Fehérjék szerkezetének predikciója, szerkezeti adatok felhasználása adatbázisok segítségével, a számítógépes molekuladinamikai modellezés alapjai

Hegedűs Tamás

tamas@hegelab.org

MTA-SE Molekuláris Biofizikai Kutatócsoport
SE Biofizikai és Sugárbiológiai Intézet

Mai témák

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

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.

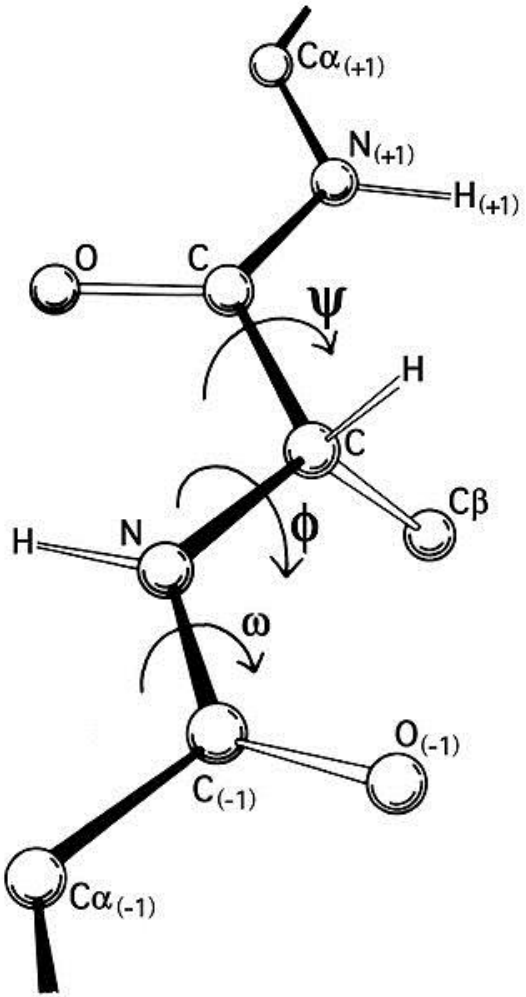
Kísérletes módszerek általában nem szolgáltatnak közvetlen információt az atomi szintű történésekről.

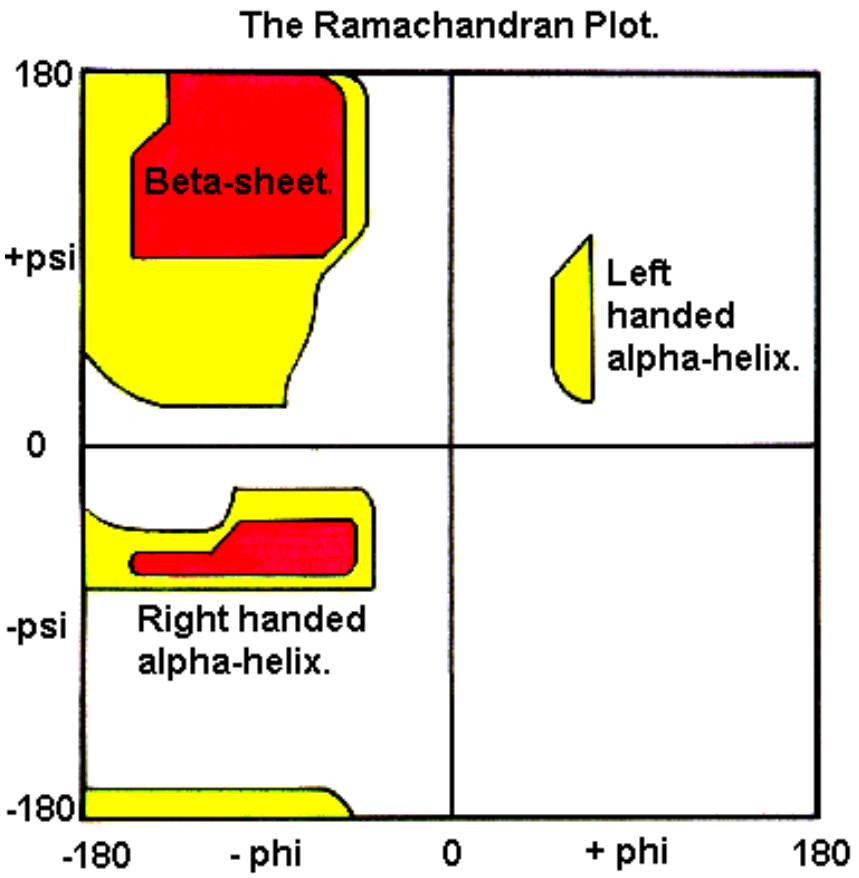
Pl. NMR és MD - igen

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
- Informatikai eszközök – biológus szempontból
- Fehérjék dinamikájának modellezése
- Fehérjék feltekeredésének szimulációja

Másodlagos szerkezeti elemek





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 pozicióra

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 proteolítikus 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:

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

Ha csak szekvencia van:

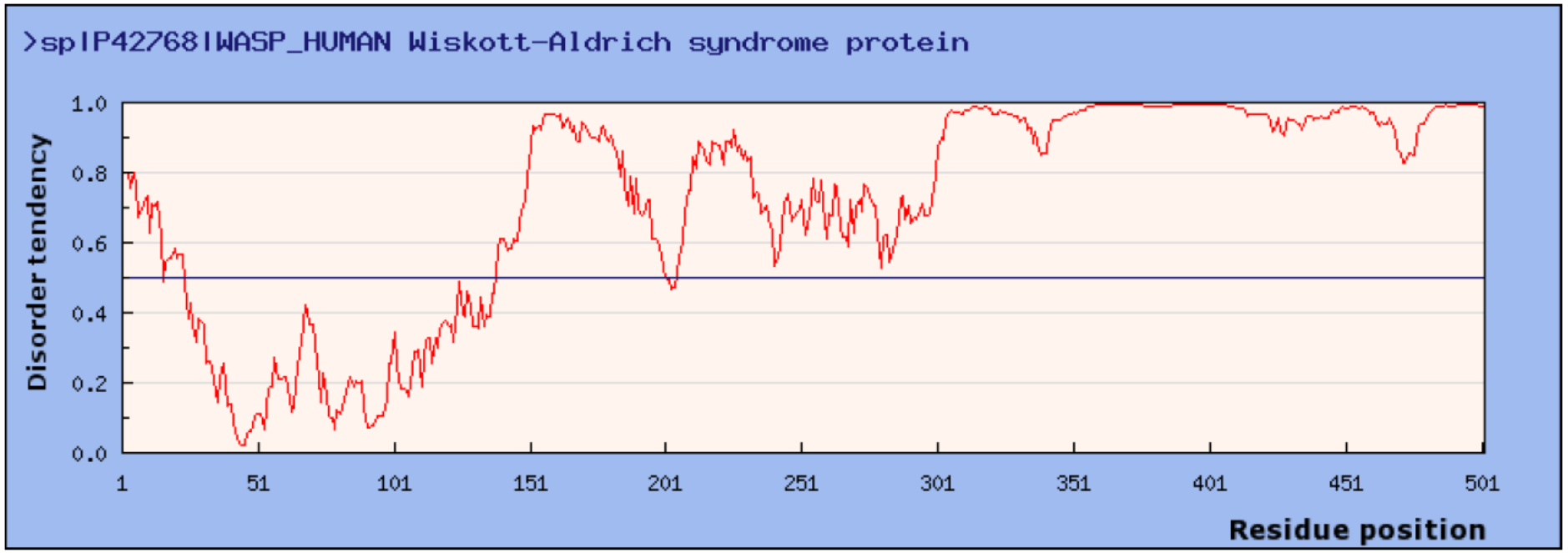
$$E_{estimated} = L \sum_{i,j} P_{ij} f_i f_j$$

Egy aminosav rendezetlensége:

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

Mészáros B. PhD dolgozat

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**
- **Informatikai eszközök – biológus szempontból**
- **Fehérjék dinamikájának modellezése**
- **Fehérjék feltekeredésének szimulációja**

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 meghatározó**

Szekvencia-illesztés

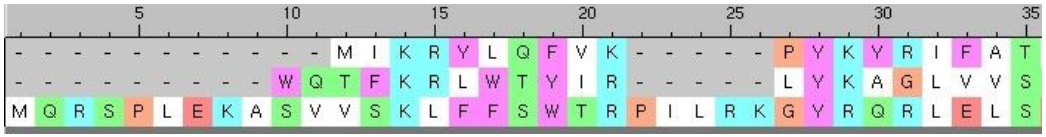
	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1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

