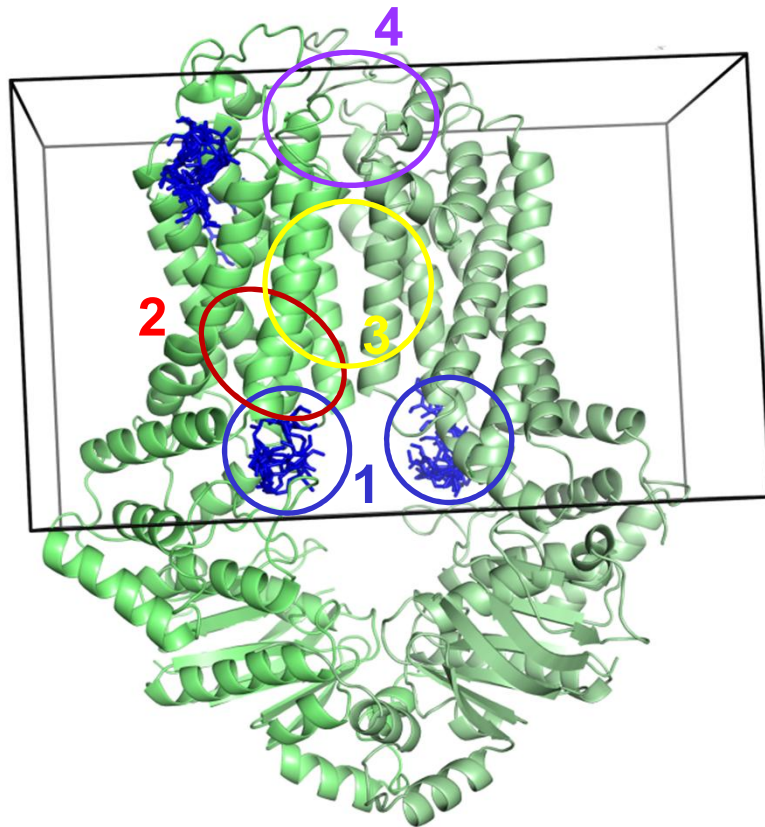
Drogkötő helyek vizsgálata *in silico* dokkolással

AutoDock Vina

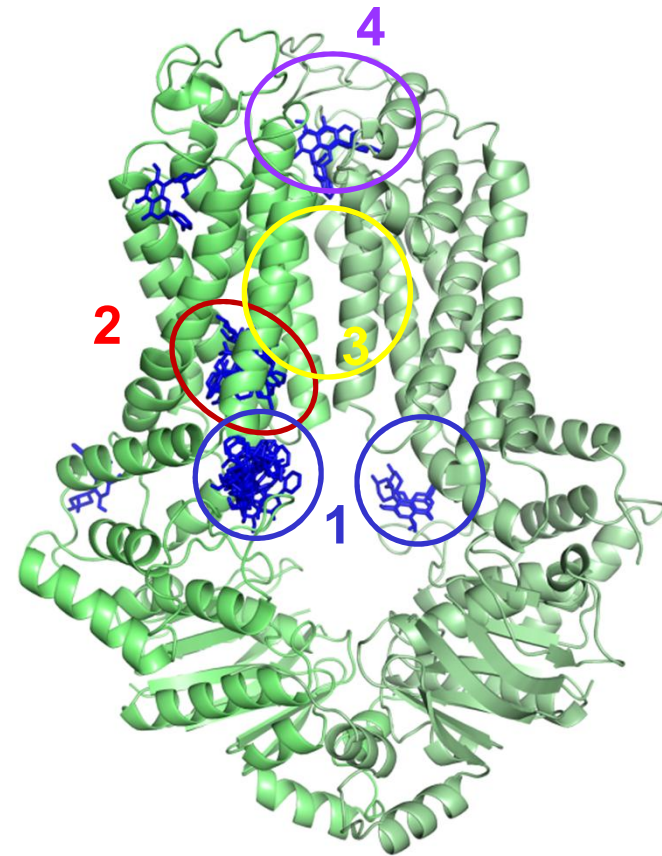
- Fix fehérje, flexibilis ligand
- Több konformáció (6 db, egyensúlyba hozott)
- Doboz; ligand odapróbálása nagyon sokszor (ebből 20/futás)
- Mindent oda lehet dokkolni... ezért szubsztrát + nemszubsztrát
- (6 ABCG2 konformáció) * (20 pozitúra) * (3 párhuzamos) *
(25 szubsztrát + 14 nem-szubsztrát) = 14 040 pozitúra
- Pozitúrák klaszterezése

Drogkötő helyek vizsgálata *in silico* dokkolással

verapamil

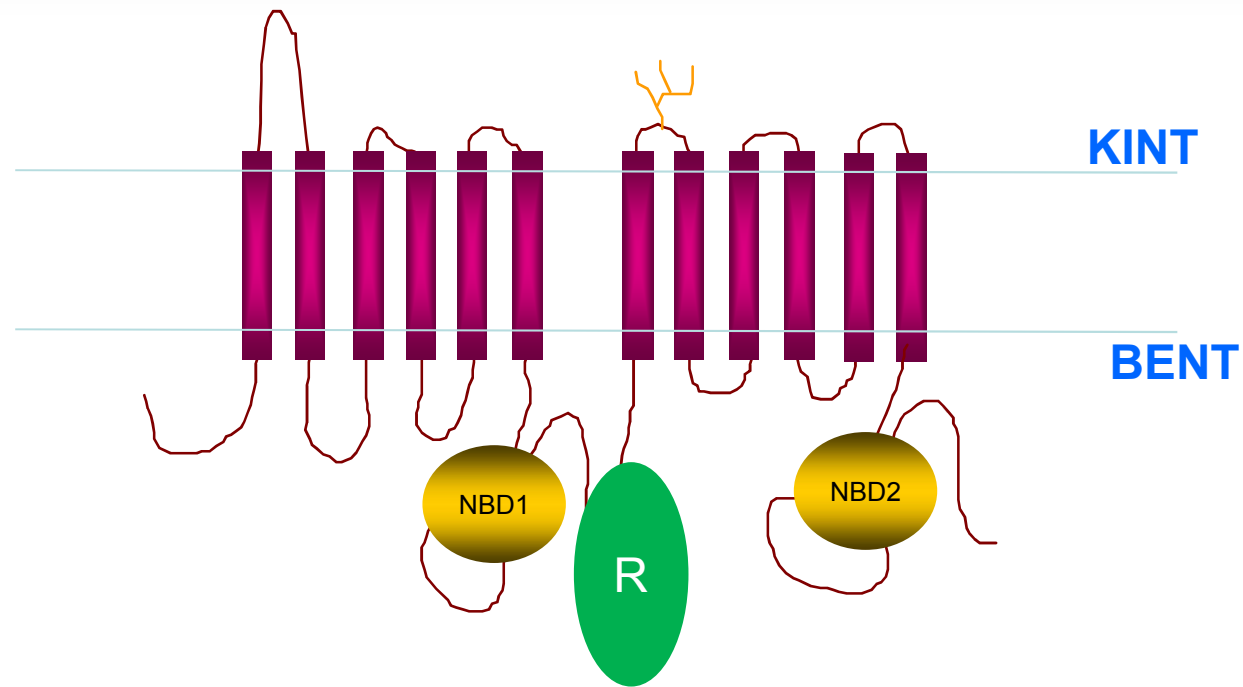


flavopiridol

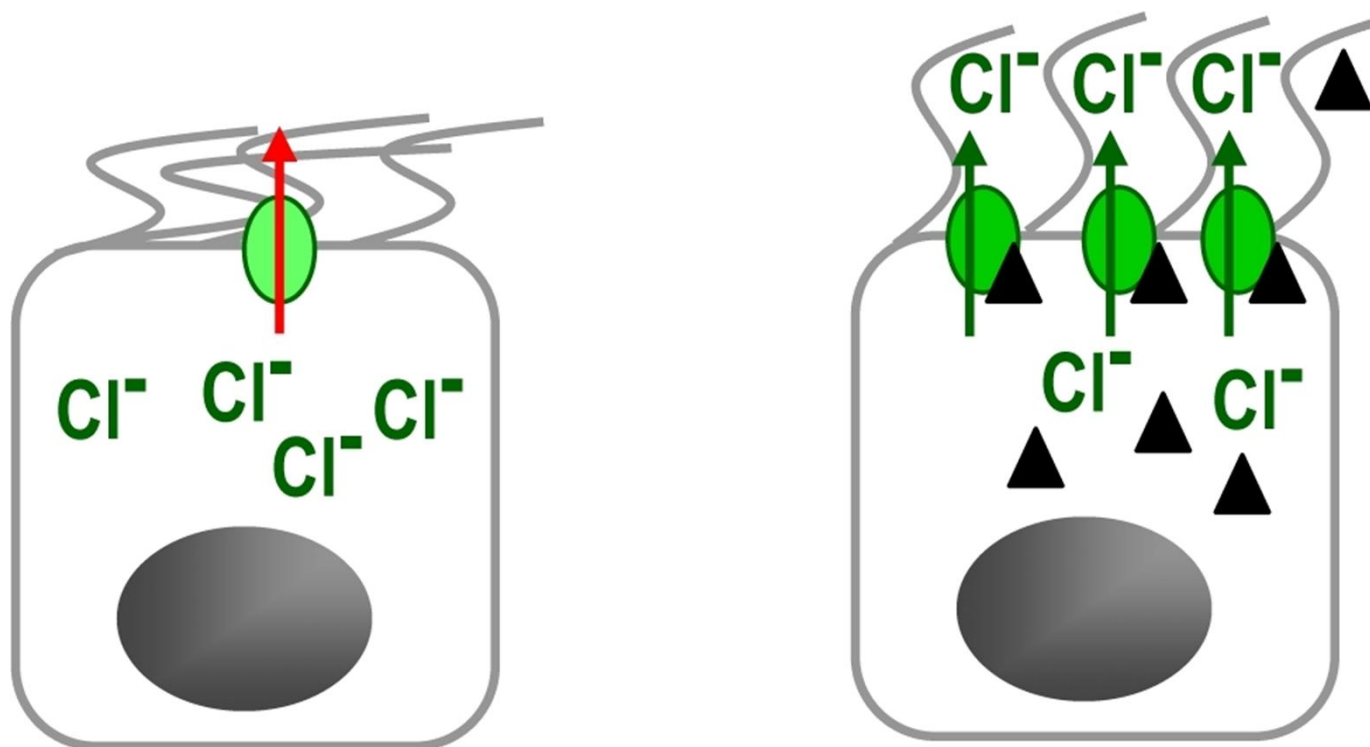


CFTR

ABCC7 (ATP Binding Cassette)

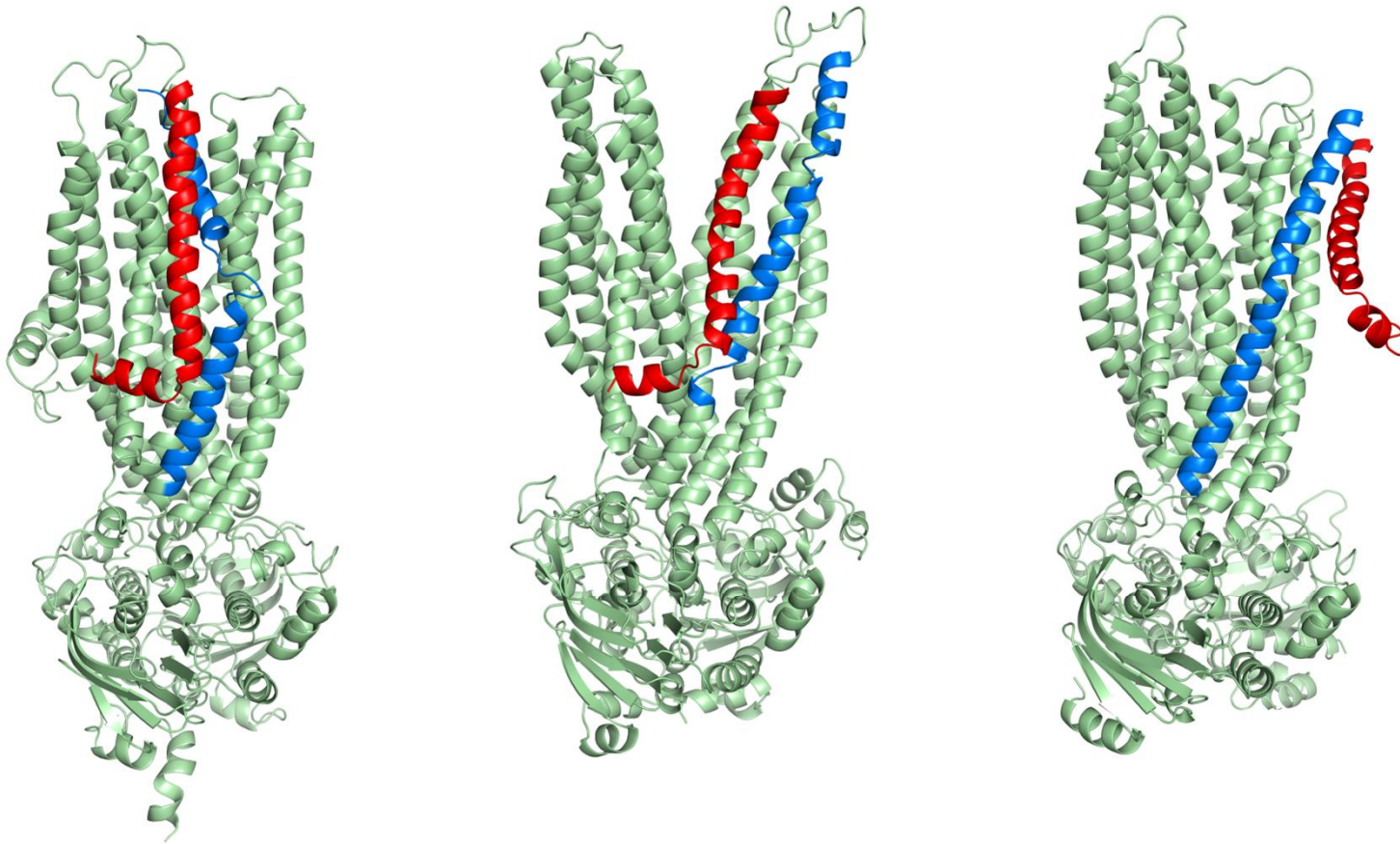


Cisztás fibrózis (CF)



