

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 jellemzése bioinformatikai eszközökkel**
- **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.

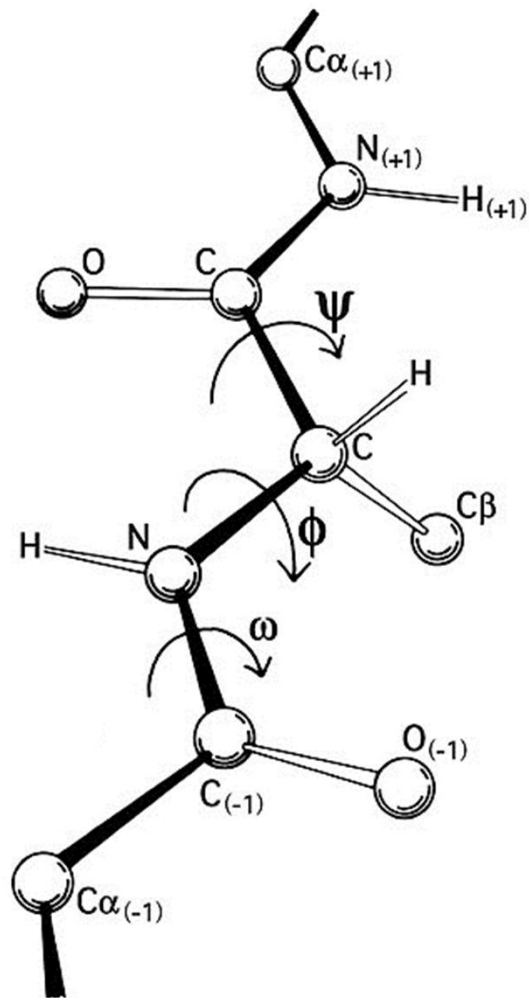
**Kísérletes módszerek általában nem
szolgáltatnak közvetlen információt az
atomi szintű történésekről.**

Pl. NMR és MD - igen

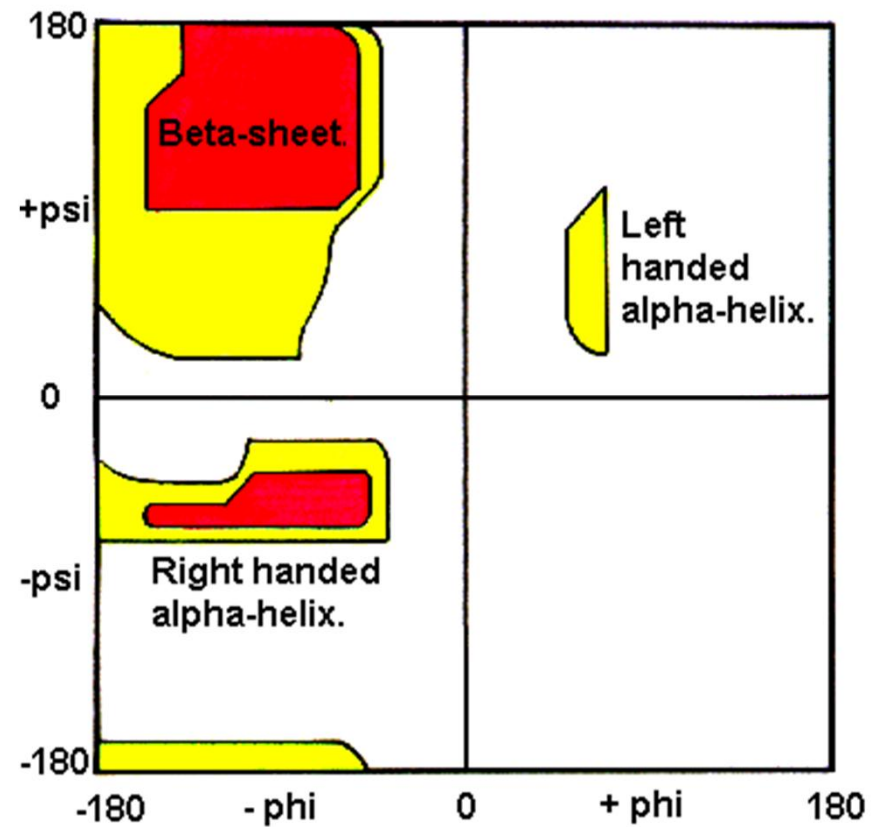
Mai témák

- **Bevezetés – szimulációk és a fehérje dinamika jelentősége**
- **Fehérjék jellemzése bioinformatikai eszközökkel**
 - Másodlagos szerkezeti mintázatok jóslása
 - Rendezetlen fehérjék
 - Funkcionális régiók azonosítása
 - A harmadlagos és negyedleges szerkezet

Másodlagos szerkezeti elemek



The Ramachandran Plot.



wikipedia

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:	K⁺ csatorna inaktiválása
Effektor:	peptid inhibitorok
Scavangers:	kazein
Összeszerelődés:	calmodesmon, F-aktin
Bemutató felület:	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:

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

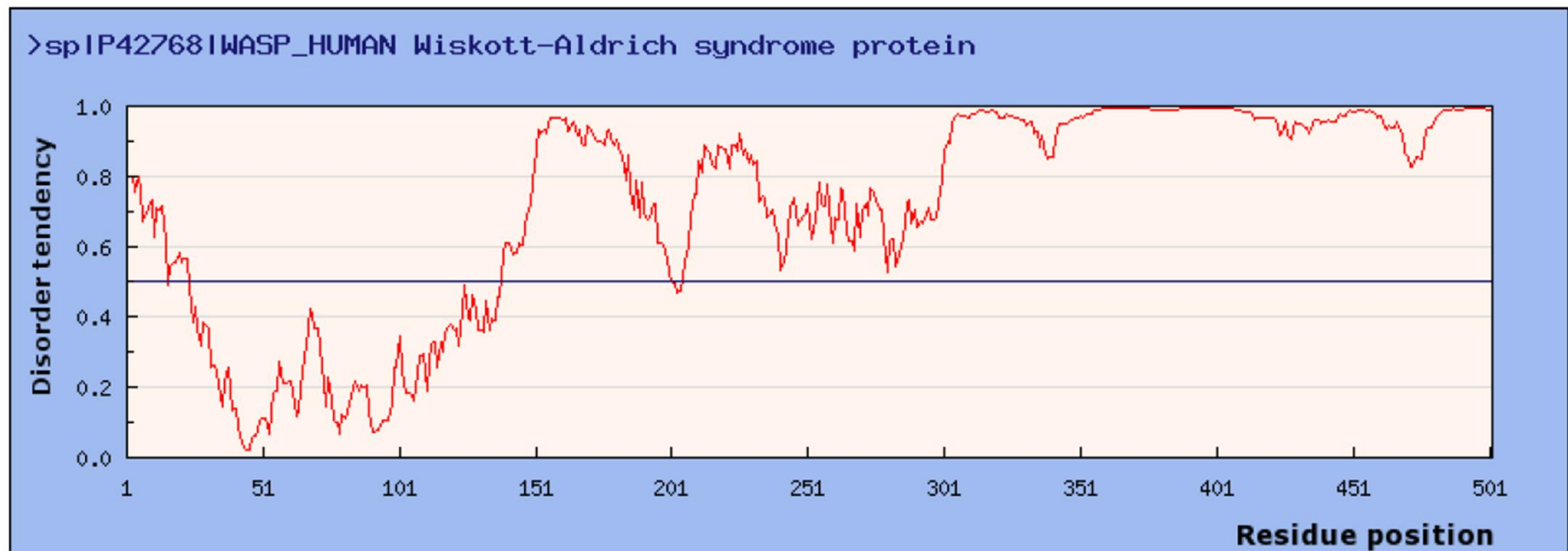
Ha csak szekvencia van:

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

Egy aminosav rendezetlensége:

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

IUPred kimenete



Mai témák

- **Bevezetés – szimulációk és a fehérje dinamika jelentősége**
- **Fehérjék jellemzése bioinformatikai eszközökkel**
 - Másodlagos szerkezeti mintázatok jóslása
 - Rendezetlen fehérjék
 - Funkcionális régiók azonosítása
 - A harmadlagos és negyedleges szerkezet

Funkcionális régiók azonosítása

Mintázat keresés (pattern search)

P-x-[STA]-x-[LIV]-[IVT]-x-[GS]-G-Y-S-[QL]-G

P.[STA].[LIV][IVT].[GS]GYS[QL]G (regular expression pattern)

Konszenzus matrix, profile (lsd. ProSite dokumentációt)

MA /M: SY='D'; M=-10,26,-29,38,34,-34,-14,-2,-33,7,-24,-23,8,-6,8,-4,0,-9,-27,-33,-19,21;

MA /M: SY='I'; M=-8,-31,-23,-35,-28,7,-32,-27,27,-24,15,13,-27,-26,-24,-23,-20,-9,25,-4,2,-27;

MA /M: SY='R'; M=-11,-12,-26,-12,-1,-13,-23,-1,-8,1,-7,-3,-8,-11,-2,8,-9,-6,-8,-22,-3,-4;

MA /M: SY='E'; M=-11,17,-27,23,29,-24,-15,-3,-27,1,-22,-20,9,-1,6,-6,3,-4,-25,-32,-17,17;

MA /M: SY='D'; M=-7,10,-23,11,2,-25,0,-6,-26,-4,-23,-18,7,-6,-5,-8,7,7,-20,-31,-17,-2;

Már mások megtették, adatbázisokba gyűjtötték ☺

ProSite (<http://prosite.expasy.org/>)

Enzimek osztályozása (EC)

Domének azonosítása (pl. Pfam: <http://pfam.sanger.ac.uk>)

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

[illegible]

Basic Local Alignment Search Tool (BLAST)

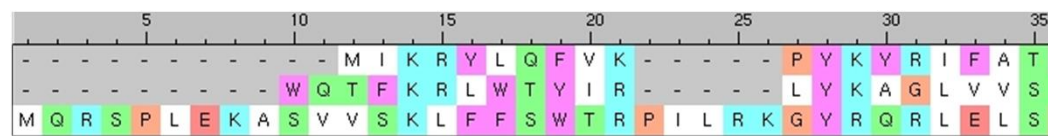
CLUSTAL W (1.83) multiple sequence alignment

Alignement – pl. ClustalW

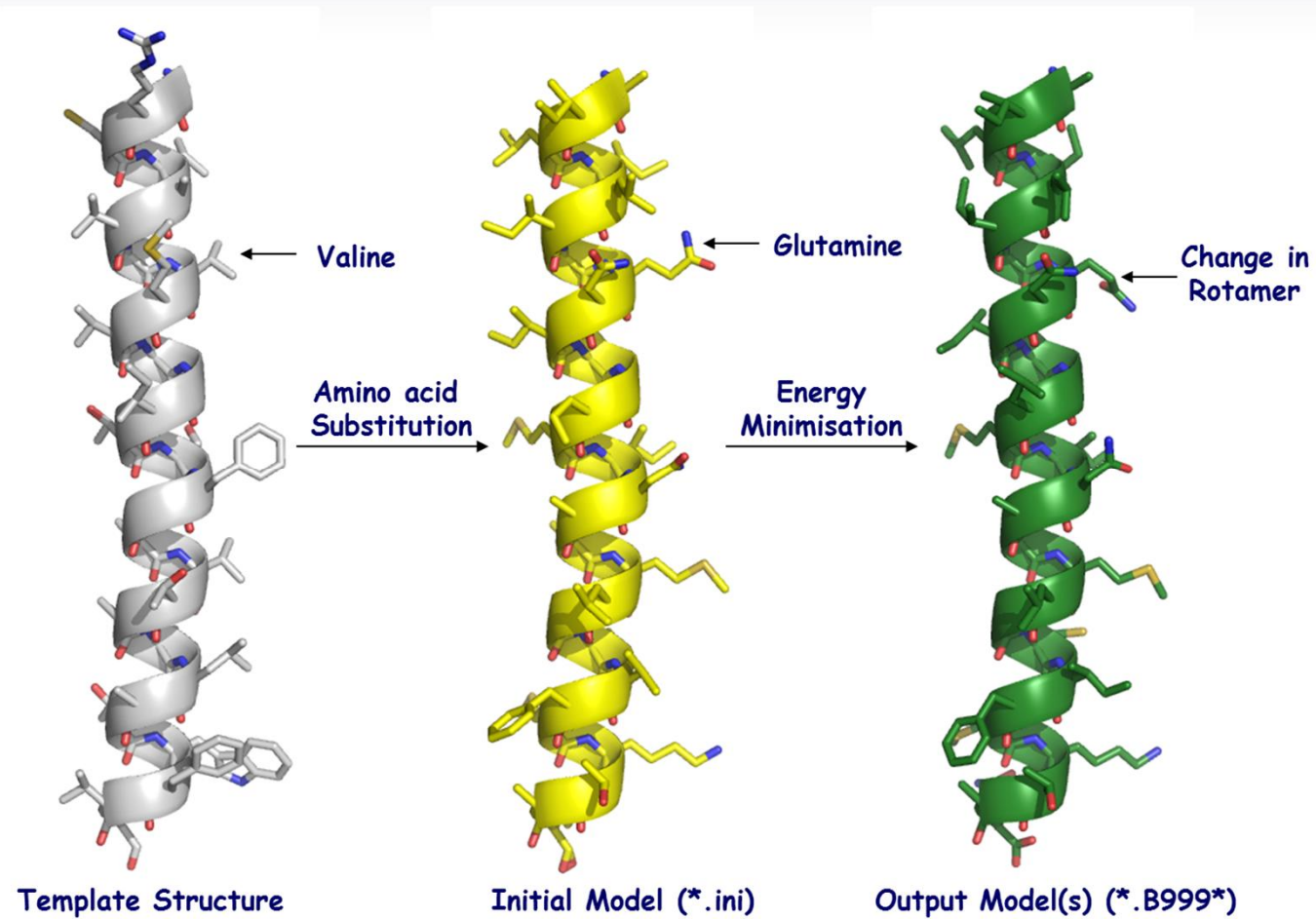
```

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

```



Homológia modellezés II.



Negyedleges szerkezet

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

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

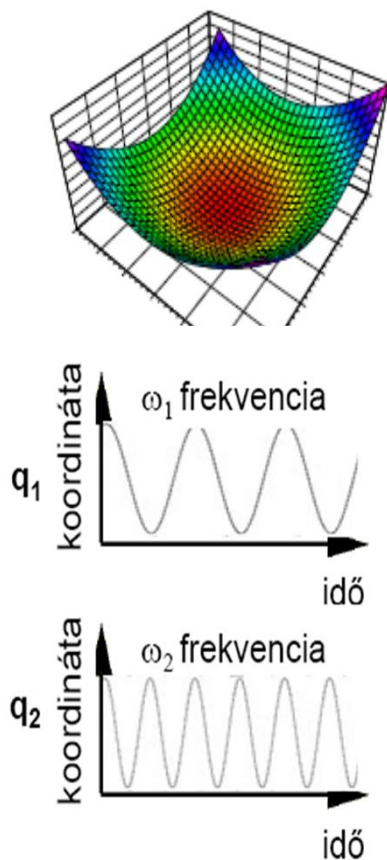
Mai témák

- **Bevezetés – a fehérje dinamika és a szimulációk jelentősége**
- **Fehérjék jellemzése bioinformatikai eszközökkel**
 - Másodlagos szerkezeti mintázatok jóslása
 - Rendezetlen fehérjék
 - Funkcionális régiók azonosítása
 - A harmadlagos és negyedleges szerkezet
- **Fehérjék dinamikájának modellezése**
- **Fehérjék feltekeredésének szimulációja**

Fehérje dinamika vizsgálata

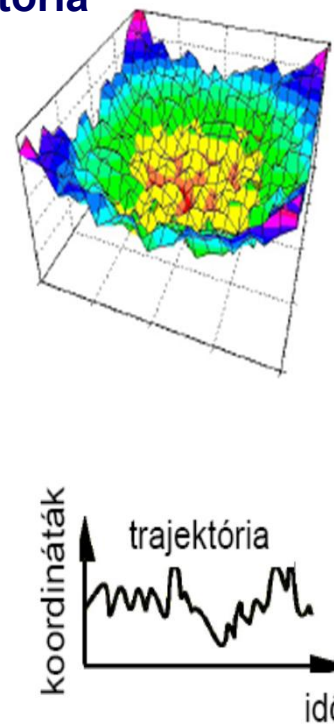
Normál-módus elemzés

- harmonikus potenciál
- analitikus mozgásegyenletek
- normál modusok



Molekuláris dinamika (MD)

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



A „force field“ - I.

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

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

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)

A „force field“ – II.

Baker (2007)

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

$$E_{\text{mbenv}} = \sum_i^{\text{natom}} \Delta G_i^{\text{ref}}(z') \quad \Delta G_i^{\text{ref}}(z') = (1 - f(z')) * (\Delta G_i^{\text{ref,chex}} - \Delta G_i^{\text{ref,water}})$$

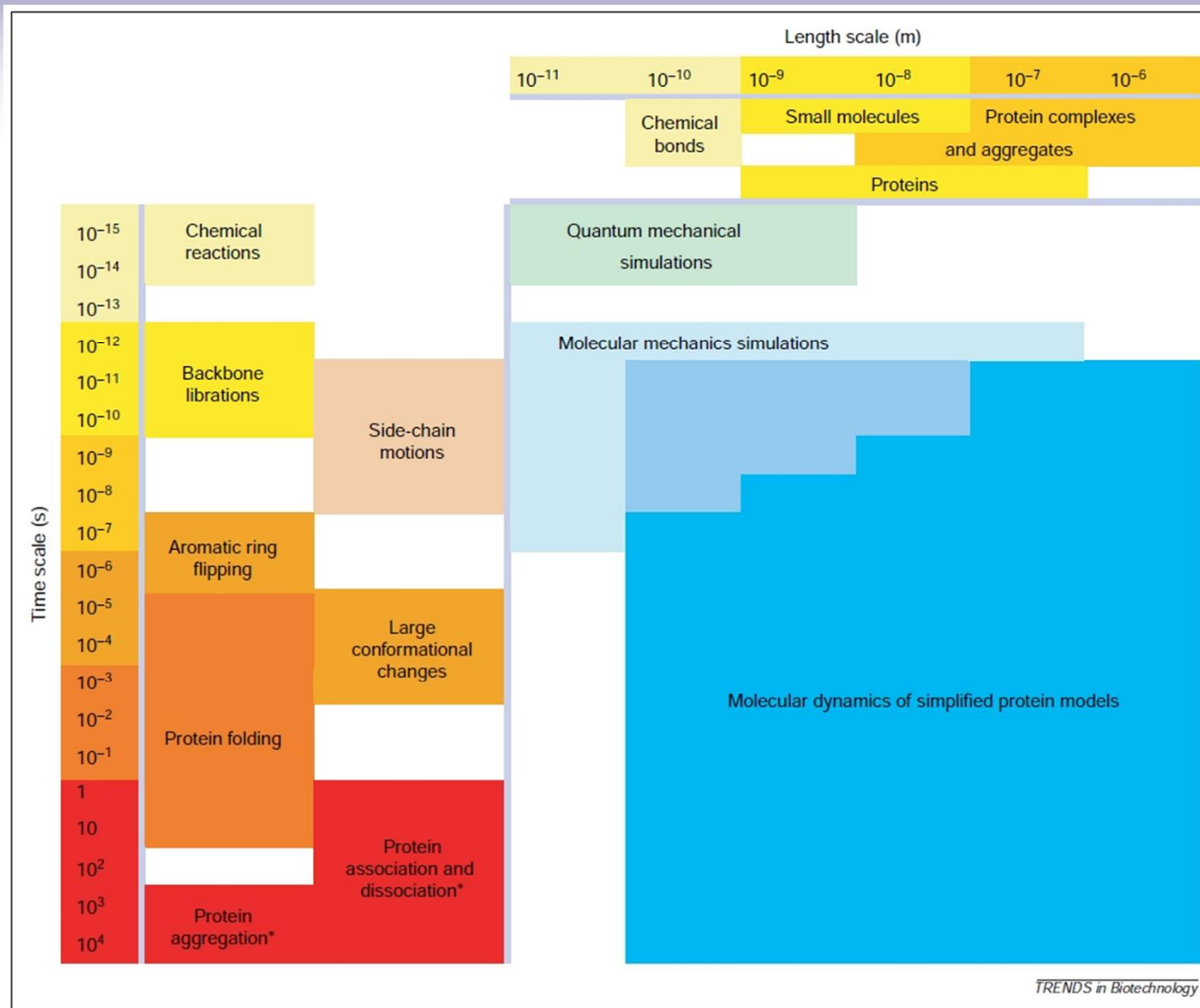
TABLE II. Atomic Contribution to the Solvation Free Energy in Water and Cyclohexane

	Water	Chex		Water	Chex
CR	-0.890	-1.350	NH3	-20.000	-1.145
CH1E	-0.187	-0.645	NC2	-10.000	-0.200
CH2E	0.372	-0.720	N	-1.000	-1.145
CH3E	1.089	-0.665	OH1	-5.920	-0.960
CR1E	0.057	-0.410	O	-5.330	-1.270
NH1	-5.950	-1.145	OC	-10.000	-0.900
NR	-3.820	-1.630	S	-3.240	-1.780
NH2	-5.450	-1.145	SH1E	-2.050	-1.855

Az MD korlátjai

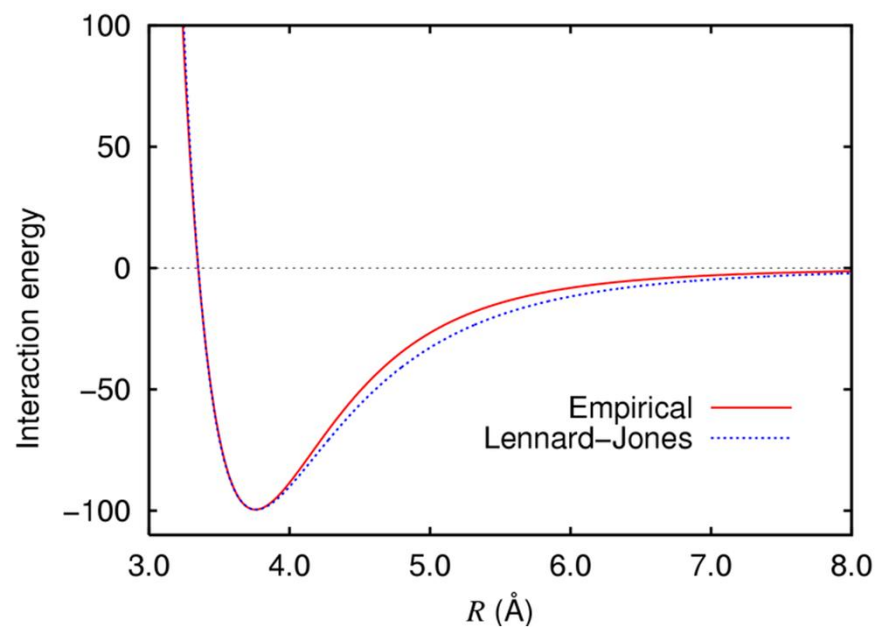
- idő (CPU, valós)
- potenciál kiszámolása a szűk keresztmetszet
- numerikus integrálás hibája
- fs-os integrációs lépések
- oldószer (explicit/implicit)
- „boundary condition”