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

Alignement – pl. ClustalW

```
CLUSTAL W (1.83) multiple sequence alignment


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



Basic Local Alignment **Search** Tool, or **BLAST**

Negyedleges szerkezet

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

PISA - Protein Interfaces, Surfaces and Assemblies
Molecular Dinamics

Kétállapotú fehérje hajtogatódás
Two-state model of protein folding

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

$$-RT \ln K = \Delta G_{protein} =$$

$$= \Delta H_{protein} + \Delta H_{solvent} + \Delta H_{protein-solvent} - T\Delta S_{protein} - T\Delta S_{solvent}$$

Fehérje-fehérje kölcsönhatások

- Hidrofóbicitás
- Elérhető felület (500-1500Å^2)
- Alak komplementeritás
- Aminosav preferenciák (4-8 atomi kontaktus)
- Evolúciósan konzerváltabb szakasz

$$\Delta_r G^\circ = \Delta_r H^\circ - T\Delta_r S^\circ$$

Fehérje-fehérje kölcsönhatás modelljei

- Kulcs-zár (lock-and-key)
- Induced fit
- Konformációs kiválasztás (conformational selection)

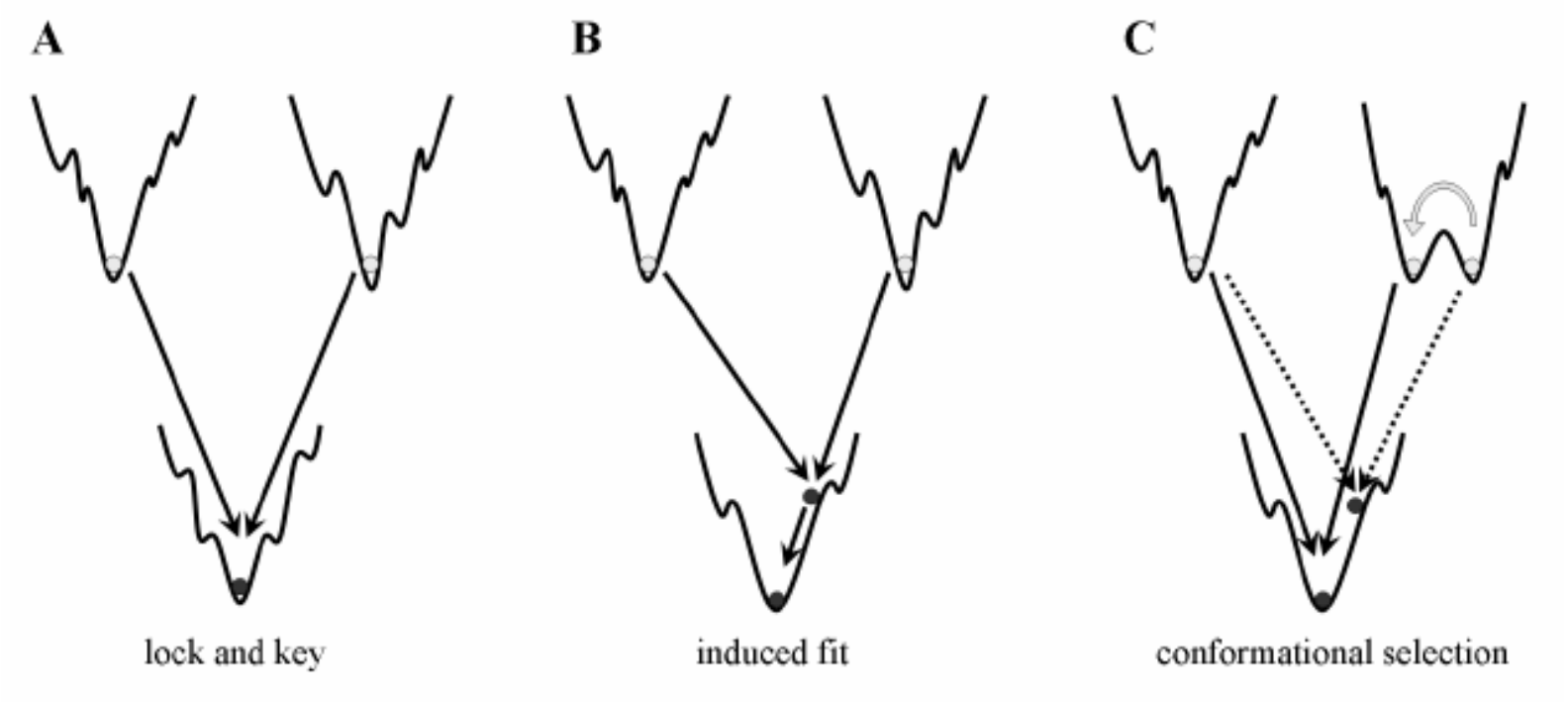


Figure 2: modes of molecular recognition in folded protein complexes

Mészáros B. PhD dolgozat

Rendezetlen fehérjék kölcsönhatásai

- 3000 $Å^2$
- Kis régiók: <100 a.a., 30 a.a.
- A kötőszakaszok 70%-a folyamatos szegmens
- Hidrofób-hidrofób kölcsönhatások
- Magasabb szekvenciális konzerváltság

Rendezetlen kötőrégiók jóslása

Rendezetlen régióban

$$S_k = \frac{1}{N} \sum_{k \neq j = b_{lower}}^{b_{upper}} s_j \qquad b_{lower} = \max(k\text{-}w_1;1) \text{ and } b_{upper} = \min(k+w_1;l)$$

Nem tudnak elegendő kedvező kölcsönhatásokat kialakítani foldinghoz

$$E_i^{\text{int},k} = \sum_{j=1}^{20} P_{ij} f_j^k(w_2)$$

Kötőpartner jelenlétében viszont igen

$$E_i^{glob} = \sum_{j=1}^{20} P_{ij} \overline{f}_{glob,j} \qquad E_i^{gain,k} = E_i^{\text{int},k} - E_i^{glob}$$

$$I_k = p_1 S_k + p_2 E_i^{\text{int},k} + p_3 E_i^{gain,k}$$

Az ANCHOR tréningje

http://anchor.enzim.hu

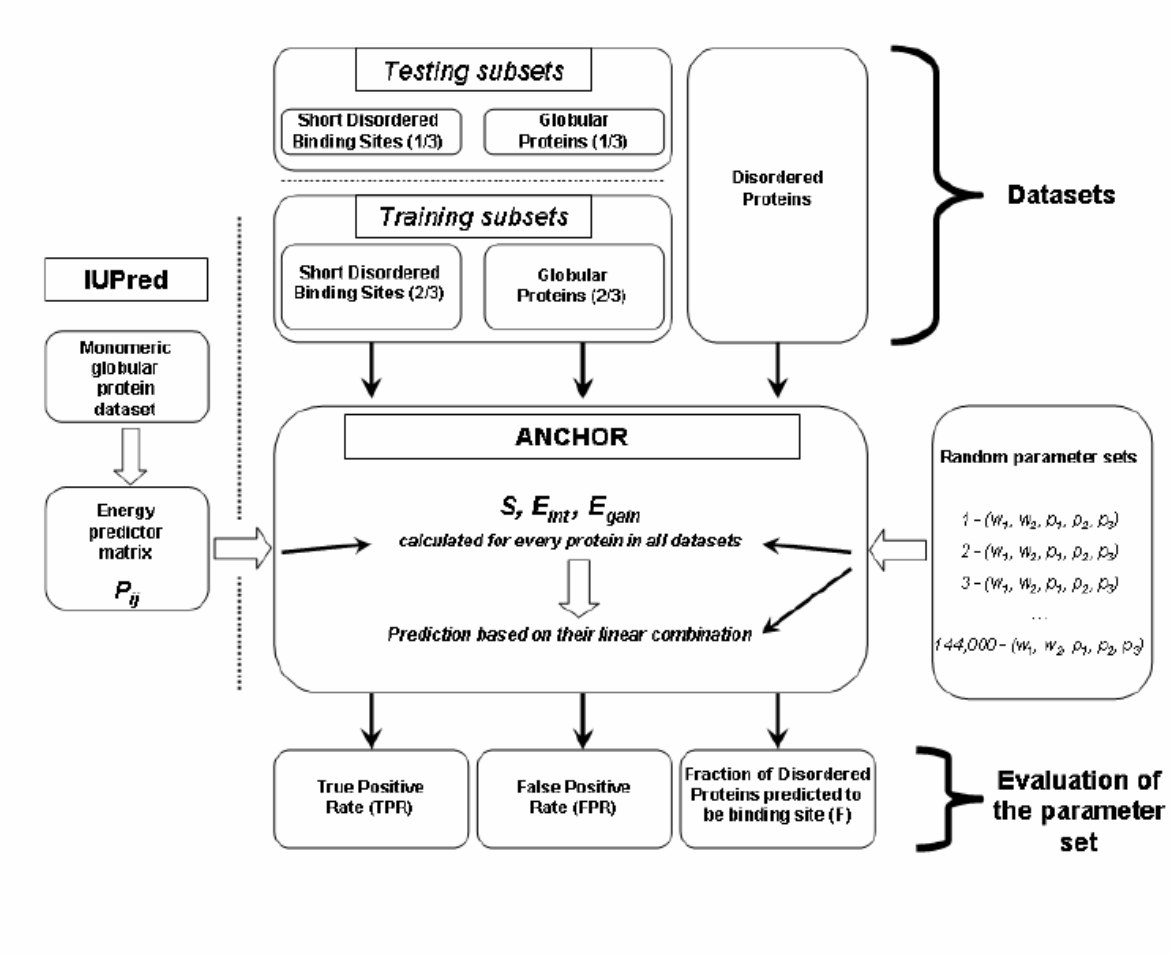


Figure 5: Outline of the training of ANCHOR

Mészáros B. PhD dolgozat

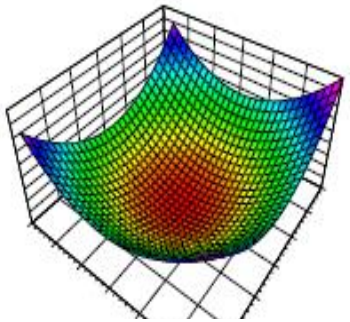
Mai témák

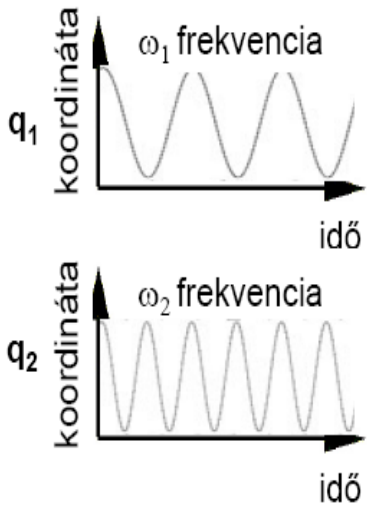
- Bevezetés – a fehérje dinamika és a szimulációk jelentősége