Full-length CFTR structures

Cryo-EM revolution

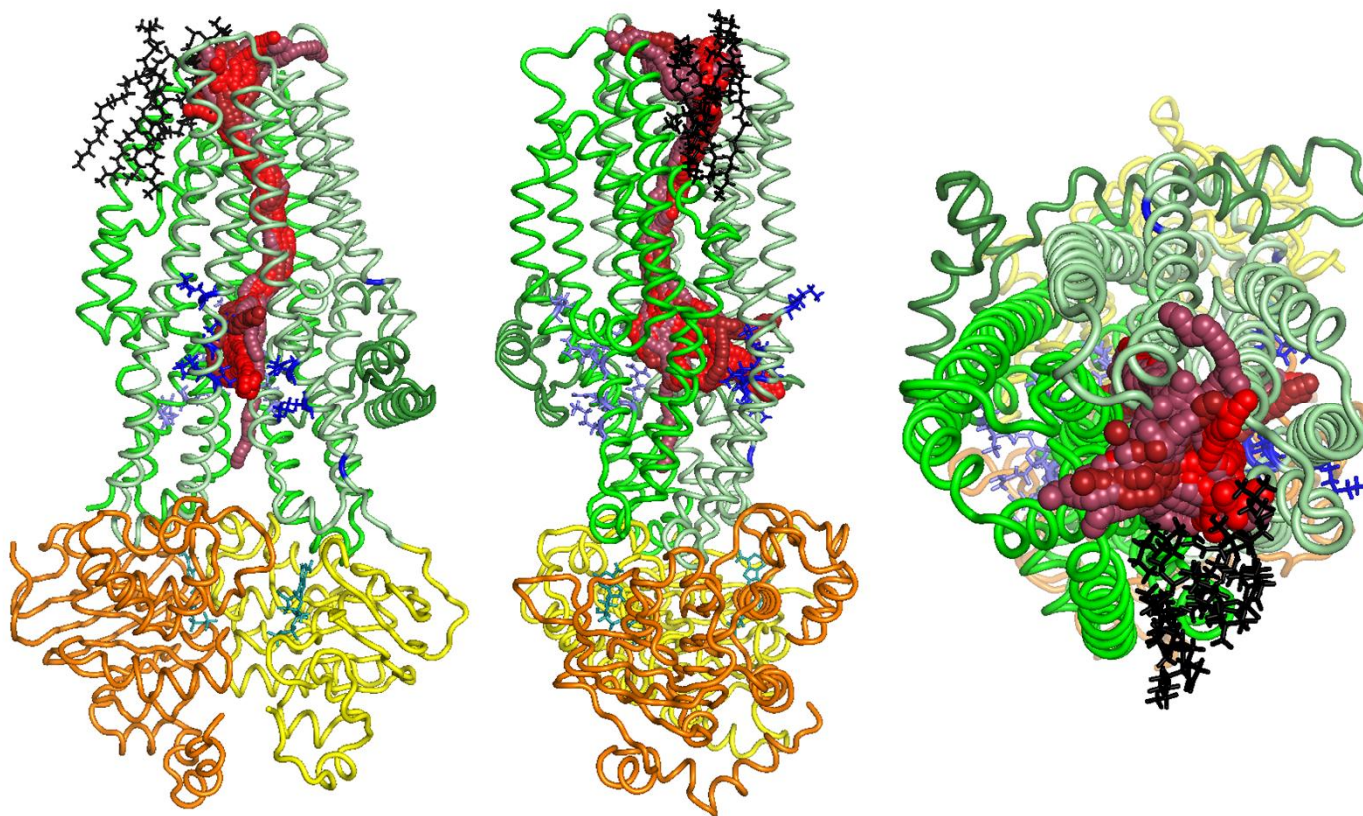


Zhang et al. (2017) Cell 170: 483-491.e8
PDBID:5W81

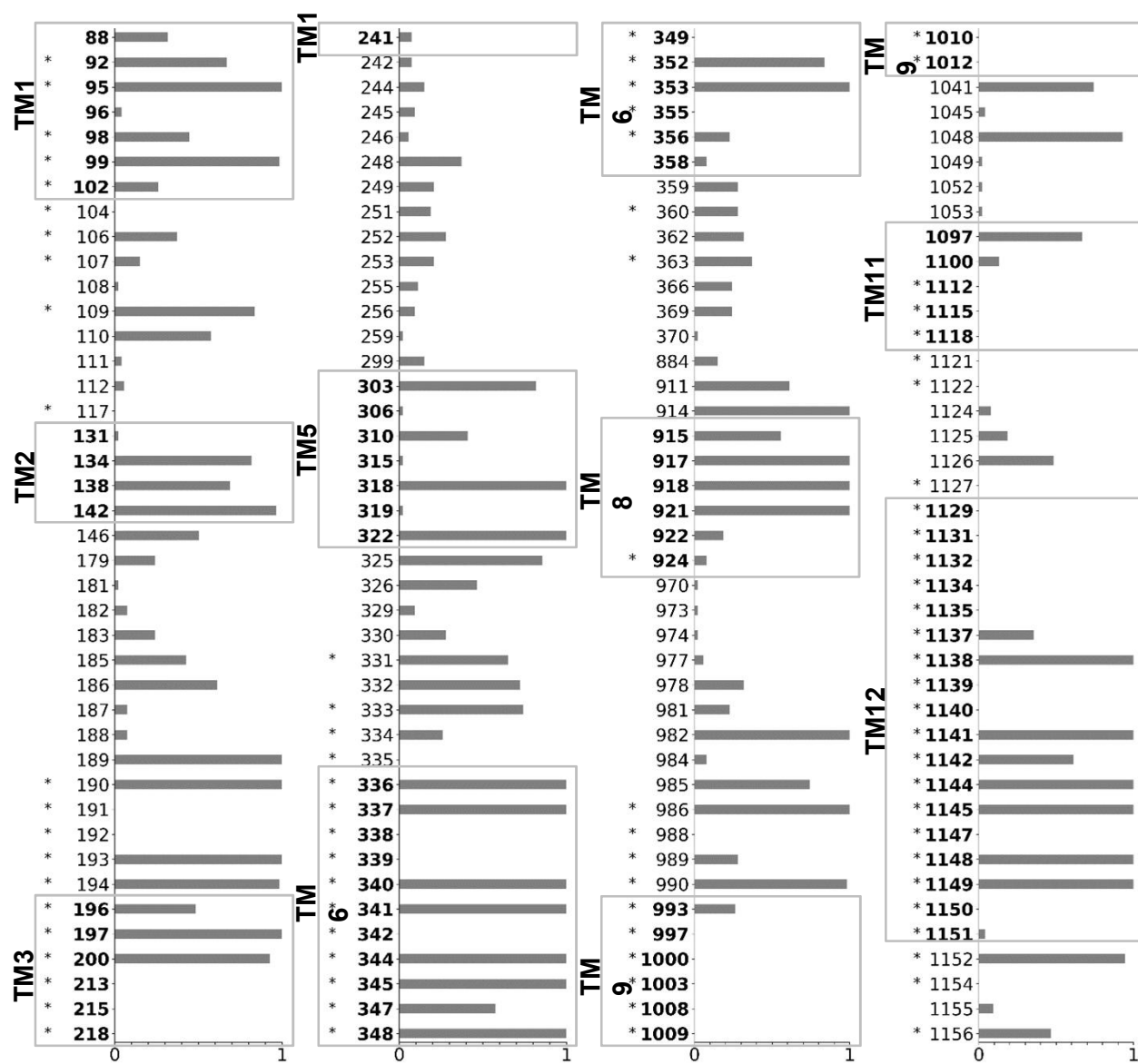
Bob Ford
University of Manchester, UK

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

A klorid csatorna azonosítása

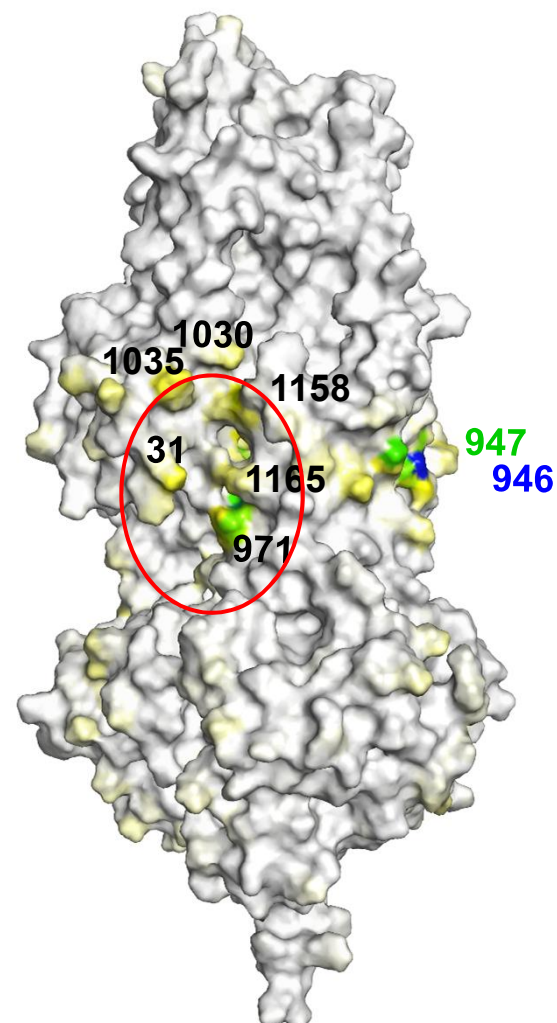
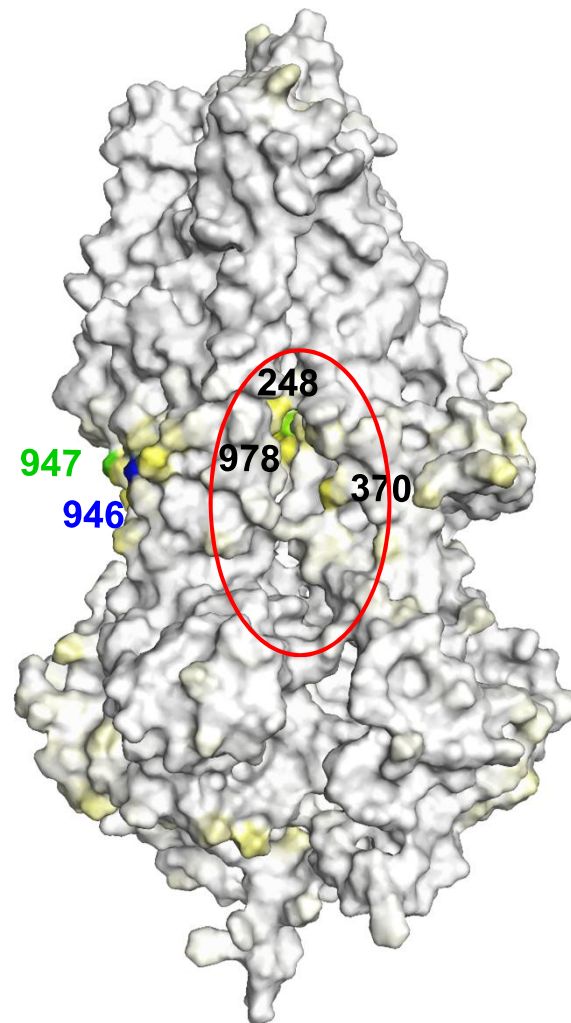