„Események” időskálája

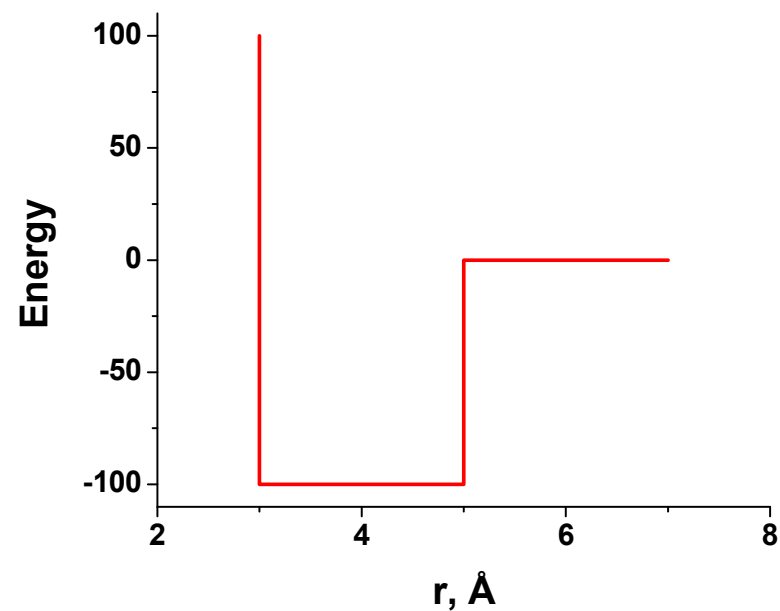


F. Ding and N.V. Dokholyan, TRENDS in Biotechnology, **23**:450 (2005)

Diszkrét Molekuláris Dinamika (DMD)



wikipedia



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

Egyszerűsített (Coarse Grain) modellek

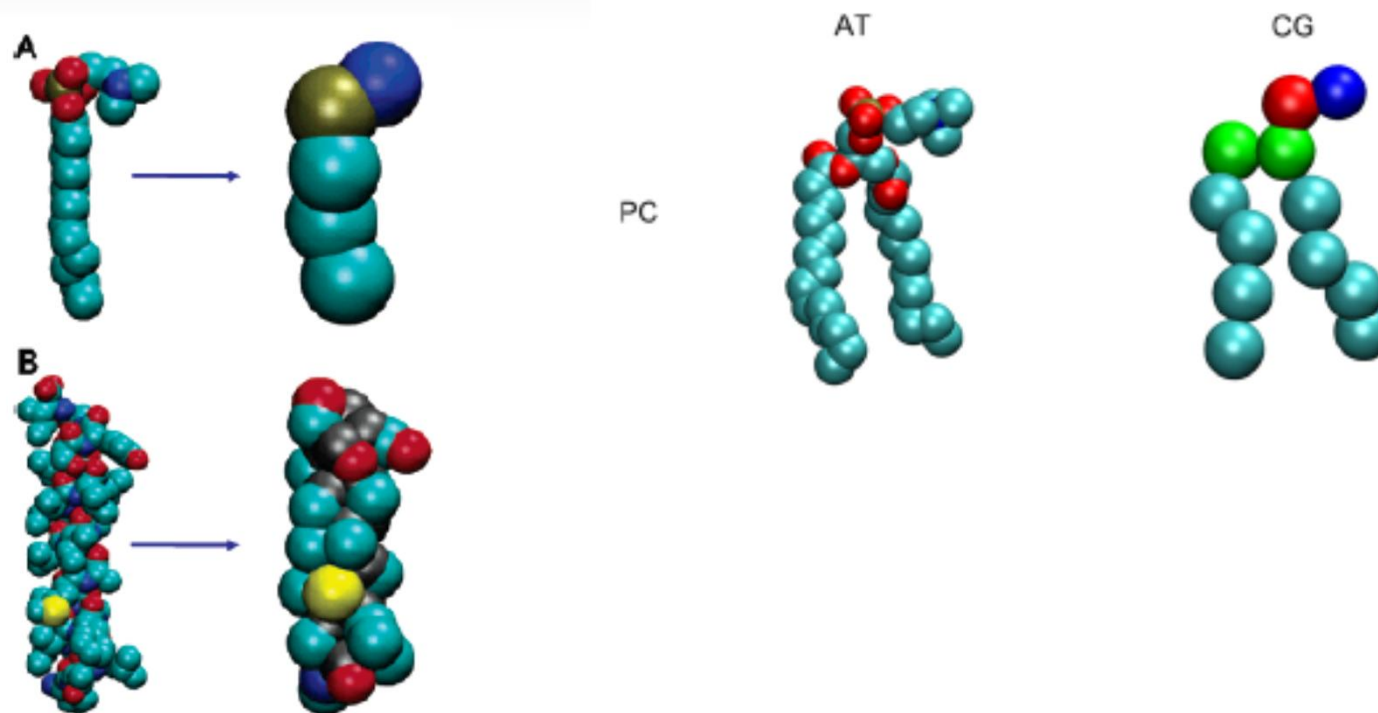
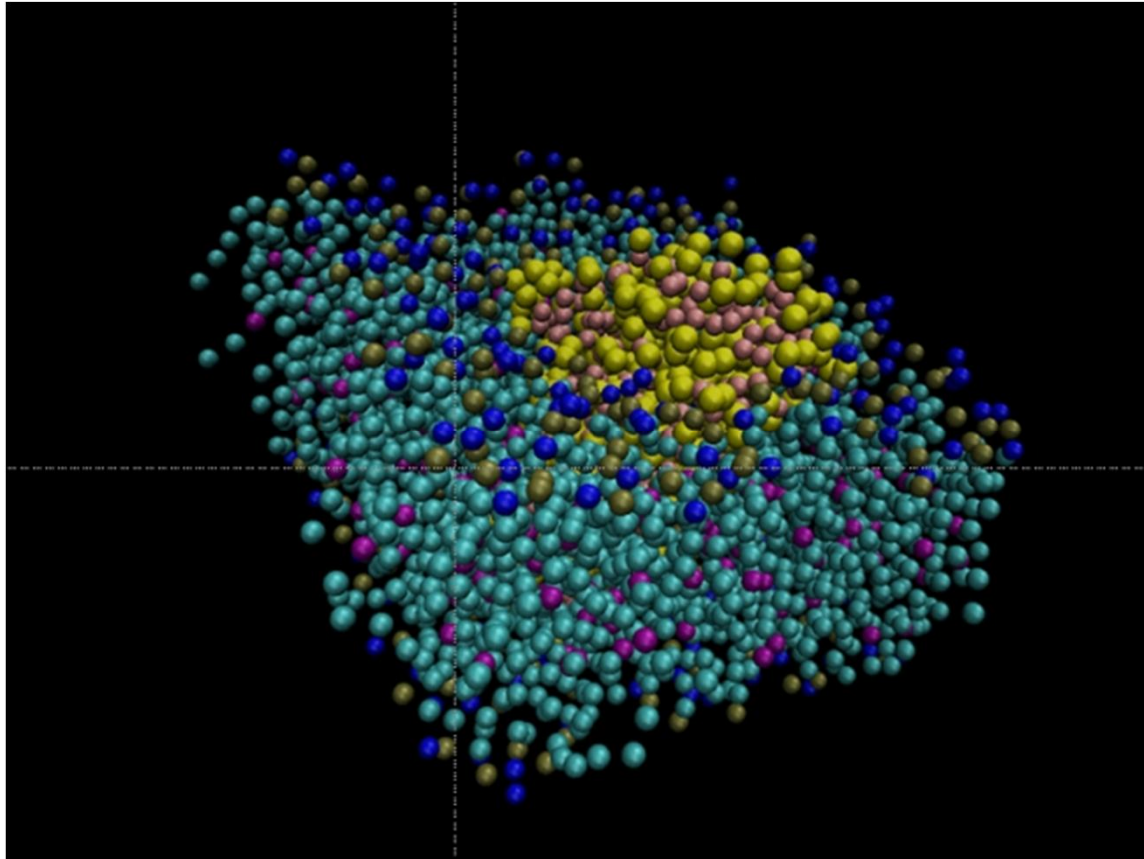


Figure 1. Atomistic (left-hand) and coarse-grained (right-hand) models compared for (A) a DPC molecule and (B) a GpA helix. Colors for atoms:

Fehérjére pl. 2 bead vagy 4+ bead modellek

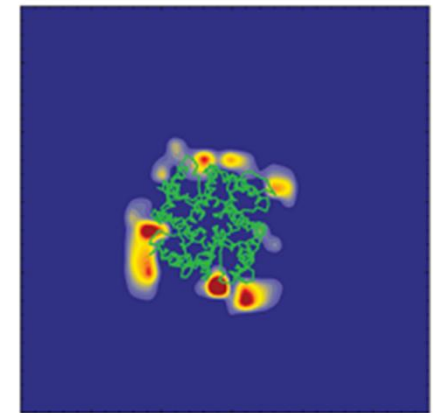
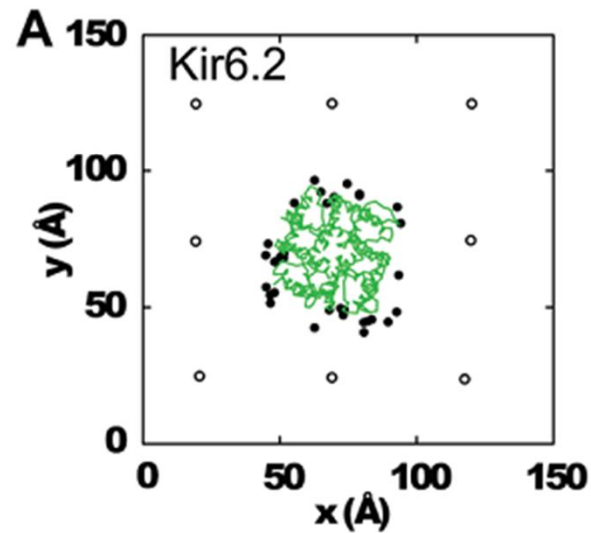
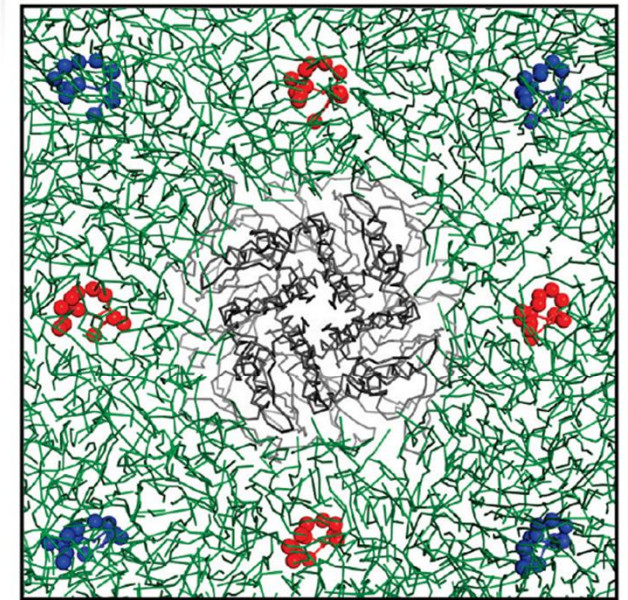
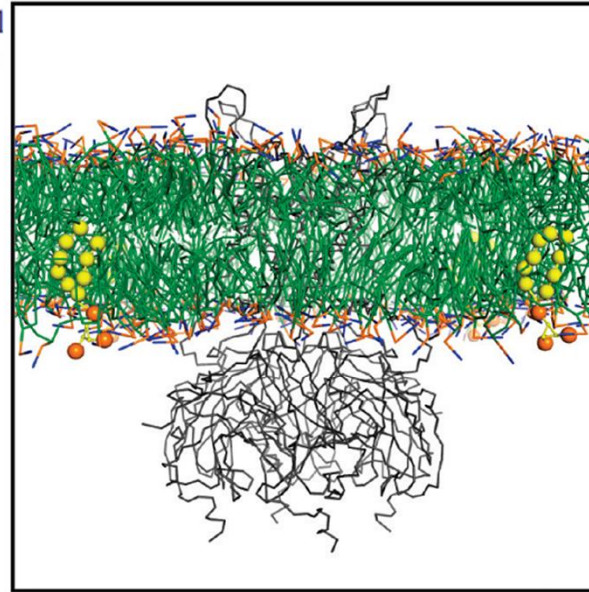
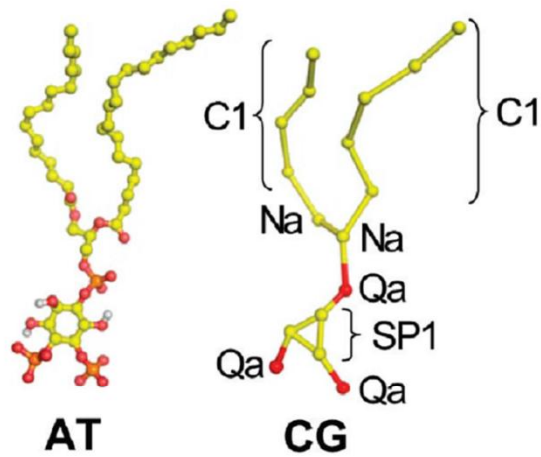
MARTINI CG force field

Kettősréteg felépülése a fehérje köré

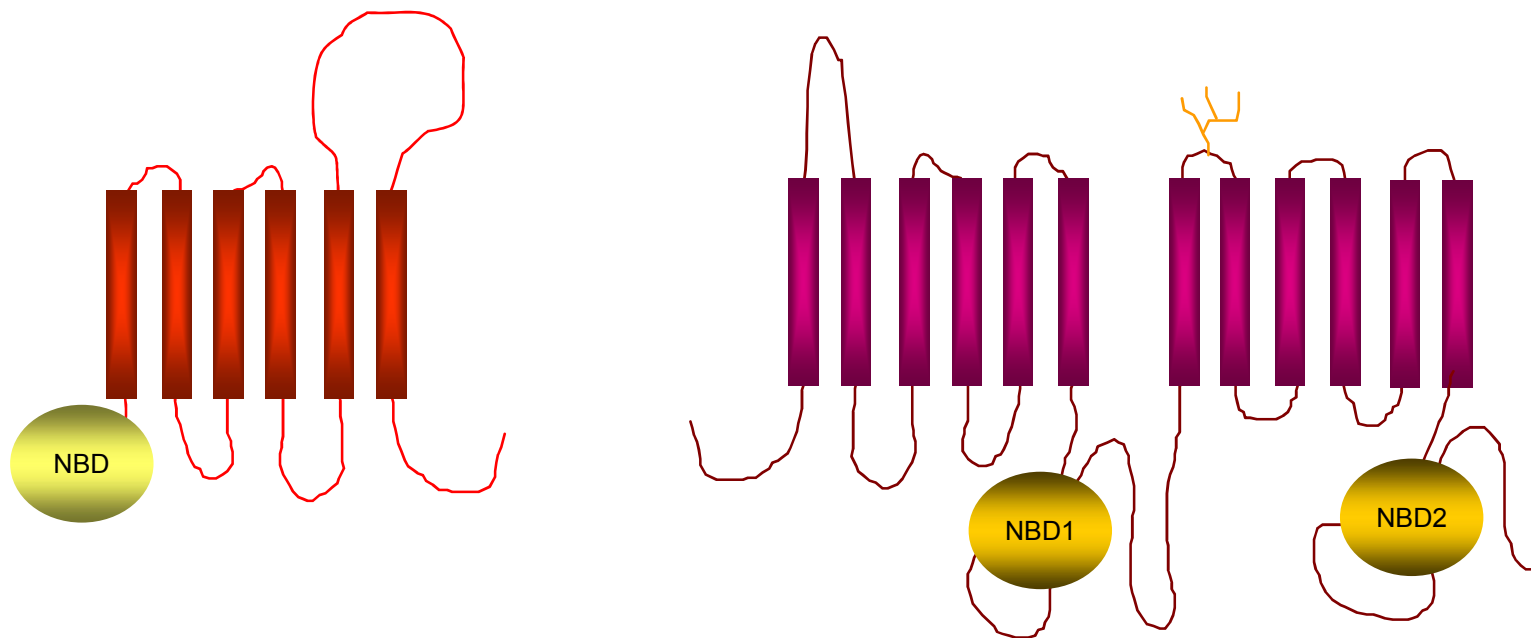


PIP2 kötődése Kir kálium csatornához

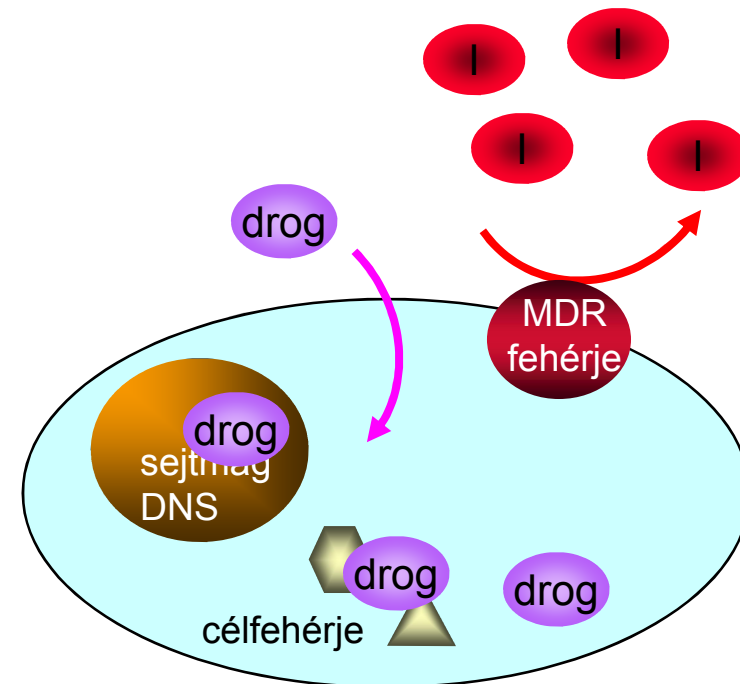
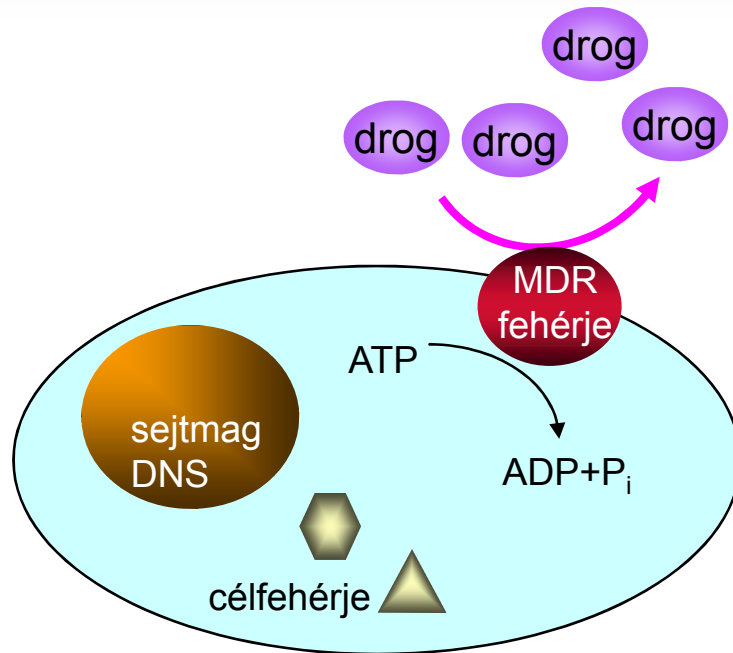
Biochemistry, Vol. 48, No. 46, 2009 1



ATP Binding Cassette (ABC) fehérjék



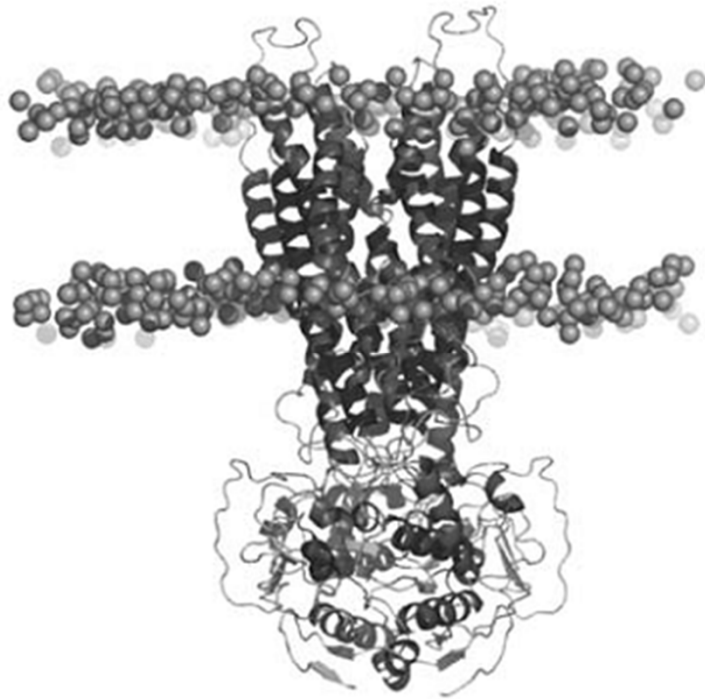
A multidrog-rezisztencia és felfüggesztése