- Fehérjék jellemzése bioinformatikai eszközökkel
  - Másodlagos szerkezeti mintázatok jóslása
  - Rendezetlen fehérjék
  - Funkcionális régiók azonosítása
  - A harmadlagos és negyedleges szerkezet
- Informatikai eszközök – biológus szempontból
  - Adatbázisok
  - programok
  - programozási nyelvek
  - operációs rendszerek
- Fehérjék dinamikájának modellezése
- Fehérjék feltekeredésének szimulációja

Fehérje dinamika vizsgálata

Normál-módus elemzés

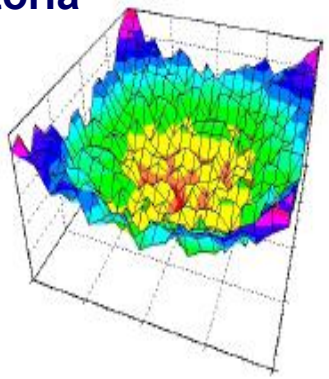
- harmonikus potenciál
- analitikus mozgásegyenletek
- normál modusok

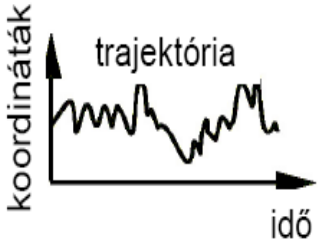




Molekuláris dinamika (MD)

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





A „force field” - I.

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

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

Lazaridis (2003)

Lazaridis (1999)

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

A „force field” – II.

Baker (2007)

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

$$E_{\text{mbenv}} = \sum_{i}^{\text{natom}} \Delta G_i^{\text{ref}}(z') \qquad \Delta G_i^{\text{ref}}(z') = (1 - f(z')) * \left(\Delta G_i^{\text{ref,chex}} - \Delta G_i^{\text{ref,water}}\right)$$

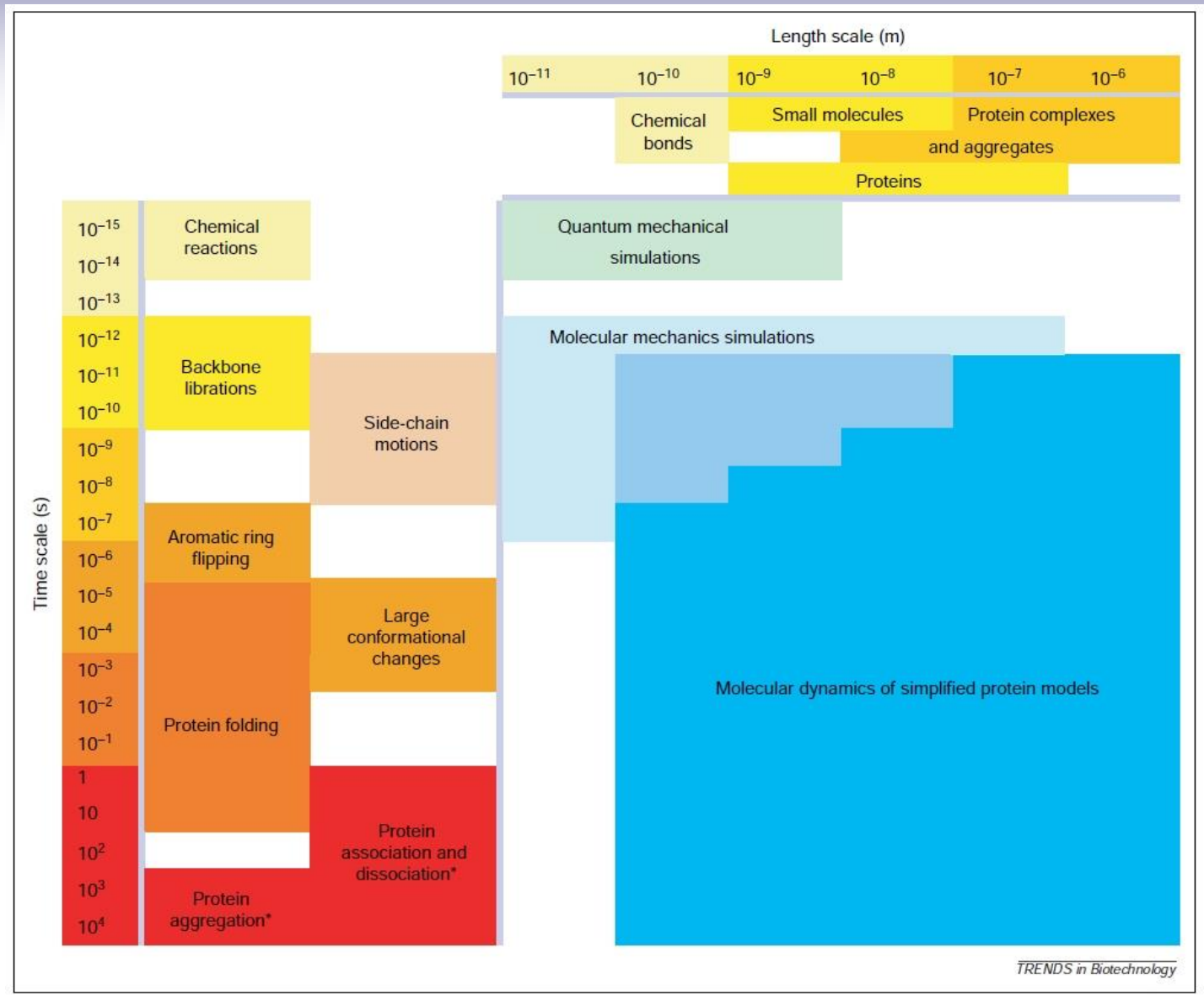
TABLE II. Atomic Contribution to the Solvation Free Energy in Water and Cyclohexane

	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

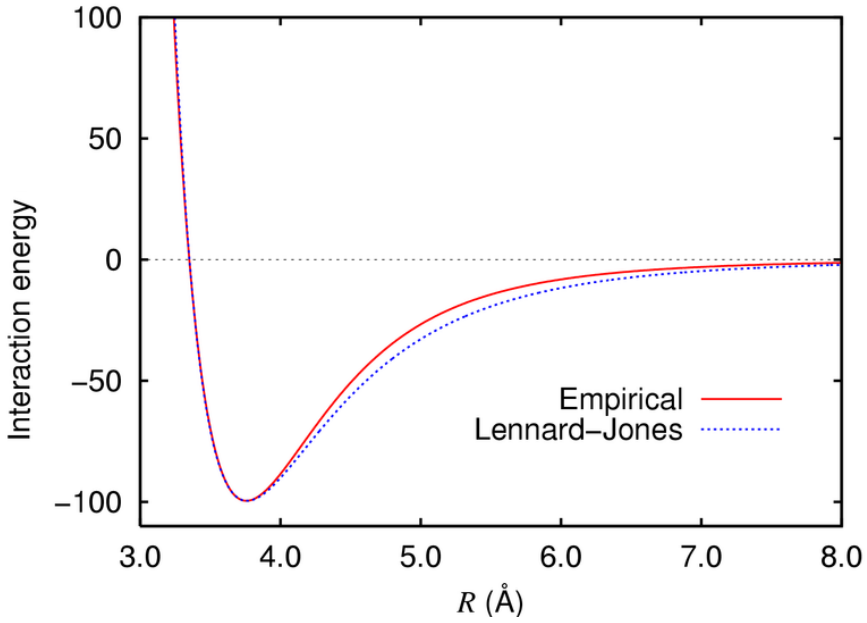
- idő (CPU, valós)
- potenciál kiszámolása a szűk keresztmetszet
- numerikus integrálás hibája
- fs-os integrációs lépések
- oldószer (explicit/implicit)
- „boundary condition”

„Események” időskálája

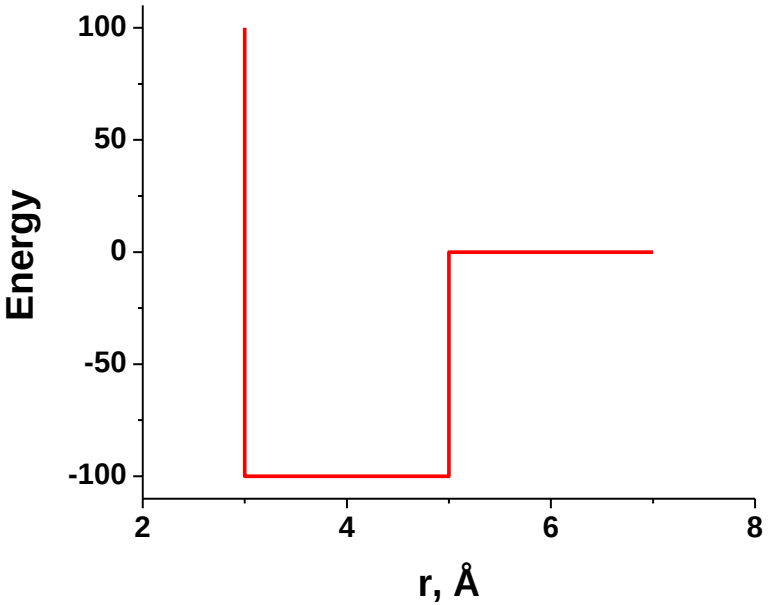


F. Ding and N.V. Dokholyan, TRENDS in Biotechnology, **23:**450 (2005)

Diszkrét Molekuláris Dinamika (DMD)



wikipedia



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

Egyszerűsített (Coarse Grain) modellek

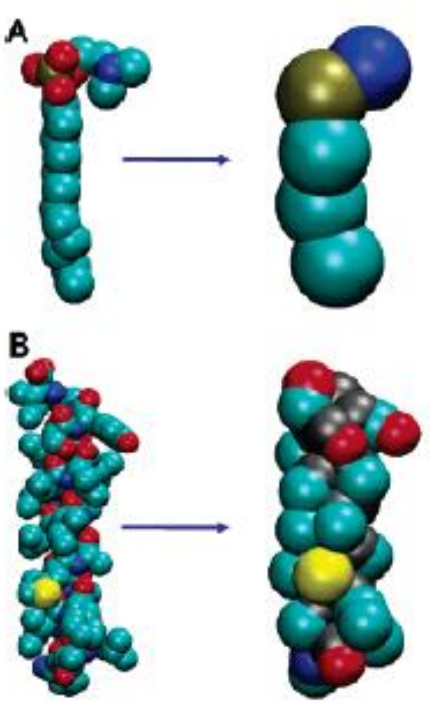
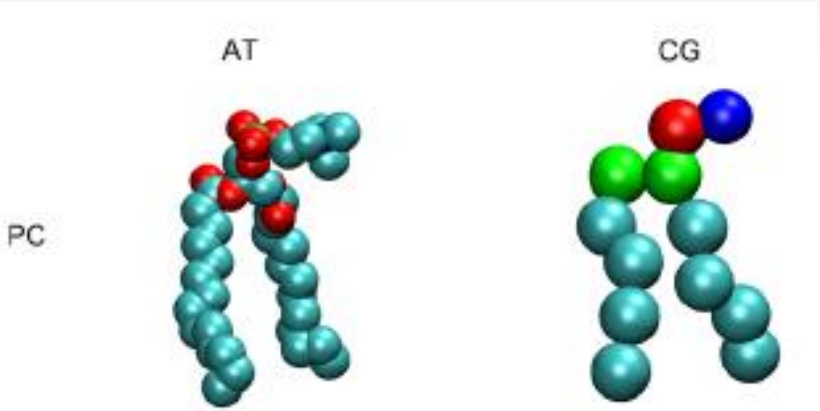


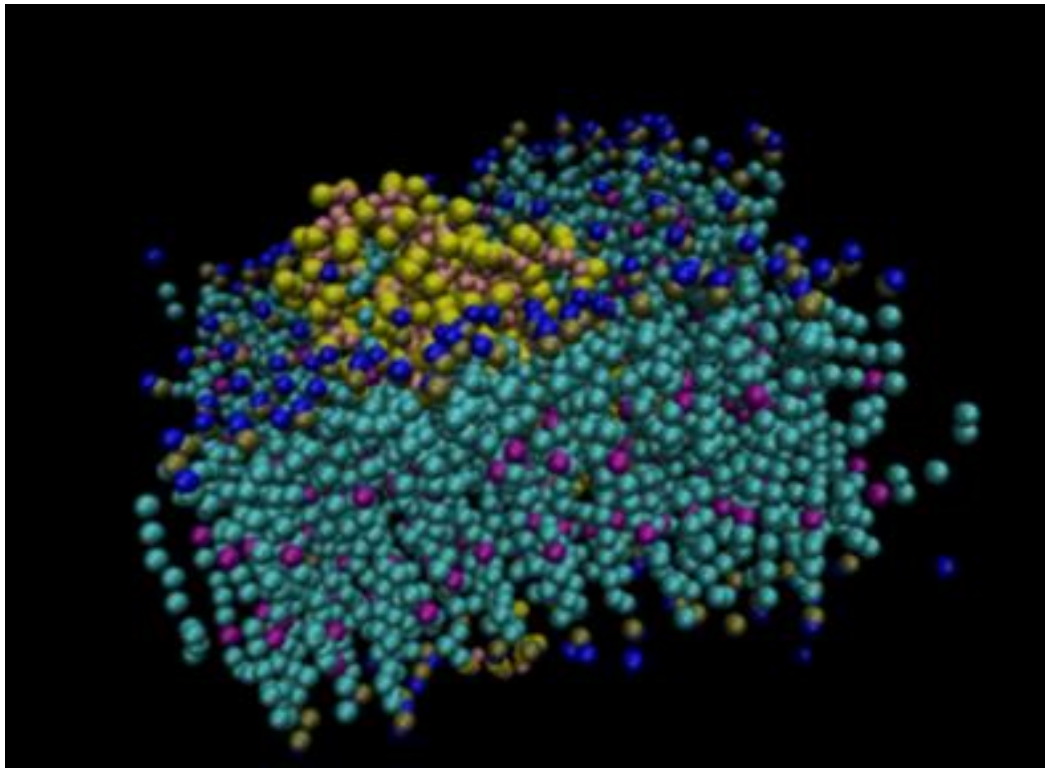
Figure 1. Atomistic (left-hand) and coarse-grained (right-hand) models compared for (A) a DPC molecule and (B) a GpA helix. Colors for atoms:



Bond, Sansom: **MARTINI**