Residue id

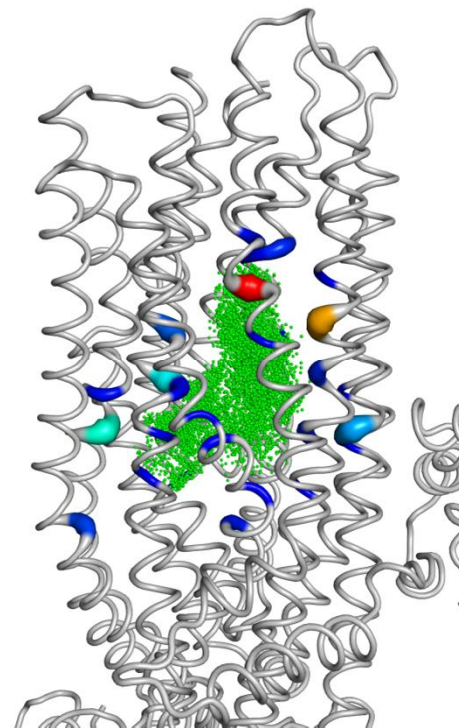
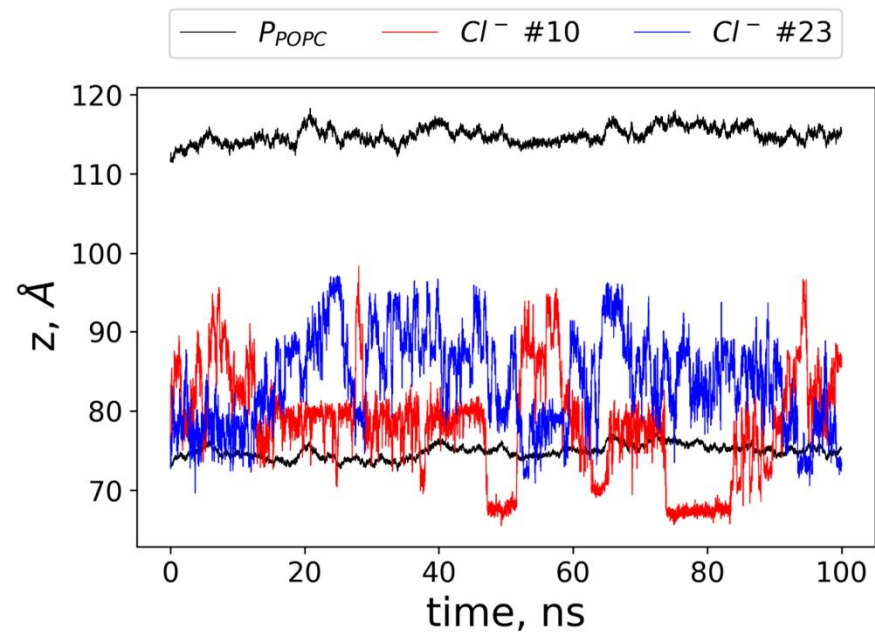


Frequency

A klorid ion belépési pontja

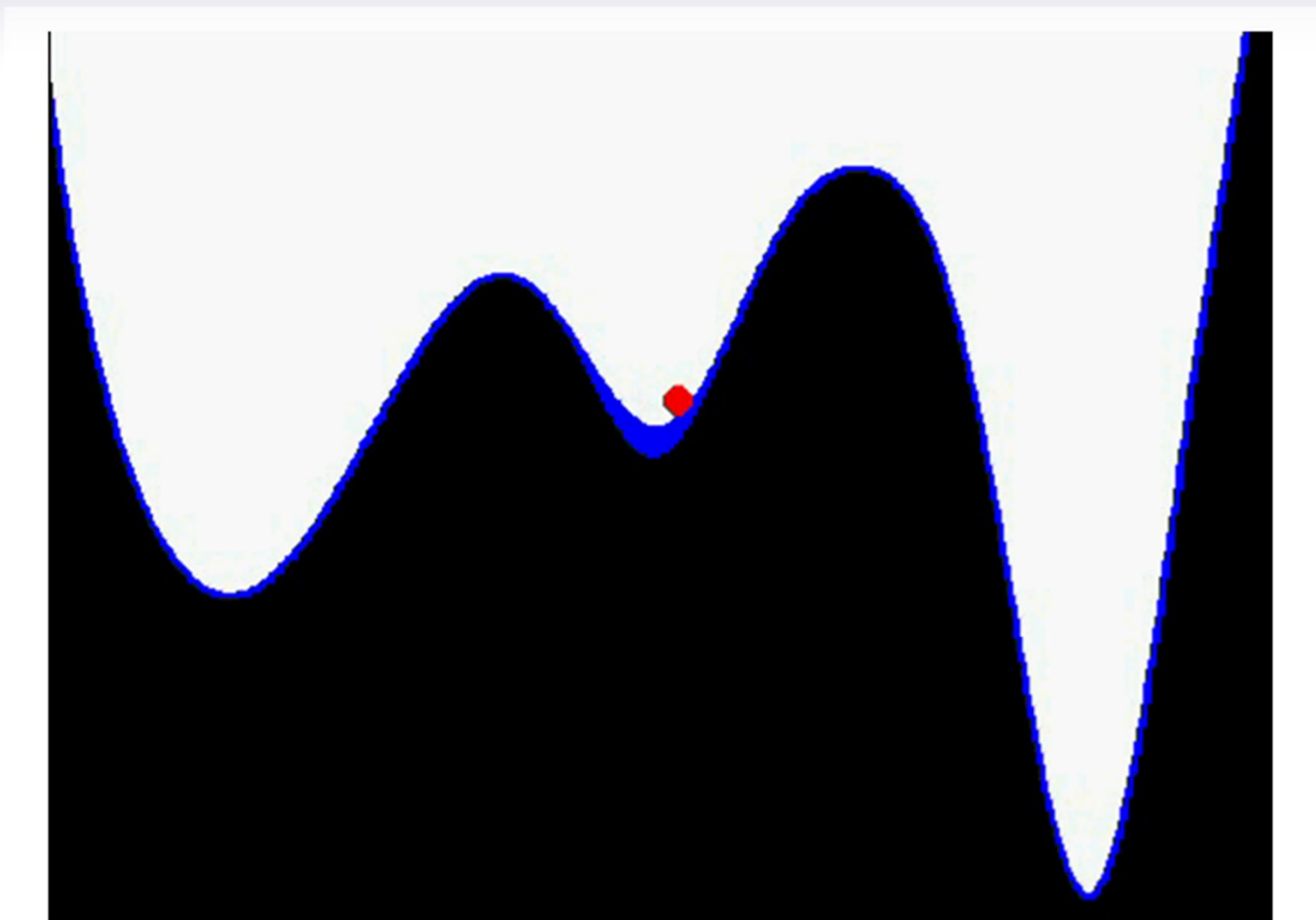


Ionok fehérjén belüli mozgása

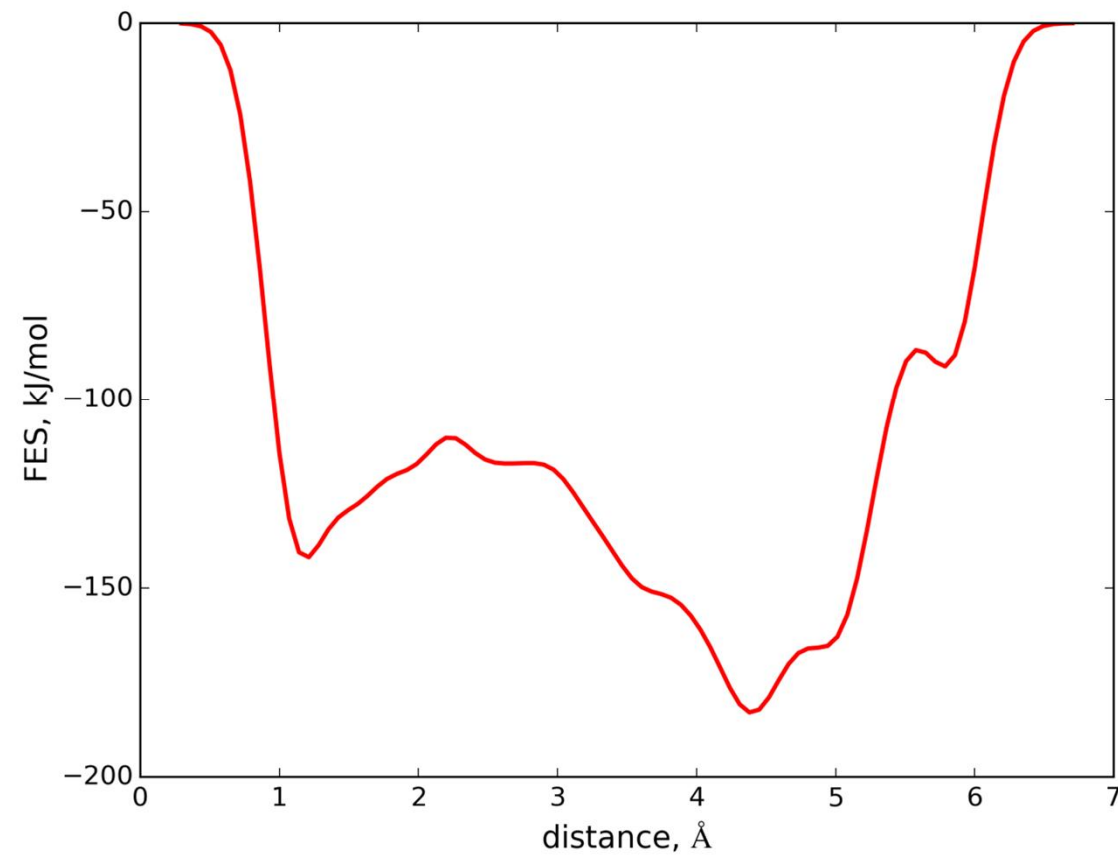


Exploring drug binding sites by biased MD simulations

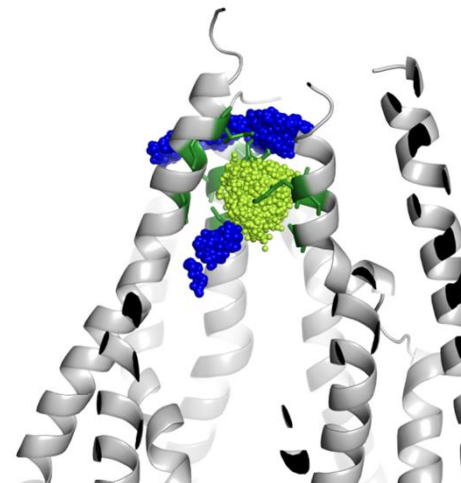
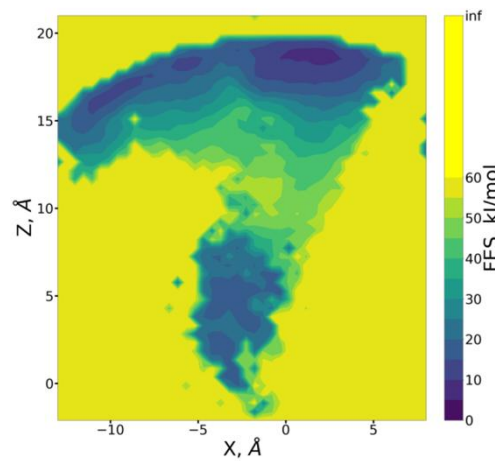
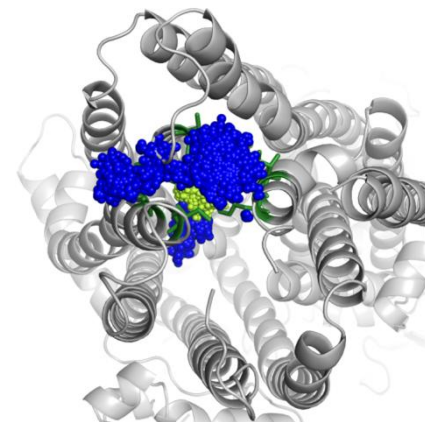
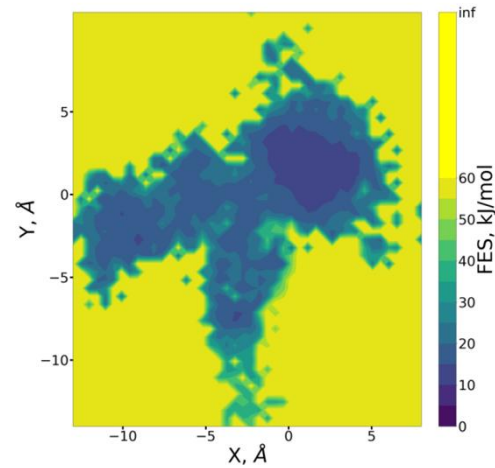
using AutoDock Vina



Szabadenergia felület - FES (Free Energy Surface)