Fehérjék konformációinak stabilitása

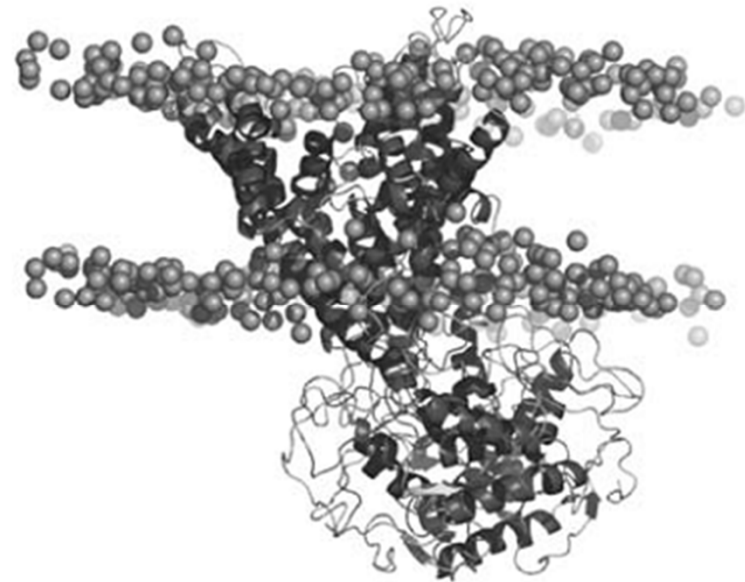
Eur Biophys J (2008) 37:403–409

B



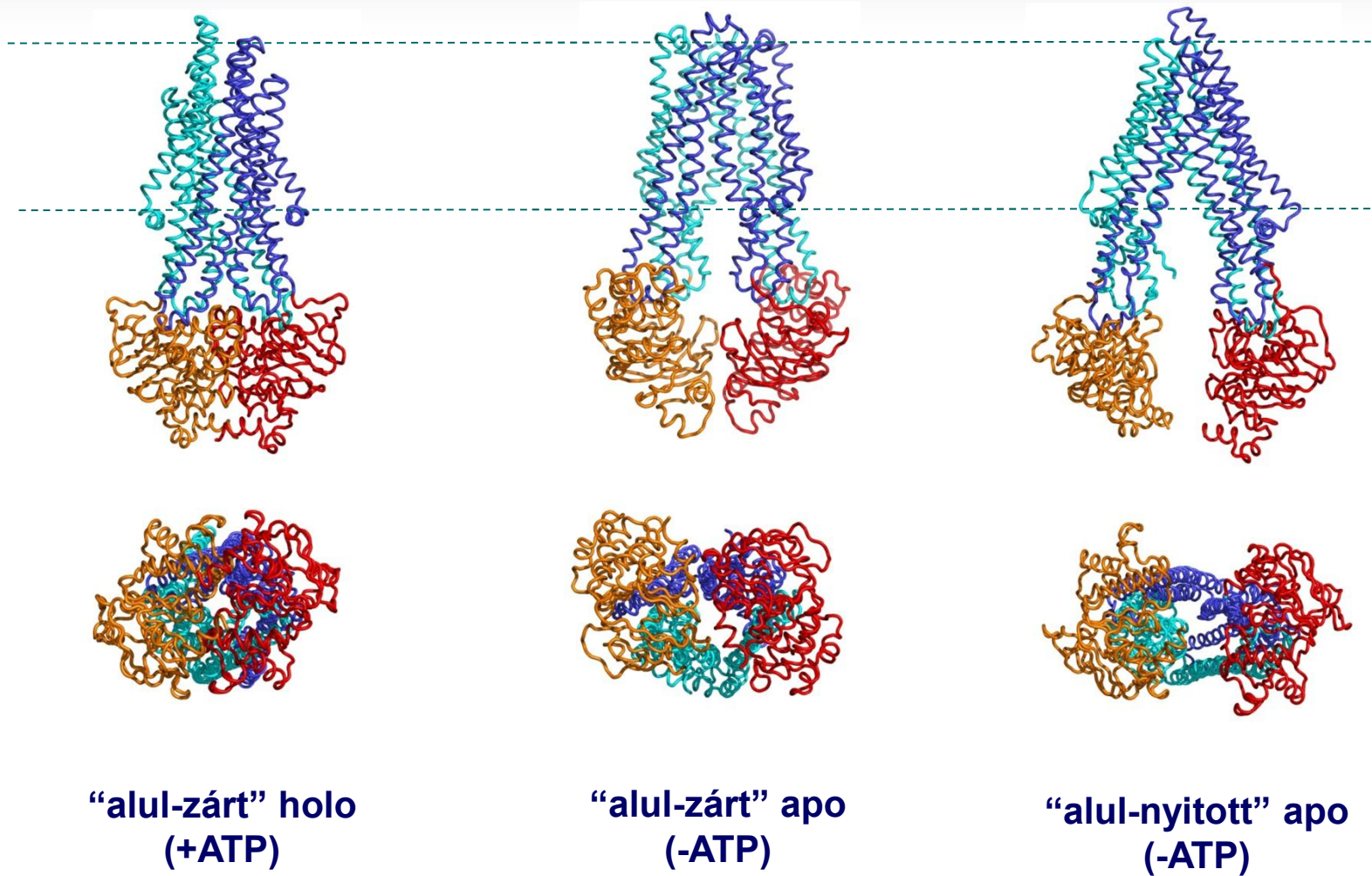
0 ns

C

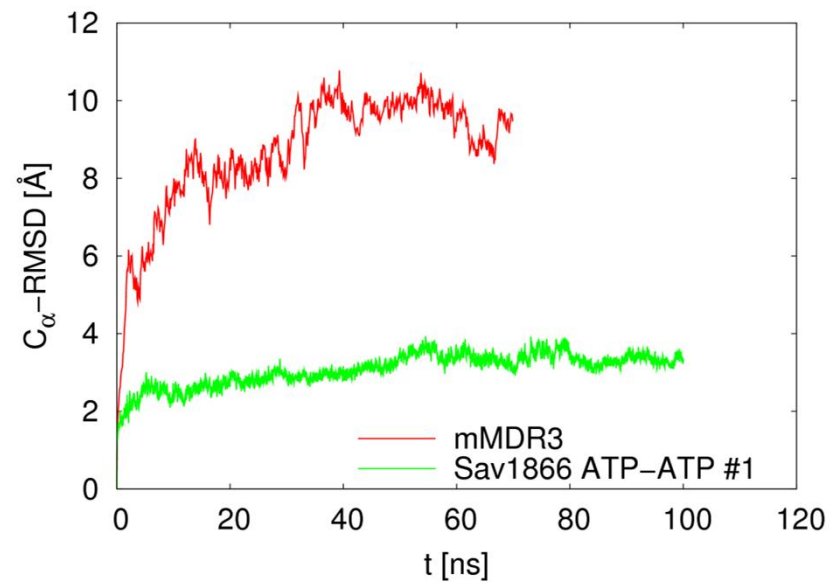
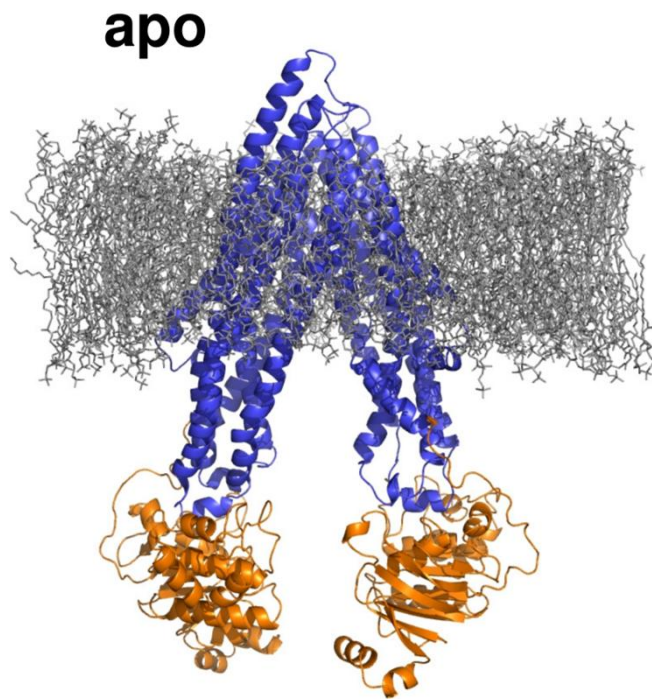


20 ns

ABC fehérjék konformációi

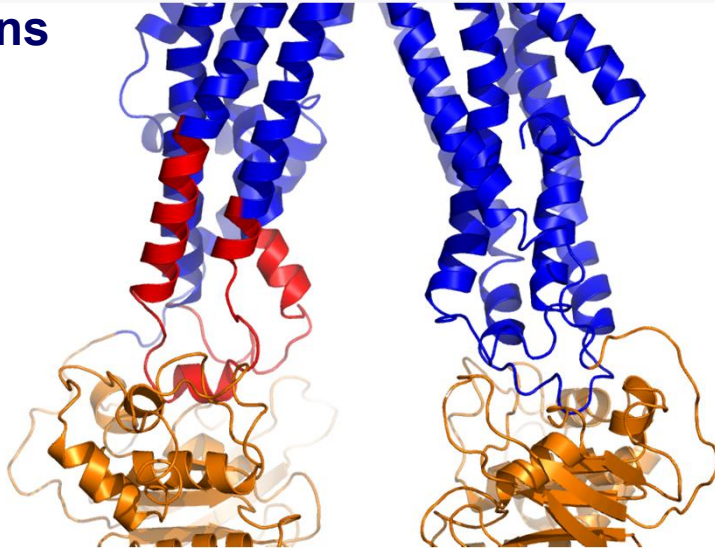


Az alul nyitott apo szerkezet nem stabil

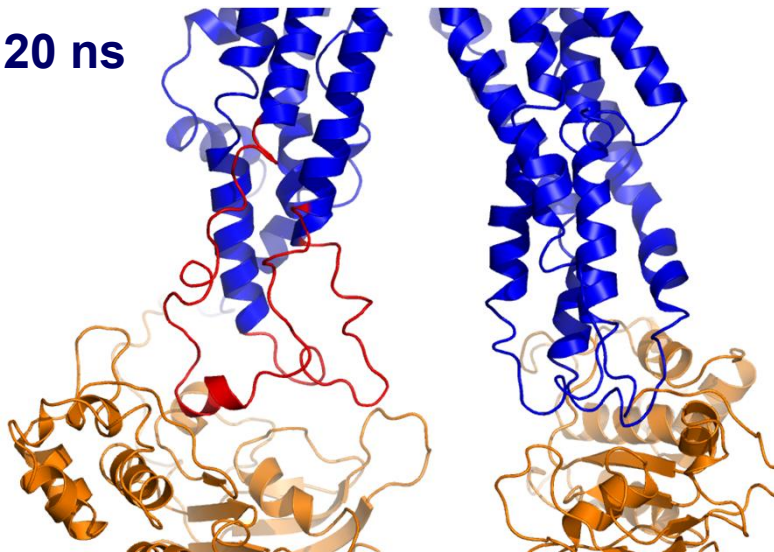


Az alul nyitott apo szerkezet nem stabil

t = 0 ns

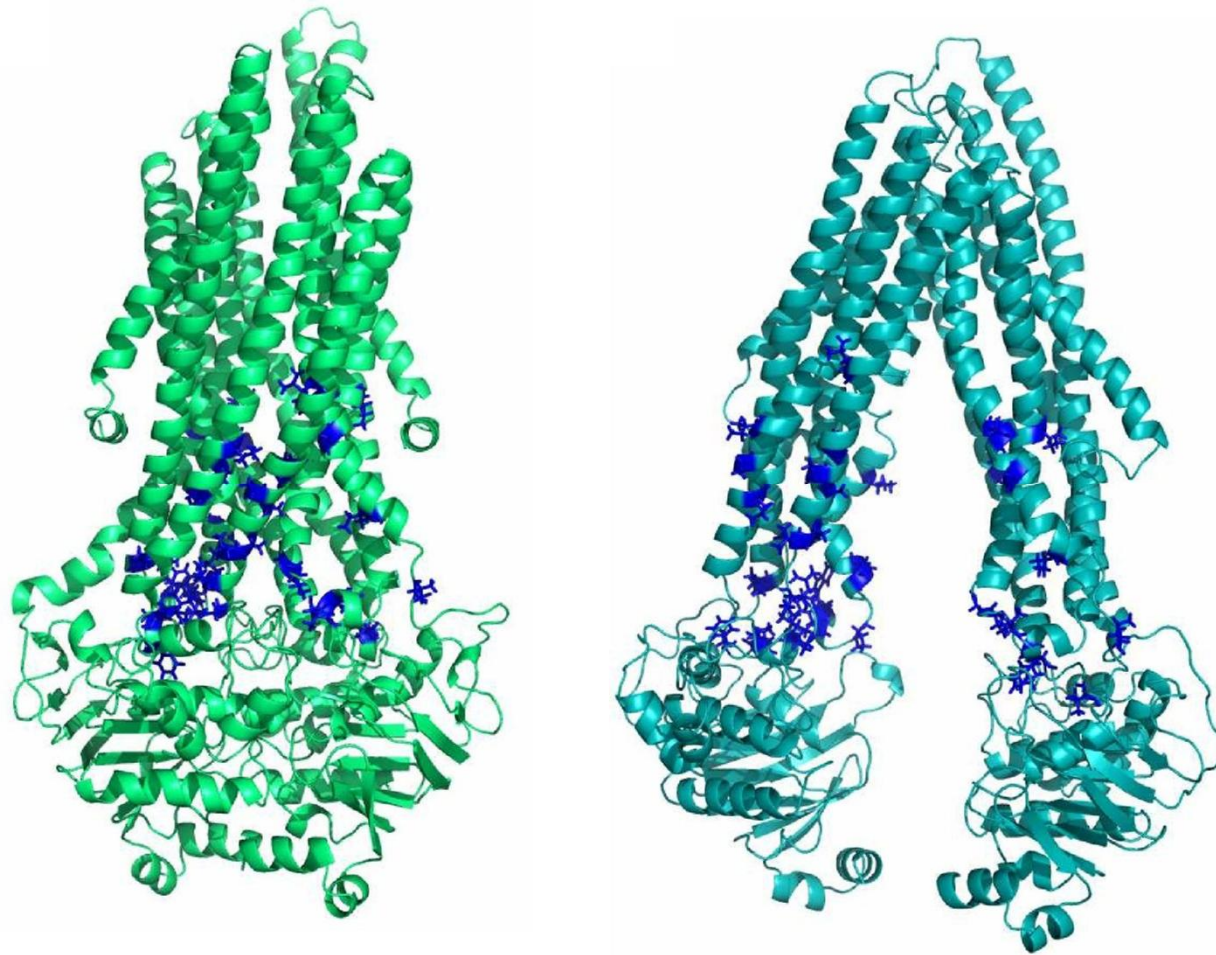


t = 20 ns



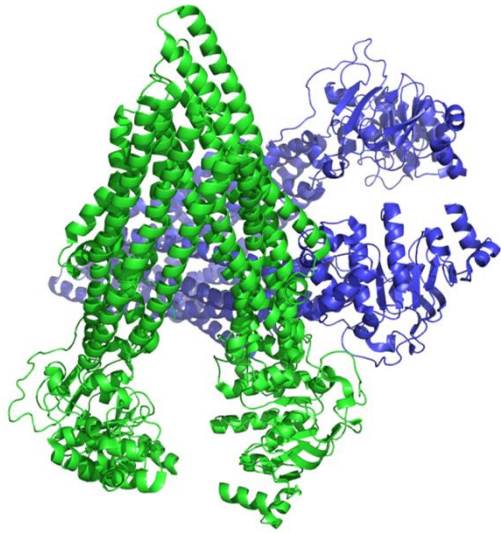
rendszer	megtartott hélixtartalom
Sav1866 ATP/ATP #1	90.04%
hMDR1 holo	91.84%
hMDR1 apo	64.30%
mMDR3	63.13%

Hidrofób aminosavak kerülnek felszínre

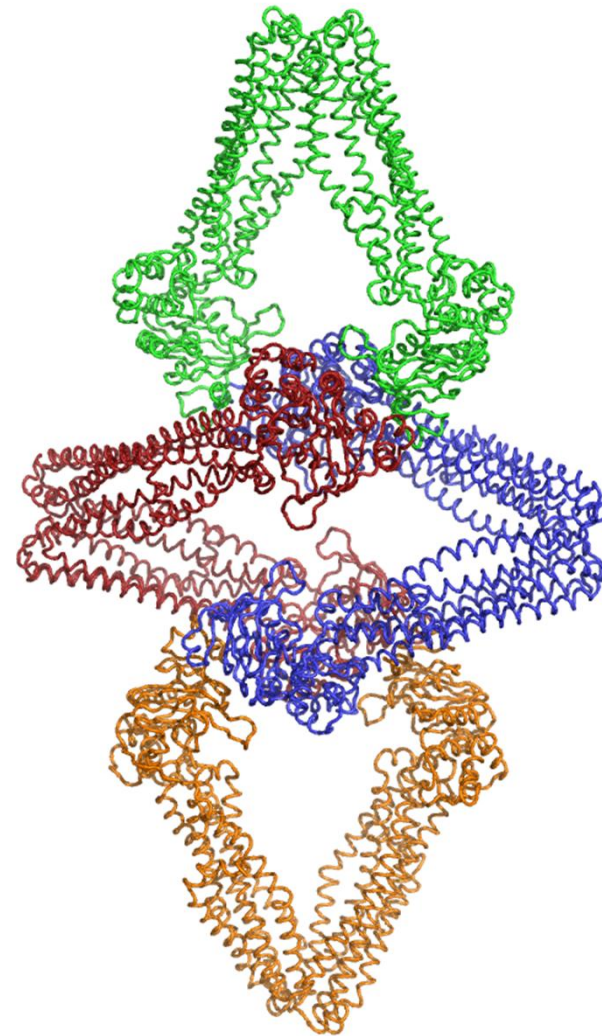


Alul nyitott apo szerkezet elemi cellája

mMDR3, PDBID:3G5U



MsbA, PDBID:3B5W

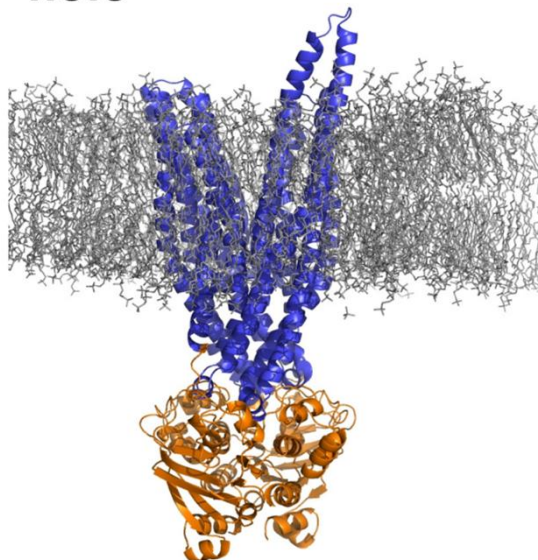


Események modellezése

- 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

Zárt-nyitott átmenet jellemzése molekuláris dinamikával

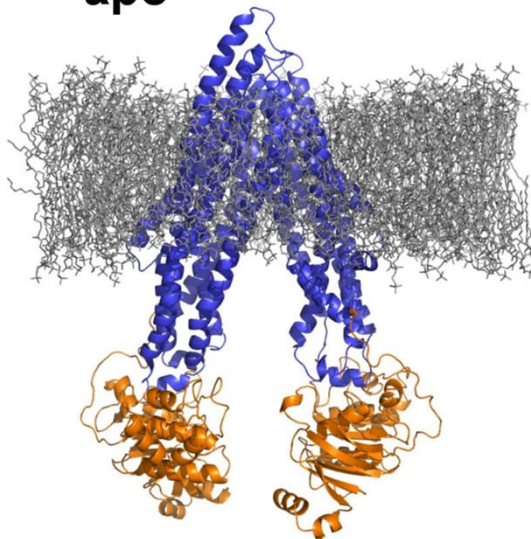
holo



hMDR1 homologia modell
(3x100 ns)