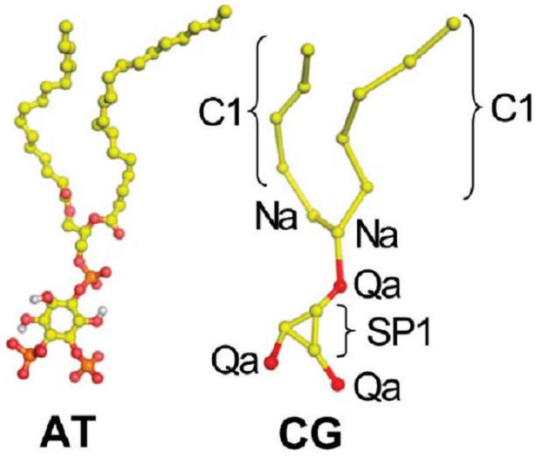
Fehérjére pl. 2 bead vagy 4+ bead modellek

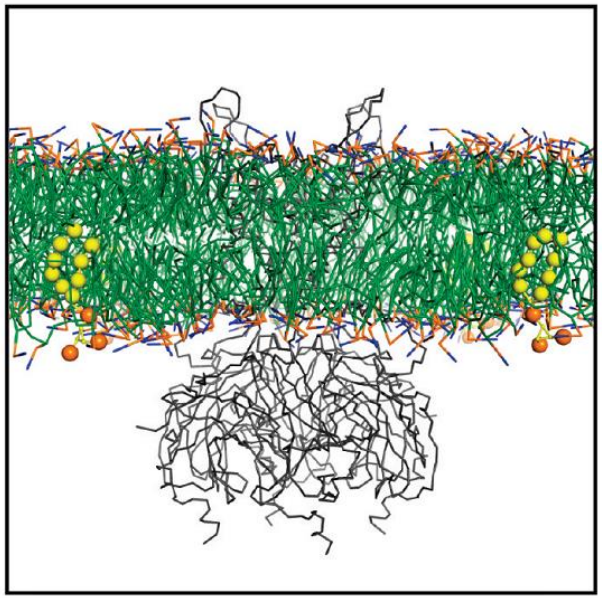
Kettősréteg felépülése a fehérje köré

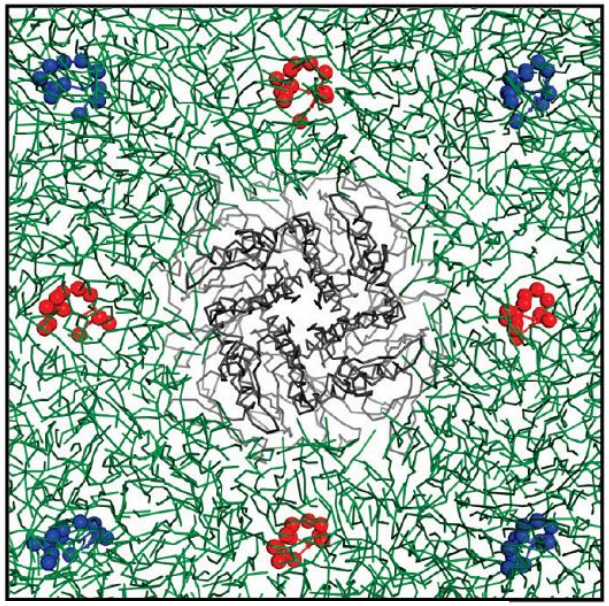


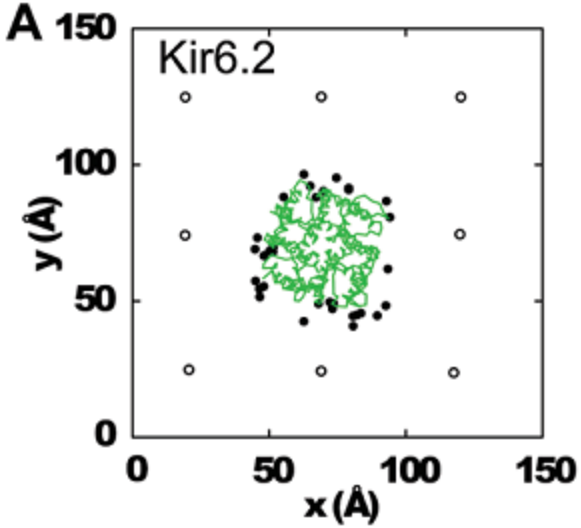
PIP2 kötődése Kir kálium csatornához

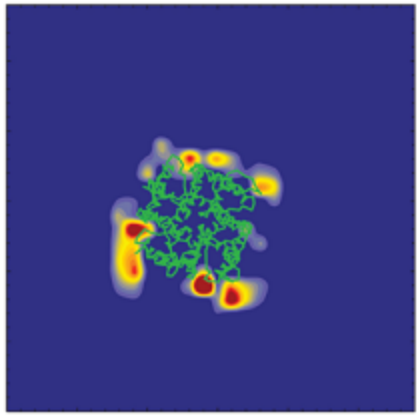
Biochemistry, Vol. 48, No. 46, 2009 1











Glikophorin A dimerizációja

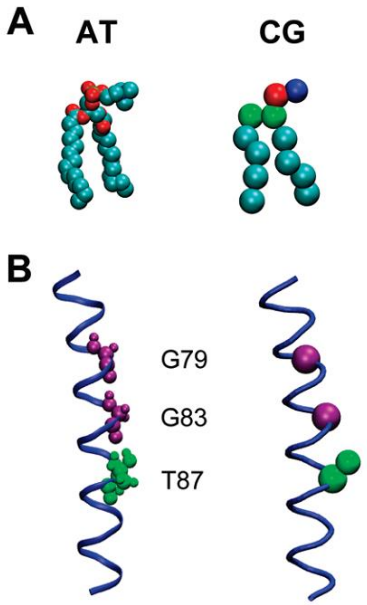


FIGURE 1. Atomistic (AT) and coarse-grained (CG) representations

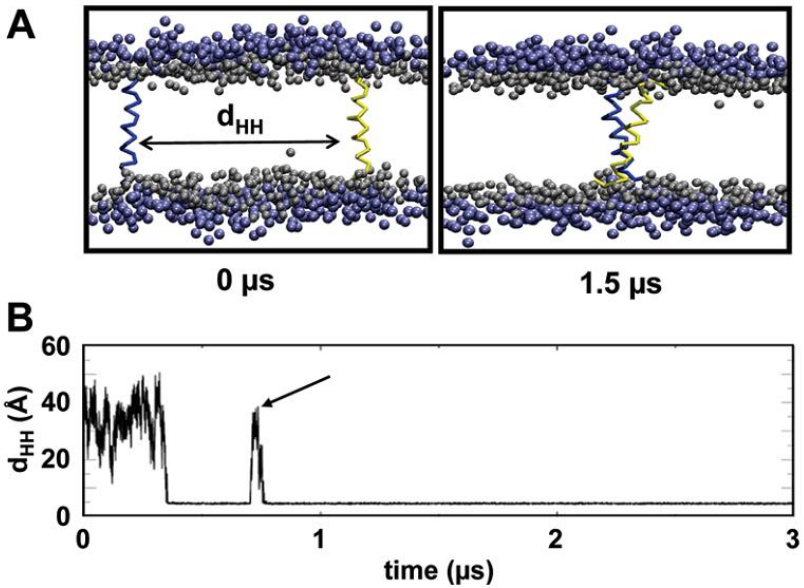
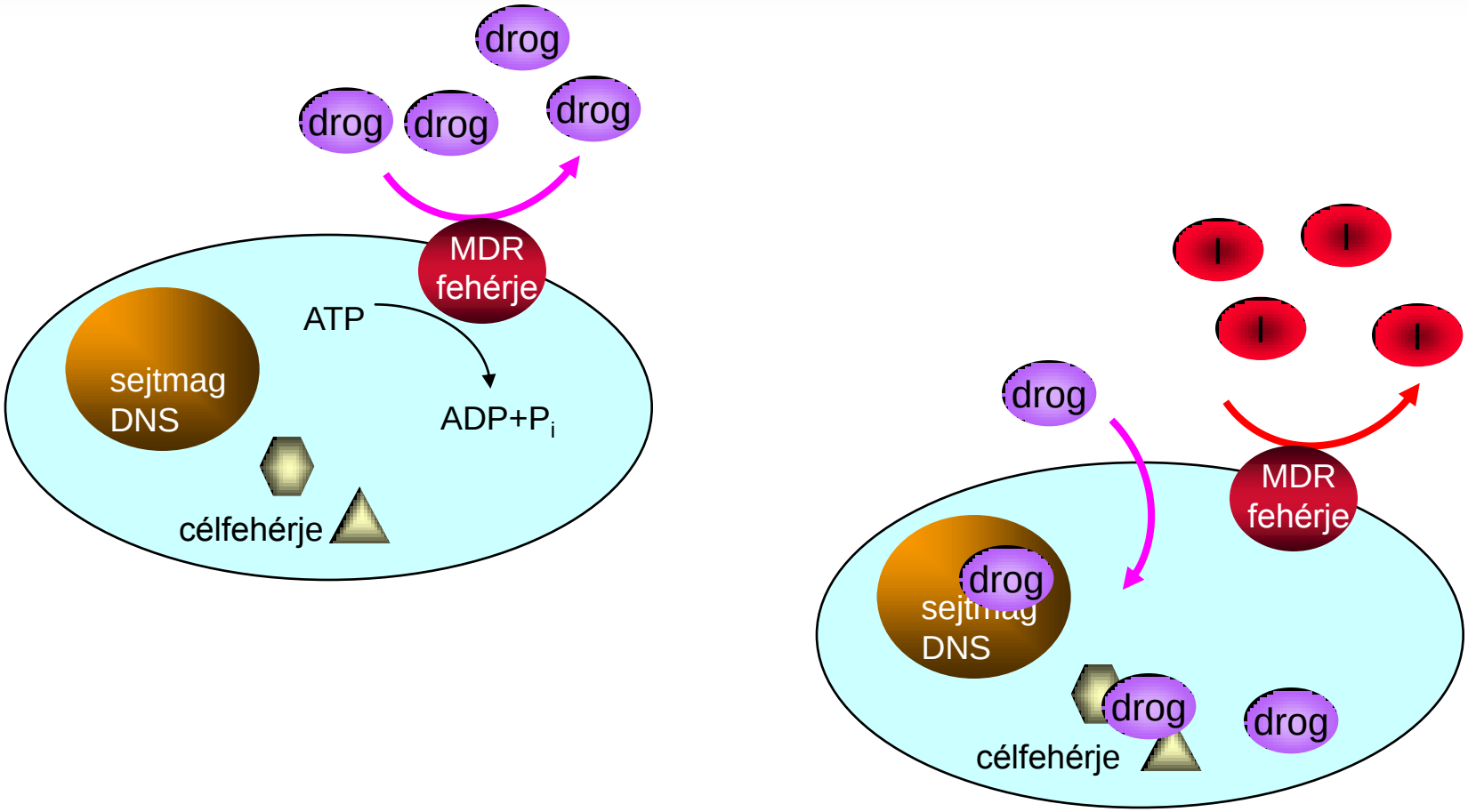
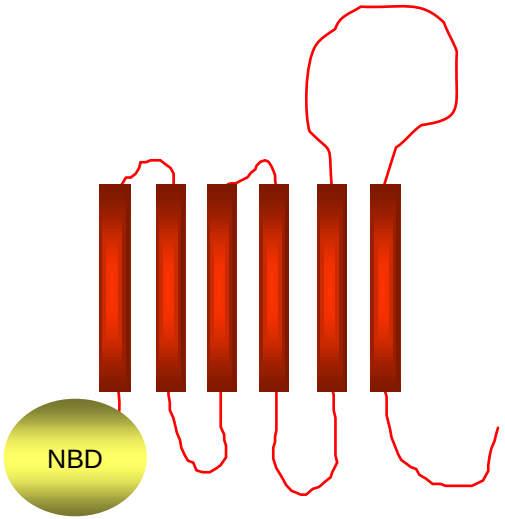


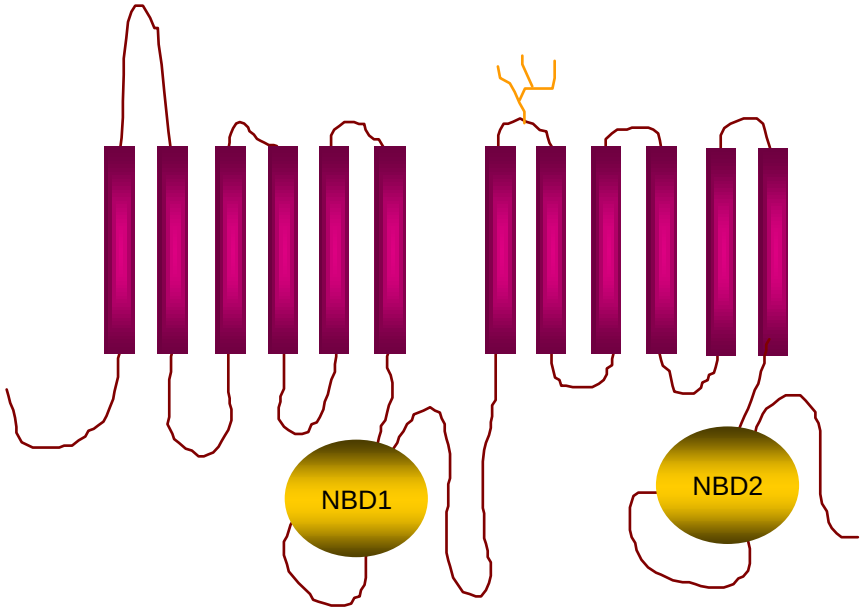
FIGURE 3. Coarse-grained GpA TM helix dimerization simulation.

A multidrog-rezisztencia és felfüggesztése



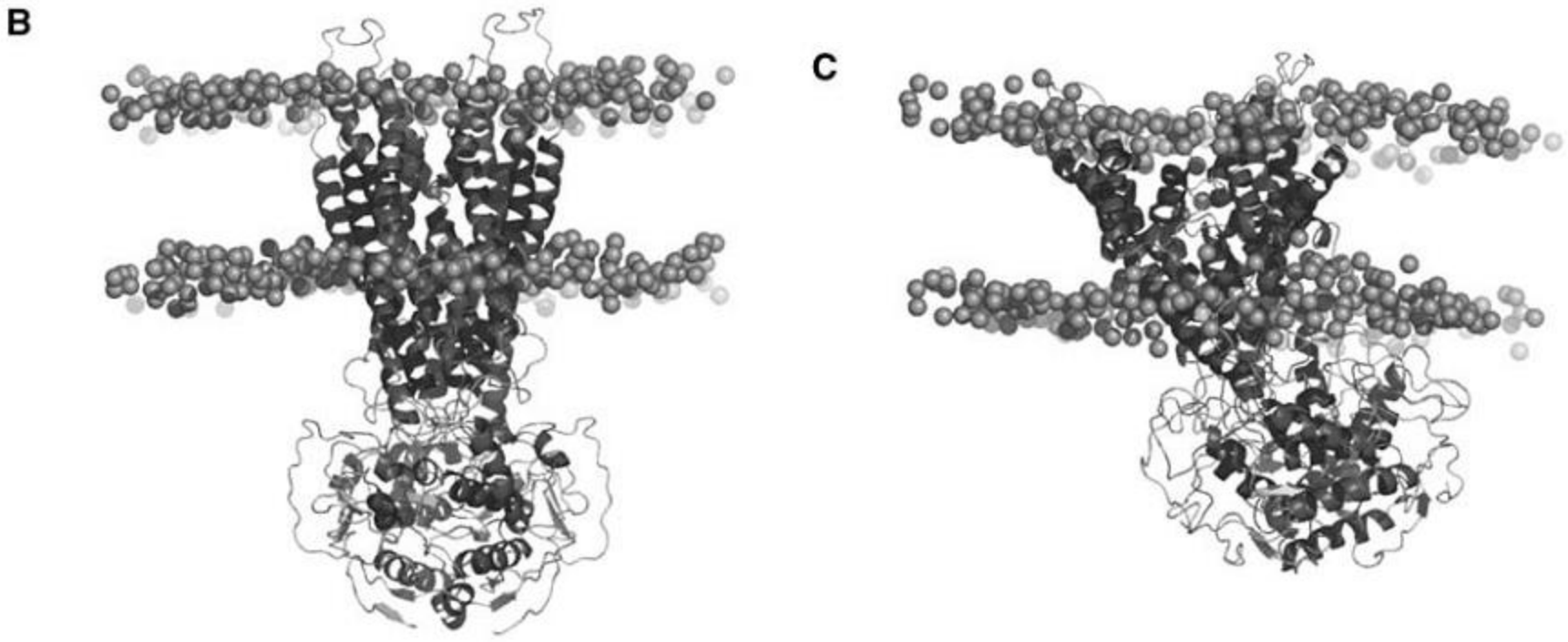
ATP Binding Cassette (ABC) fehérjék





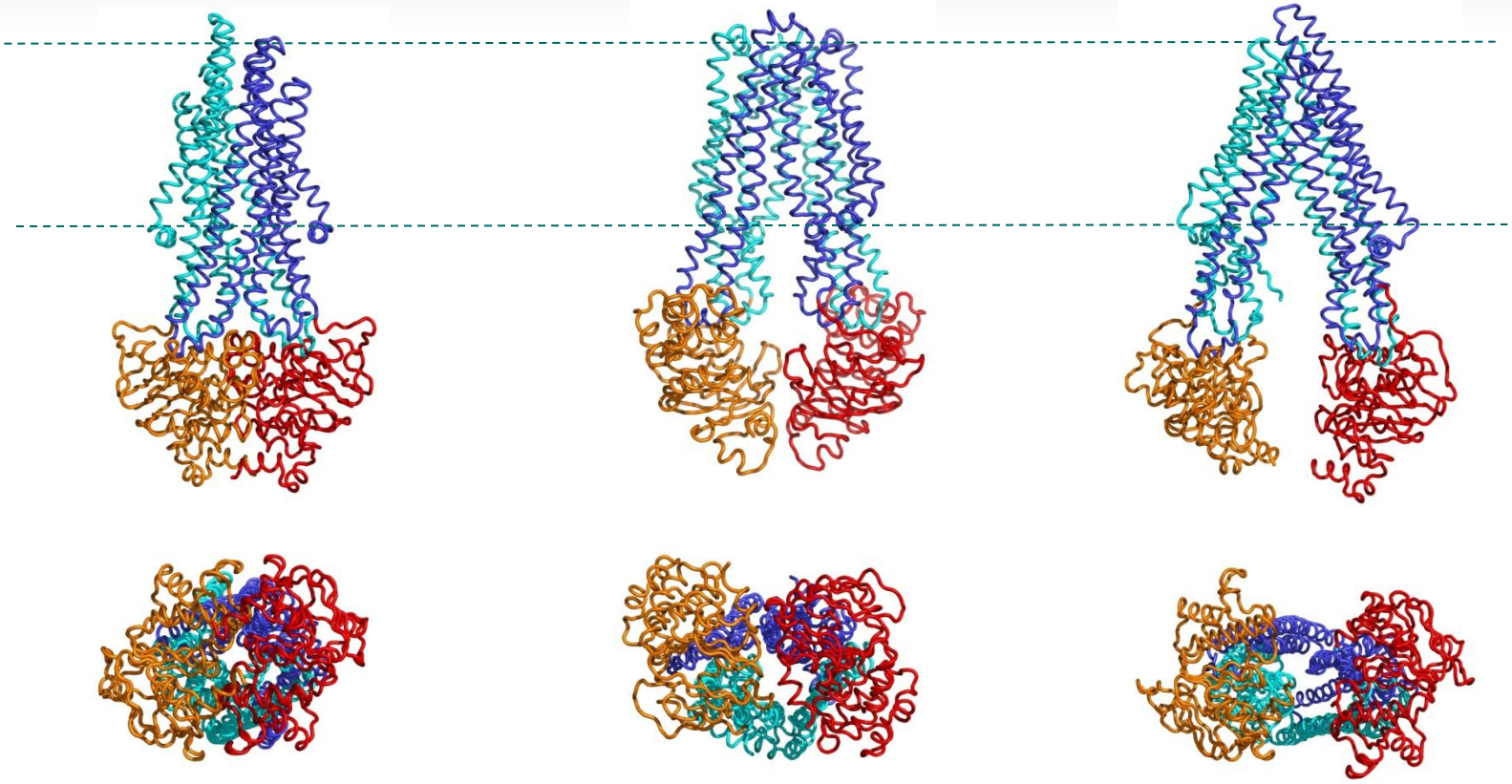
Fehérjék konformációinak stabilitása

Eur Biophys J (2008) 37:403–409



0 ns **20 ns**

ABC fehérjék konformációi

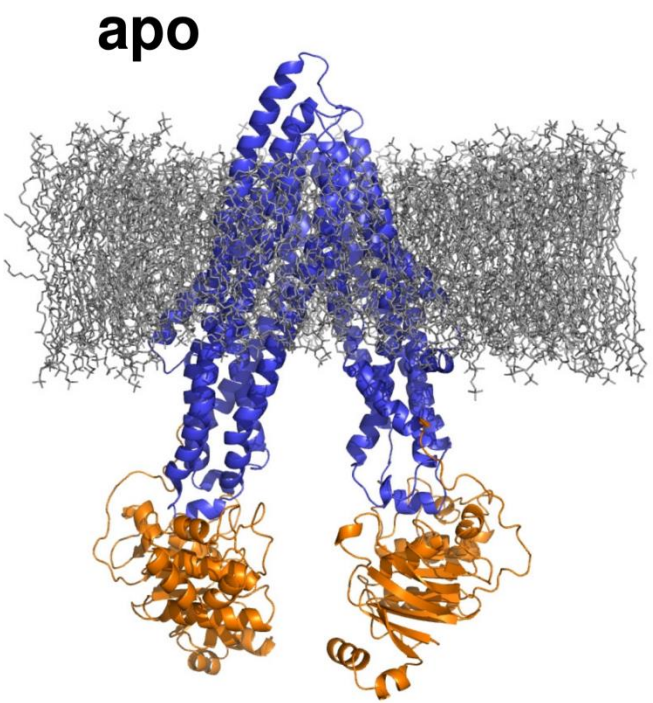


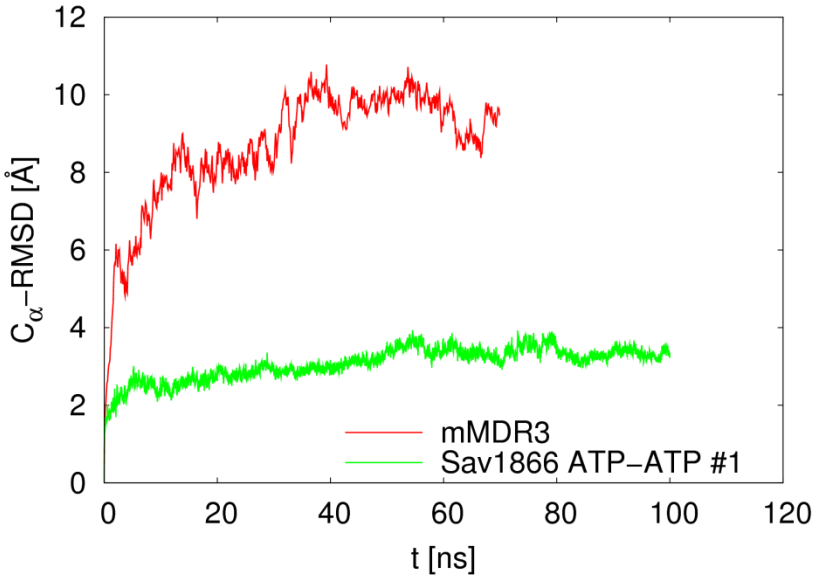
“alul-zárt” holo (+ATP)

“alul-zárt” apo (-ATP)

“alul-nyitott” apo (-ATP)

Az alul nyitott apo szerkezet nem stabil

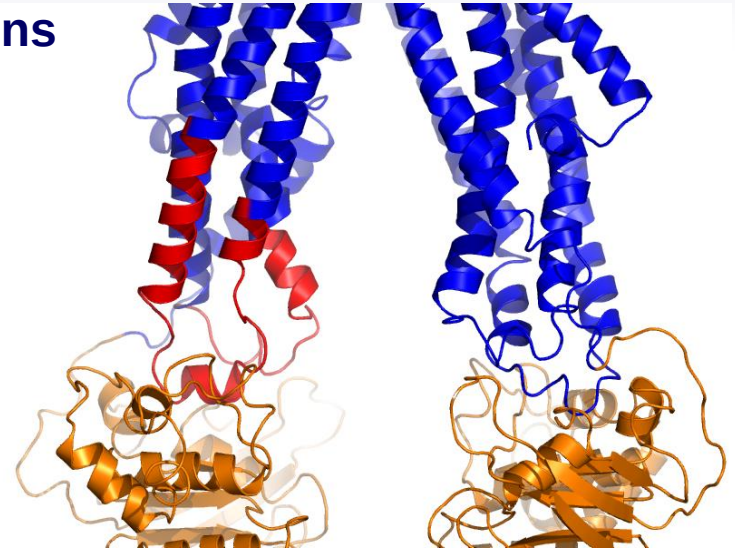




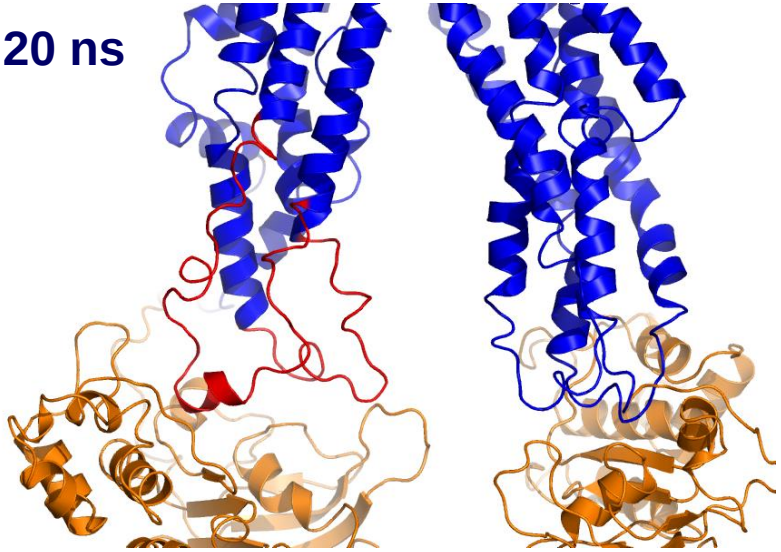
Gyimesi *et al.* BBA 2012

Az alul nyitott apo szerkezet nem stabil

t = 0 ns

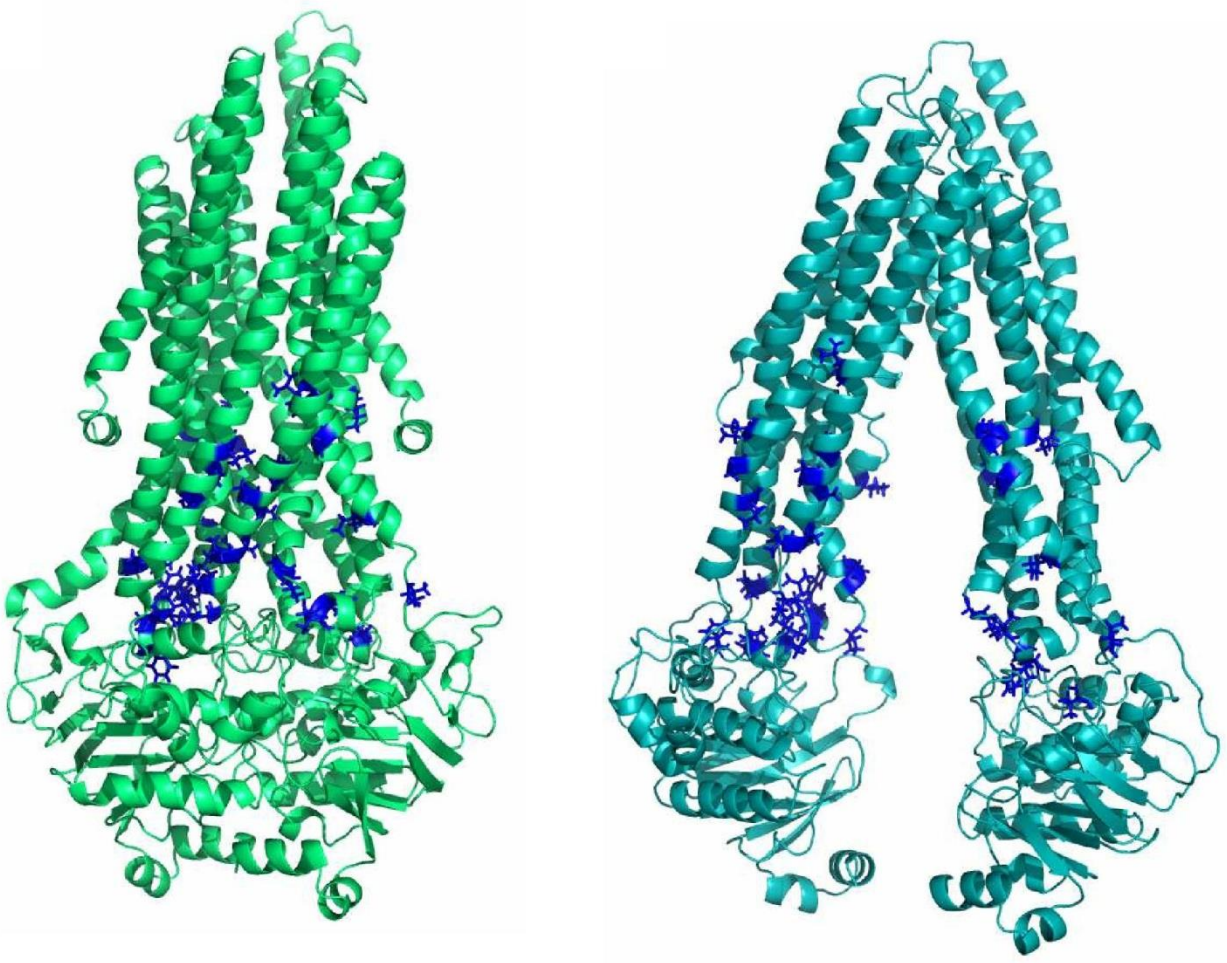


t = 20 ns



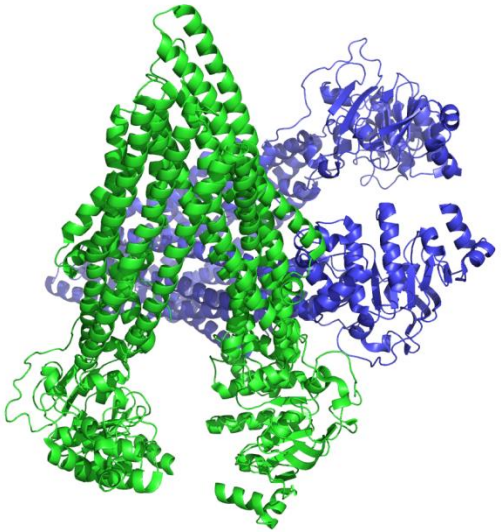
rendszer	megtartott hélixtartalom
Sav1866 ATP/ATP #1	90.04%
hMDR1 holo	91.84%
hMDR1 apo	64.30%
mMDR3	63.13%

Hidrofób aminosavak kerülnek felszínre

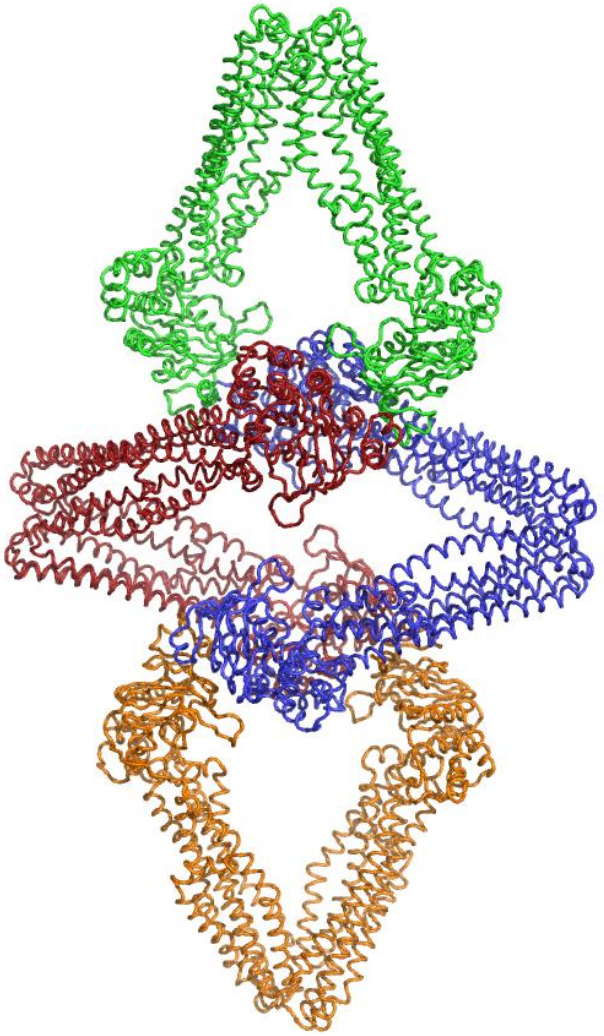


Alul nyitott apo szerkezet elemi cellája

mMDR3, PDBID:3G5U



MsbA, PDBID:3B5W

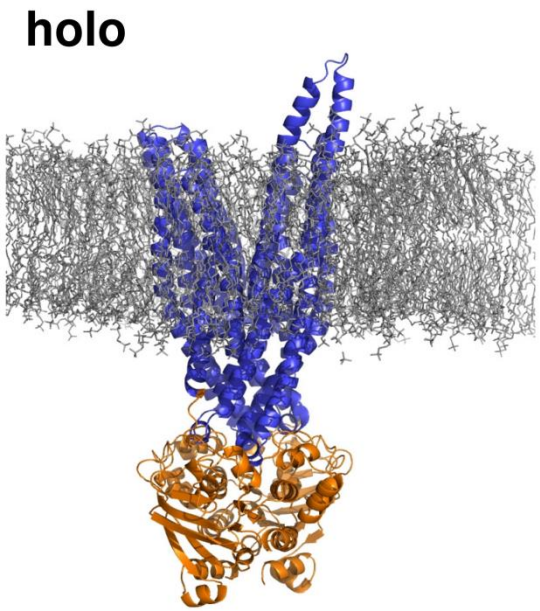


Események modellezése

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

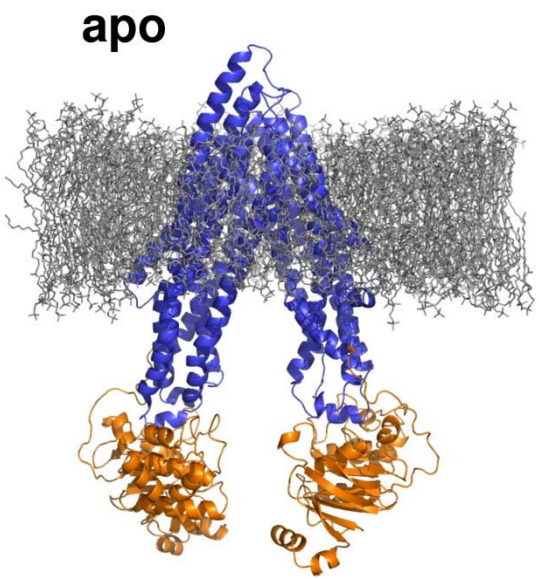
  Pl. targeted MD

Zárt-nyitott átmenet jellemzése molekuláris dinamikával



hMDR1 homologia modell
(3x100 ns)

M. Wiese modellje



hMDR1 homologia modell

molekularis dinamika trajektóriák

⇩

Esszenciális dinamika

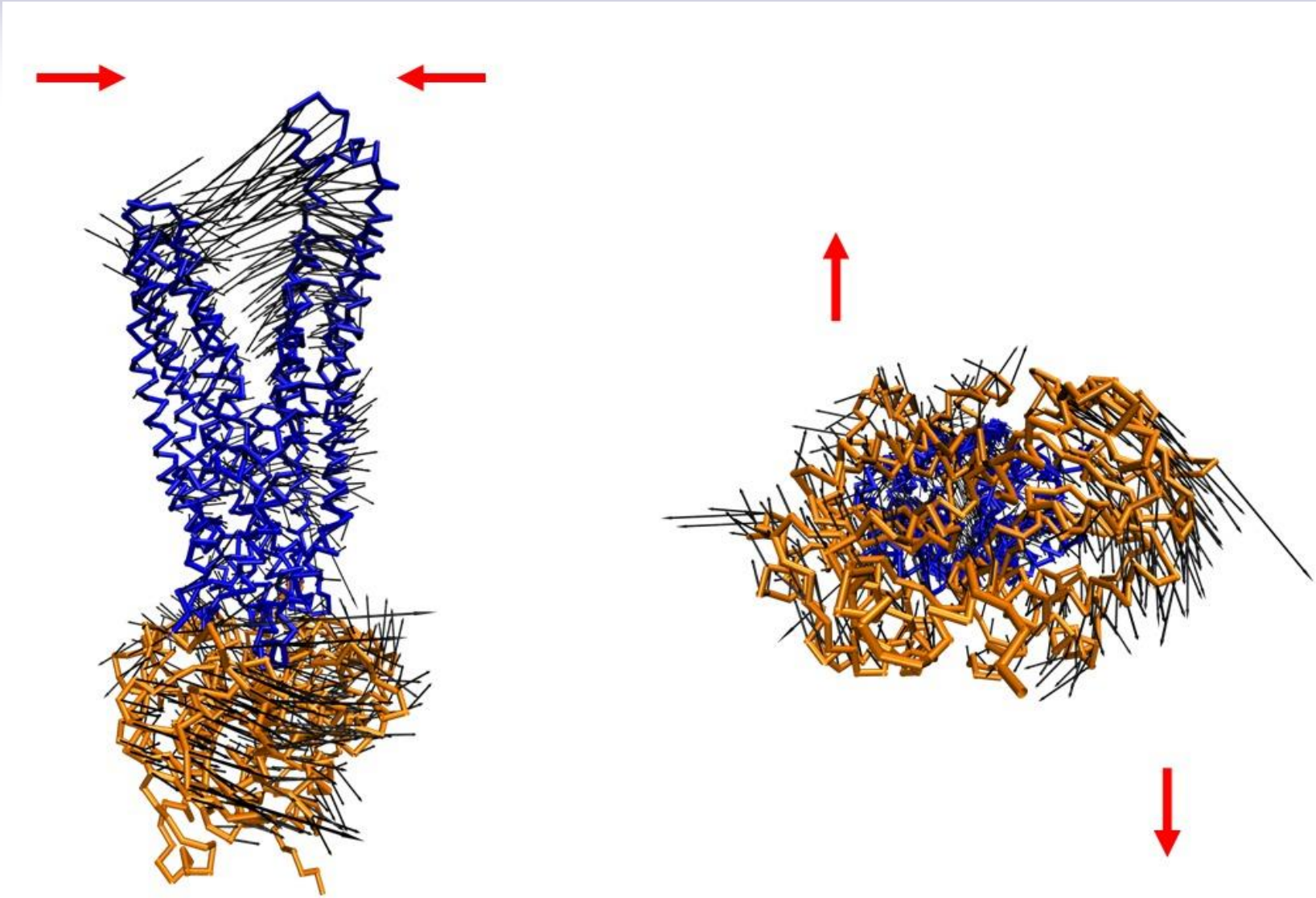
⇩

módus kiválasztás

⇩

kollektív mozgások