A szűk keresztmetszet jellemzése metadinamikával

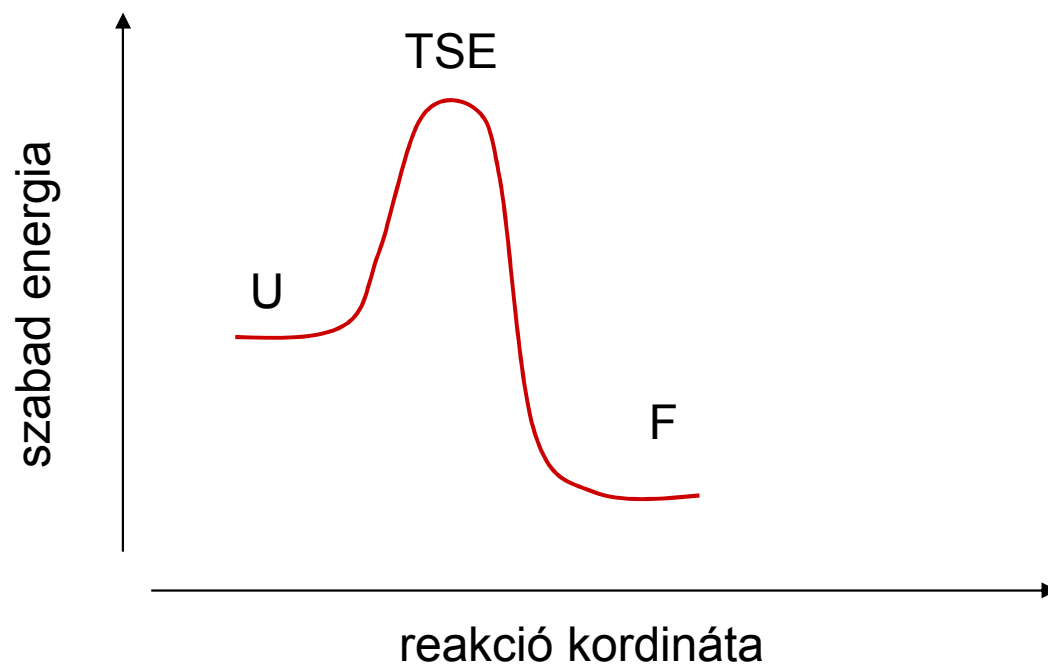


Mai témák

- Bevezetés – a fehérje dinamika és a szimulációk jelentősége
- Fehérjék jellemzése bioinformatikai eszközökkel
- Fehérjék dinamikájának modellezése
- Fehérjék feltekeredésének szimulációja

Kétállapotú fehérje feltekeredés

Two-state model of protein folding



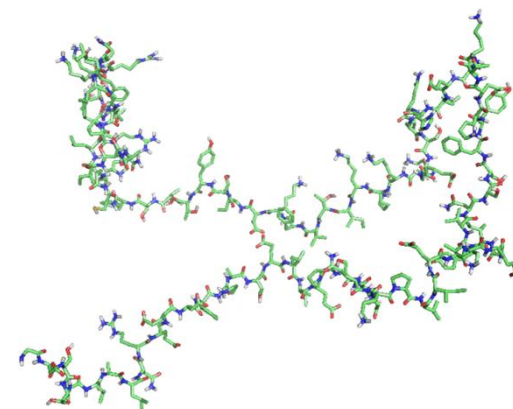
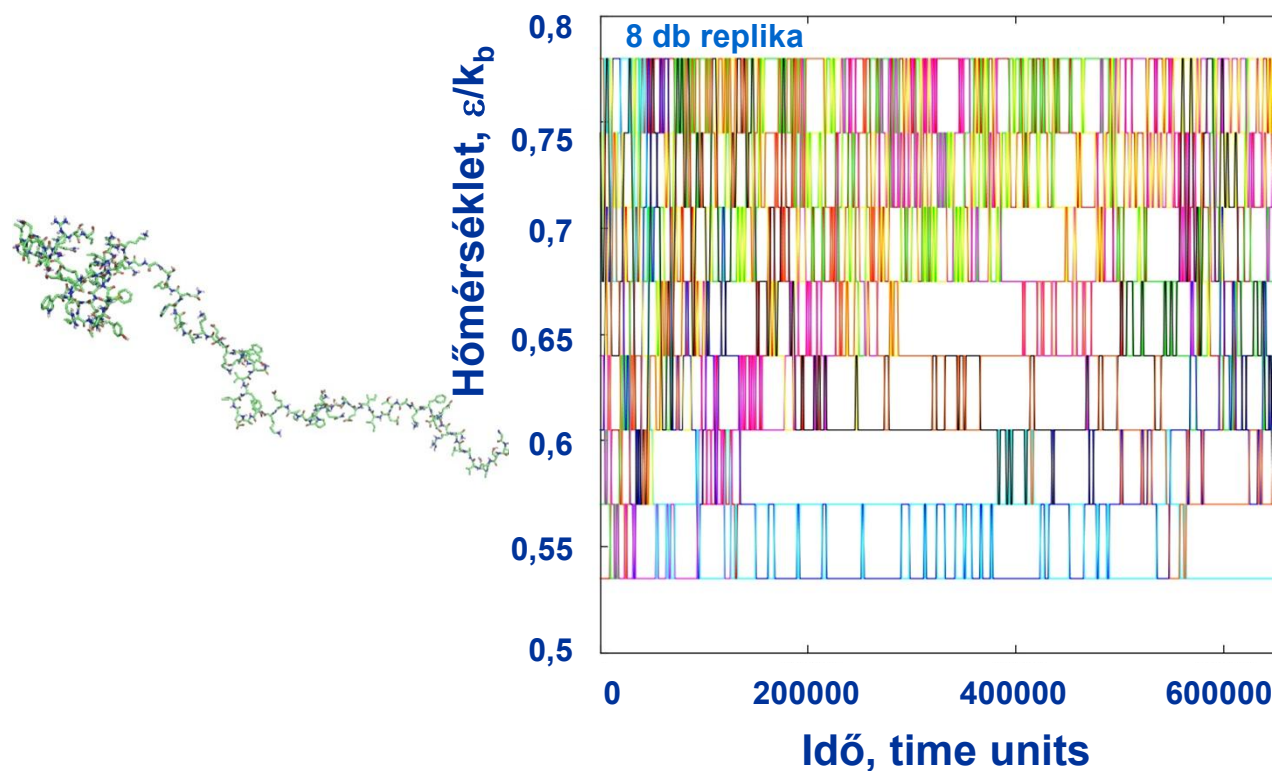
nukleáció

Levinthal paradoxon

Fehérje feltekeredés szimulációja

All atom force-field:

Potenciál függvény számolása erőforrásigényes
Reprezentatív konformációs mintavételezés problémás
Umbrella sampling, [replica exchange](#).



Összefoglalás

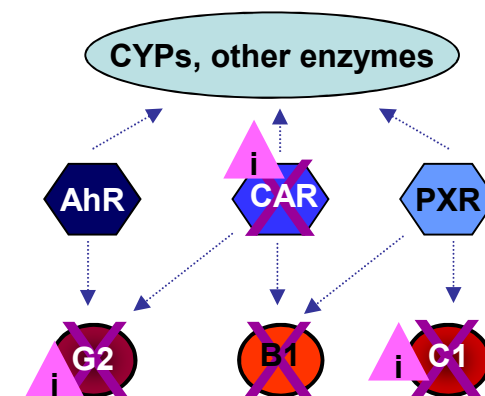
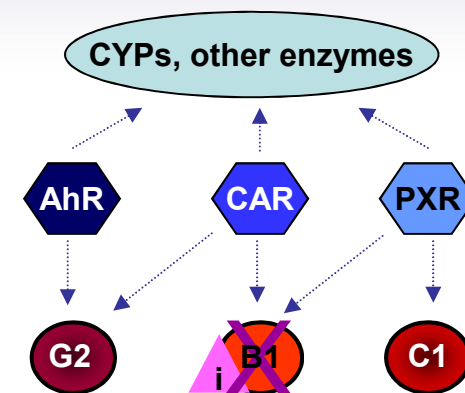
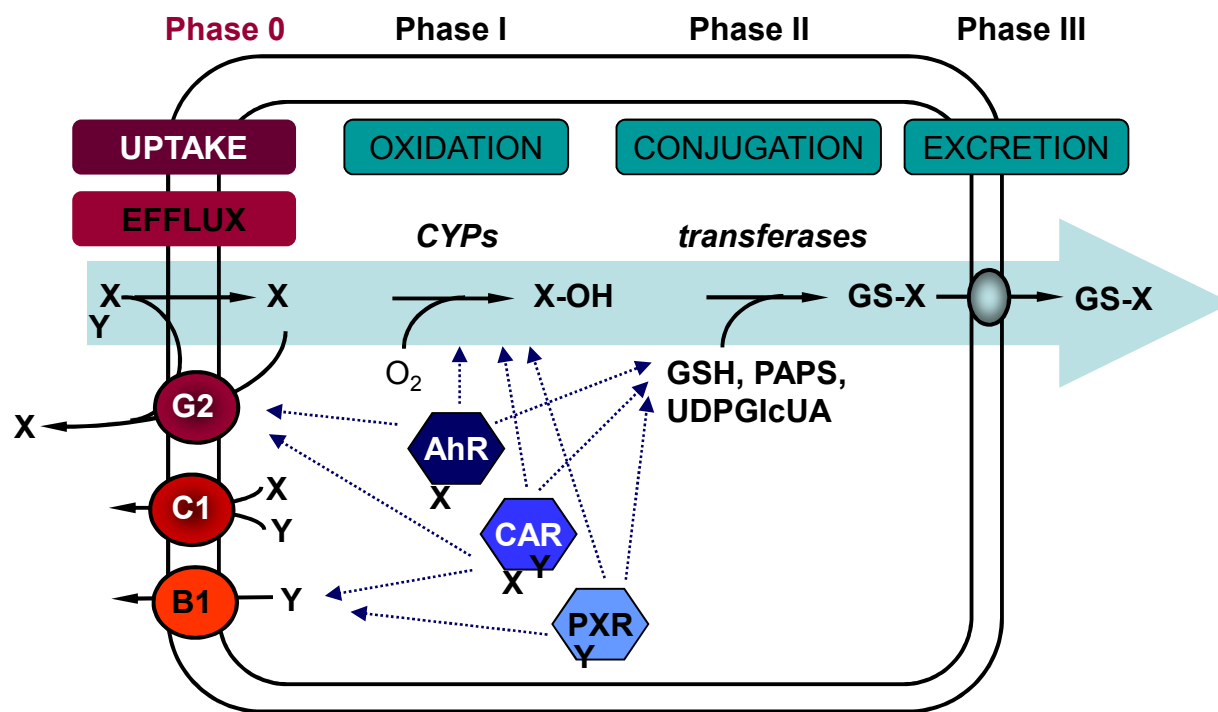
- **Bevezetés – a fehérje dinamika és a szimulációk jelentősége**
- **Fehérjék jellemzése bioinformatikai eszközökkel**
- **Fehérjék dinamikájának modellezése**
- **Fehérjék feltekeredésének szimulációja**

Hálózatok – fehérje hálózatok

- Rendszerbiológia
- Fehérjék-fehérjék kölcsönhatási hálózata
- Gének-fehérjék-drogok kölcsönhatása
- Kapcsolatok aminosavak mozgásaiban

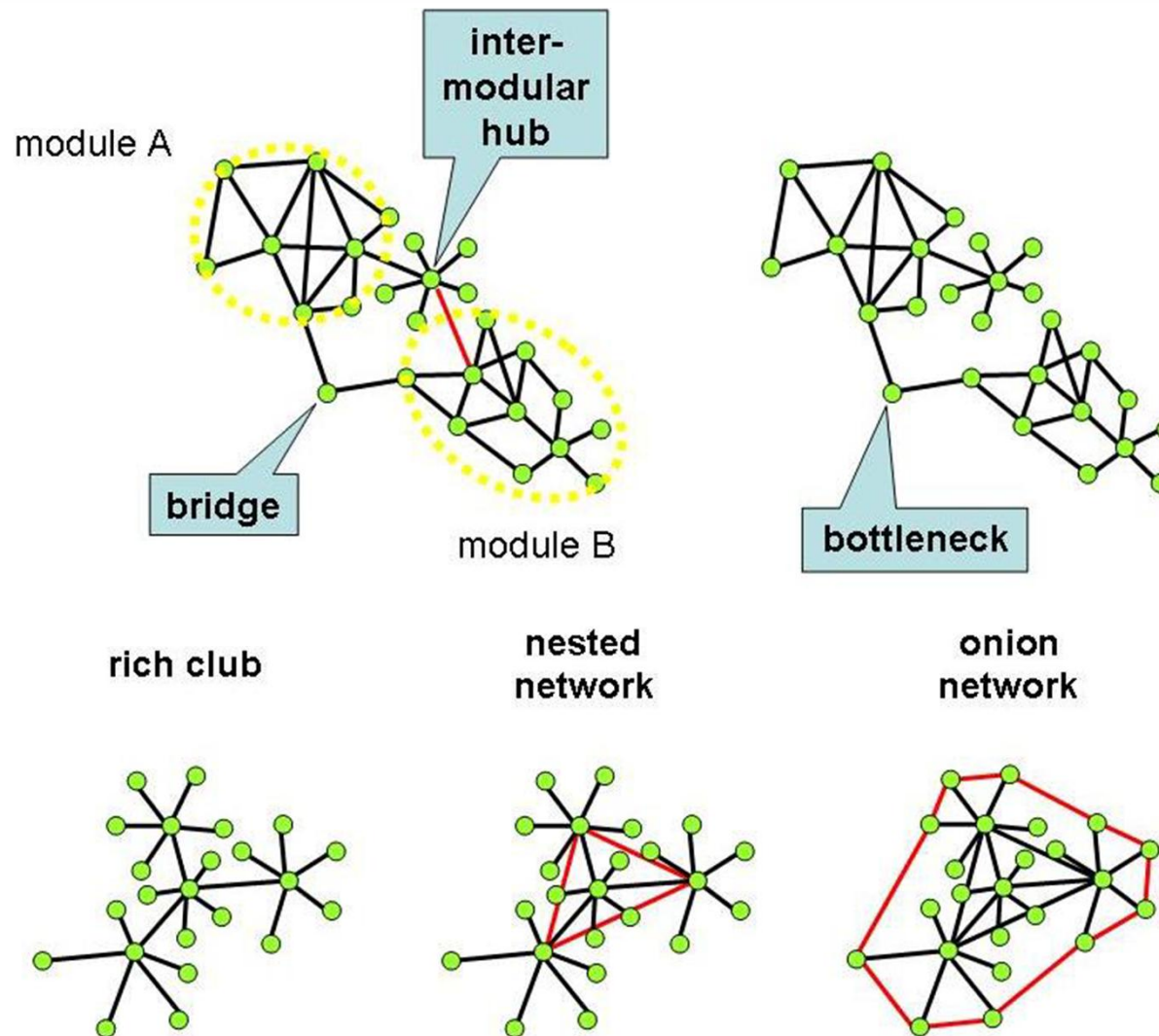
Csermely P. *et al.* 2012, <http://arxiv.org/abs/1210.0330>