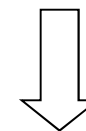
M. Wiese modellje

apo

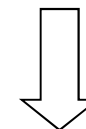


hMDR1 homologia modell

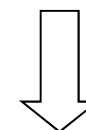
**molekularis dinamika
trajektóriák**



**esszenciális
dinamika**

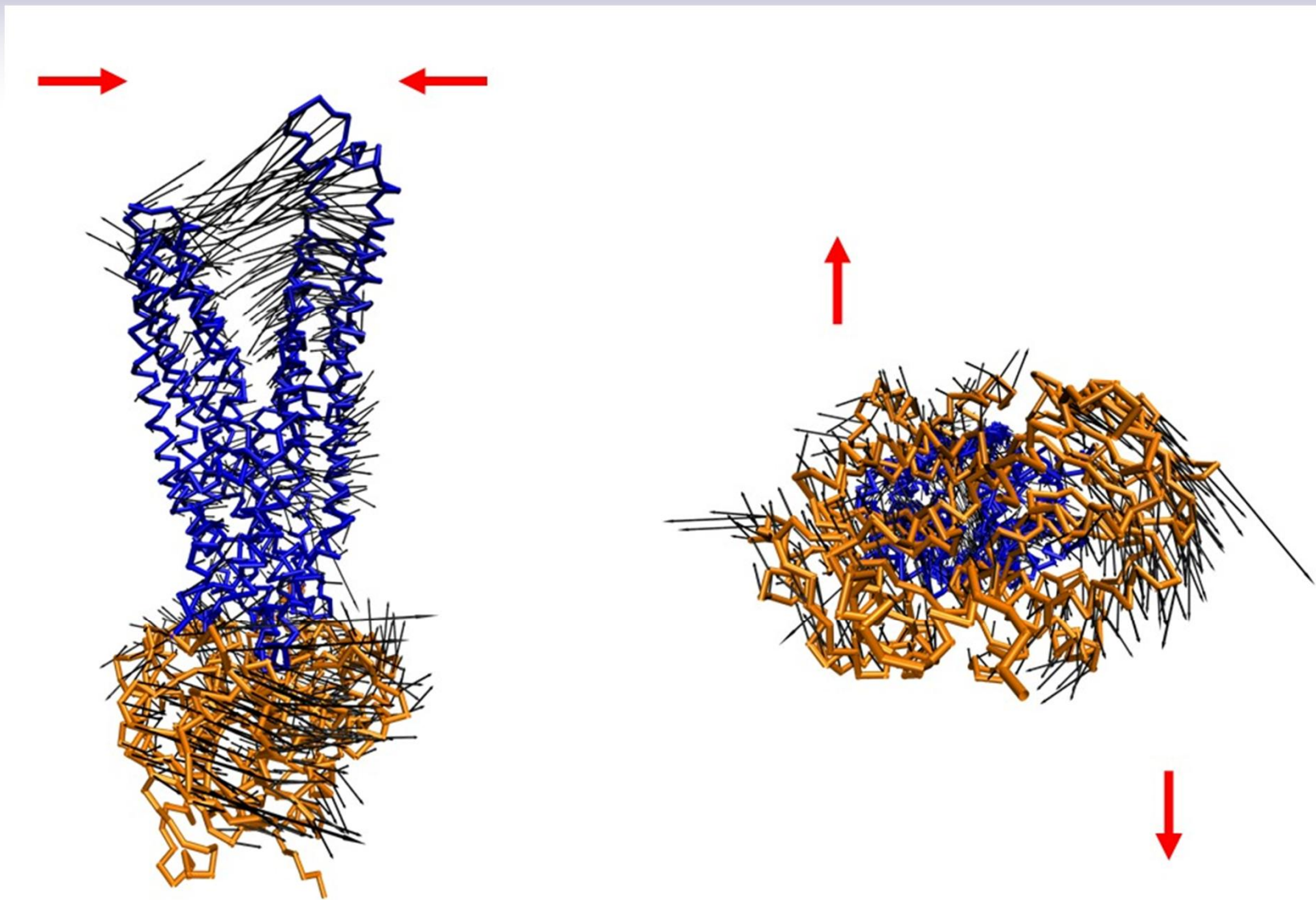


módus kiválasztás



kollektív mozgások

Zárt-nyitott átmenet



Mai témák

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

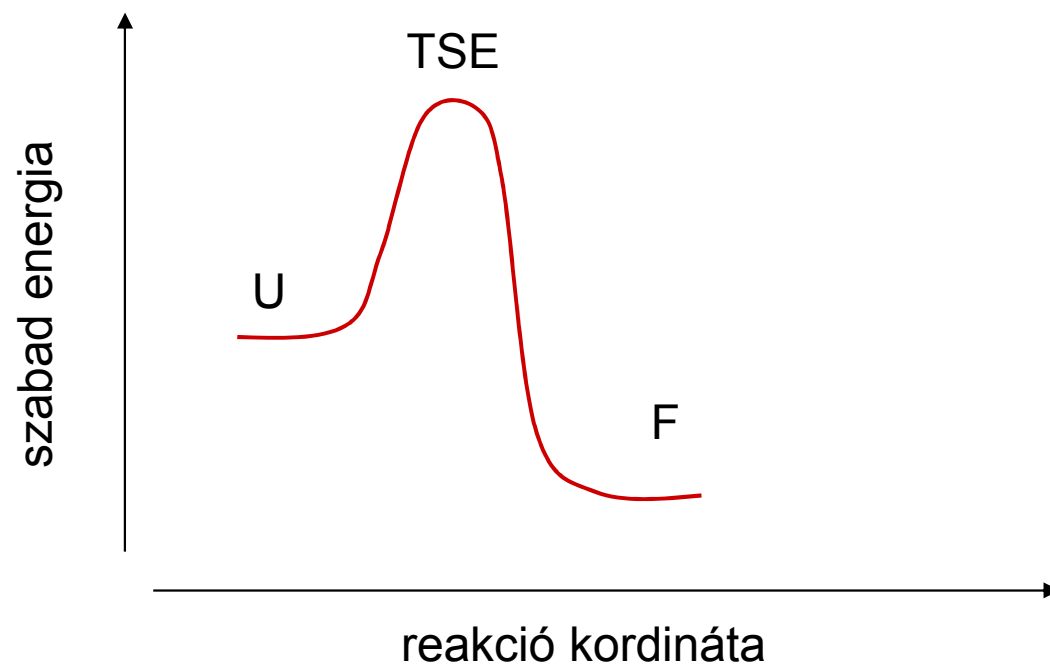
Kétállapotú fehérje hajtogatódás

Two-state model of protein folding

$$G_{\text{protein}} = H_{\text{protein}} - TS_{\text{protein}}$$

$$\begin{aligned} -RT \ln K &= \Delta G_{\text{protein}} = \\ &= \Delta H_{\text{protein}} + \Delta H_{\text{solvent}} + \Delta H_{\text{protein-solvent}} - T\Delta S_{\text{protein}} - T\Delta S_{\text{solvent}} \end{aligned}$$

Fehérje feltekeredés



nukleáció

Levinthal paradoxon

Fehérje stabilitás I.

Konformációs stabilitást elősegítik:

- Hidrofób kölcsönhatások
- Intramolekuláris H-híd kötések
- Intramolekuláris ionos kölcsönhatások
- Intramolekuláris van der Waals kölcsönhatások
- Intramolekuláris diszulfid hidak

Destabilizáló tényezők:

- H-híd az oldószerrel
- Van der Waals kölcsönhatás az oldószerrel
- Az ionos csoportok solvatációja
- entrópia

Fehérje stabilitás II.

A fehérje stabilitás nem ér el maximális értéket.

Erre utalnak:

**termofil baktériumok fehérjéi;
igen stabil, tervezett fehérjék.**

Ennek okai lehetnek:

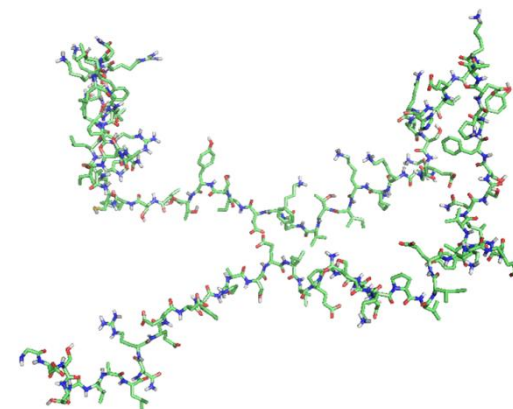
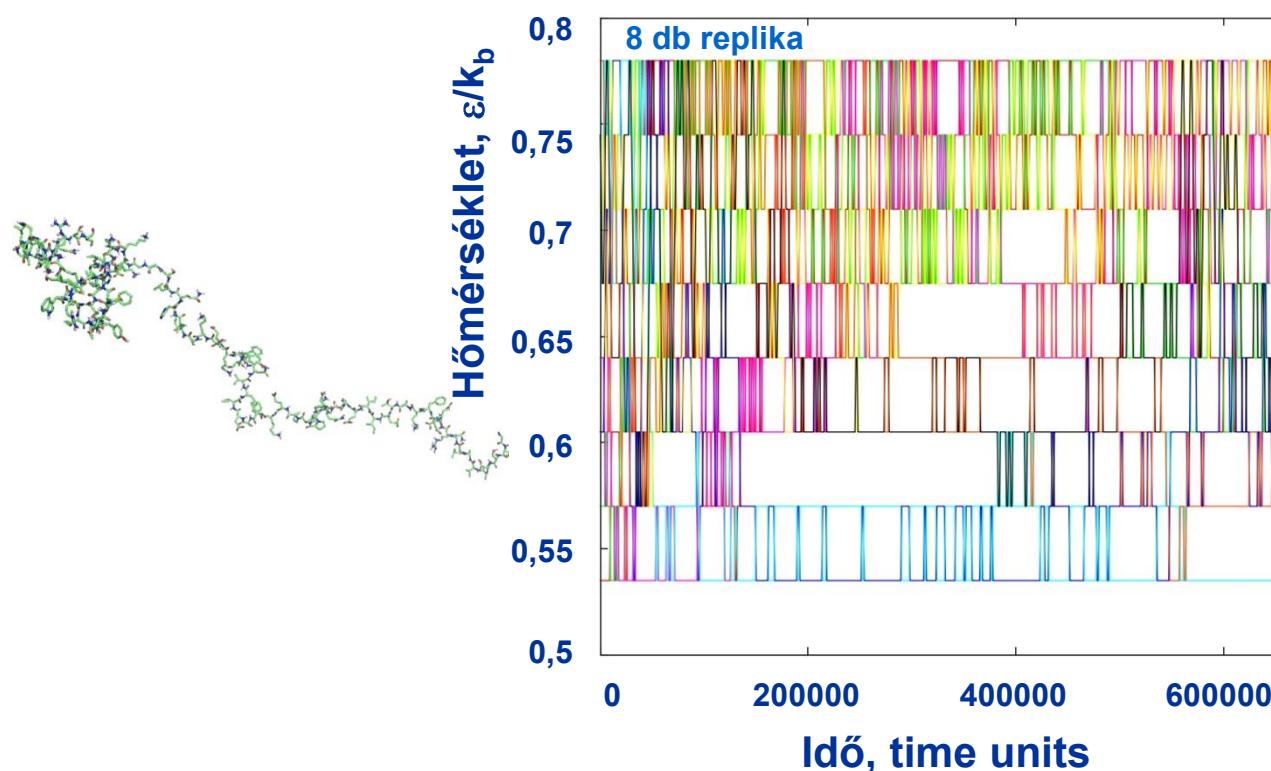
**az evolúció nem igényel stabilabb fehérjét mint a funkció önmaga;
a fehérjéknek le is kell bomlaniuk;
a funkcióhoz flexibilitás szükséges.**

**Folding szimulációk során mi az abszolút energiaminimumot
(maximum stabilitást) keressük.**

Fehérje feltekeredés szimulációja

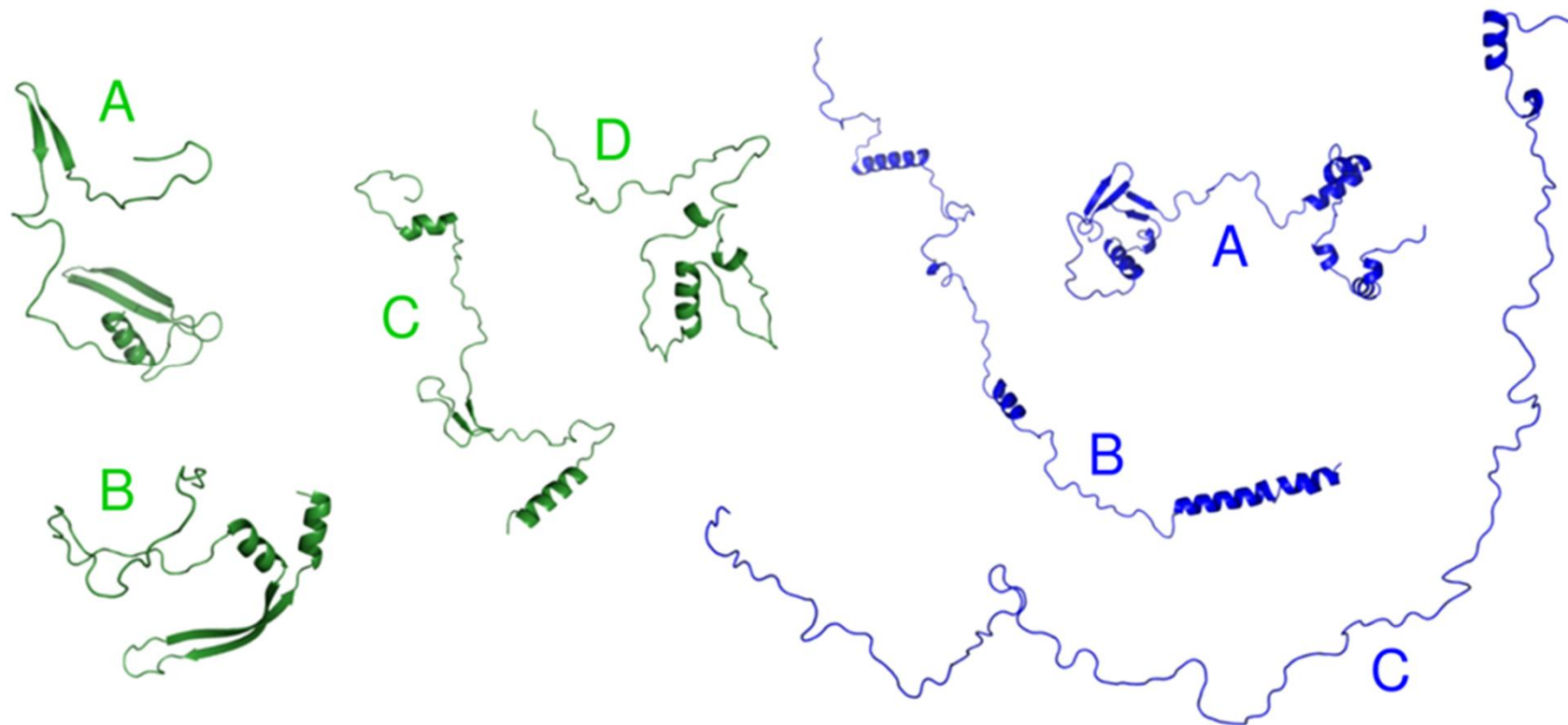
All atom force-field:

Potenciál függvény számolása erőforrásigényes
Reprezentatív konformációs mintavételezés problémás
Umbrella sampling, replica exchange.

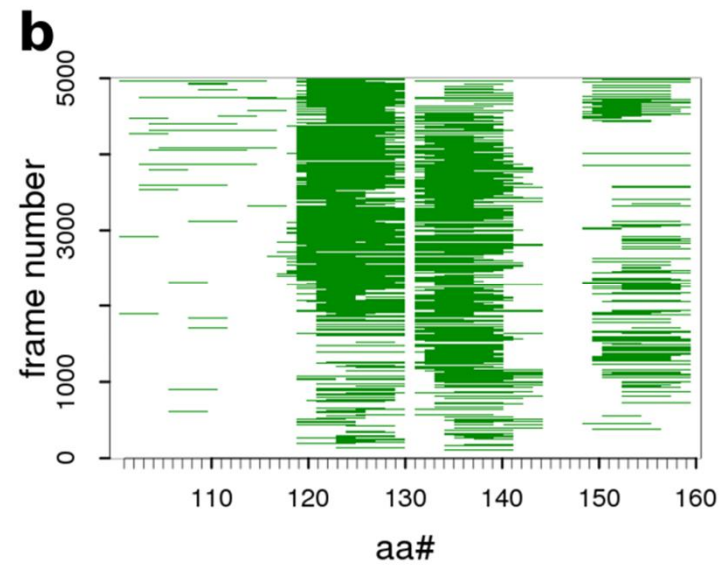
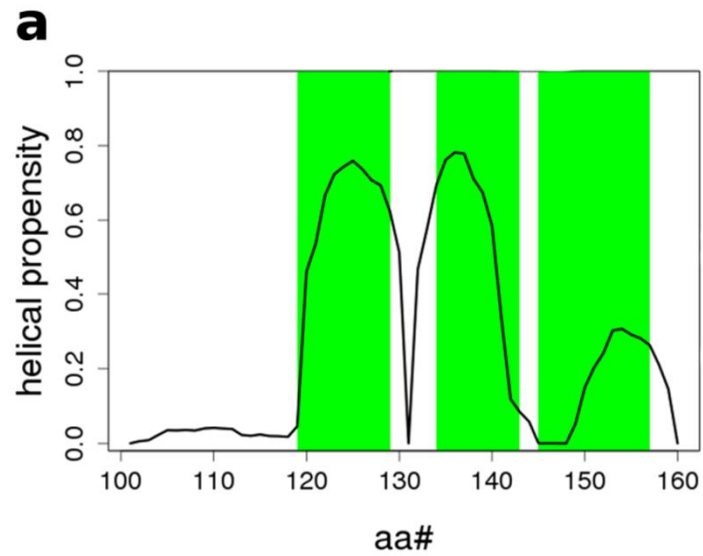


Rendezetlen fehérjék rendezettsége

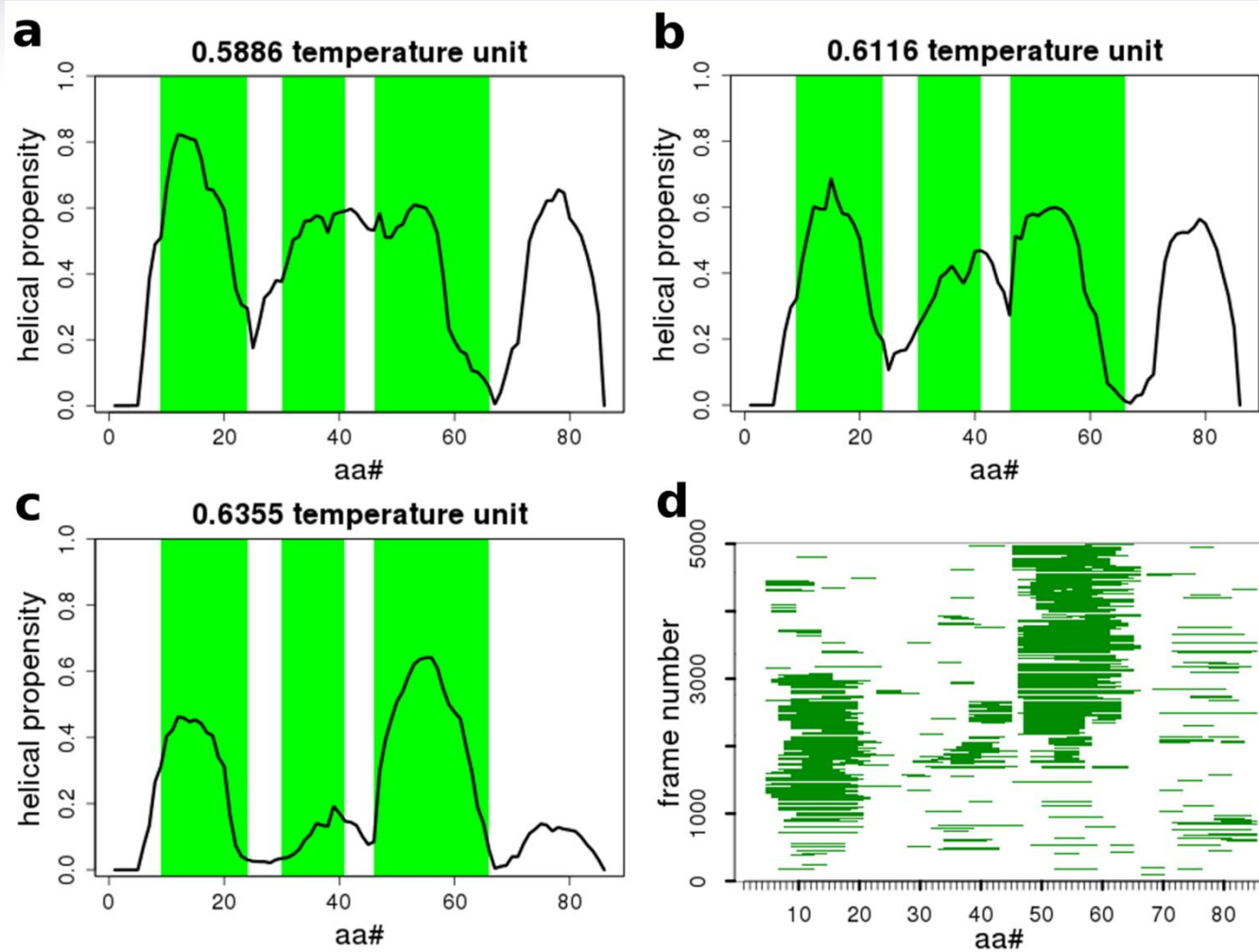
PreSMo: Prestructured Motif



PreSMo jóslás I.

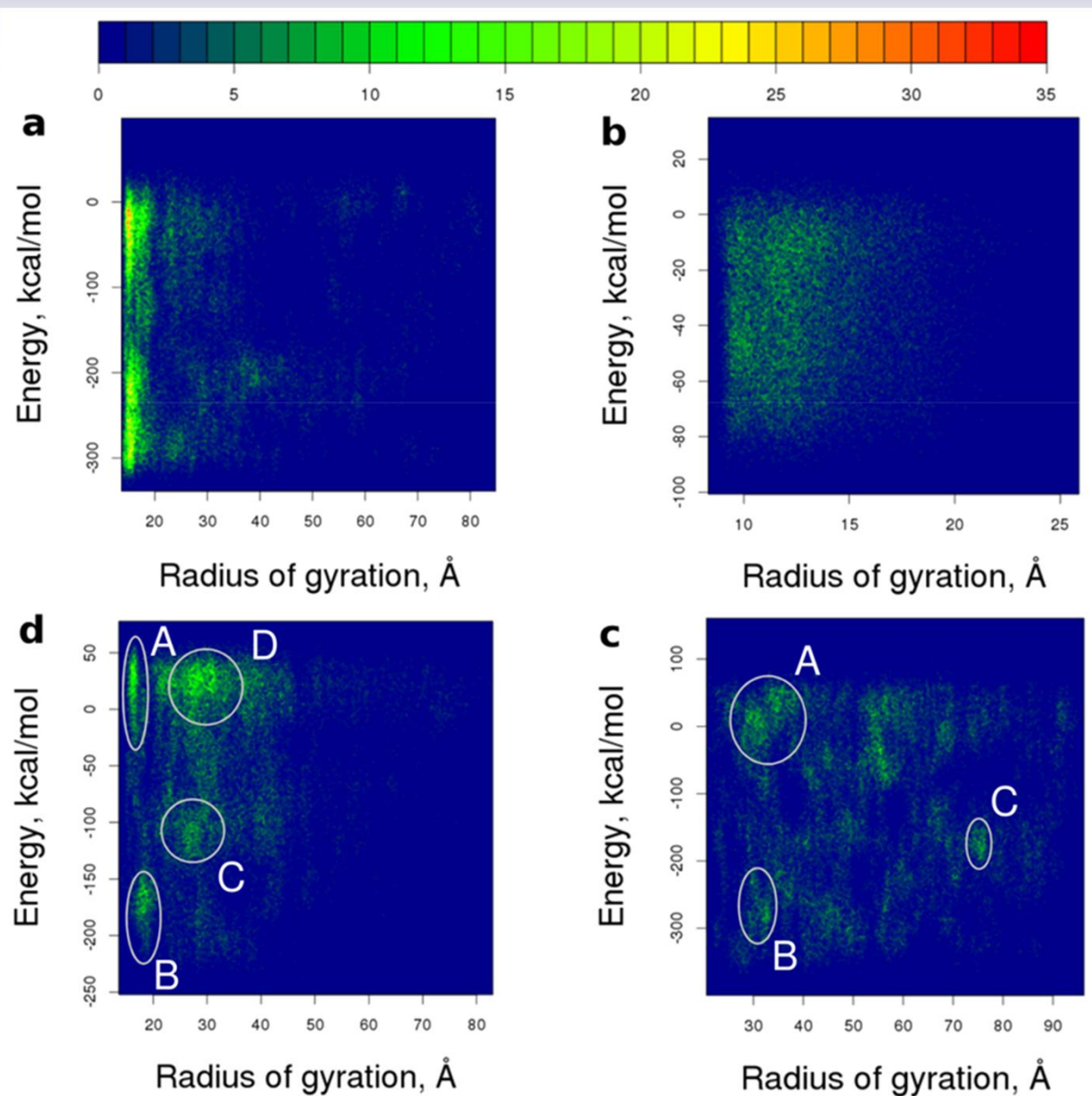


PreSMo jóslás II.