Zárt-nyitott átmenet



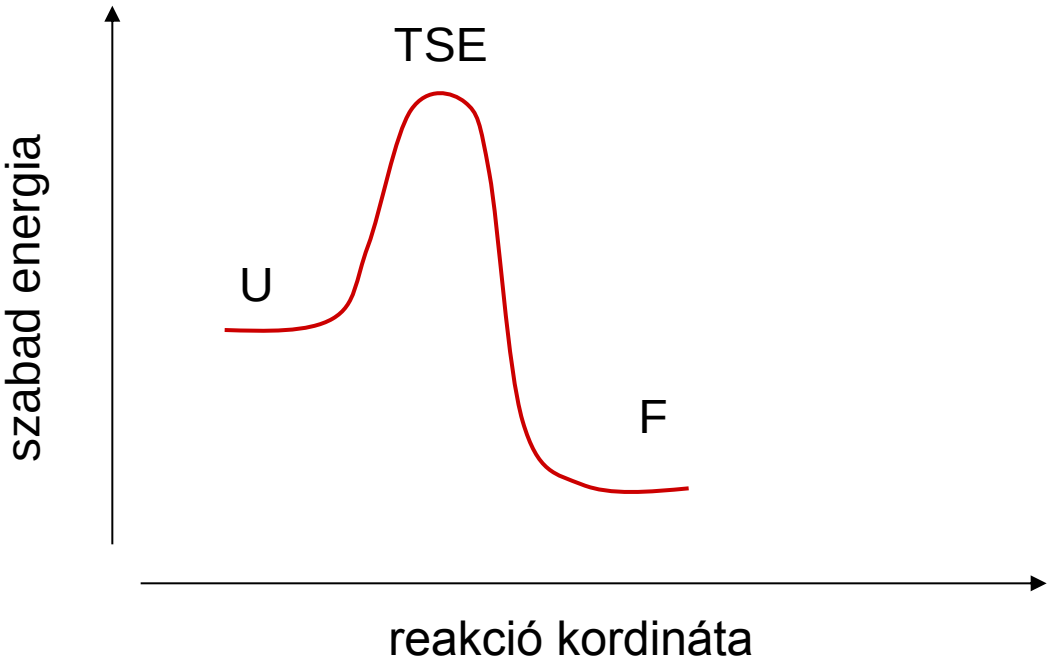
Mai témák

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

Fehérje feltekeredés

Levinthal paradoxon

nukleáció



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 szolvatációja
- entrópia

Fehérje stabilitás II.

A fehérje stabilitás nem ér el maximális értéket.

Erre utalnak:

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

Ennek okai lehetnek:

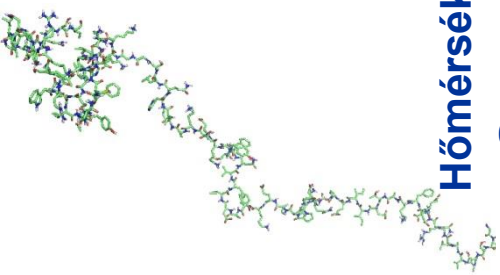
- **az evolúció nem igényel stabilabb fehérjét mint a funkció önmaga;**
- **a fehérjéknek le is kell bomlaniuk;**
- **a funkcióhoz flexibilitás szükséges.**

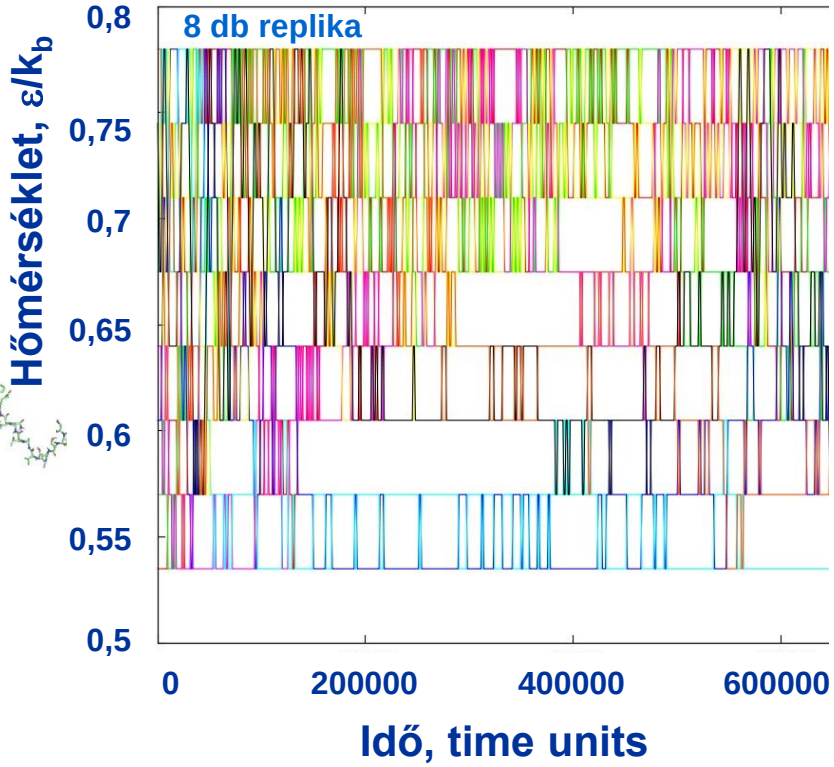
Folding szimulációk során mi az abszolút energiaminimumot (maximum stabilitást) keressük.

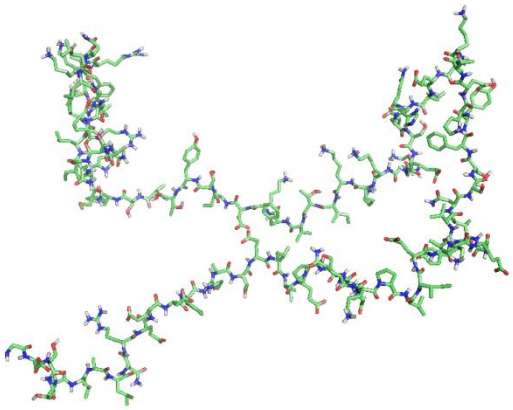
Fehérje feltekeredés szimulációja – I.

All atom force-field:

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

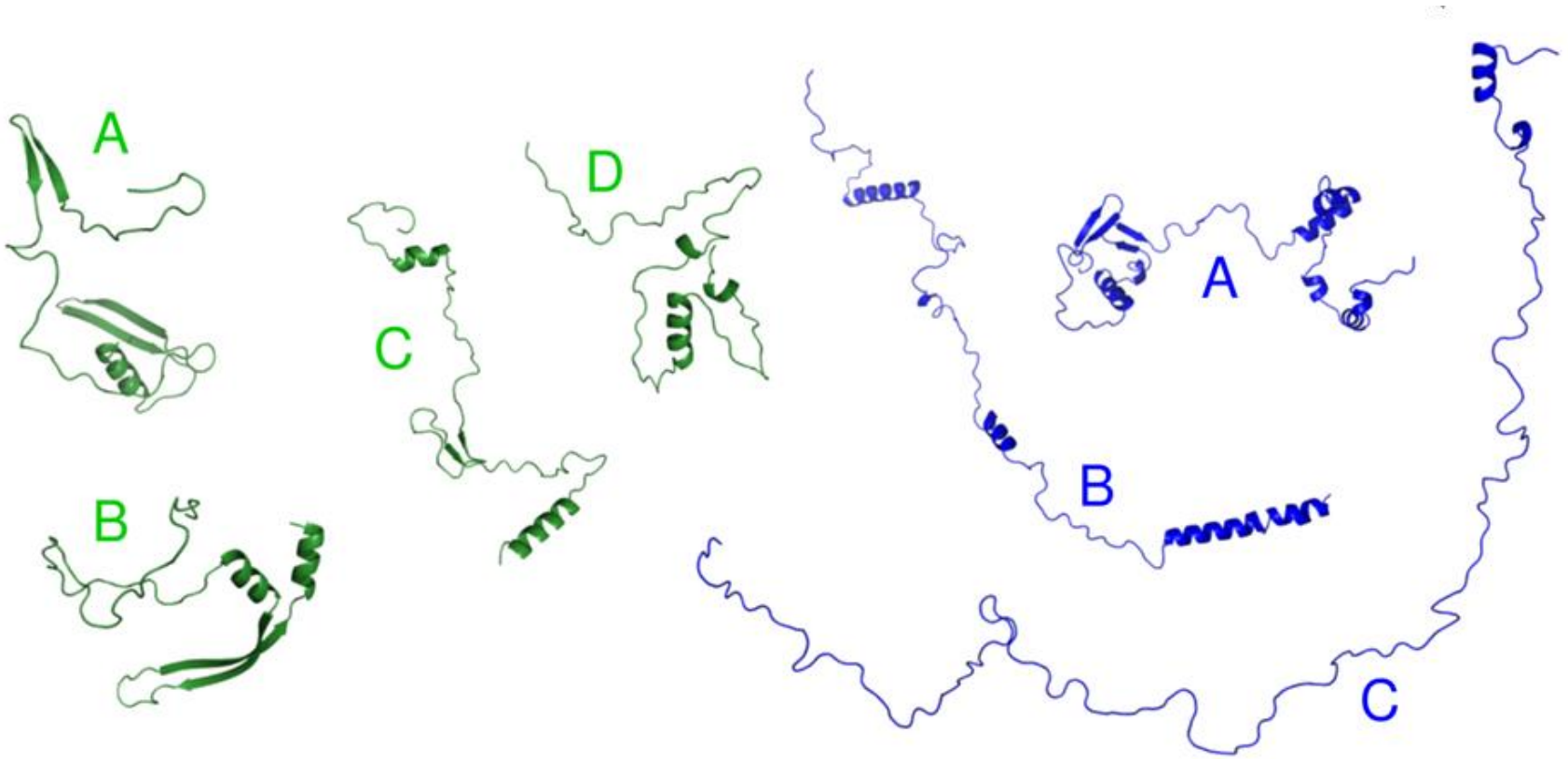




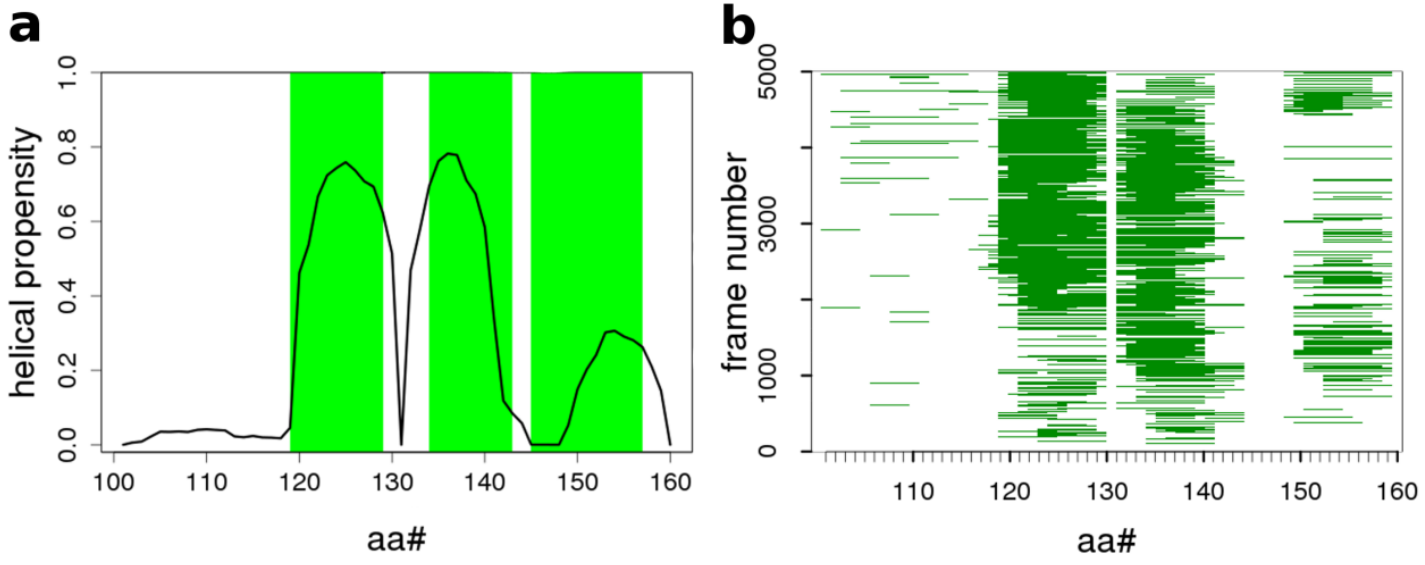


Rendezetlen fehérjék rendezettsége

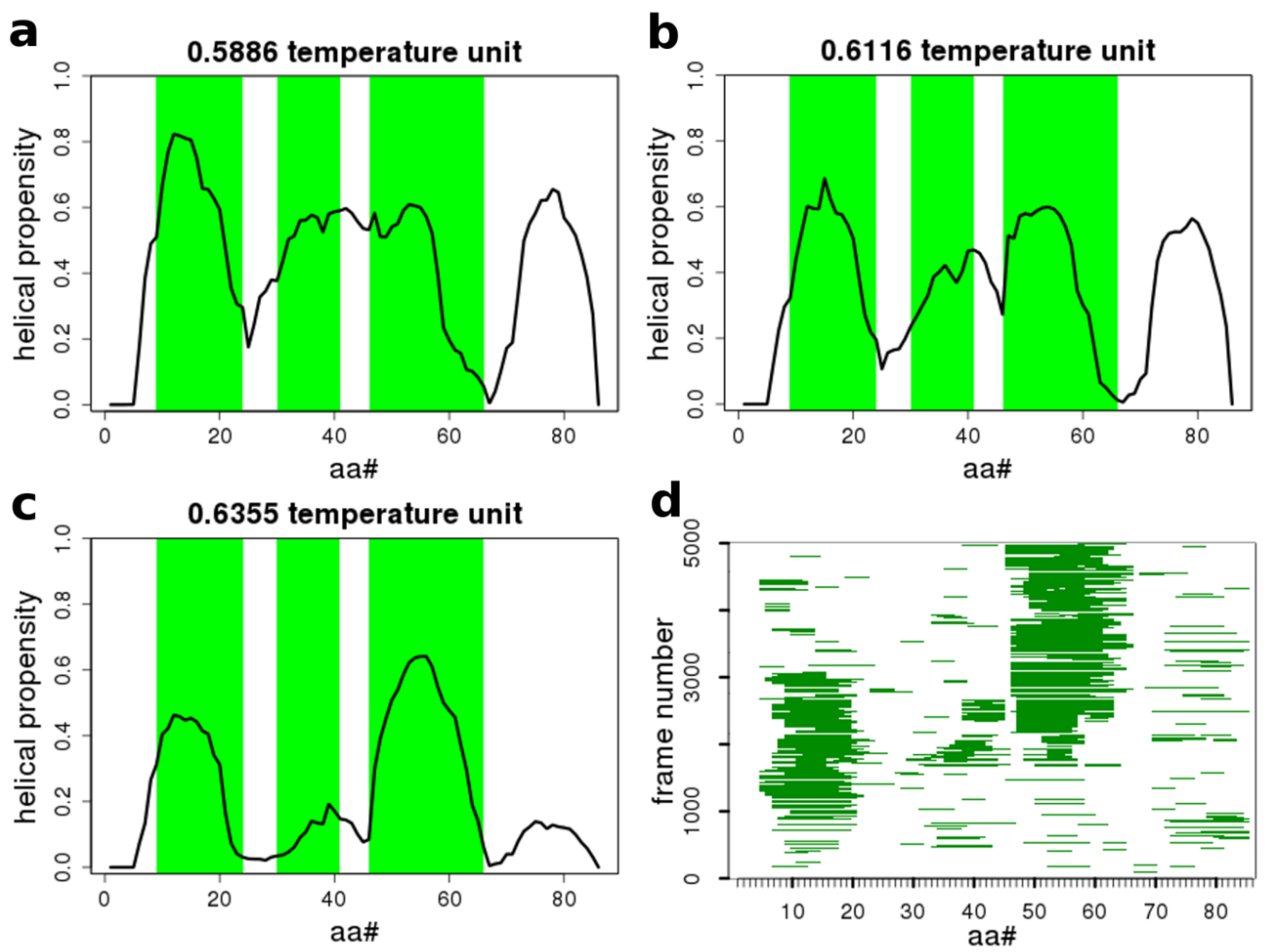
PreSMo: Prestructured Motif



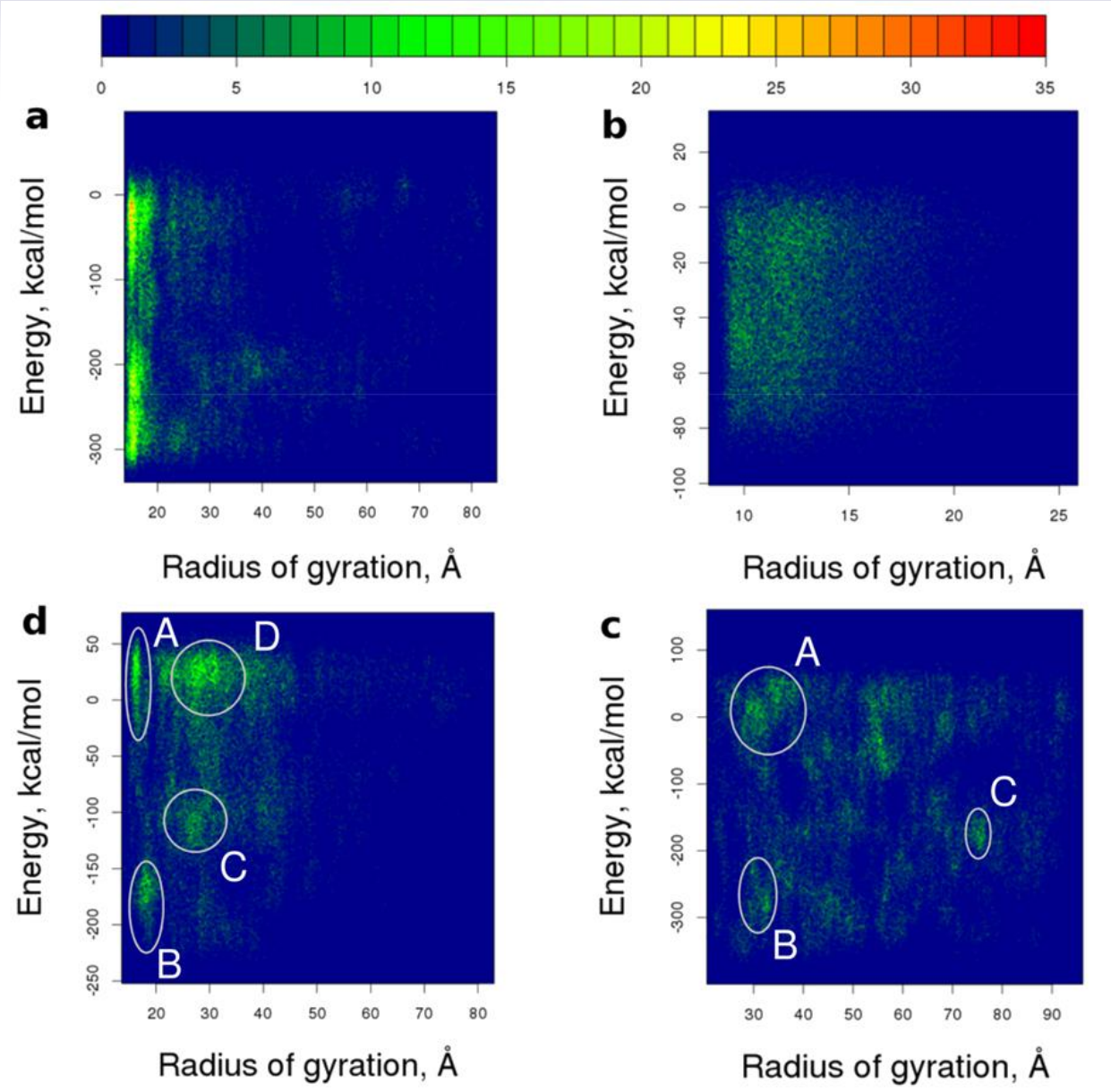
PreSMo jóslás I.



PreSMo jóslás II.



Potenciál felület
DoS (Density of States)



Összefoglalás

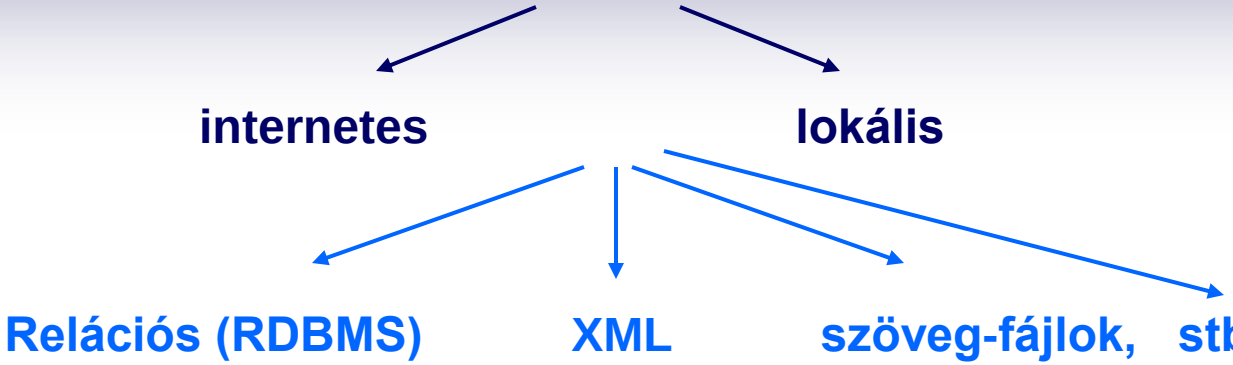
- Bevezetés – a fehérje dinamika és a szimulációk jelentősége
- Fehérjék jellemzése bioinformatikai eszközökkel
- Informatikai eszközök – biológus szempontból
- Fehérjék dinamikájának modellezése
- Fehérjék feltekeredésének szimulációja

Mai témák

- **Informatikai eszközök – biológus szempontból**

  adatbázisok
  programok
  programozási nyelvek

Adatbázisok I.



Internetes adatbázisok 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

Adatbázisok II.

Lokális adatbázisok:

- **RDMBS**
- **fájlok**

Előnyei:

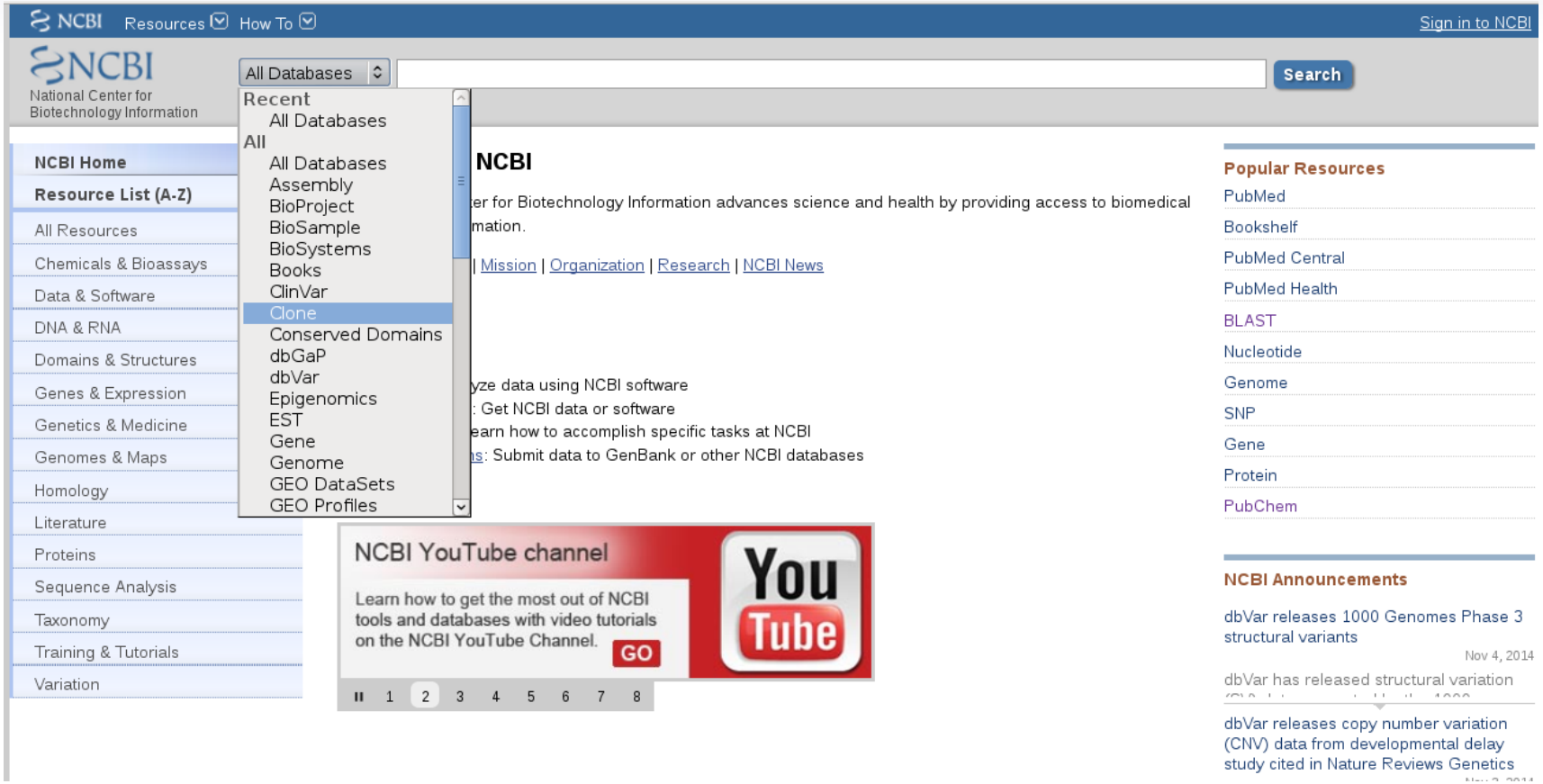
- **lokális**
- **gyors elérés**
- **adott verzió (kézirat!)**
- **„akármilyen” szoftverrel használható**

Hátrányai:

- **lokális**
- **erőforrás-igény**
- **hozzáértés-igény**

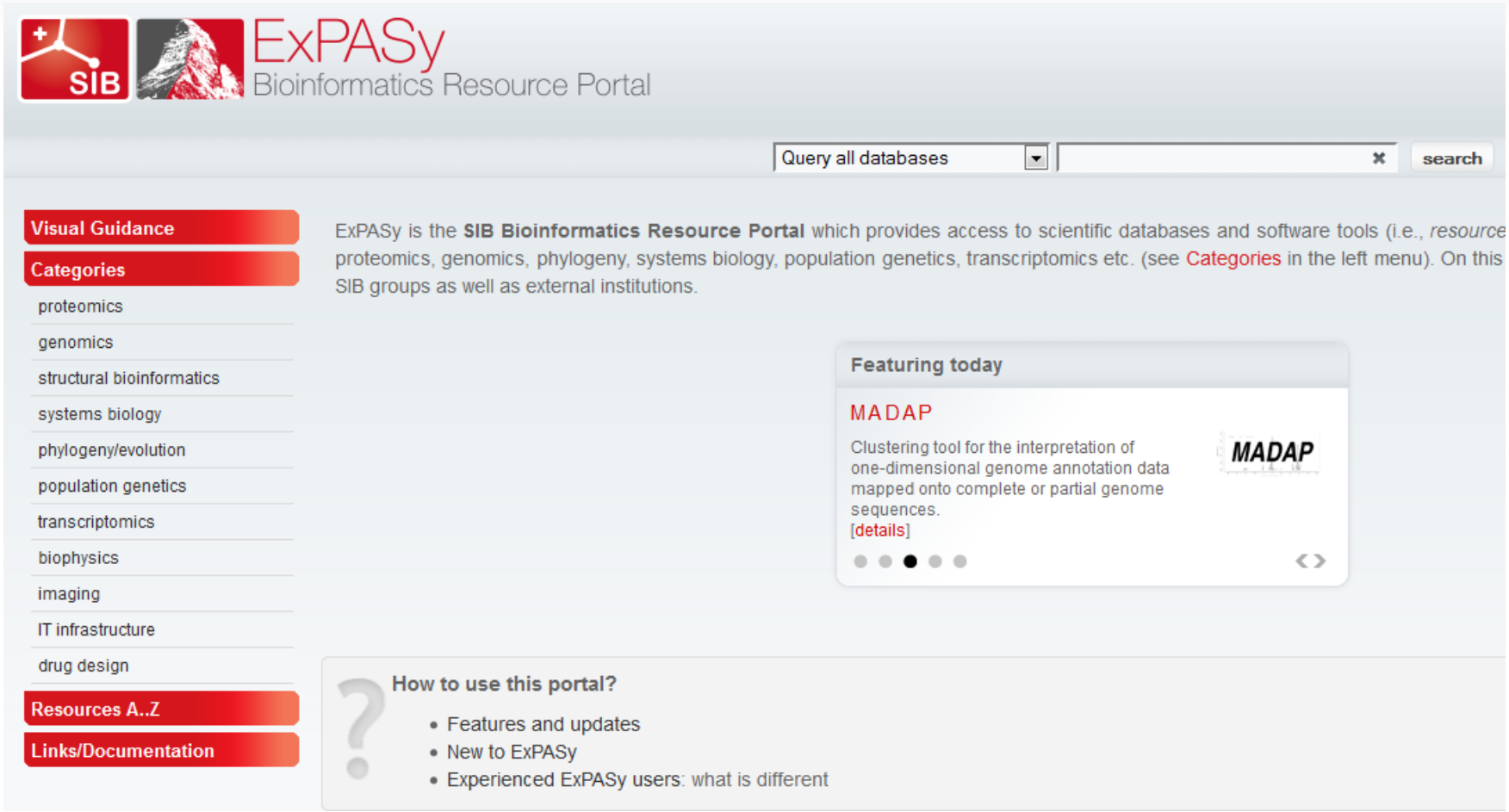
Adatbázisok III.

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



Adatbázisok III/b.

http://expasy.org/



Szekvencia fájl formátumok

FASTA