A kemoimmun (ChI) rendszer (sejtszintű immunitás)

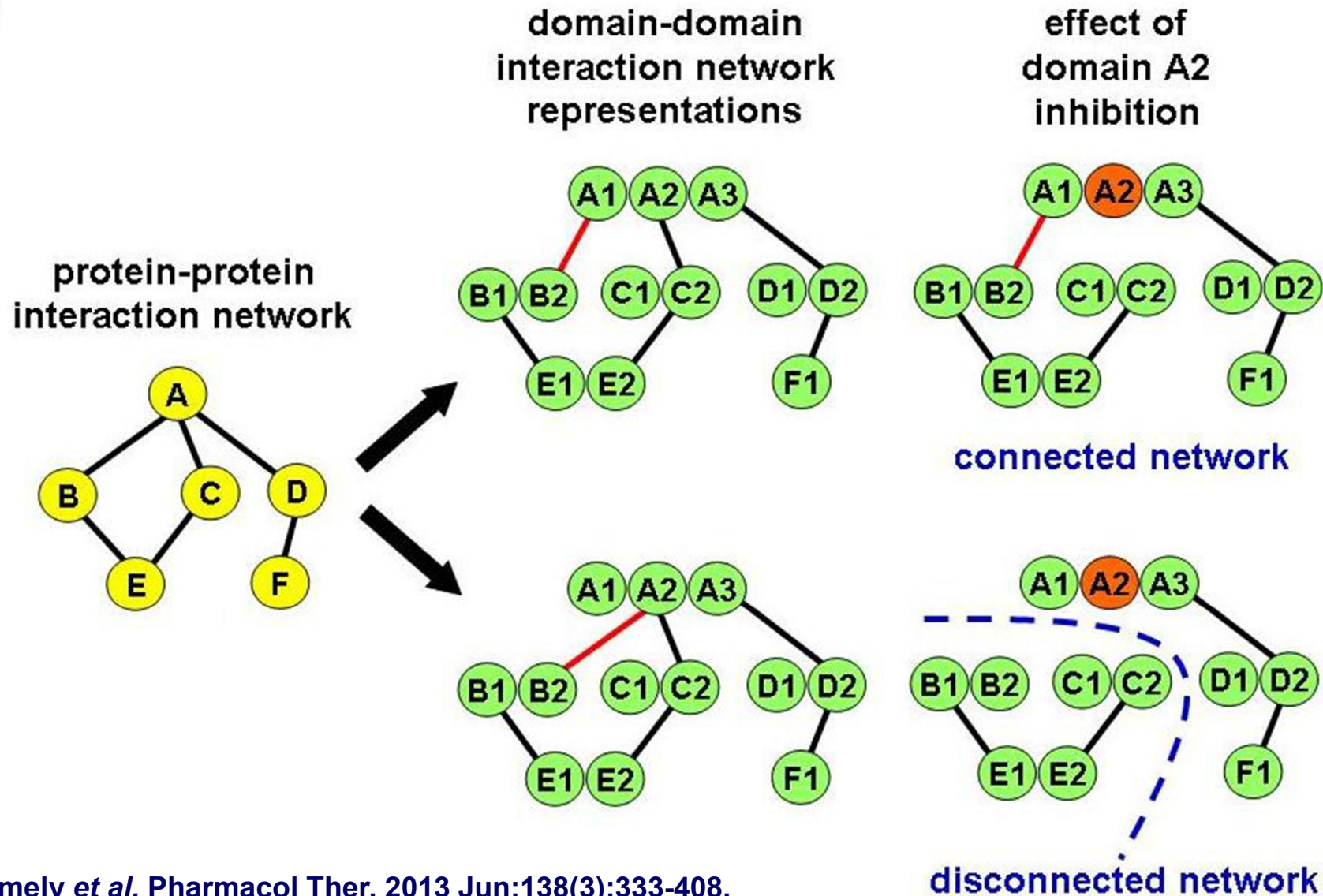


multi-target drugs

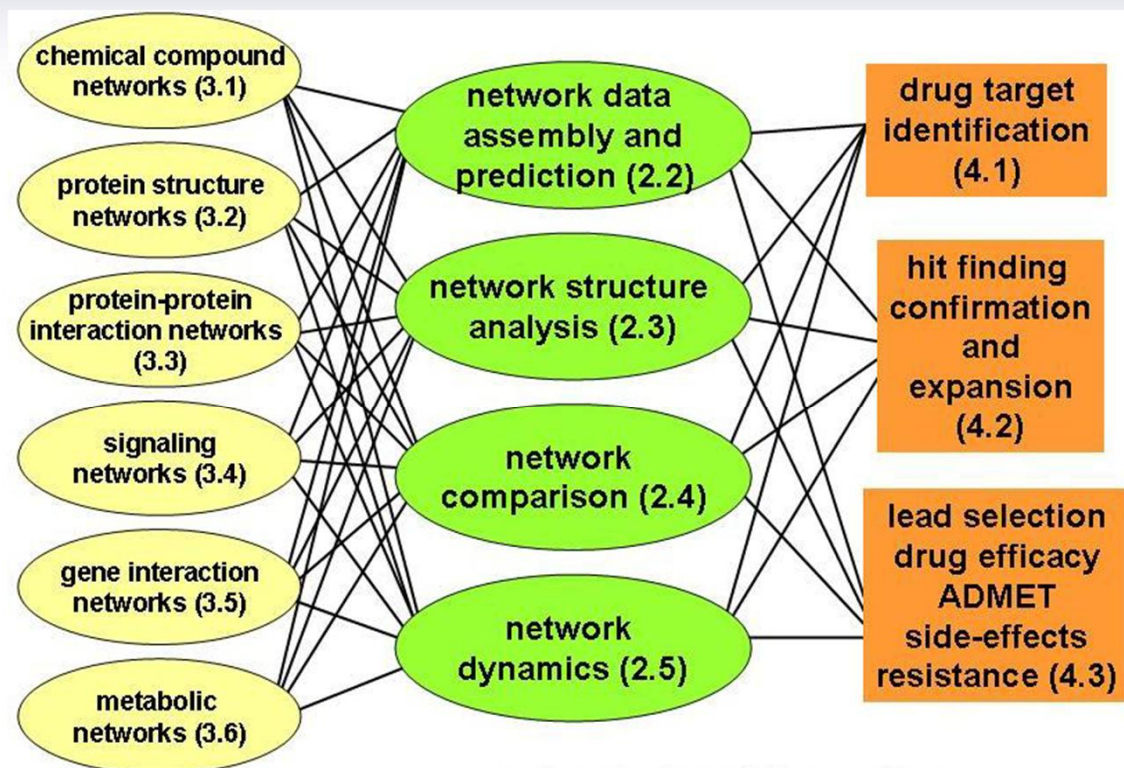
Hálózatok szerkezete



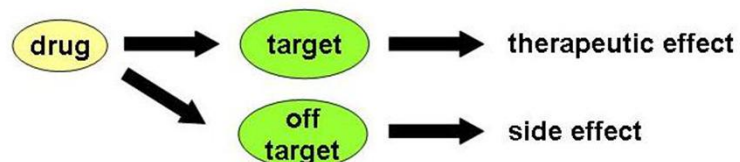
Fehérje-fehérje kölcsönhatási hálózatok



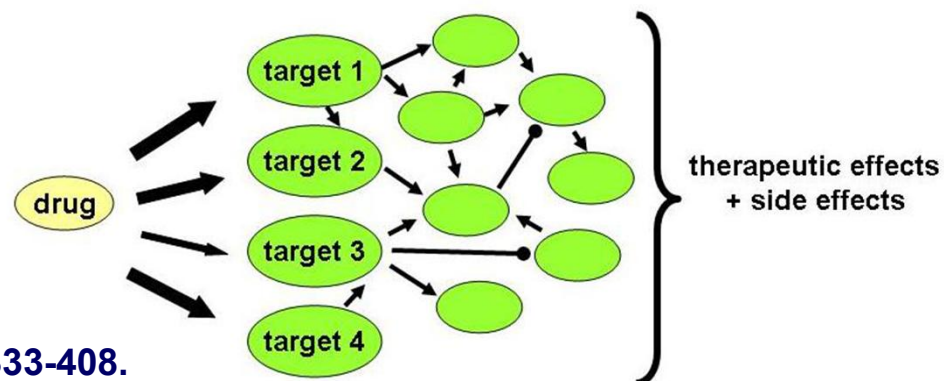
Hálózatok és gyógyszertervezés



classic view of drug action

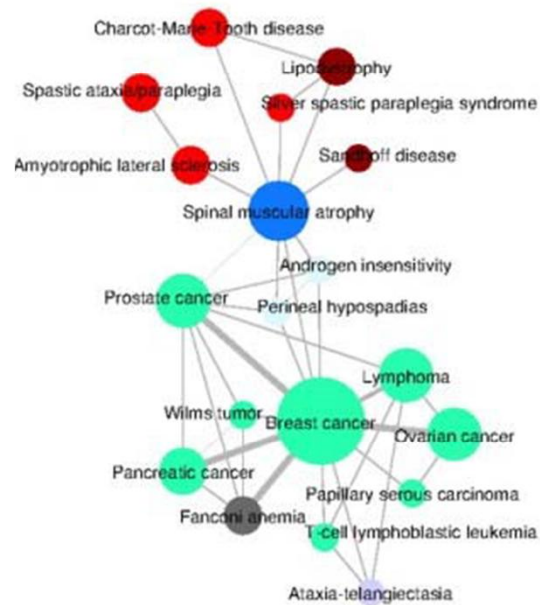


network view of drug action



Network assembly

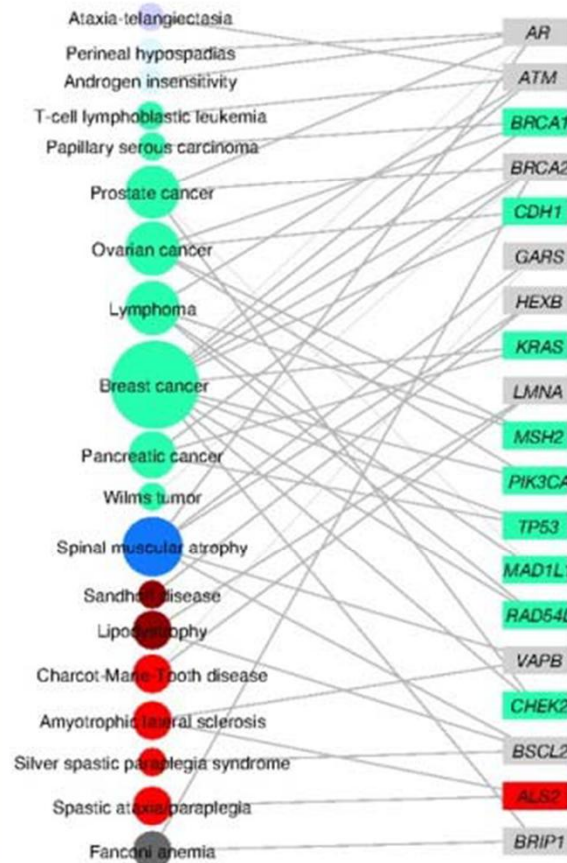
*Human Disease Network
(HDN)*



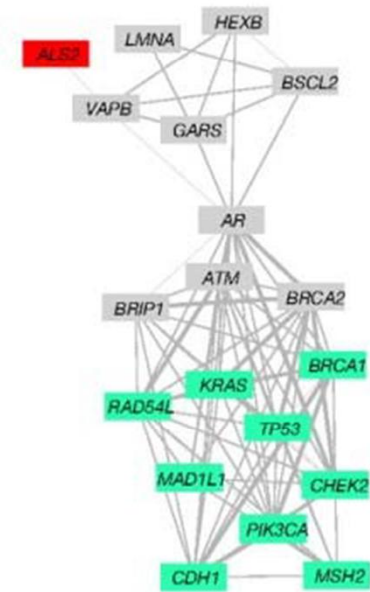
DISEASOME

disease phenotype

disease genome

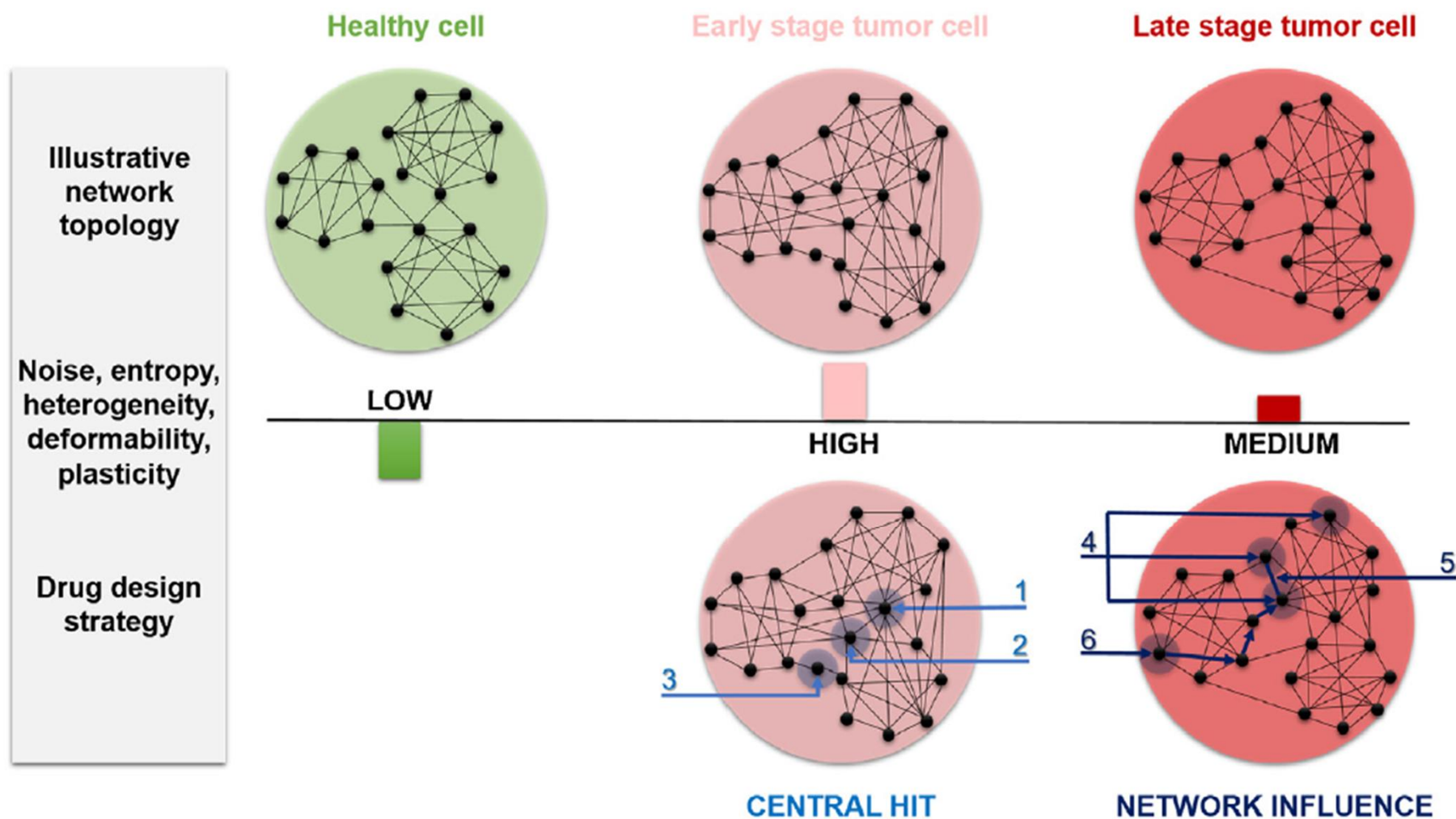


*Disease Gene Network
(DGN)*



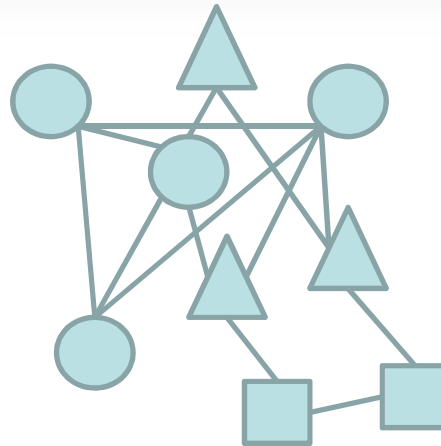
Rákos sejtek hálózatai

D.M. Gyurkó et al. / Seminars in Cancer Biology 23 (2013) 262–269

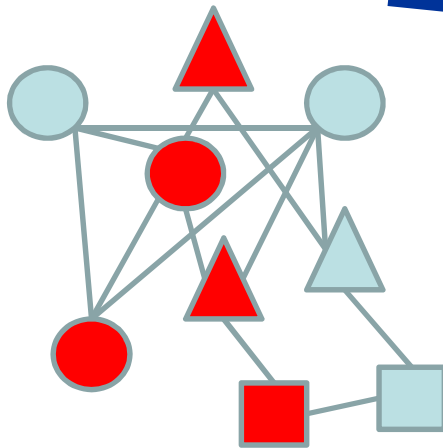


A kemoimmunitási hálózat működésének feltérképezése

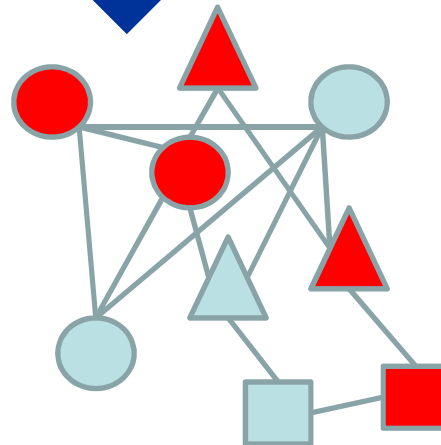
s_i - sejtvonali vagy drog i vagy...



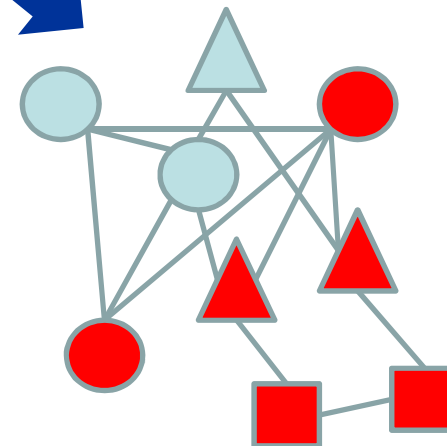
s_1



s_2



s_3



Pipeline of analysis

- Human samples treated with drugs
- NCBI Gene Expression Omnibus (GEO) database