Potenciál felület

DoS (Density of States)



Összefoglalás

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

Köszönetnyilvánítás

www.hegelab.org

Hedvig Tordai
Kinga András
Kristóf Jakab

Hajnalka Sarankó
Attila Tóth
Dániel Szöllősi

Balázs Sarkadi

MTA TTK, Institute of Enzymology
MTA-SE Molecular Biophysics Research Group

Miklós Kellermayer

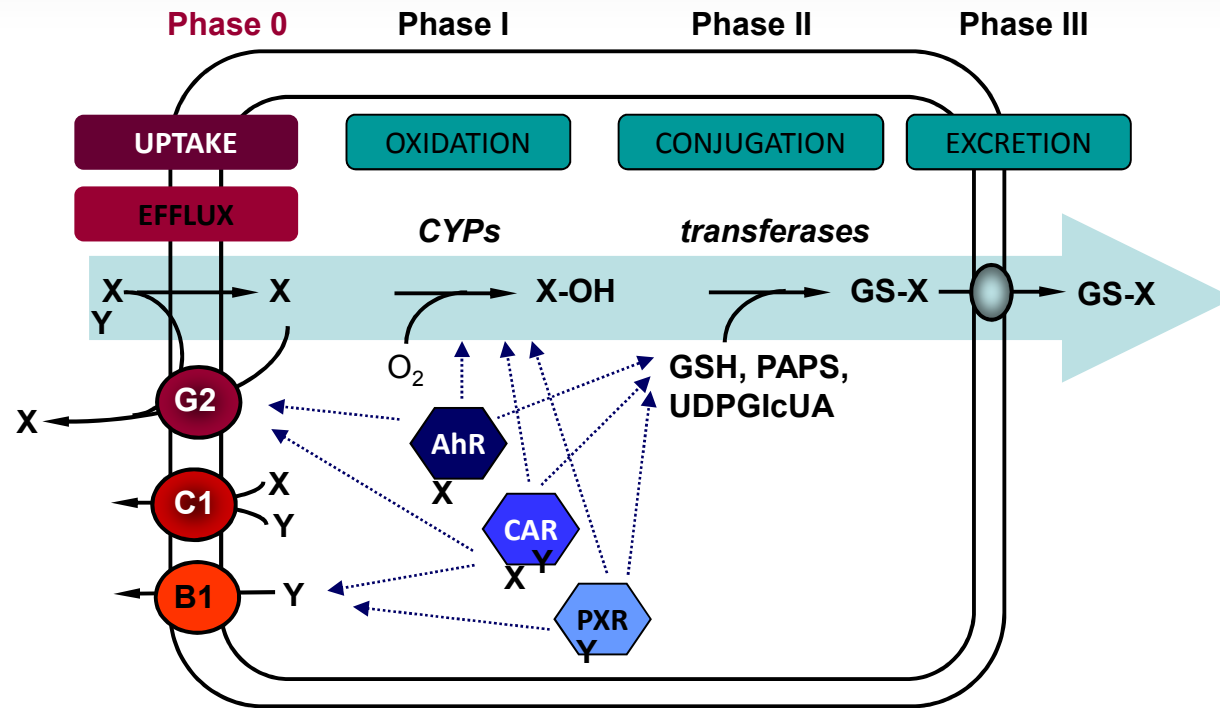
SE Department of Biophysics and Radiation Biol.
MTA-SE Molecular Biophysics Research Group

www.biomembrane.hu

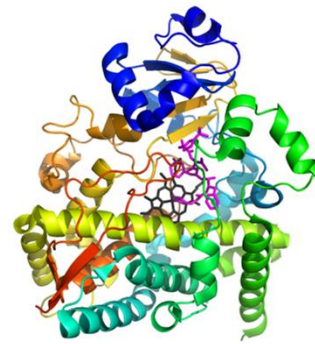
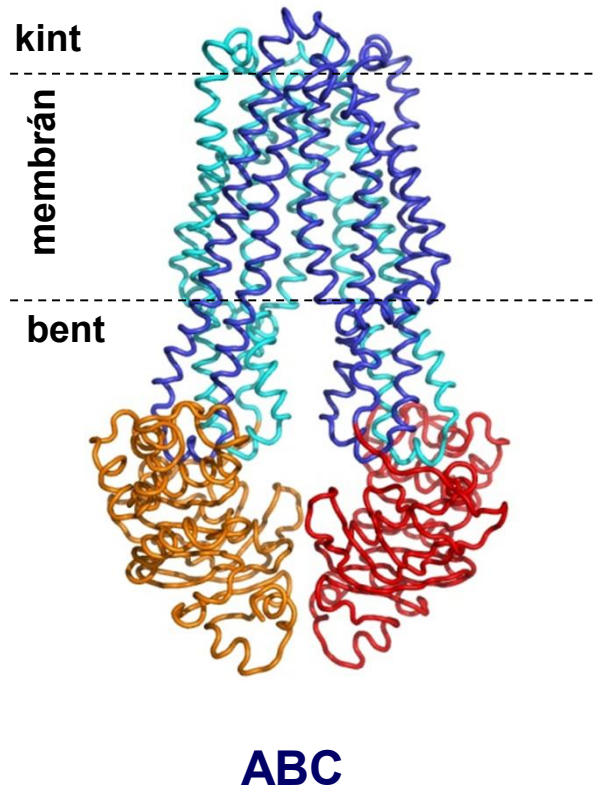


NKFIH K 111678

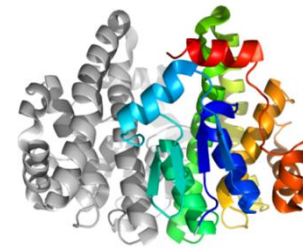
The chemoimmune (ChI) system



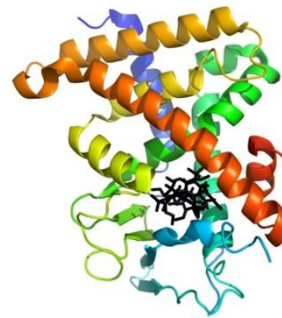
Általános mechanizmus a multidrog-felismerésre?



CYP3A4 (PDBID:3NXU)



GSTA1 (PDBID:3IK9)

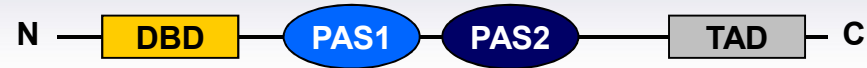


PXR LBD (PDBID:1ILH)

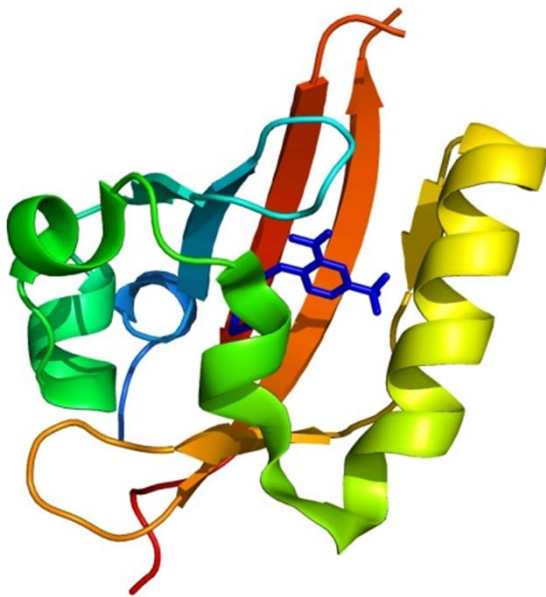


AhR PAS-B (homológia modell)

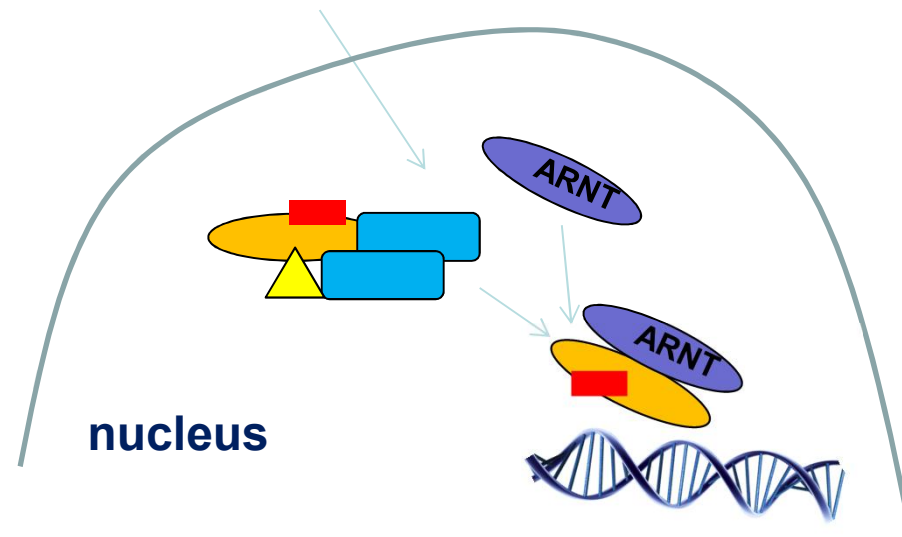
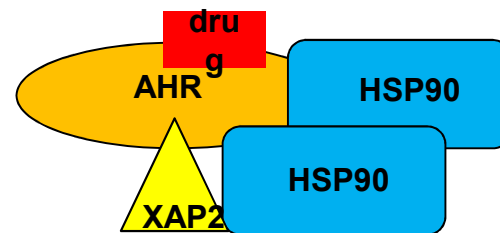
Aryl Hydrocarbon Receptor (AHR)



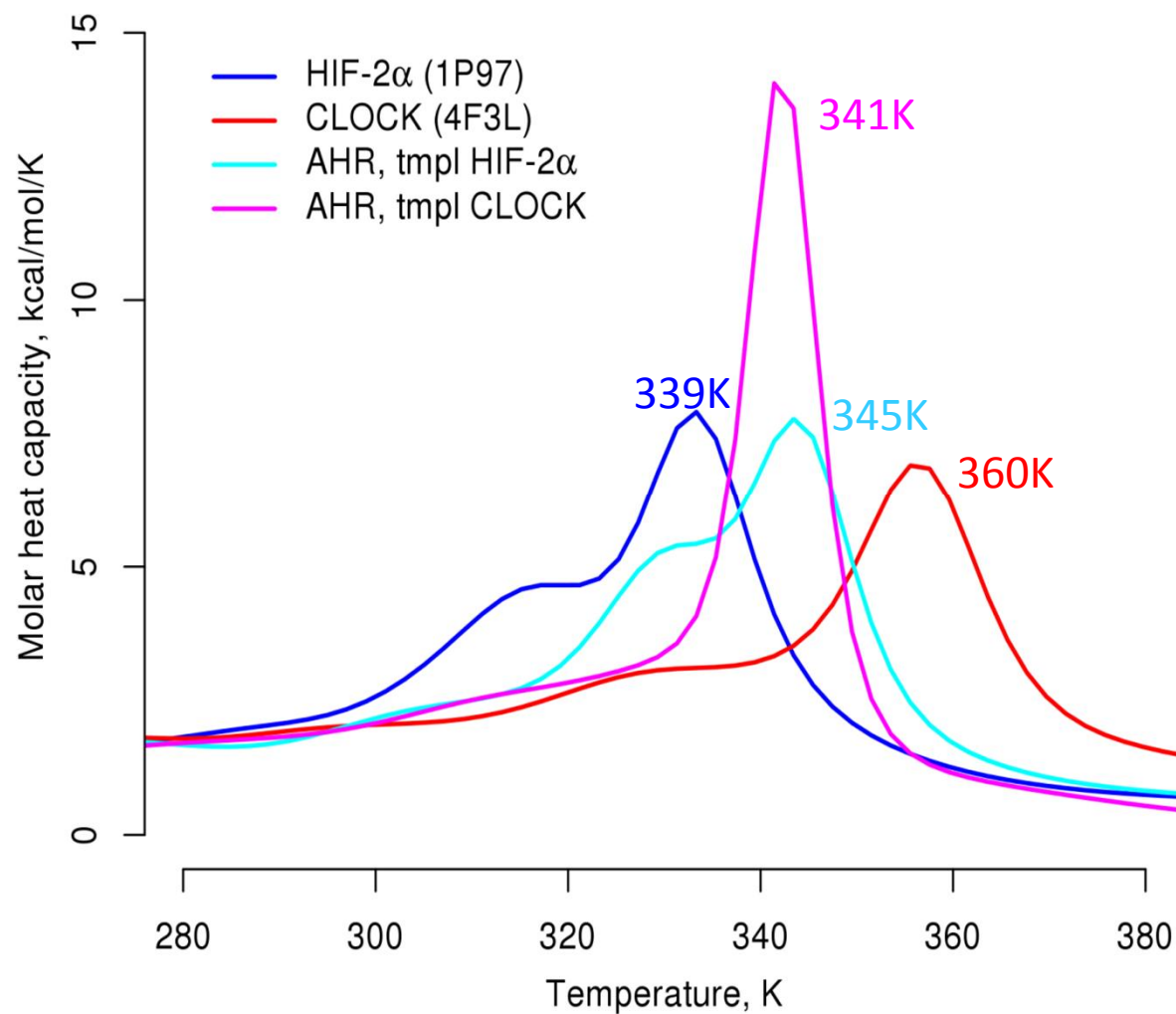
Per/Arnt/Sim



HIF2 α (PDBID:3F1P)

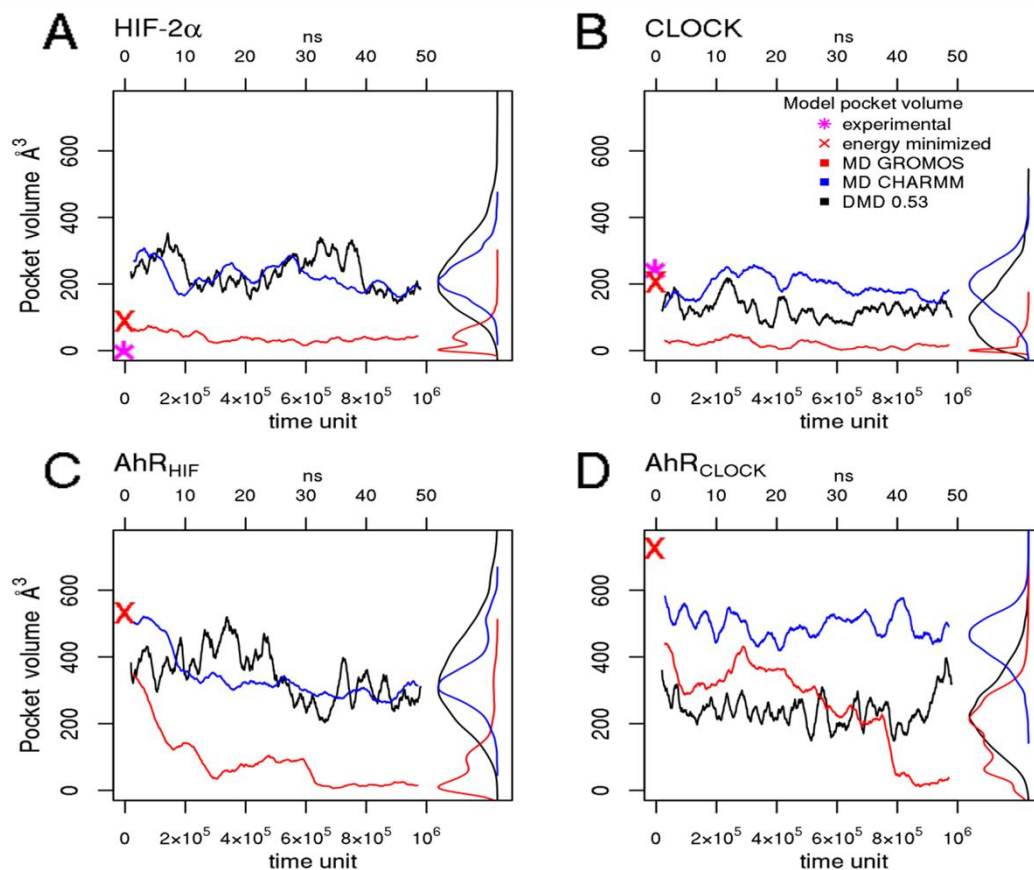


Replica Exchange DMD to characterize dynamics

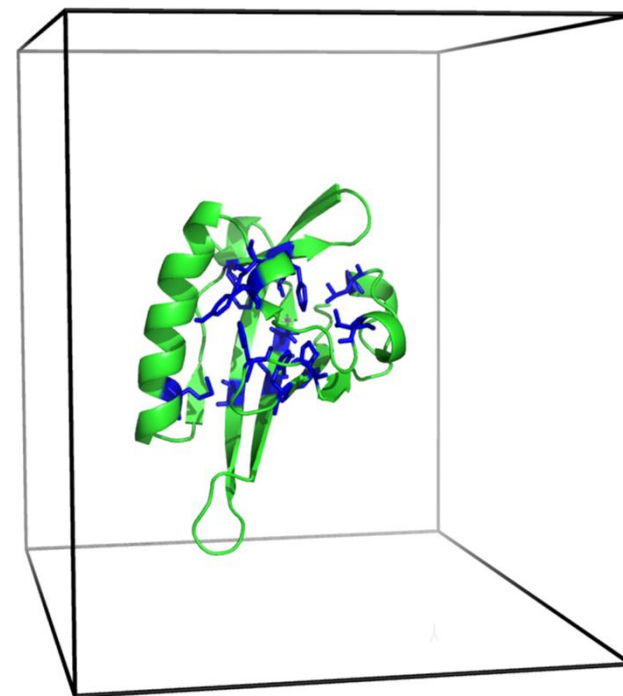


Dokkolás AhR PAS-B konformációs sokasághoz

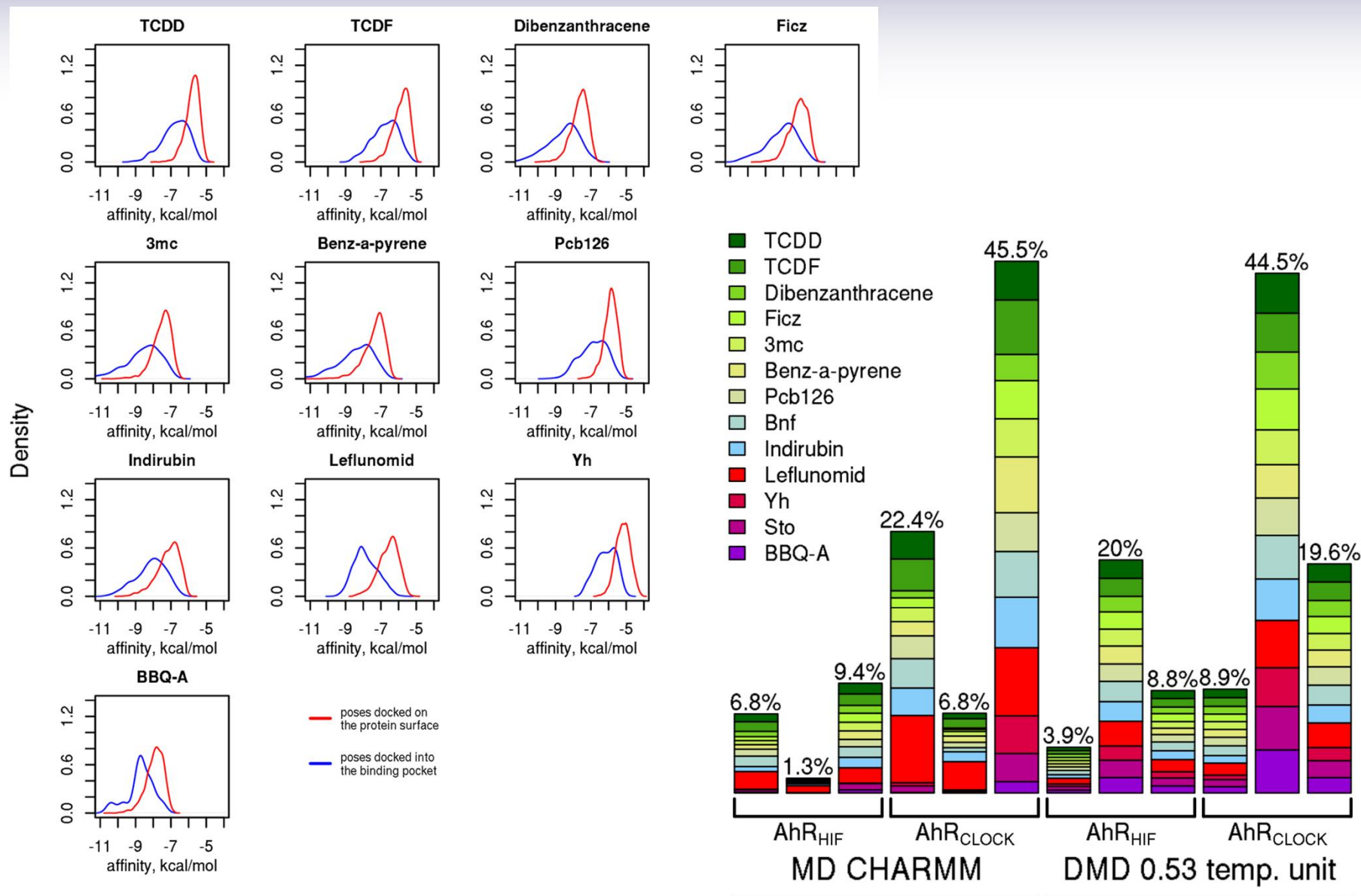
Molekuláris Dinamika (MD)



Dokkolás

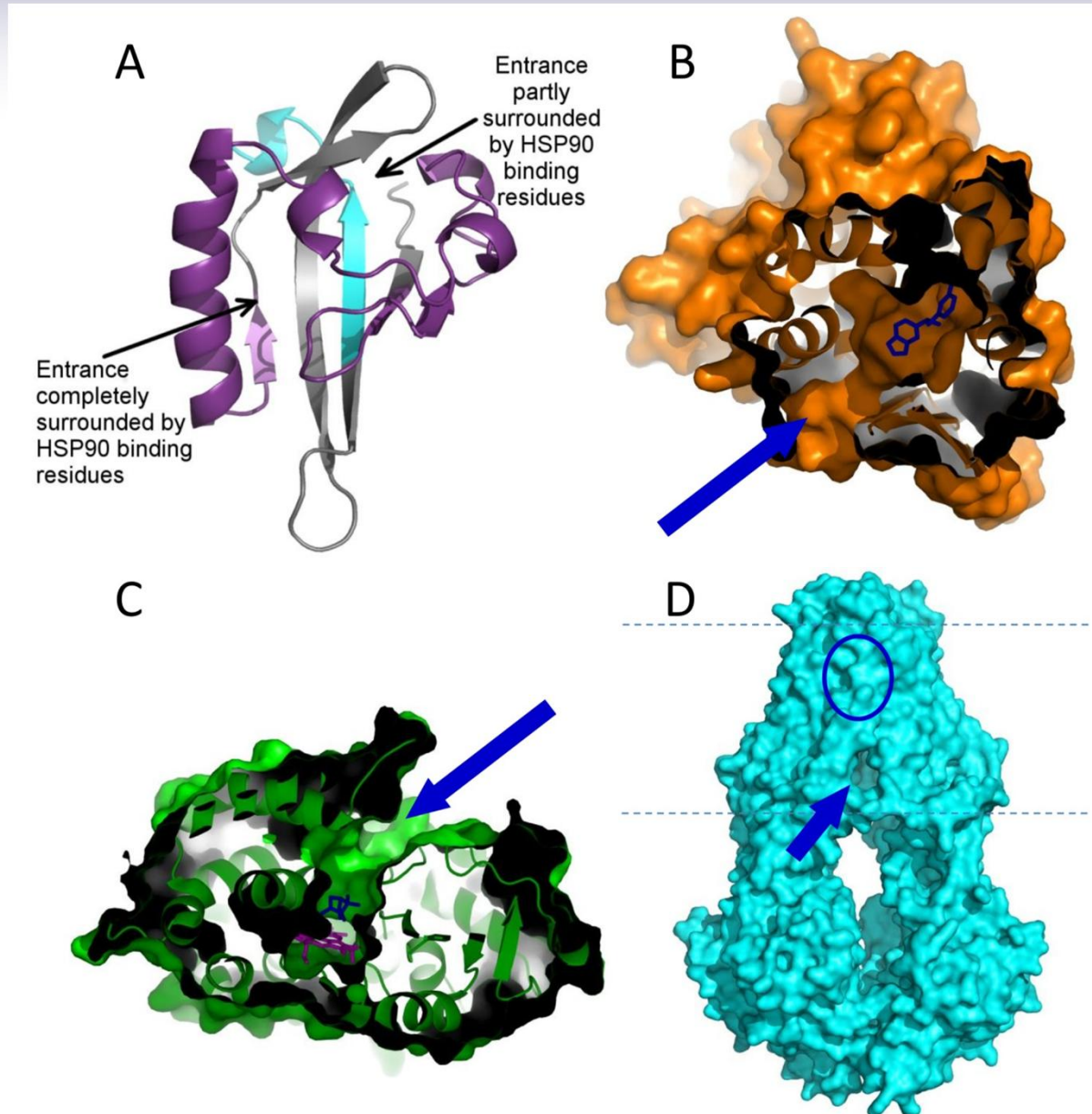


Eltérő affinitású ligandumok egyformán kötődnek



Kötőzseb és elérési útvonal együtt...