```
>CFTR_HUMAN | P13569 | Cystic fibrosis transmembrane conductance regulator…
MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLSEKLEREWDRE
LASKKNPKLINALRRCFFWRFMFYGIFLYLGEVTKAVQPLLLGRIIASYDPDNKEERSIA
IYLGIGLCLLFIVRTLLLHPAIFGLHHIGMQMRIAMFSLIYKKTLKLSSRVLDKISIGQL
VSLLSNNLNKFDEGLALAHFVWIAPLQVALLMGLIWELLQASAFCGLGFLIVLALFQAGL
GRMMMKYRDQRAGKISERLVITSEMIENIQSVKAYCWEEAMEKMIENLRQTELKLTRKAA
... DTRL
```

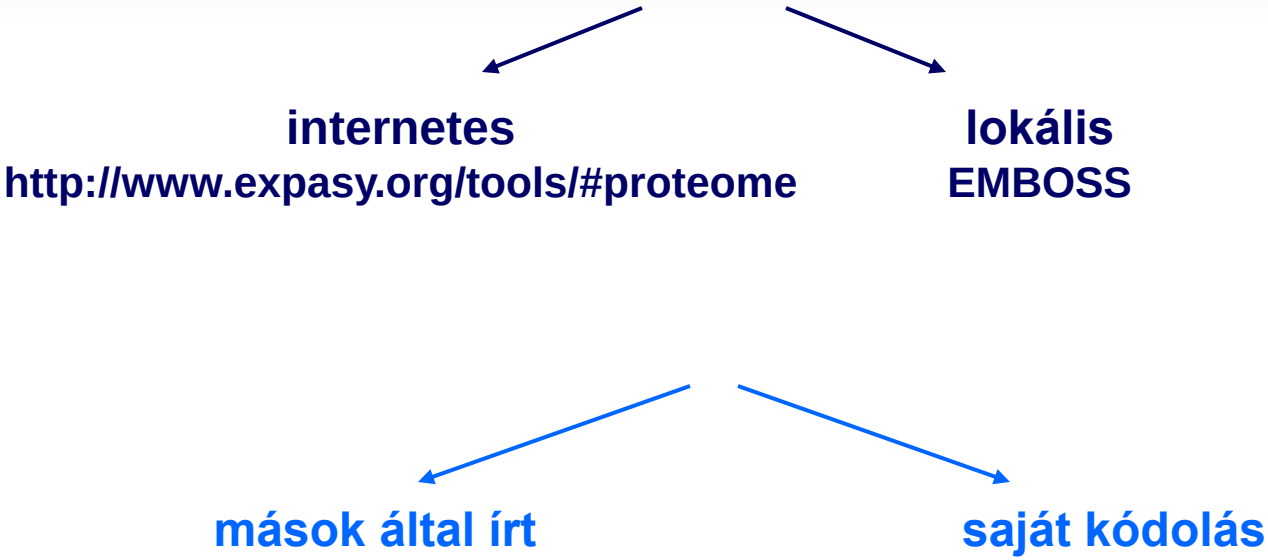
PIR