- Preprocessed (by GEO) data & quality check

	our interest	example(s)
DataSet (GDS) ¹	180 (1 335 human)	Anti-cancer agent sapphyrin PCI-2050 effect on lung cancer cell line: dose response (GDS2499)
Experiment ²	883 (2786 cont.+treat.)	treatment: 1) Actinomycin-D 5 ug/ml 2-3) Sapphyrin PCI-2050 1.25 ug/ml, 2.5 ug/ml
Tissue/cell	132	lung cancer cell line, MCF-7, HUVEC, primary fetal astrocytes, tumor biopsies ...
Drug or xenobiotic	222	actinomycin D, sapphyrin PCI-2050, thapsigargin, tunicamycin, doxorubicin ...
Microarray platform (GPL)	26	Affymetrix - Human Genome U133 Plus 2.0 Array (GPL570)

¹Collection of coherent experiments (by GEO)

²One celltype, one agent, one timecourse, one dose

Pipeline of analysis

- Calculate the expression changes
 - Discretization

	Expression change (fold)	Discret value
upregulated	>2x	1
downregulated	<0,5x	-1
no change	-	0
no probe on chip	-	2

– Vectors

[illegible]

Data

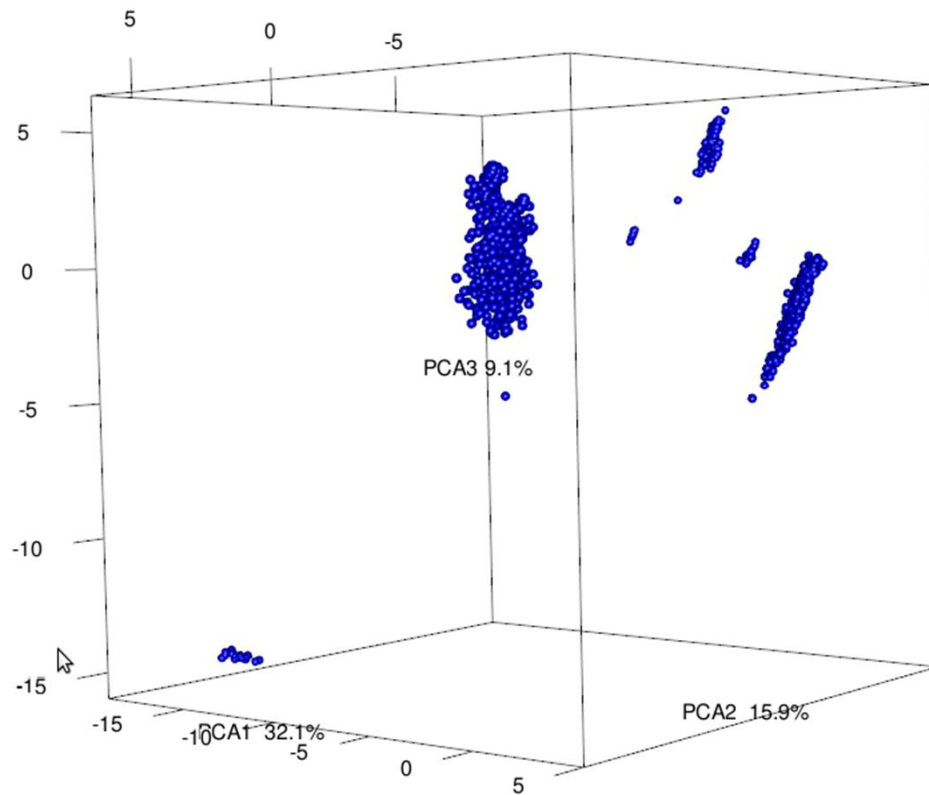
Expression-
changes

Expression-
patterns

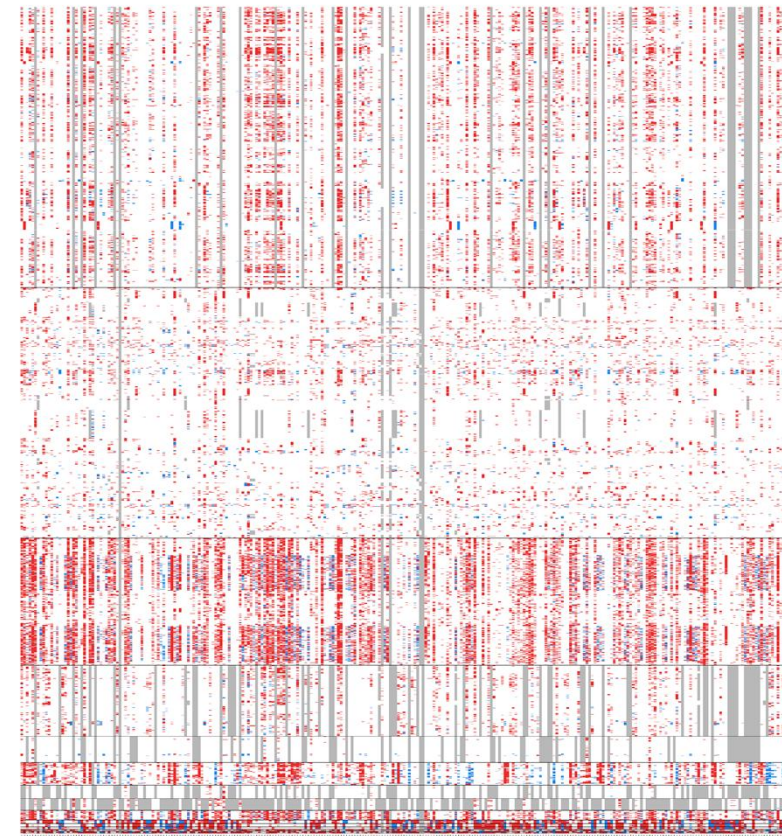
Biochemical
pathways

Whole dataset

PCA analysis (the first 3 component)



Heatmap (result of clustering)



experiments (883)

proteins (280)