Szollosi et al. PLOS ONE
visszaküldve



Szollosi és Thomas és Tamás

Összefoglalás

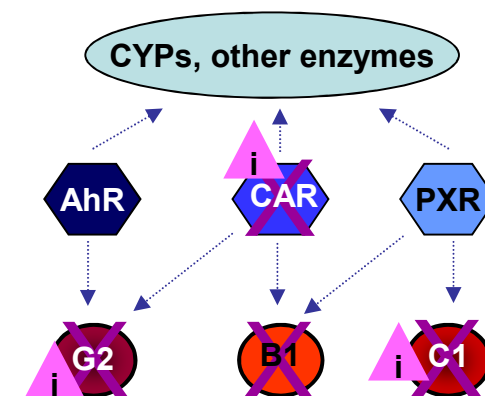
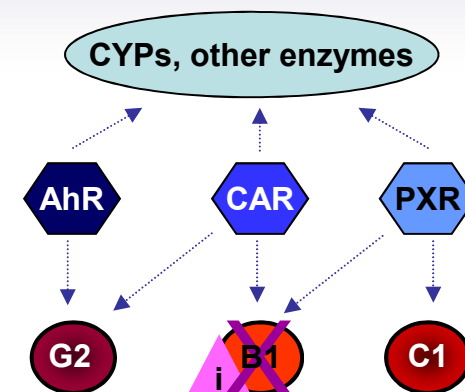
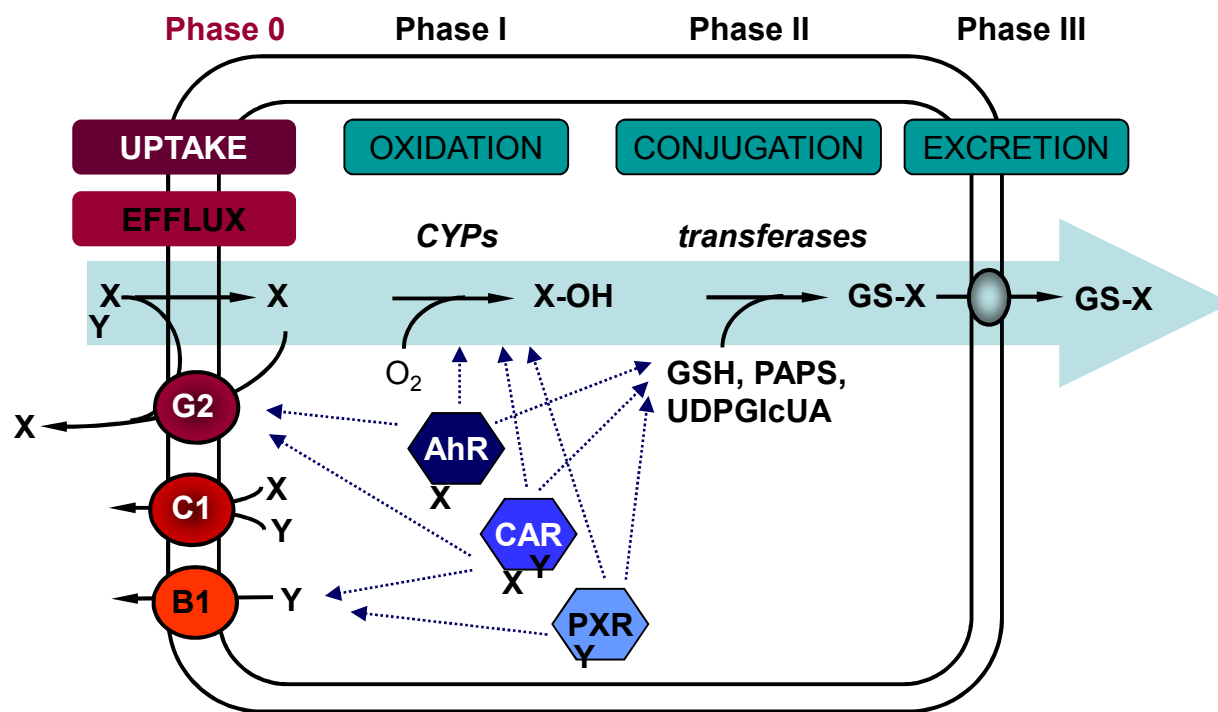
- Az AhR nagyfokú mozgékonyága hozzájárulhat a multidrog-kötéshez.
- A ligand bekötése által okozott mozgékonyág csökkenése vezethet a nukleáris transzlokációért felelős konformációs változásokhoz.
- A Hsp90 partnernek hármass szerepe lehet:
 1. a hidrofób kötőzseb elrejtése és felfedése;
 2. ligand (elő)szűrés;
 3. a mozgékonyág csökkenéséből fakadó alloszterikus kölcsönhatások propagálása.
- Az AhR PAS domén dinamikájának jellemzése nem csak a templáttól, de a *force field*-től is jelentősen függ.
- A multidrog-bekötődést nem pusztán a kötőzseb határozza meg.

Hálózatok – fehérje hálózatok

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

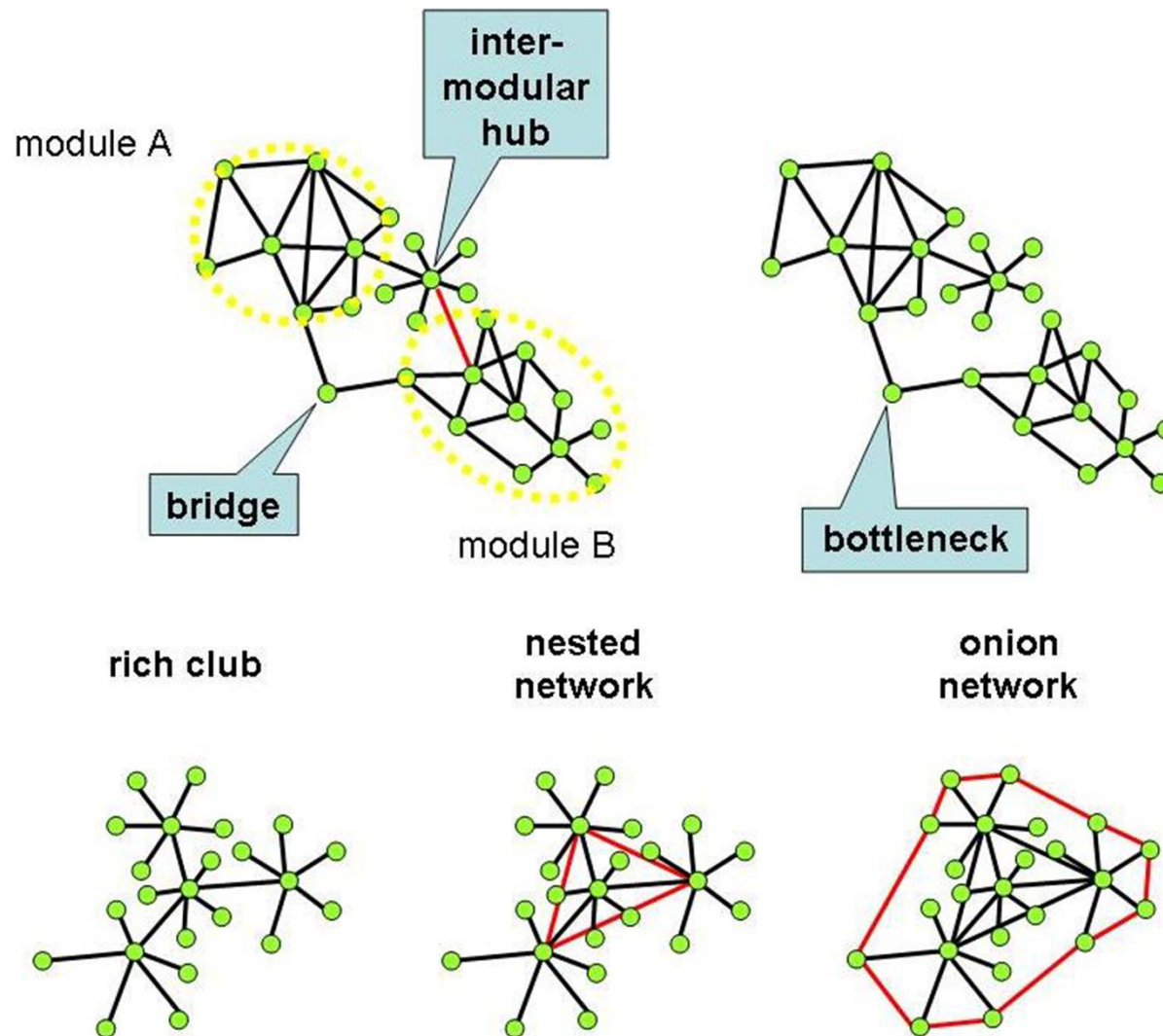
Csermely P. *et al.* 2012, <http://arxiv.org/abs/1210.0330>

A kemoimmun (ChI) rendszer (sejtszintű immunitás)