```
>P1;CRAB_ANAPL
ALPHA CRYSTALLIN B CHAIN (ALPHA(B)-CRYSTALLIN).
  MDITIHNPLI RRPLFSWLAP SRIFDQIFGE HLQESELLPA SPSLSPFLMR
  SPIFRMPSWL ETGLSEMRLE KDKFSVNLDV KHFSPEELKV KVLGDMVEIH
  GKHEERQDEH GFIAREFNRK YRIPADVDPL 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
```

Programok, inter-fészek☺



REST (Representational state transfer)

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

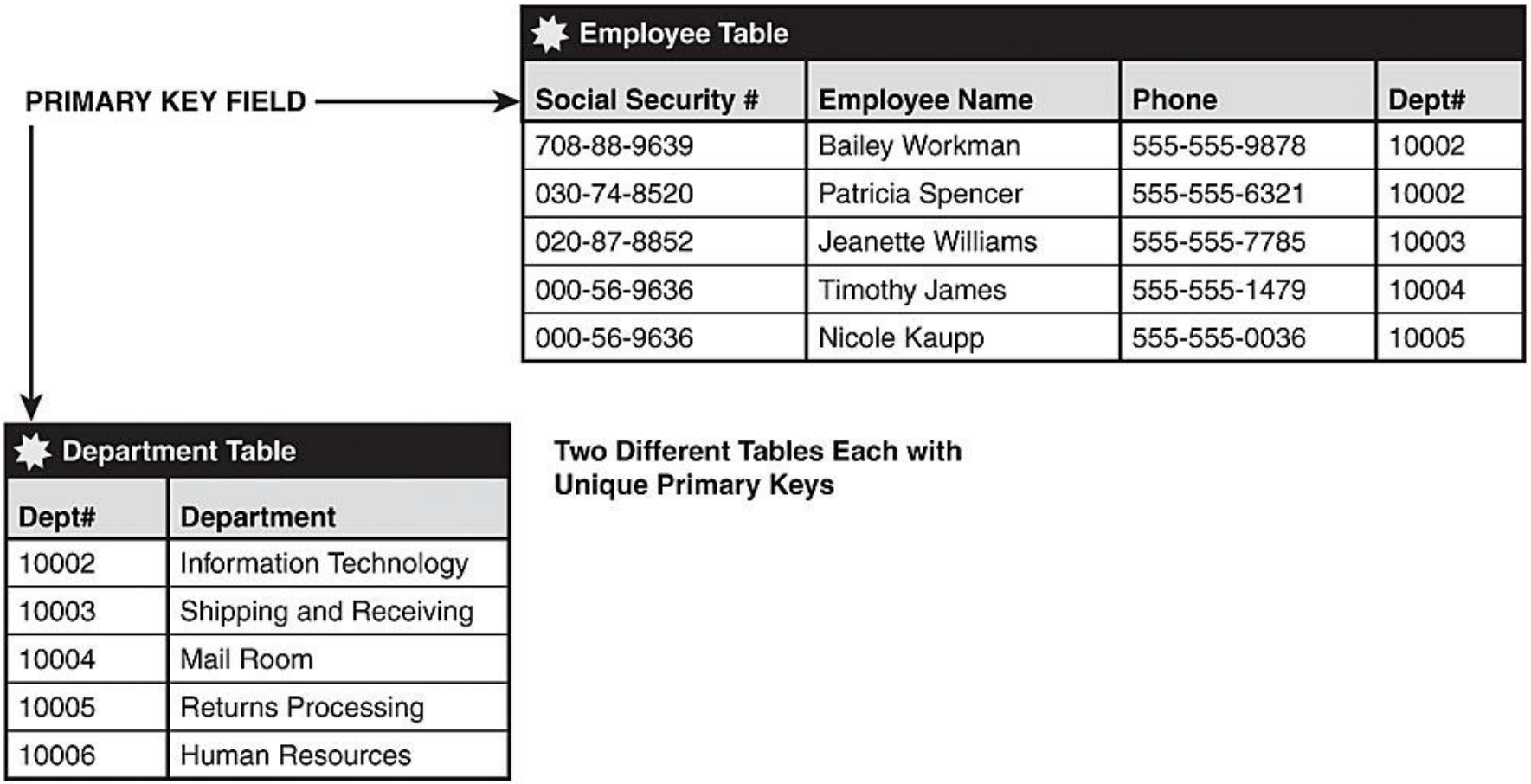
Saját programok – programozási nyelvek

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

GUI

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

RDMS


PRIMARY KEY FIELD
Employee Table
Social Security #
Employee Name
Phone
Dept#
708-88-9639
Bailey Workman
555-555-9878
10002
030-74-8520
Patricia Spencer
555-555-6321
10002
020-87-8852
Jeanette Williams
555-555-7785
10003
000-56-9636
Timothy James
555-555-1479
10004
000-56-9636
Nicole Kaupp
555-555-0036
10005
Department Table
Dept#
Department
10002
Information Technology
10003
Shipping and Receiving
10004
Mail Room
10005
Returns Processing
10006
Human Resources
Two Different Tables Each with
Unique Primary Keys

SQL vs. ORM

A)