Data

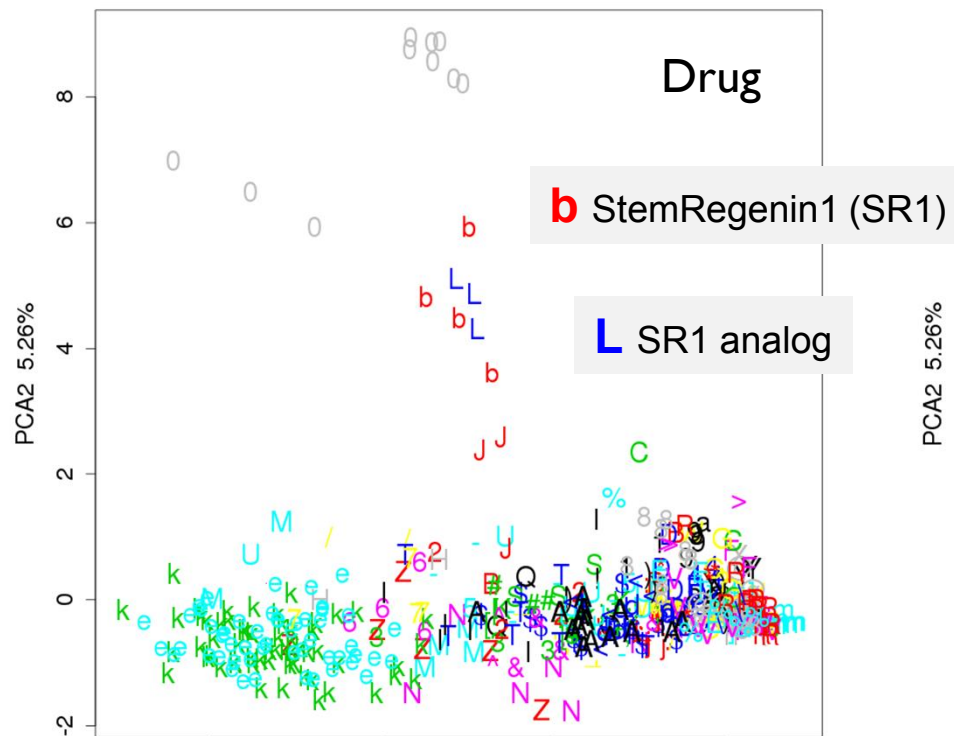
Expression-
changes

Expression-
patterns

Biochemical
pathways

PCA analysis

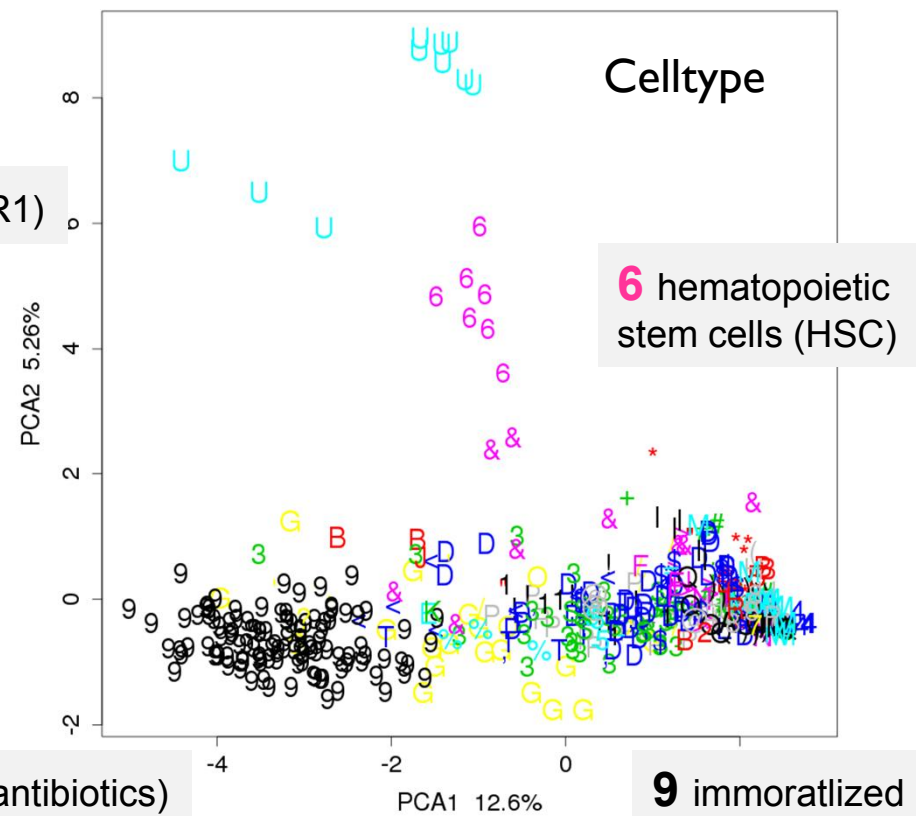
0 BPDE (carcinogen)



e thapsigargin (SERCA inhibitor)

k tunicamycin (antibiotics)

U normal lung WI-38 fibroblasts



Avoiding common pitfalls when clustering biological data

Tom Ronan, Zhijie Qi, Kristen M. Naegle*

www.SCIENCESIGNALING.org 14 June 2016 Vol 9 Issue 432 re6

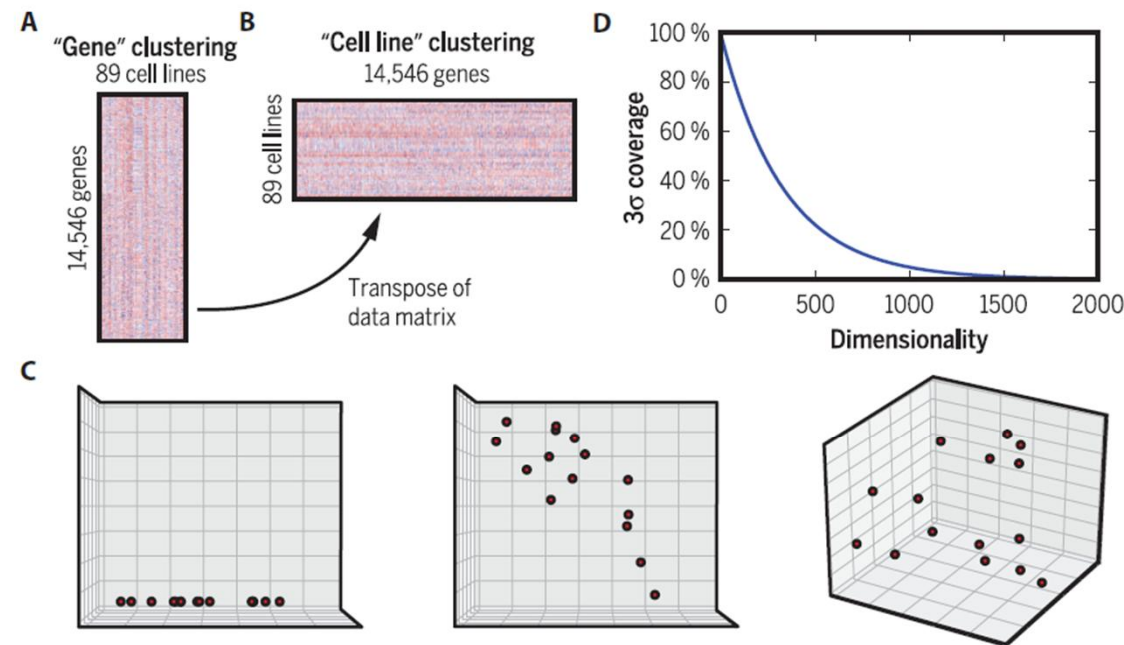


Fig. 1. Determining the dimensionality of a clustering problem. (A and B) Representation of the mRNA clustering problem consisting of >14,000 mRNAs measured across 89 cell lines. Data are from Lu *et al.* (6). When the mRNAs are clustered, the mRNAs are the objects and each cell line represents a feature, resulting in an 89-dimensional problem (A). When attempting to classify normal and tumor cell lines using gene expression, the cells lines are the objects and each mRNA is a feature, resulting in a clustering problem with thousands of dimensions (B). (C) Effect of dimensionality on sparsity. (D) Effect of dimensionality on coverage of the data based on SD from the mean.

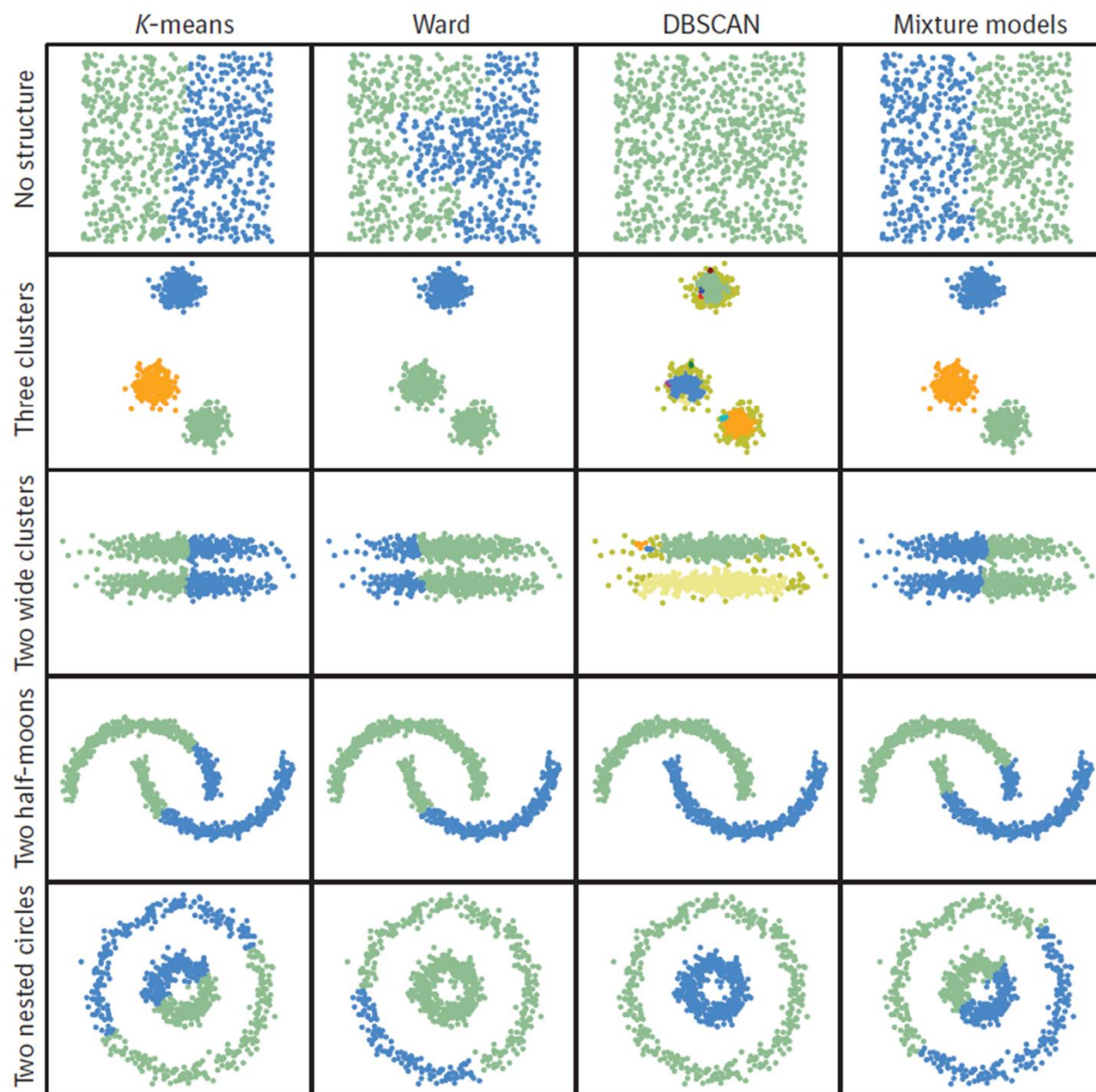
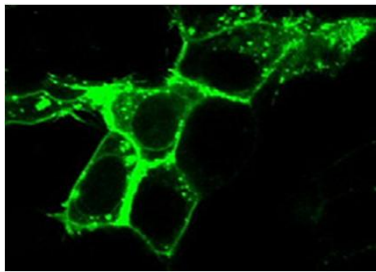


Fig. 4. The effect of algorithm on clustering results. Four toy data sets (<https://github.com/knaegle/clusteringReview>) demonstrate the effects of different types of clustering algorithms on various known structures in two-dimensional data.