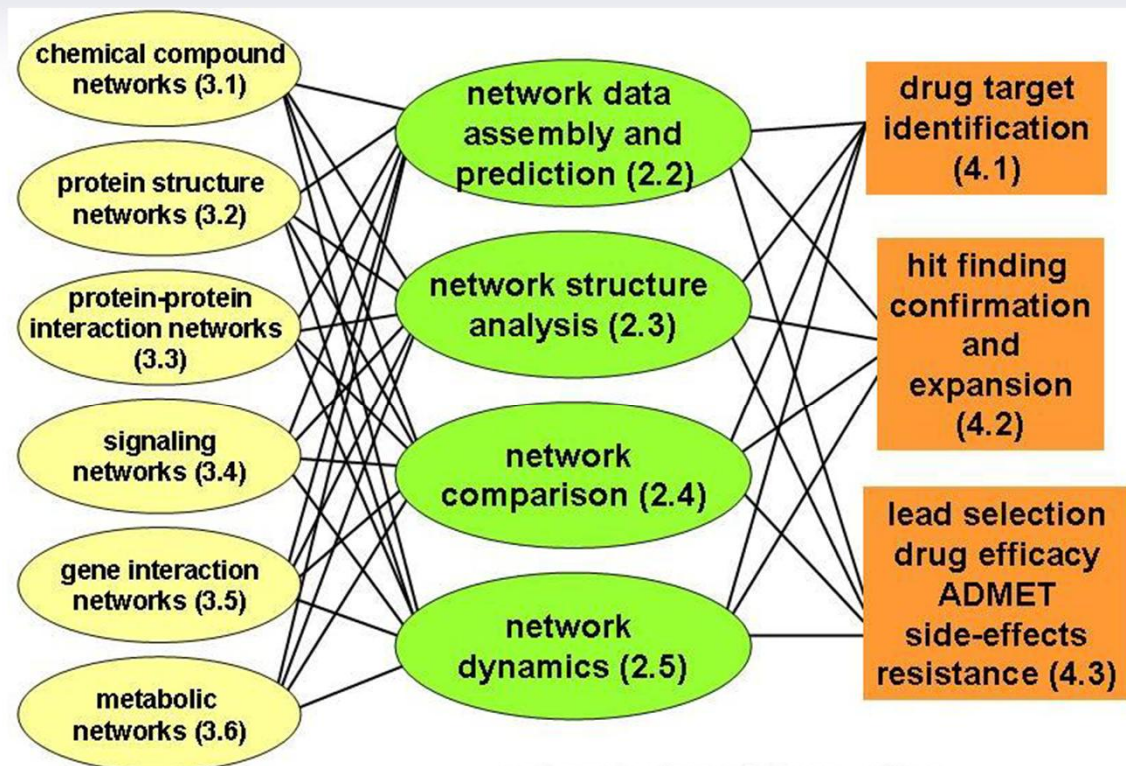


multi-target drugs

Hálózatok szerkezete



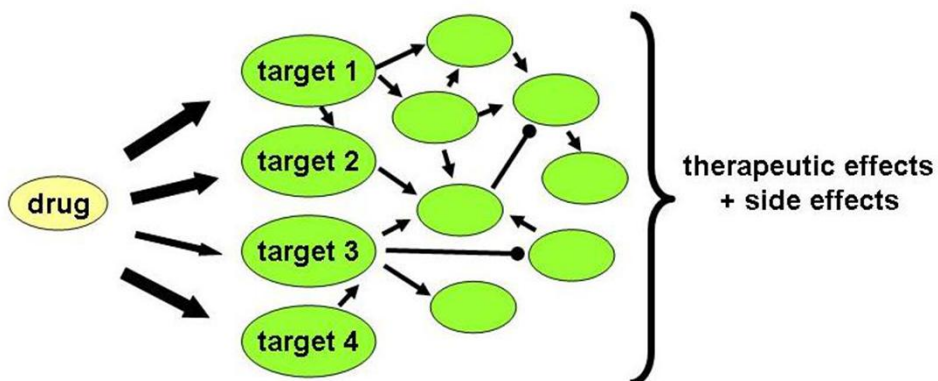
Hálózatok és gyógyszertervezés



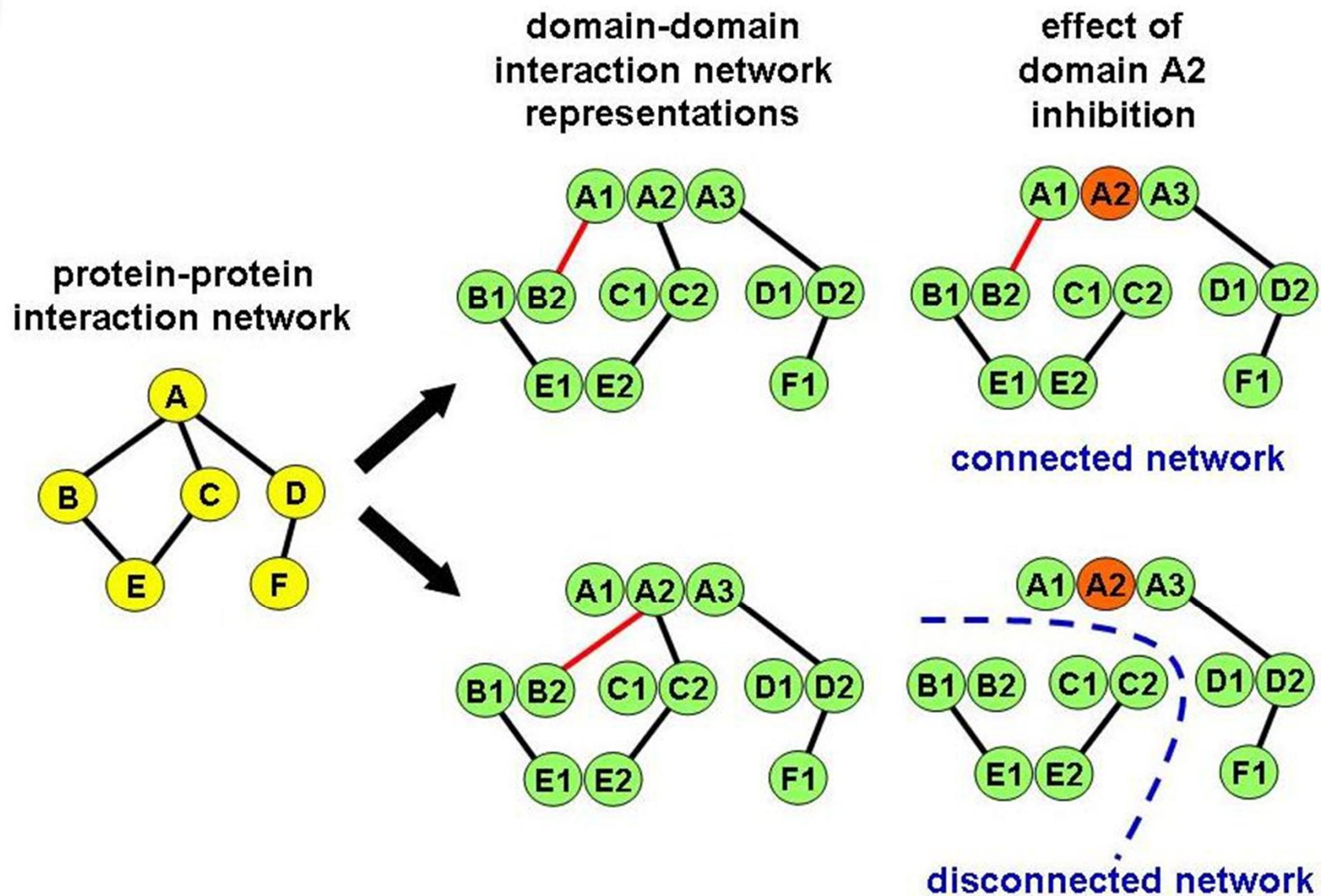
classic view of drug action



network view of drug action

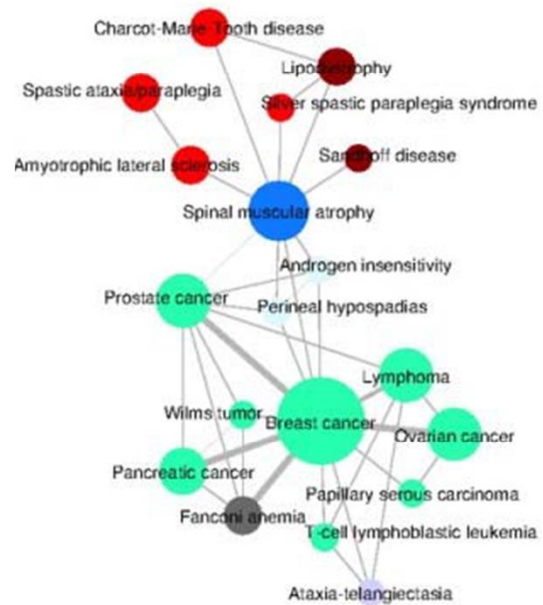


Fehérje-fehérje kölcsönhatási hálózatok



Network assembly

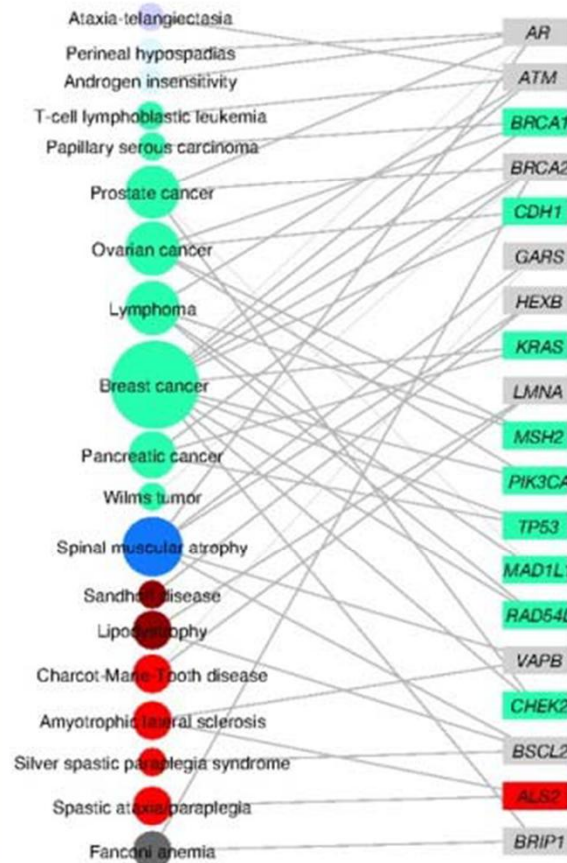
*Human Disease Network
(HDN)*



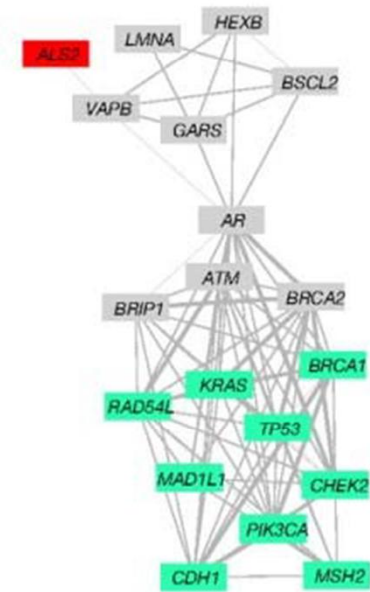
DISEASOME

disease phenome

disease genome

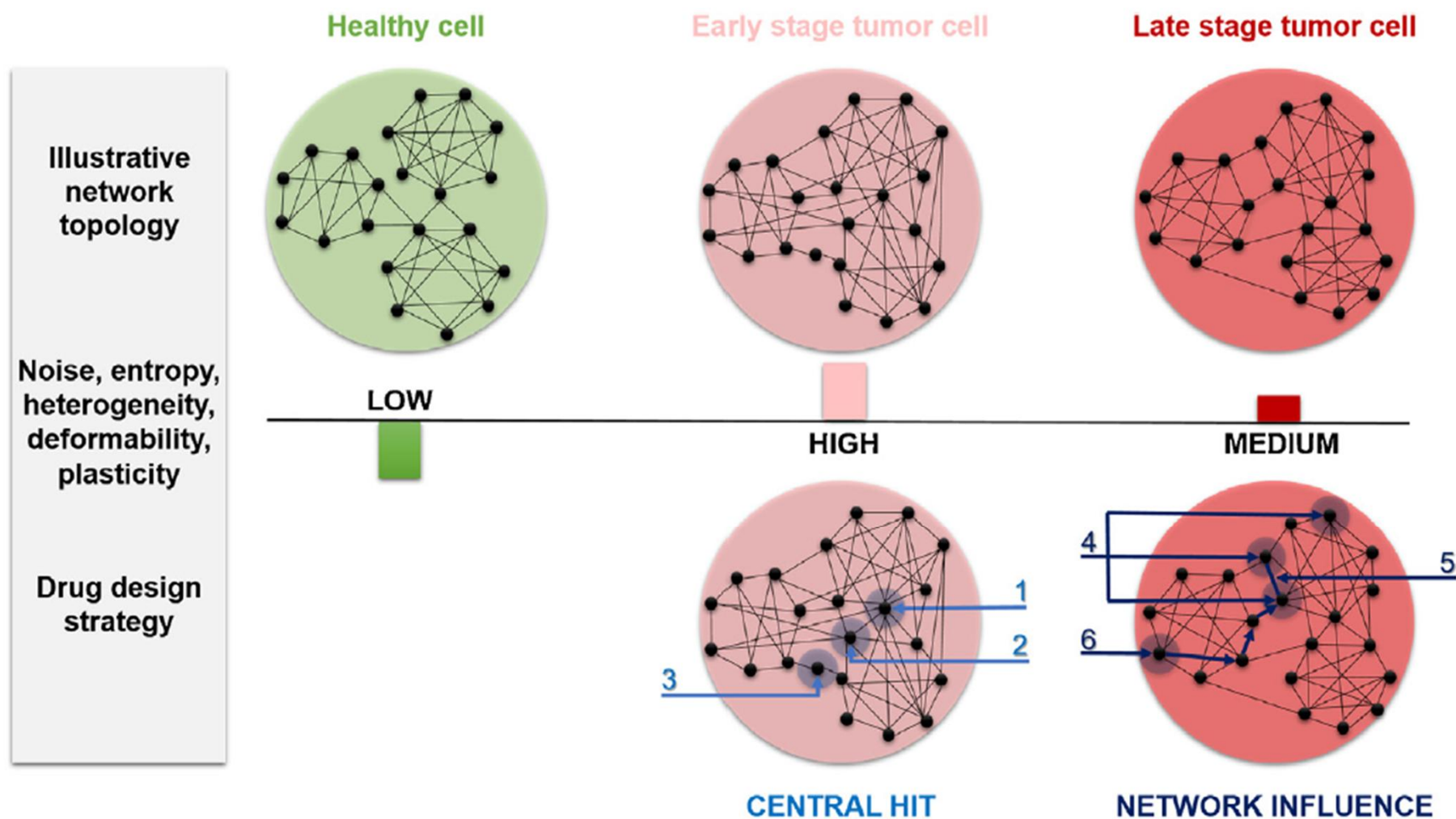


*Disease Gene Network
(DGN)*



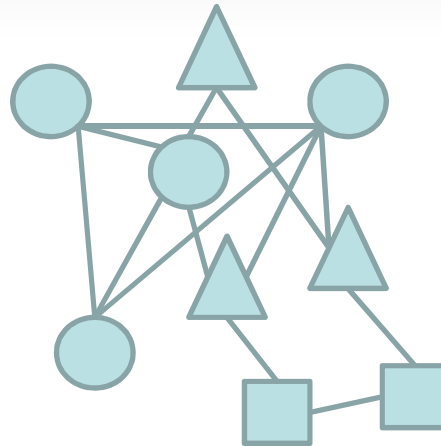
Rákos sejtek hálózatai

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

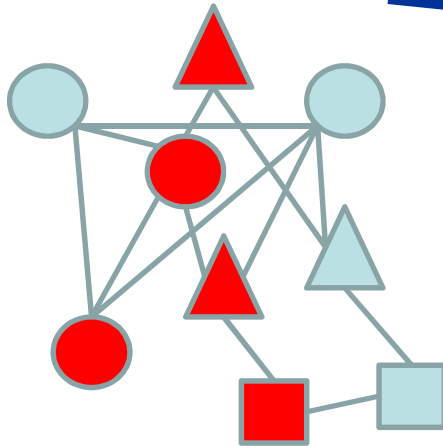


Hálózatok működésének feltérképezése

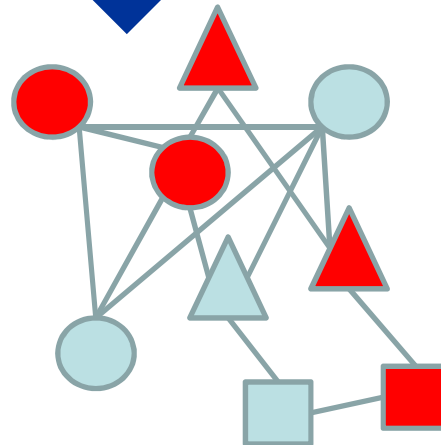
s_i - sejtvonali vagy drog_i vagy...



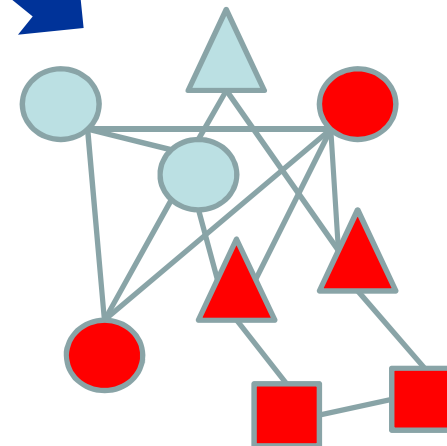
s_1



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

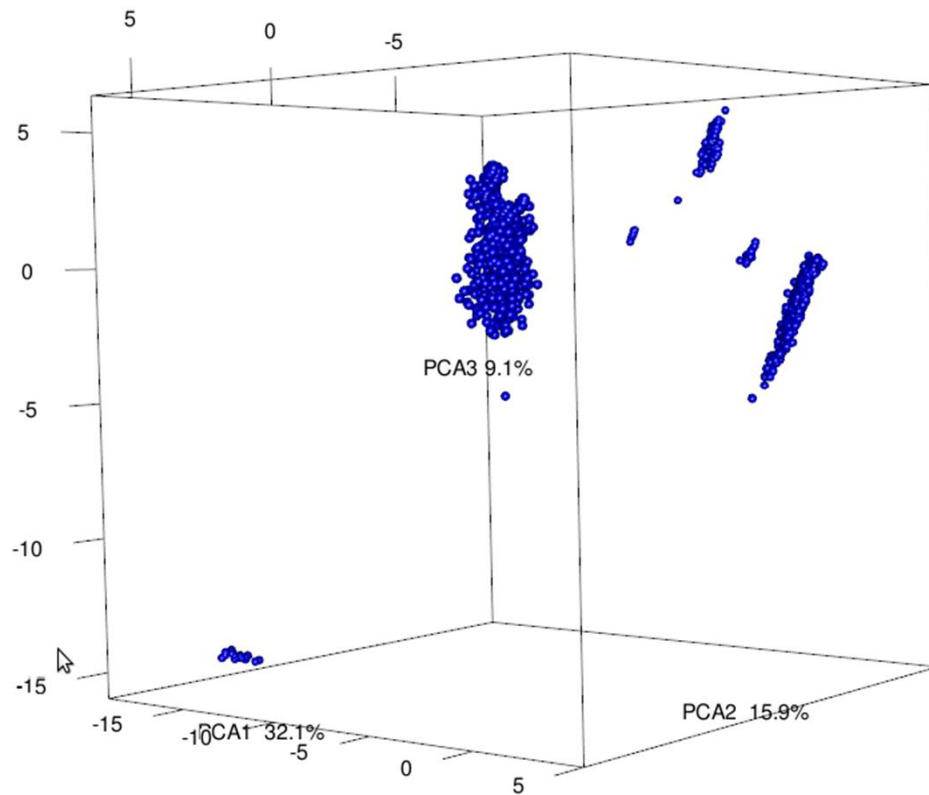
Expression-
changes

Expression-
patterns

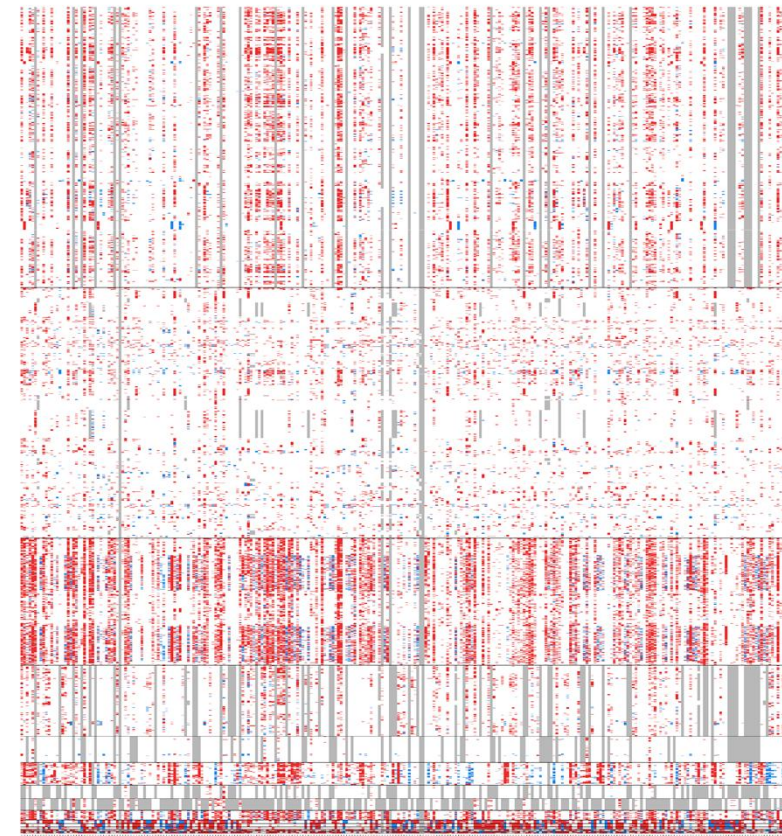
Biochemical
pathways

Whole dataset

PCA analysis (the first 3 component)



Heatmap (result of clustering)



experiments (883)

proteins (280)

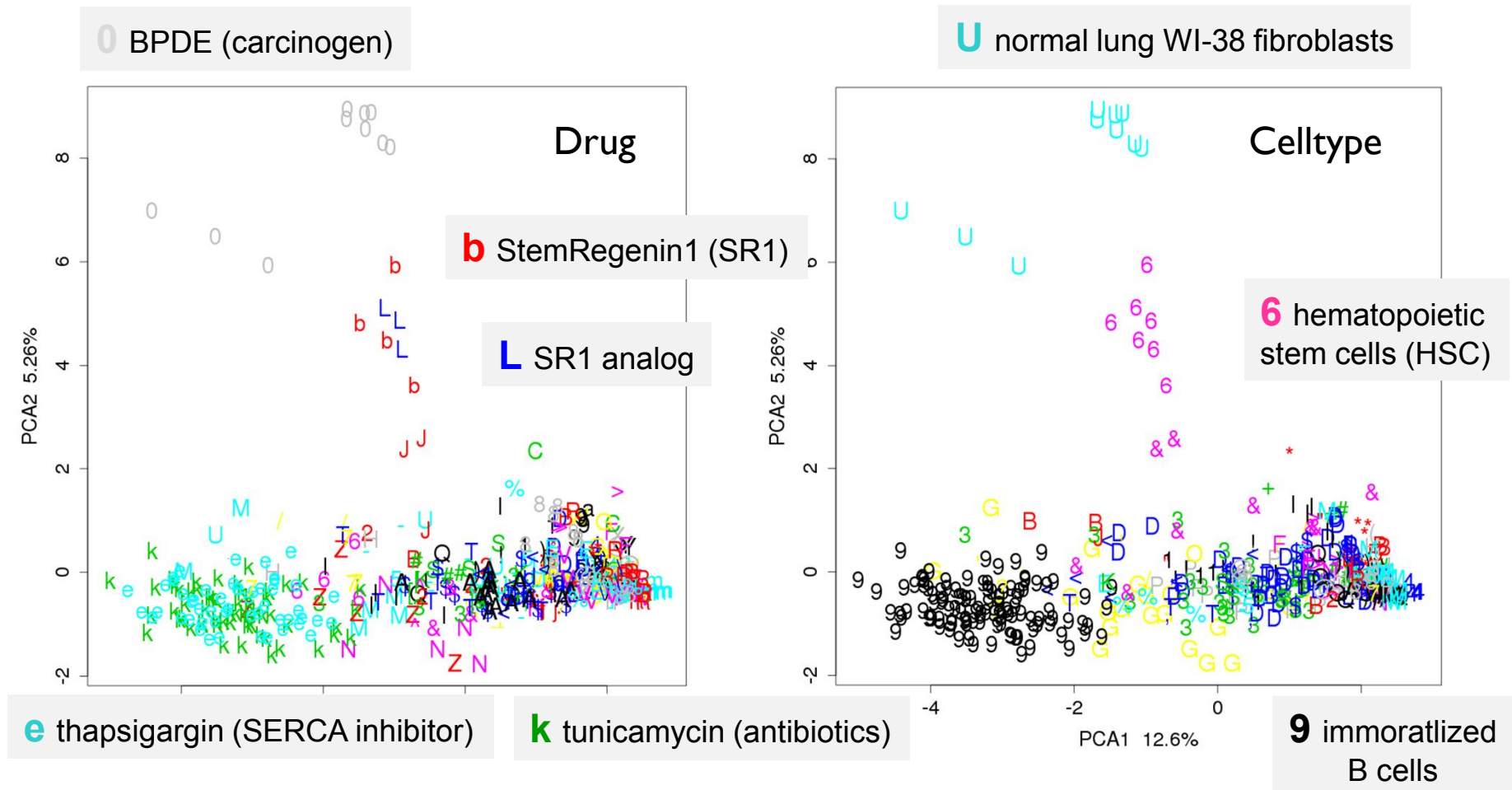
Data

Expression-
changes

Expression-
patterns

Biochemical
pathways

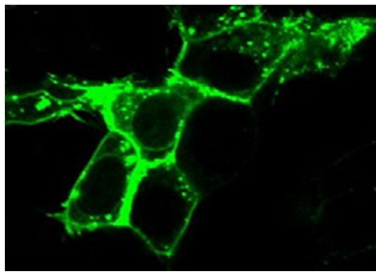
PCA analysis



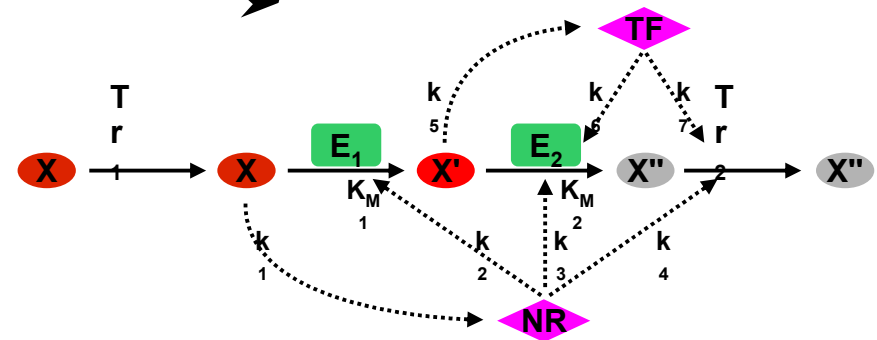
A ChI rendszer kinetikai modellezése



kísérletes kutatások

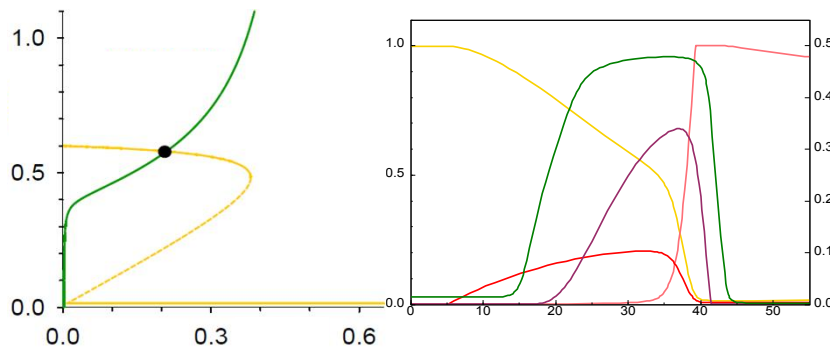


fiziológia



hálózati diagram

fázissík analízis



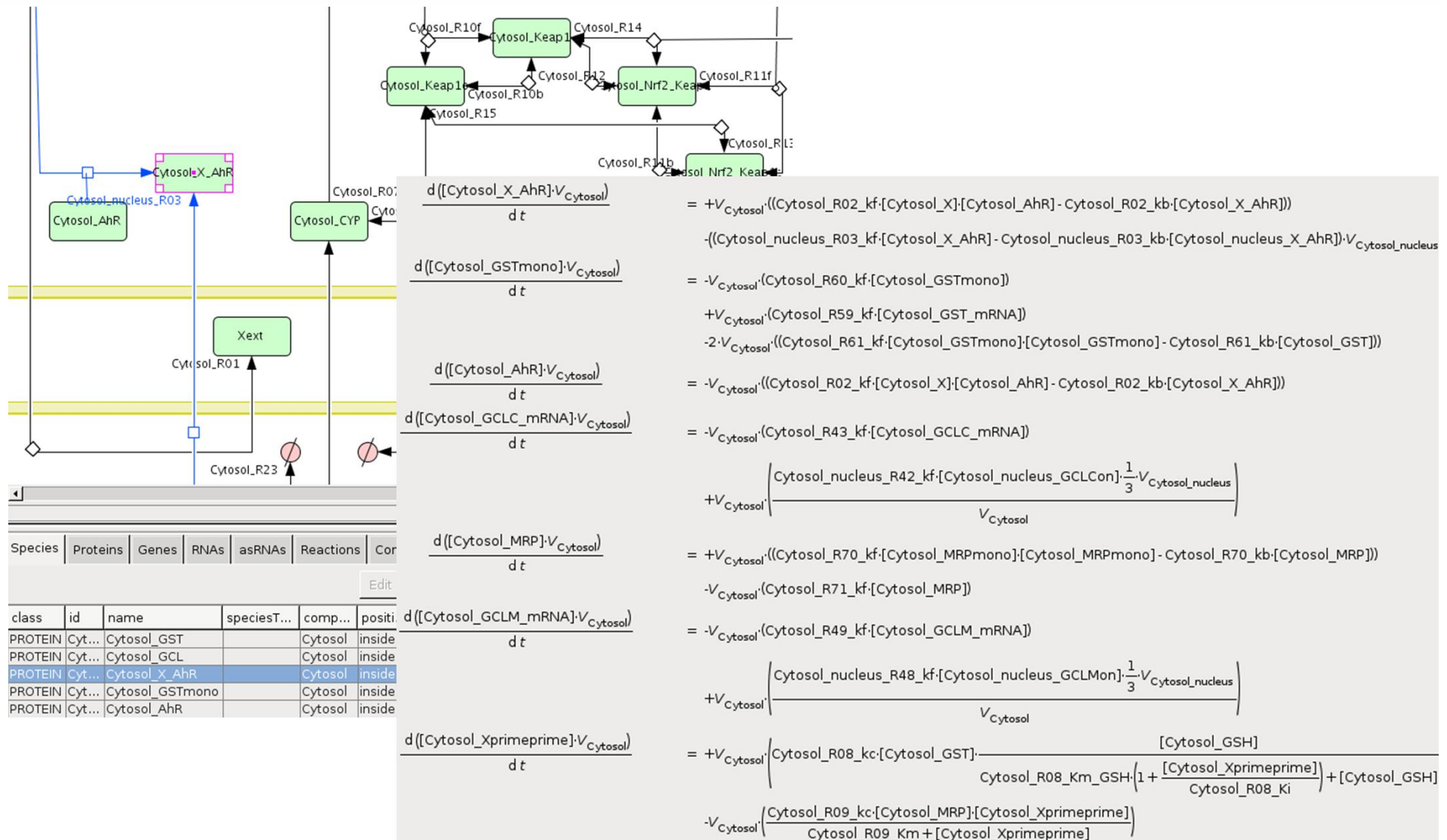
numerikus szimulációk



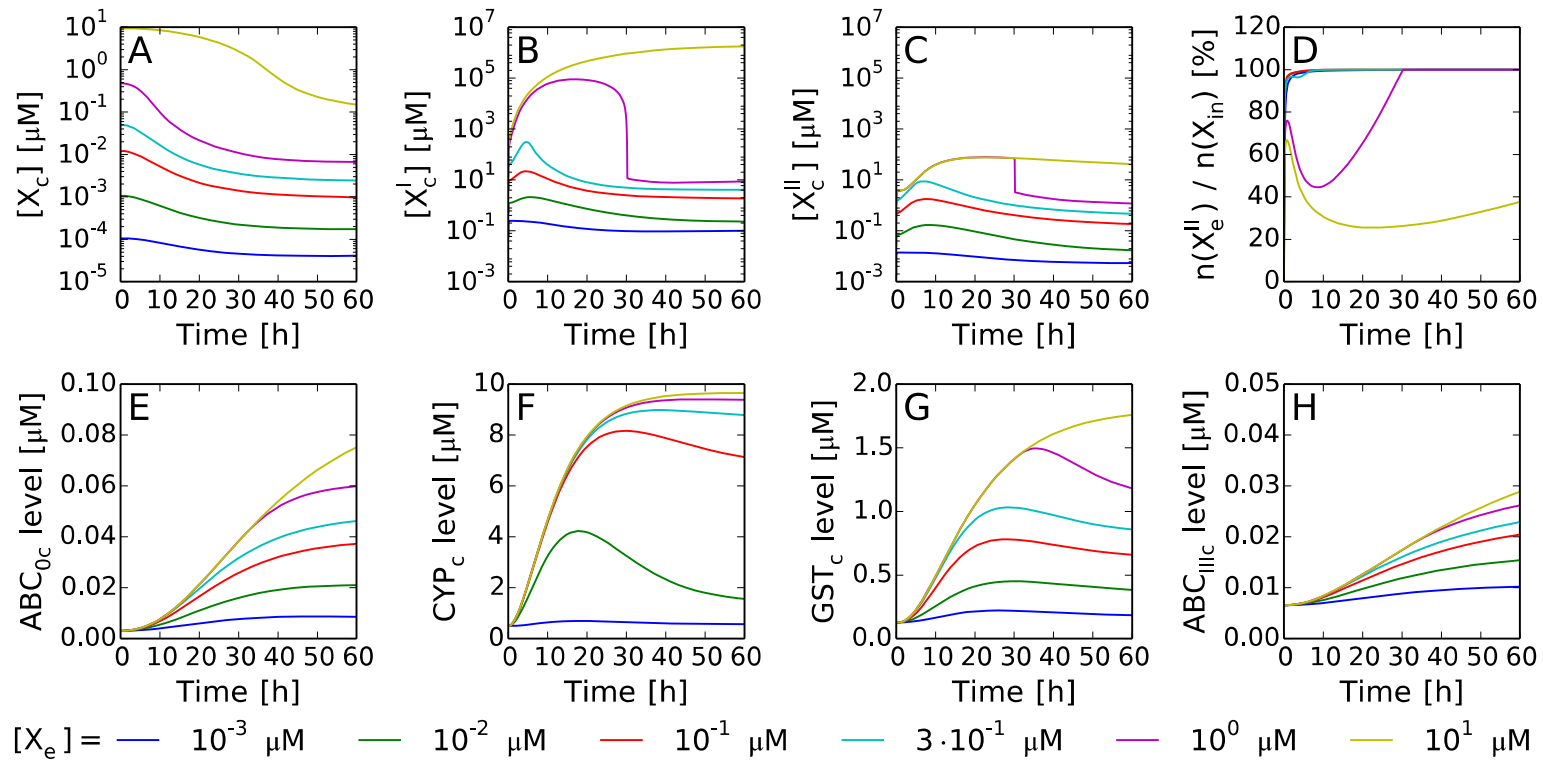
$$\begin{aligned}\frac{dX'}{dt} &= k_1 X - \frac{k_2 E_2 X'}{K_{M2} + X'} \\ \frac{dX''}{dt} &= \dots \\ \frac{dNR}{dt} &= \dots \\ \frac{dTF}{dt} &= \dots\end{aligned}$$