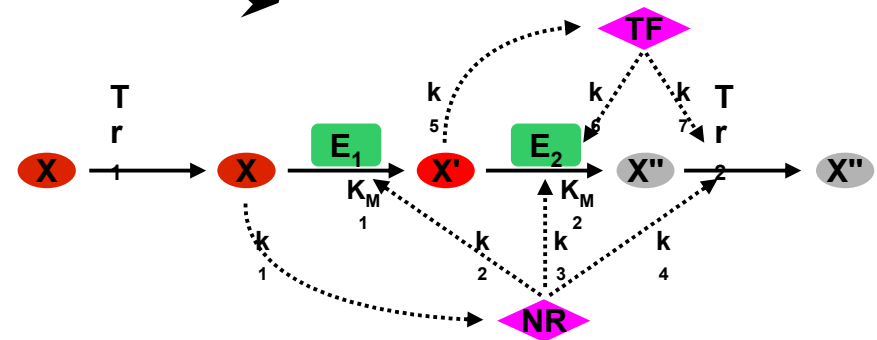
A ChI rendszer kinetikai modellezése



kísérletes kutatások

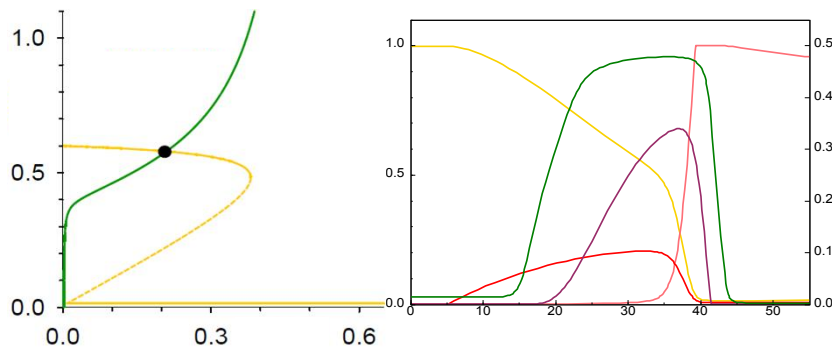


fiziológia



hálózati diagram

fázissík analízis



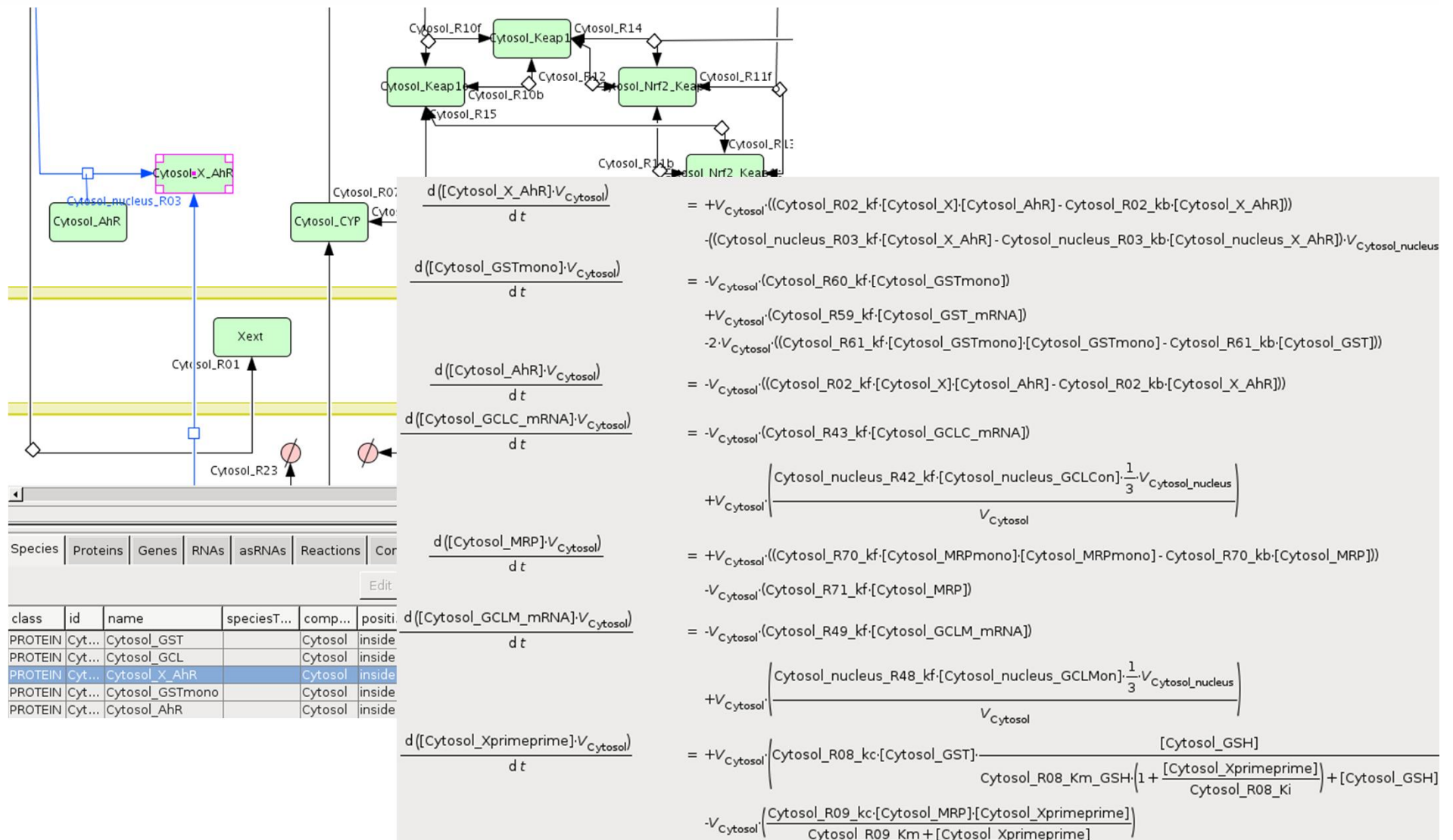
numerikus szimulációk



$$\begin{aligned}\frac{dX'}{dt} &= k_1 X - \frac{k_2 E_2 X'}{K_{M2} + X'} \\ \frac{dX''}{dt} &= \dots \\ \frac{dNR}{dt} &= \dots \\ \frac{dTF}{dt} &= \dots\end{aligned}$$

matematikai modell

Mathematical modeling to understand relationships in the chemodense network



Összefoglalás


- Fehérjék dinamikájának, feltekeredésének modellezése
- Hálózatok, gének, fehérjék, drogok
- tamas@hegelab.org

Informatikai "szemlélet"

- Adat
- Logika
- Megjelenítés

Biológiai adatbázisok

<http://www.ncbi.nlm.nih.gov/>



National Center for
Biotechnology Information

[Resources](#) ☒ [How To](#) ☒

[Sign in to NCBI](#)

All Databases

Recent

All

All Databases

Assembly

BioProject

BioSample

BioSystems

Books

ClinVar

Clone

Conserved Domains

dbGaP

dbVar

Epigenomics

EST

Gene

Genome

GEO DataSets

GEO Profiles

NCBI

er for Biotechnology Information advances science and health by providing access to biomedical
mation.

[Mission](#) | [Organization](#) | [Research](#) | [NCBI News](#)

alyze data using NCBI software
: Get NCBI data or software
earn how to accomplish specific tasks at NCBI
s: Submit data to GenBank or other NCBI databases

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[PubMed Health](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI Announcements

dbVar releases 1000 Genomes Phase 3
structural variants
Nov 4, 2014

dbVar has released structural variation
(SV) data from developmental delay
study cited in Nature Reviews Genetics
Nov 3, 2014

dbVar releases copy number variation
(CNV) data from developmental delay
study cited in Nature Reviews Genetics
Nov 3, 2014

NCBI YouTube channel

Learn how to get the most out of NCBI
tools and databases with video tutorials
on the NCBI YouTube Channel.

GO



1

2

3

4

5

6

7

8

Biológiai adatbázisok

<http://expasy.org/>



ExPASy

Bioinformatics Resource Portal

Query all databases



search

Visual Guidance

Categories

- proteomics
- genomics
- structural bioinformatics
- systems biology
- phylogeny/evolution
- population genetics
- transcriptomics
- biophysics
- imaging
- IT infrastructure
- drug design

Resources A..Z

Links/Documentation

ExPASy is the **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software tools (i.e., *resource* proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see **Categories** in the left menu). On this SIB groups as well as external institutions.

Featuring today

MADAP

Clustering tool for the interpretation of one-dimensional genome annotation data mapped onto complete or partial genome sequences.

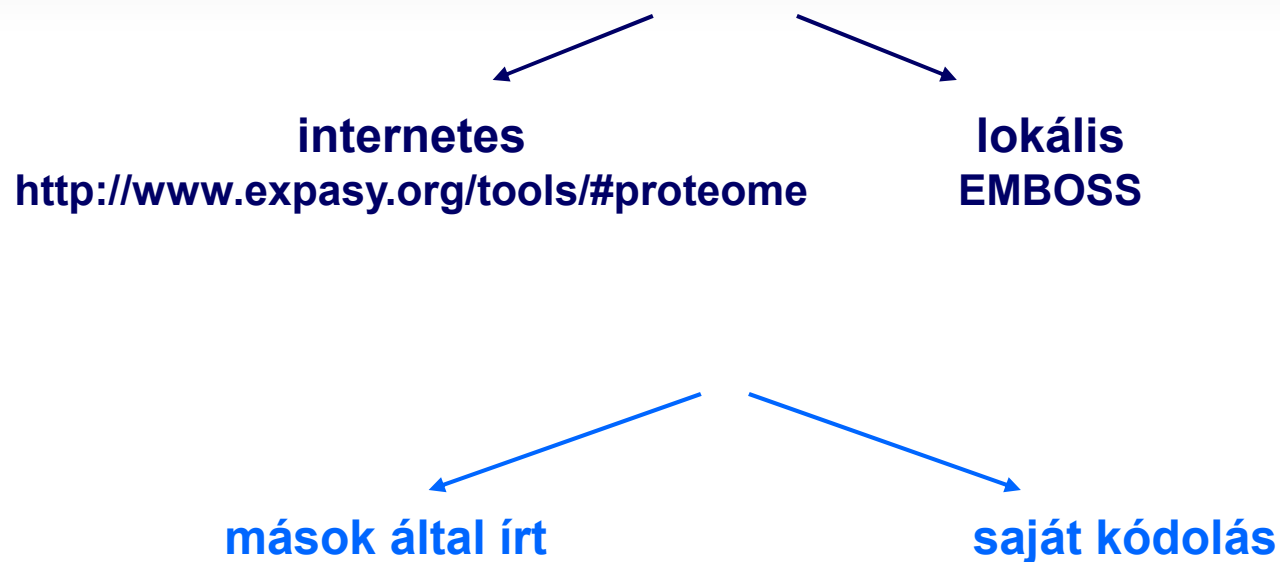
[\[details\]](#)



How to use this portal?

- Features and updates
- New to ExPASy
- Experienced ExPASy users: what is different

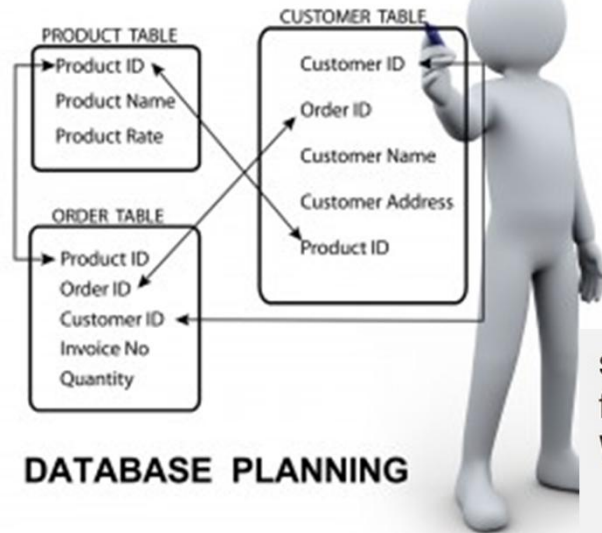
Programok, inter-fészek[☺]



REST (Representational state transfer)

[http://pubchem.ncbi.nlm.nih.gov/rest/pug/<input specification>/<operation specification>/\[<output specification>\]\[?<operation_options>\]](http://pubchem.ncbi.nlm.nih.gov/rest/pug/<input specification>/<operation specification>/[<output specification>][?<operation_options>])

RDBMS



select A.id, A.lname, B.type, B.number, C.target, C.duration
 from contact A, phones B, phonestx C
 Where A.id = B.id and B.number = C.number

id	lname	type	number	target	duration
g9	Moschetti	home	1-900-555-1212	1-222-707-7070	23
g10	Kalan	work	1-999-444-9999	1-222-907-7071	7
g9	Moschetti	work	1-800-989-2231	1-987-707-7072	9
g9	Moschetti	home	1-900-555-1212	1-222-707-7071	7
g9	Moschetti	home	1-777-999-1212	1-222-807-7070	23
g9	Moschetti	home	1-777-999-1212	1-222-807-7071	7
g10	Kalan	work	1-999-444-9999	1-222-907-7070	23
g9	Moschetti	home	1-900-555-1212	1-222-707-7072	9
g10	Kalan	work	1-999-444-9999	1-222-907-7072	9
g9	Moschetti	home	1-777-999-1212	1-222-807-7072	9

How to turn this into a list of names –
each with a list of numbers, **each of those** with a list of target
 numbers?

ORM

<p>A)</p> <pre>CREATE TABLE `protein` (`protein_id` int(11) NOT NULL AUTO_INCREMENT, `gene_name` varchar(300) DEFAULT NULL, `acc` varchar(300) DEFAULT NULL, `entry_name` varchar(300) DEFAULT NULL, `organism` varchar(300) DEFAULT NULL, PRIMARY KEY (`protein_id`), UNIQUE KEY `acc` (`acc`)) ENGINE=InnoDB AUTO_INCREMENT=12 DEFAULT CHARSET=latin1\$\$</pre>	<p>B)</p> <pre>class Protein(Base): __tablename__ = "protein" protein_id = Column(Integer, primary_key=True, autoincrement=True) gene_name = Column(String(300)) acc = Column(String(300), unique=True) entry_name = Column(String(300)) organism = Column(String(300)) # relationships bindings=relationship("ProteinBinding", backref="protein") ...</pre>
<p>C)</p> <pre>SELECT `protein`.`protein_id`, `protein`.`gene_name`, `protein`.`acc`, `protein`.`entry_name`, `protein`.`organism` FROM `drugdb`.`protein` WHERE `protein`.`acc`='075469';</pre>	<p>D)</p> <pre>result=DBSession.query(Protein).filter(Protein.acc=='075469').all() for a in result: print str(a.protein_id)+'\t' \ +str(a.gene_name)+'\t' \ +str(a.acc)+'\t' \ +str(a.entry_name)+'\t' \ +str(a.organism)+'\n'</pre> <p>print result[0]</p>

Informatikai "szemlélet"

- **Adat** - RDBMS, ORM
- **Logika** - Python
- **Megjelenítés** - browser, templating

KÖSZÖNÖM A FIGYELMET!

➤ tamas@hegelab.org