matematikai modell

Mathematical modeling to understand relationships in the chemodetox network

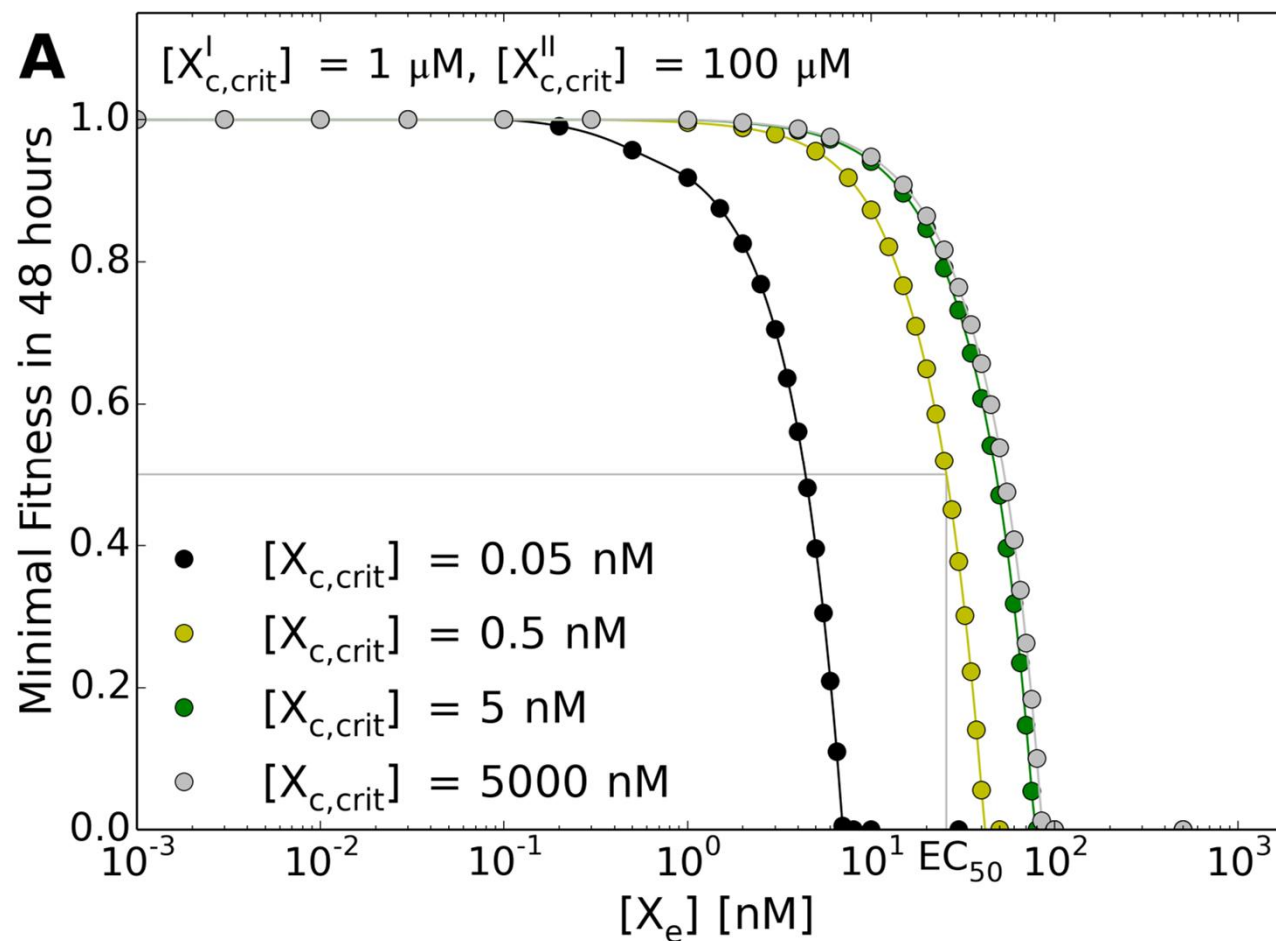


A rendszer kinetikai modellezése



A rendszer kinetikai modellezése

Ismeretlen vegyület mit csinál?



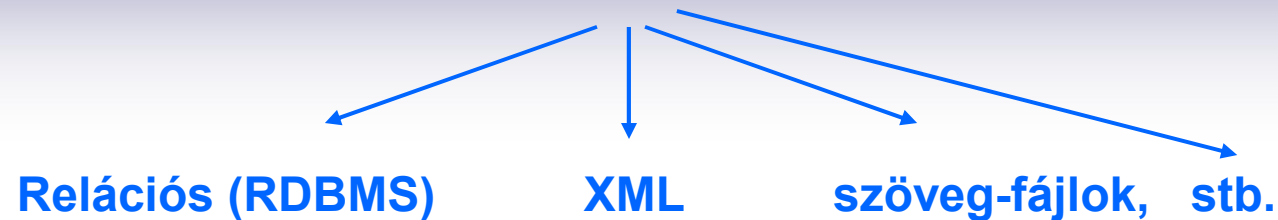
Összefoglalás

- **Bevezetés – a fehérje dinamika és a szimulációk jelentősége**
- **Fehérjék jellemzése bioinformatikai eszközökkel**
- **Fehérjék dinamikájának modellezése**
- **Fehérjék feltekeredésének szimulációja**
- **Informatikai eszközök – biológus szempontból**
- **Hálózatok, gének, fehérjék, drogok**

Informatikai "szemlélet"

- Adat
- Logika
- Megjelenítés

Adatbázisok formátuma



FASTA

```
>CFTR_HUMAN | P13569 | Cystic fibrosis transmembrane conductance regulator...
MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLSEKLEREWDR
LASKKNPKLINALRRCCFFWRFMFYGIFLYLGEVTKAVQPLLLGRIIASYDPDNKEERSIA
IYLGIGLCLLFIVRTLHHPAIFGLHHIGMQMRIAMFSLIYKTKLSSRVLDKISIGQL
VSLLSNNLNKFDEGLALAHFVWIAPLQVALLMGLIWELLQASAFGLGFLIVLALFQAGL
GRMMKYRDQRAGKISERLVITSEMIENIQSVKAYCWEEAMEKMIENLRQTELKLTRKAA
... DTRL
```

PIR

```
>P1;CRAB_ANAPL
ALPHA CRYSTALLIN B CHAIN (ALPHA(B)-CRYSTALLIN) .
MDITIHNPLI RRPLFSWLAP SRIFDQIFGE HLQESSELLPA SPSLSPFLMR
SPIFRMPSWL ETGLSEMRLE KDKFSVNLDV KHFSPEELKV KVLGDMVEIH
GKHEERQDEH GFIAREFNRK YRIPADVDP L TITSSLSLDG VLTVSAPRKQ
SDVPERSIPI TREEKPAIAG AQRK*
```

Szerkezeti, pdb fájl formátum

```

HEADER      MEMBRANE PROTEIN                      26-OCT-07   3B60
TITLE      CRYSTAL STRUCTURE OF MSBA FROM SALMONELLA TYPHIMURIUM WITH
...
ATOM       1  N   TRP A  10      104.628 -32.601  66.108  1.00205.48      N
ATOM       2  CA  TRP A  10      104.119 -32.609  64.706  1.00205.48      C
ATOM       3  C   TRP A  10      103.171 -31.436  64.470  1.00205.48      C
ATOM       4  O   TRP A  10      102.922 -30.633  65.393  1.00205.48      O
ATOM       5  CB  TRP A  10      103.367 -33.919  64.430  1.00205.48      C
ATOM       6  CG  TRP A  10      102.940 -34.096  62.995  1.00205.48      C
ATOM       7  CD1 TRP A  10      103.750 -34.347  61.925  1.00205.48      C
ATOM       8  CD2 TRP A  10      101.605 -34.018  62.477  1.00205.48      C
ATOM       9  NE1 TRP A  10      103.004 -34.430  60.774  1.00205.48      N
ATOM      10  CE2 TRP A  10      101.684 -34.229  61.083  1.00205.48      C
ATOM      11  CE3 TRP A  10      100.349 -33.784  63.055  1.00205.48      C
ATOM      12  CZ2 TRP A  10      100.555 -34.220  60.256  1.00205.48      C
ATOM      13  CZ3 TRP A  10          99.224 -33.775  62.232  1.00205.48      C
ATOM      14  CH2 TRP A  10          99.338 -33.990  60.847  1.00205.48      C
ATOM      15  N   GLN A  11      102.764 -31.247  63.200  1.00205.36      N
ATOM      16  CA  GLN A  11      101.723 -30.228  63.006  1.00205.36      C
ATOM      17  C   GLN A  11      102.262 -28.816  63.134  1.00205.36      C

```

Adatbázisok fizikai lokalizációja



internetes

lokális

Előnyei:

- Mások tartják karban (frissítés és annotálás)
- Máshol foglal erőforrásokat
- Általában több helyen elérhető (hardware hiba toleráns)

Hátrányai:

- Mások tartják karban
- Adott eszköztár
- Lassú elérés

Előnyei:

- lokális
- gyors elérés
- adott verzió (kézirat!)
- „akármilyen” szoftverrel használható

Hátrányai:

- lokális
- erőforrás-igény
- hozzáértés-igény

Biológiai adatbázisok

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

The screenshot displays the NCBI (National Center for Biotechnology Information) homepage. At the top, there is a navigation bar with links for 'Resources' and 'How To', and a 'Sign in to NCBI' link. Below this, the NCBI logo and name are visible. A search bar is located on the right side of the header. On the left, a vertical navigation menu lists various categories: 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, and Variation. A dropdown menu is open under 'All Databases', showing a list of recent databases including All Databases, Assembly, BioProject, BioSample, BioSystems, Books, ClinVar, Clone, Conserved Domains, dbGaP, dbVar, Epigenomics, EST, Gene, Genome, GEO DataSets, and GEO Profiles. The main content area features the NCBI logo and a brief description of the center's mission. Below this, there are links for 'Mission', 'Organization', 'Research', and 'NCBI News'. A section titled 'Popular Resources' lists various tools and databases: PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. Another section titled 'NCBI Announcements' contains news items, such as 'dbVar releases 1000 Genomes Phase 3 structural variants' dated Nov 4, 2014, and 'dbVar has released structural variation (SV) data from developmental delay study cited in Nature Reviews Genetics' dated Nov 3, 2014. At the bottom, there is a 'NCBI YouTube channel' banner with a 'GO' button and a video player interface.

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.

[\[details\]](#)

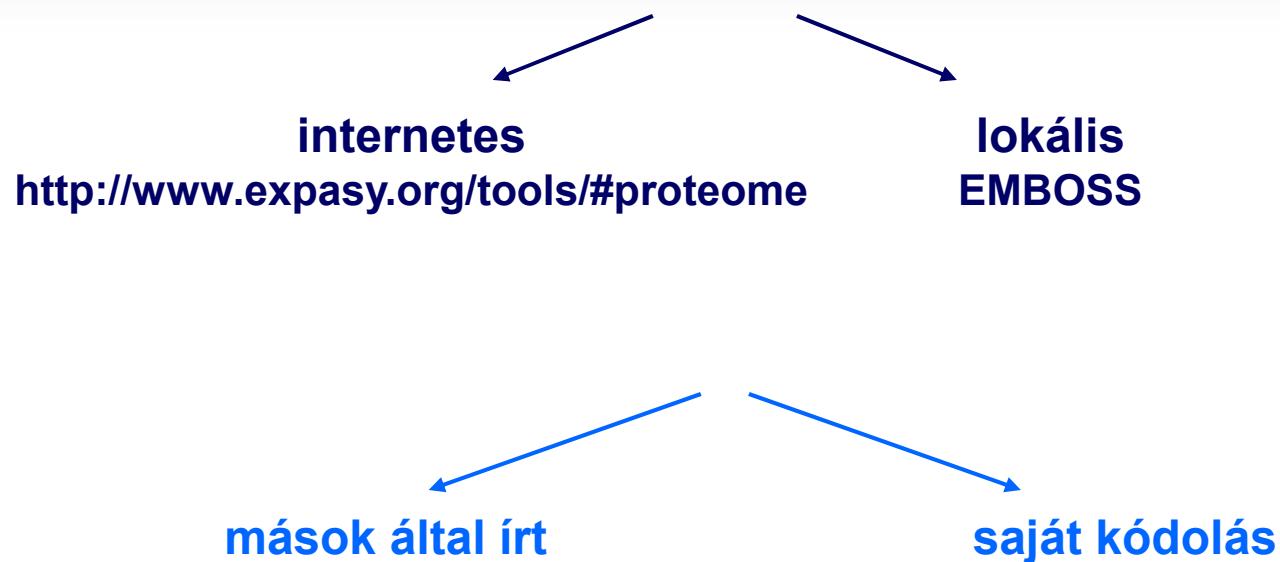
MADAP



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>])

Saját programok – programozási nyelvek

C/C++:	lassú fejlesztés ha sebesség kell; mégis ritkán tanácsolt
Java:	lassú fejlesztés; általában a szükségelt csomag beta ☹
Script nyelv:	igen gyors fejlesztés bizonyos feladatokhoz igen lassú

- Könyvtárak
- Olvashatóság, dokumentálhatóság
- Objektum orientáltság
- Több fejlesztő: subversion vagy hasonló megoldások
- Adatbázis kezelés (Big Data); ORM!
- Egyéni száj-íz

A Kiválasztott szkriptnyelv

Java

```
public class HelloWorld
{
    public static void main (String[] args)
    {
        System.out.println("Hello!");
    }
}
```

```
int    myCounter = 0;
String myString = String.valueOf(myCounter);
if (myString.equals("0")) ...
```

Python

```
print "Hello, world!"
```

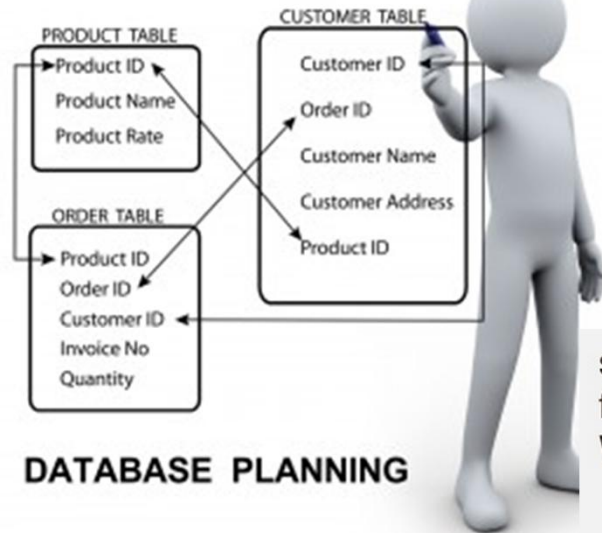
```
myCounter = 0
myString = str(myCounter)
if myString == "0": ...
```

Perl

```
sub foo {
    my $arg1 = shift;
    my @rest_of_args = @_;
    ...do stuff...
}
```

```
def foo (arg1, rest_of_args):
    ...do stuff...
```

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>

Prezentáció – cmdIn vs GUI

Iceweasel

File Edit View History Bookmarks Tools Help

Circular chromosome... x Nuclear organization ... x file:///home...t_b6_g1.html x

file:///home/hegedus/projects/babu_abcprt/gergonek/out_b6_g1.html

rl compare syntax few line of code3

Save2Mendeley HTML R CPP Python BioInfo PubMed pubcrawler EnHu Webster MyDb Maps Időjárás ToRead Irodalmazo HPC En ChemInfo lipids

ABCB6 PPI data compared to that of ABCG1

Pathway	Proteins in both the pathway and experiment	Proteins only in the pathway
ABC-family proteins mediated transport (11) [reactome.org]	CFTR_HUMAN*	ABCA7_HUMAN APOA1_HUMAN PEX19_HUMAN PEX3_HUMAN ABCG8_HUMAN ABCG5_HUMAN ABCG1_HUMAN ABCG4_HUMAN ABCB6_HUMAN ABCA4_HUMAN
Activated NOTCH1 Transmits Signal to the Nucleus (16) [reactome.org]	RL40_HUMAN	DLL1_MOUSE DLL4_MOUSE NOTC1_MOUSE JAG1_MOUSE JAG2_MOUSE ADA10_MOUSE NICA_MOUSE PEN2_MOUSE CNTN1_MOUSE NICA_RAT PEN2_RAT DNER_MOUSE ITCH_MOUSE DLK1_MOUSE NUMB_MOUSE
Activated NOTCH1 Transmits Signal to the Nucleus (11) [reactome.org]	RL40_HUMAN	DL_DROME SERR_DROME NOTCH_DROME PSN_DROME NICA_DROME PEN2_DROME Q9V393_DROME DTX_DROME SUDX_DROME NUMB_DROME
Activation of BAD and translocation to mitochondria (6) [reactome.org]	1433E_HUMAN	BAD_HUMAN CANB1_HUMAN PP2BC_HUMAN BID_HUMAN BCL2_HUMAN
Activation of Chaperone Genes by ATF6-alpha (10) [reactome.org]	ENPL_HUMAN GRP78_HUMAN	NFYC_HUMAN NFYA_HUMAN NFYB_HUMAN ATF4_HUMAN ATF6A_HUMAN DDIT3_HUMAN CALR_HUMAN XBP1_HUMAN
Activation of DNA fragmentation factor (12) [reactome.org]	H10_HUMAN HMGB1_HUMAN H14_HUMAN H12_HUMAN HMGB2_HUMAN H11_HUMAN	H13_HUMAN DFFA_HUMAN DFFB_HUMAN IMA1_HUMAN IMB1_HUMAN CASP3_HUMAN
Activation of Matrix Metalloproteinases (15) [reactome.org]	TRY1_HUMAN	MMP1_HUMAN TIMP1_HUMAN MMP9_HUMAN TIMP2_HUMAN MMP14_HUMAN MMP2_HUMAN COI1_HUMAN MMP3_HUMAN MMP7_HUMAN MMP8_HUMAN MMP10_HUMAN MMP11_HUMAN FURIN_HUMAN MMP13_HUMAN
Adaptive Immune System (214) [reactome.org]	PSMD2_HUMAN PSMD1_HUMAN PRS6A_HUMAN* PDIA3_HUMAN CLH1_HUMAN	PSMD3_HUMAN CALX_HUMAN PSD7_HUMAN PRS8_HUMAN PHAG1_HUMAN CSK_HUMAN CD3E_HUMAN CD3G_HUMAN CD3D_HUMAN CD3Z_HUMAN LCK_HUMAN CD4_HUMAN PTPRC_HUMAN ZAP70_HUMAN LAT_HUMAN GRAP2_HUMAN LCP2_HUMAN ITK_HUMAN PLCG1_HUMAN FYB_HUMAN NCK1_HUMAN WASP_HUMAN TRAF1_HUMAN SHIP1_HUMAN PTEN_HUMAN PDPK1_HUMAN KPCT_HUMAN CAR11_HUMAN BCL10_HUMAN RIPK2_HUMAN MALT1_HUMAN TRAF6_HUMAN UBE2N_HUMAN UB2V1_HUMAN TAB2_HUMAN M3K7_HUMAN NEMO_HUMAN IKKA_HUMAN IKKB_HUMAN IKBA_HUMAN NFKB1_HUMAN TF65_HUMAN CD28_HUMAN CD80_HUMAN CD86_HUMAN PK3CA_HUMAN MTOR_HUMAN RICTR_HUMAN LST8_HUMAN SIN1_HUMAN M3K8_HUMAN M3K14_HUMAN GRB2_HUMAN VAV_HUMAN FYN_HUMAN RAC1_HUMAN

[win7 [Running] ...] LXTerminal Iceweasel

11:26 US