```
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$$
```

B)

```
#------------------------------------
class Protein(Base):
    __tablename__   = "protein_binding"

    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")
```

C)

```
SELECT
`protein`.`protein_id`,
`protein`.`gene_name`,
`protein`.`acc`,
`protein`.`entry_name`,
`protein`.`organism`
FROM `drugdb`.`protein`
WHERE `protein`.`acc`='075469';
```

D)

```
results=DBSession.query(Protein).filter(Protein.acc=='O75469').all()
for protein in results: print protein
```

Hálózatok – fehérje hálózatok

- **Rendszerbiológia**
- **Fehérjék-fehérjék kölcsönhatási hálózata**
- **Gének-fehérjék-drogok kölcsönhatása**

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

Hálózatok – Gráfok

vertices/nodes, edges

paths, distance, degree

subgraphs, hubs

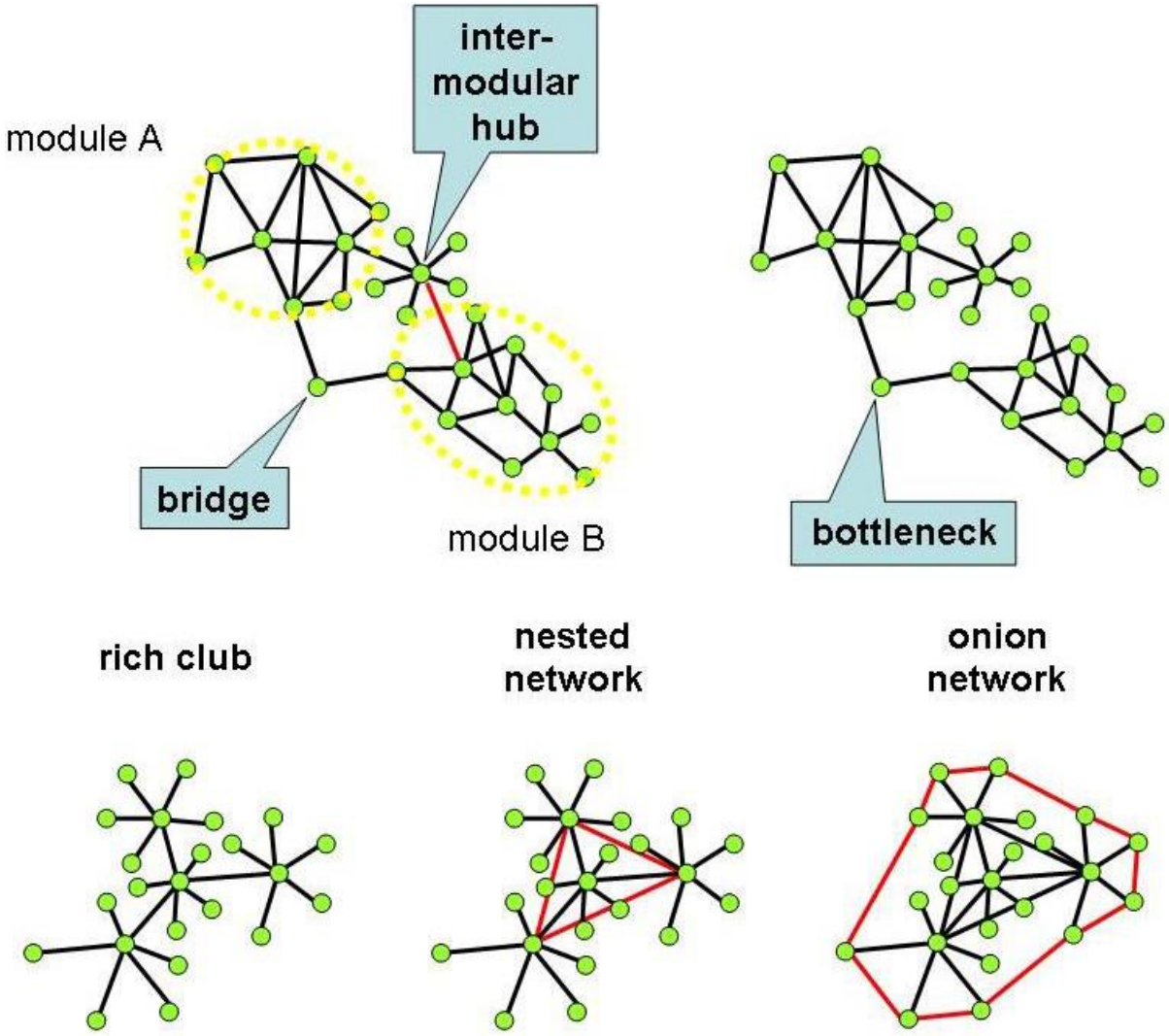
scale-free network, small-world

Teljes redukció

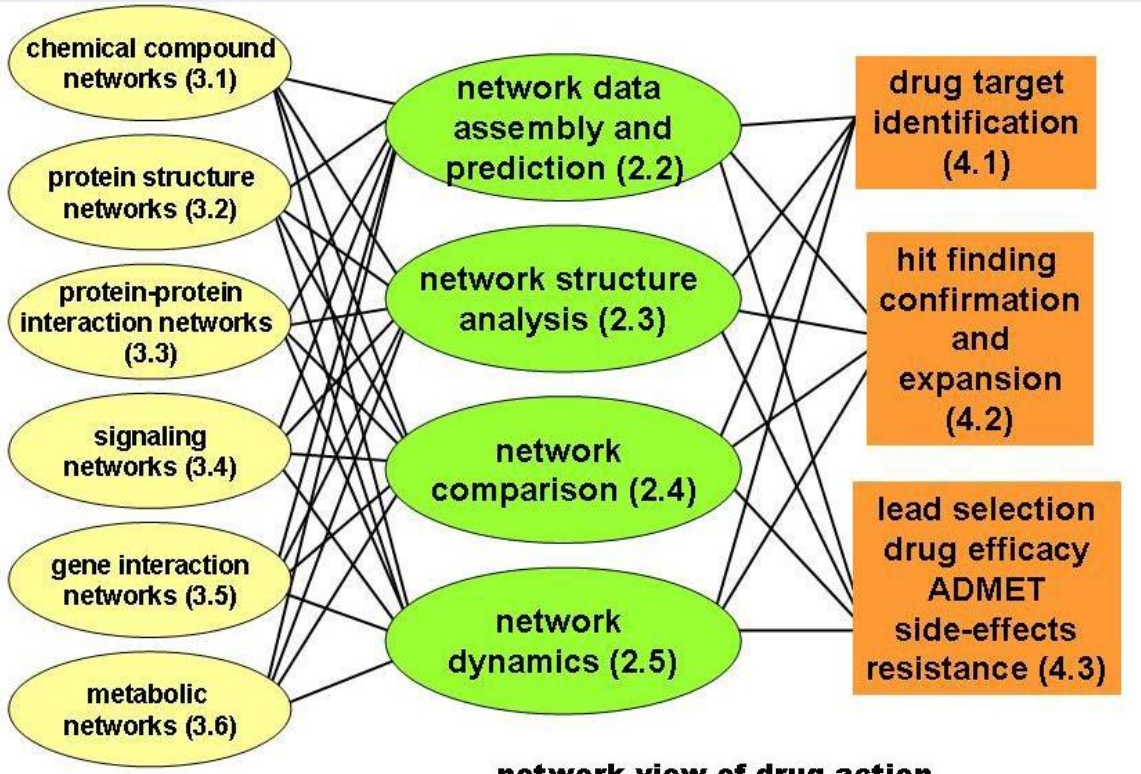
Hálózati leírás

„Mindent-bele”

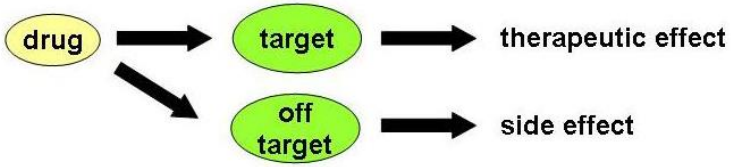
Hálózatok szerkezete



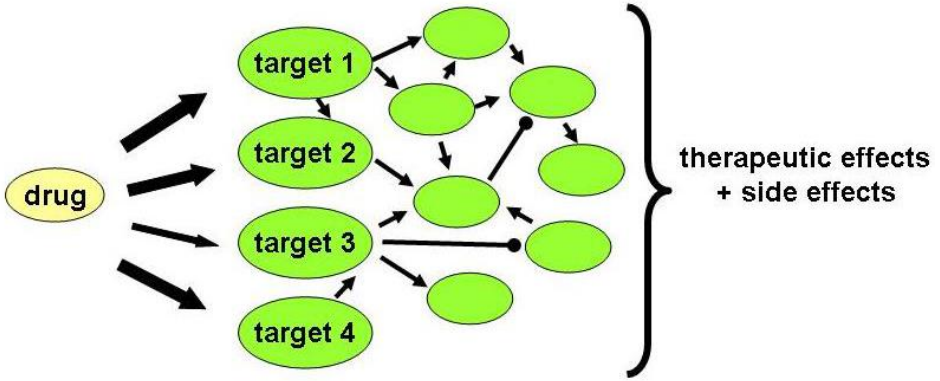
Hálózatok és gyógyszertervezés



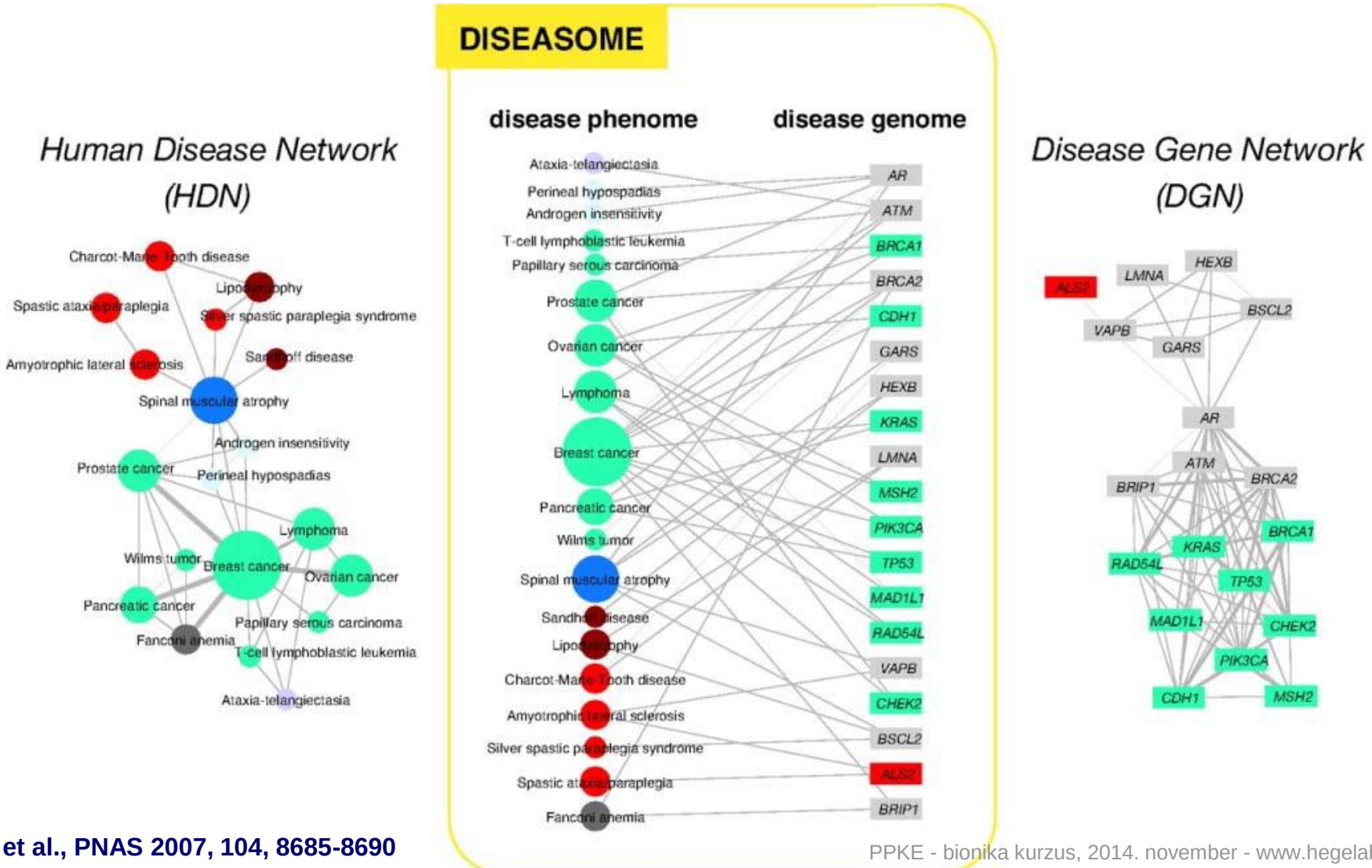
classic view of drug action



network view of drug action

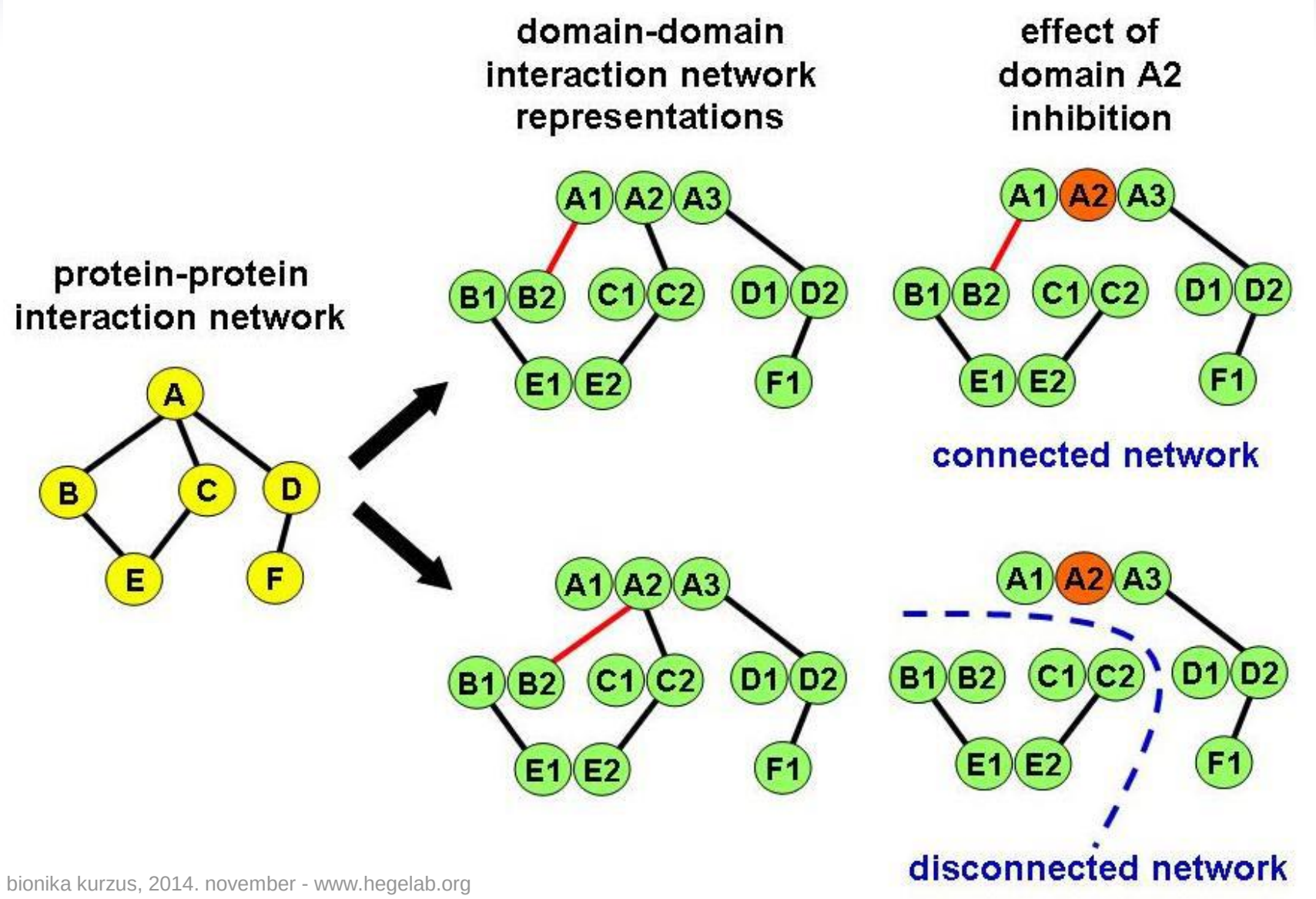


Network assembly

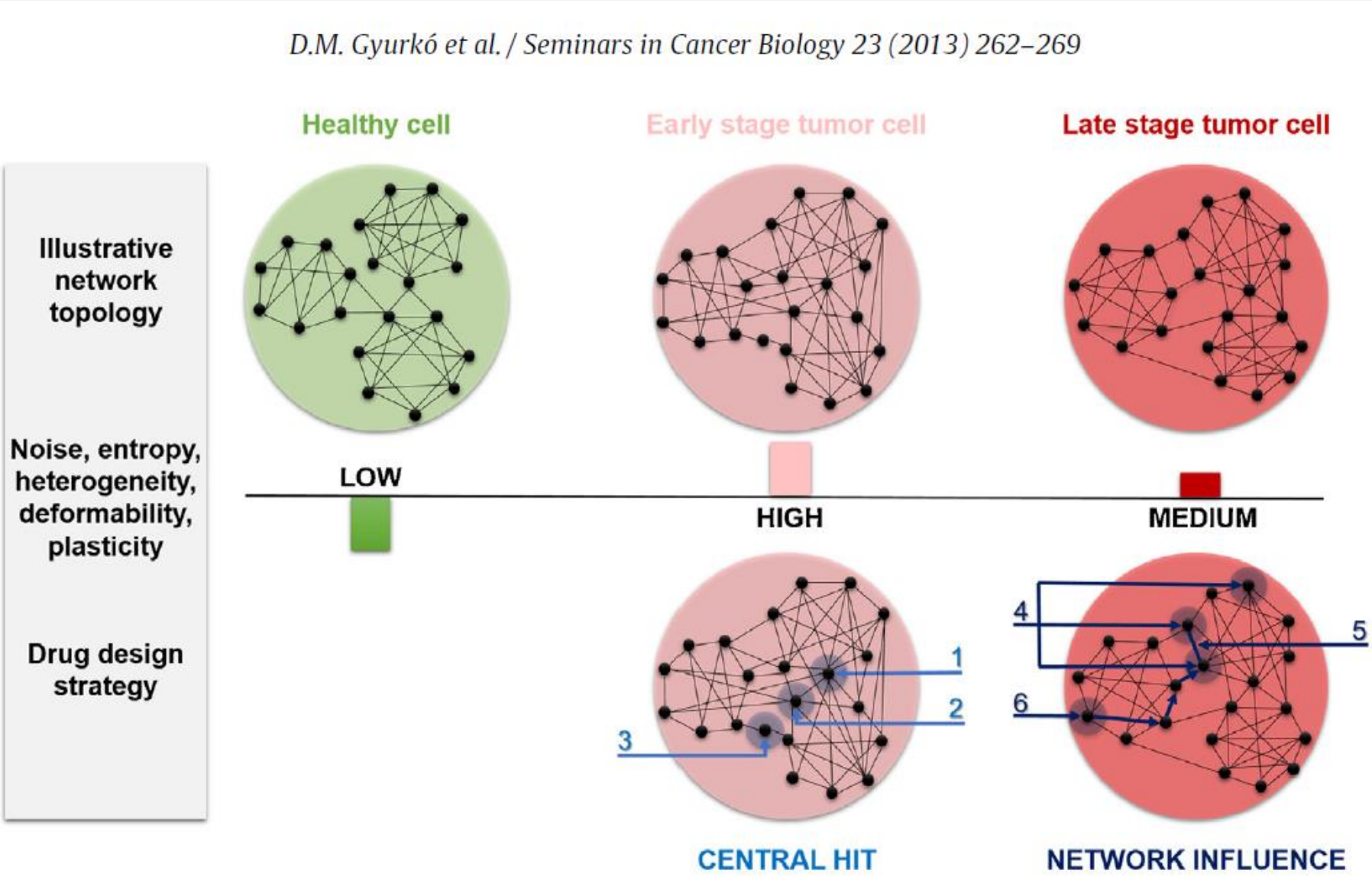


Goh et al., PNAS 2007, 104, 8685-8690

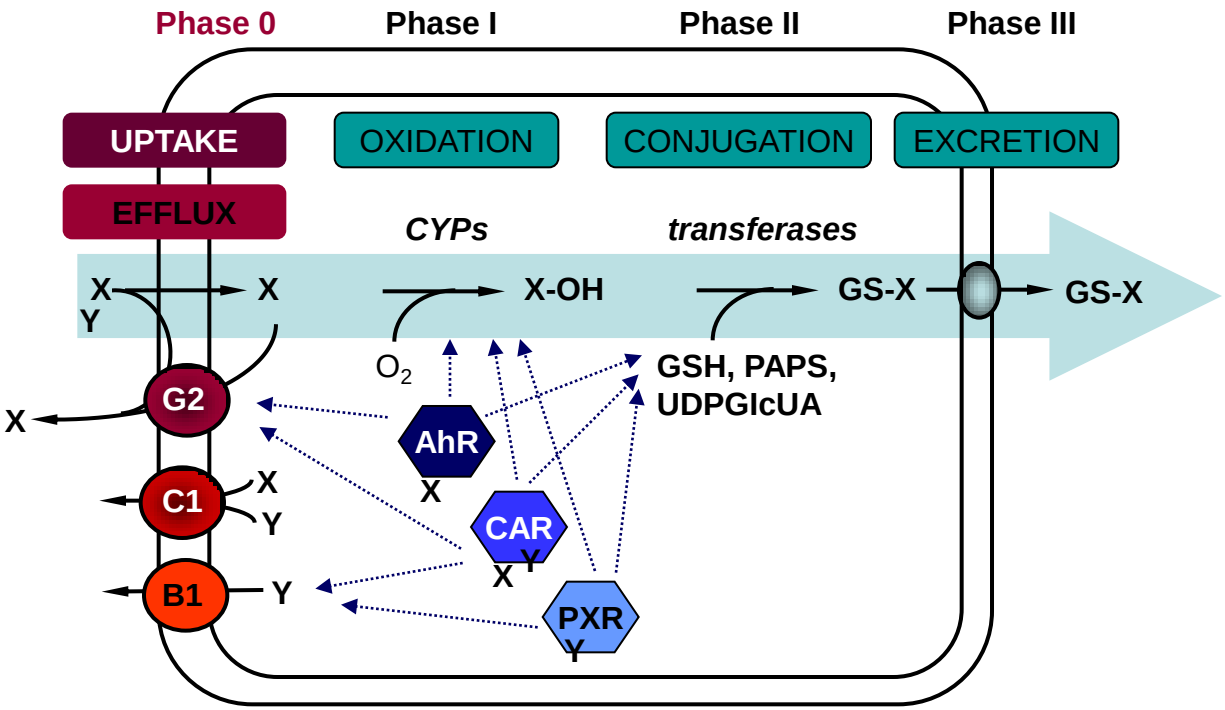
Fehérje-fehérje kölcsönhatási hálózatok

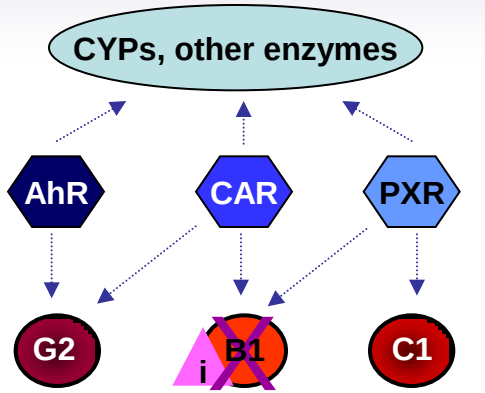


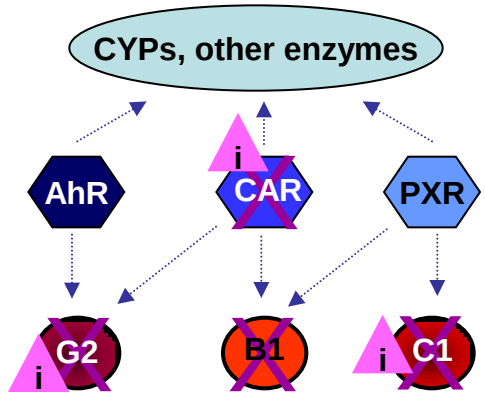
Rákos sejtek hálózatai



A sejt vegyvédelme (sejtszintű immunitás)



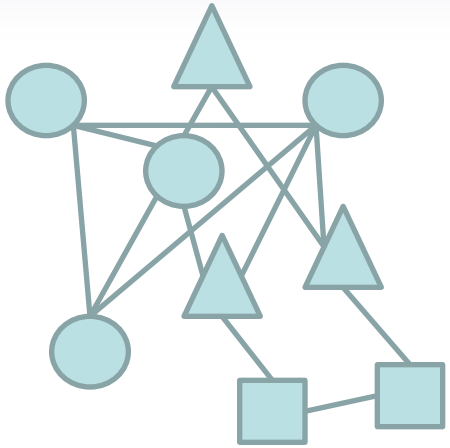


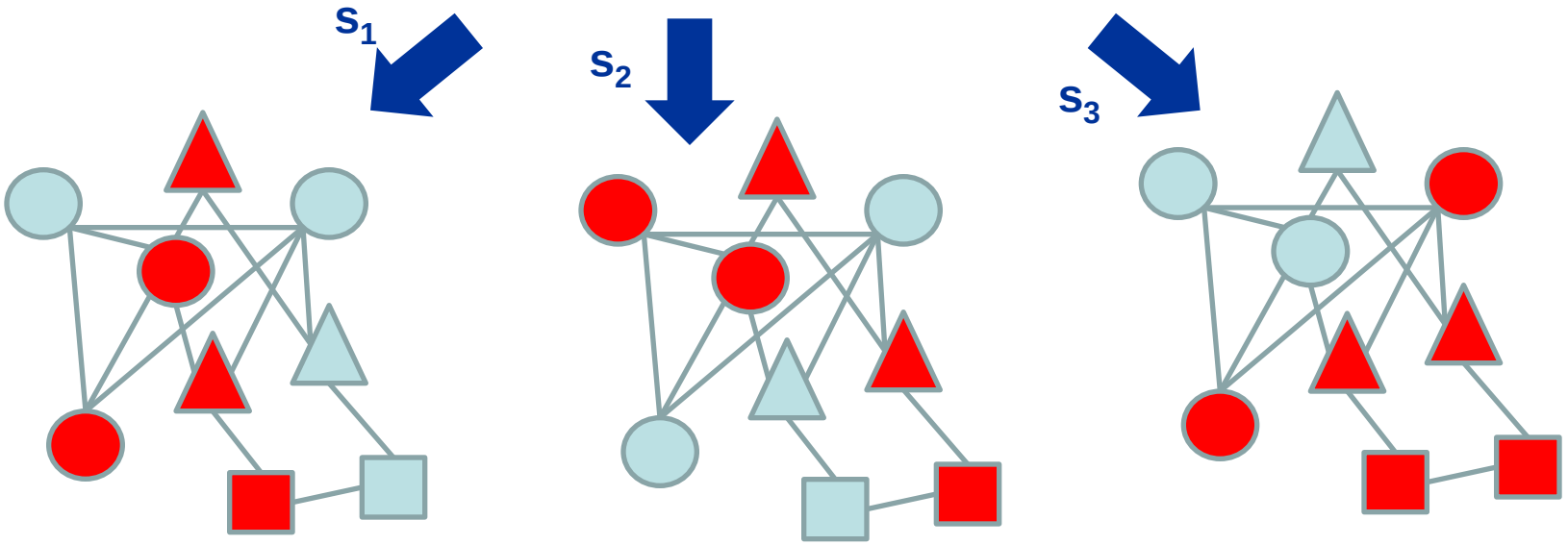


multi-target drugs

Hálozatok működésének feltérképezése

s_i - sejtvonal$_i$ vagy drog$_i$ vagy...





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)

[1]Collection of coherent experiments (by GEO)
[2]One celltype, one agent, one timecourse, one dose

Pipeline of analysis

- Calculate the expression changes
  - Discretization

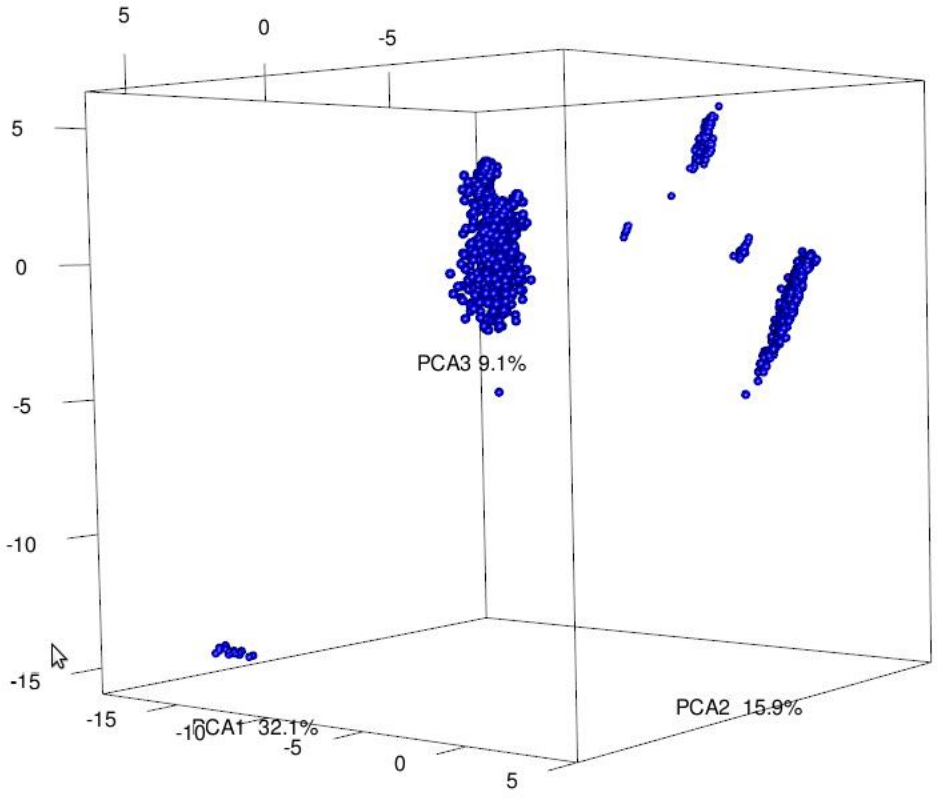
	Expression change (fold)	Discret value
upregulated	**>2x**	**1**
downregulated	**<0,5x**	**-1**
no change	-	0
no probe on chip	-	2

  - Vectors

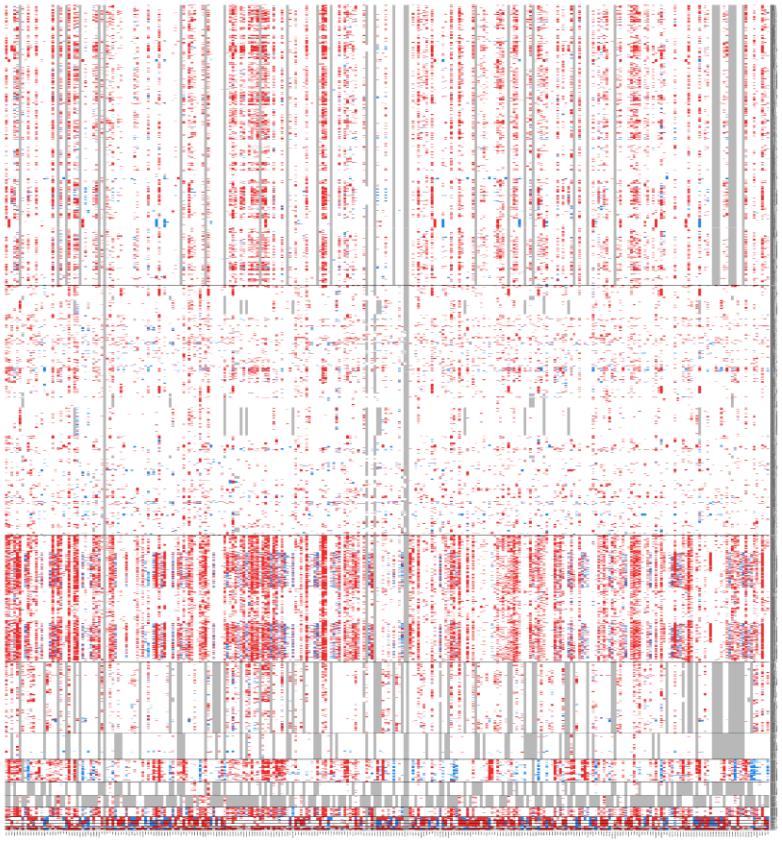
experiment	ABCA1	ABCA8	ABCB1	ABCB11	ABCB4	ABCB5	ABCC1	ABCC2	ABCC3	ABCC4	ABCC5	ABCC6	ABCD2	ABCG2	ABHD10	ABL1	ADH1A	ADH1B	ADH1C	ADH4	ADH6	AHR	AHRR	AKAP13	AKR1A1	AKR1C1	AKR1C2	ALDH16A1	(...)
GDS1249_1	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	**-1**	0	0	**1**	0	0	0	0	0	**1**	0	0	0	0	(...)
GDS1249_2	**1**	**1**	**1**	0	0	0	0	0	0	**1**	0	1	0	0	**-1**	0	0	**1**	0	0	**1**	0	0	**1**	0	**1**	0	0	(...)
GDS1249_3	**1**	**1**	0	0	0	**1**	**1**	0	**1**	**1**	**1**	**1**	0	0	0	0	0	**1**	0	**1**	**1**	0	0	**1**	0	**1**	**1**	0	(...)
(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)

Whole dataset

PCA analysis (the first 3 component)



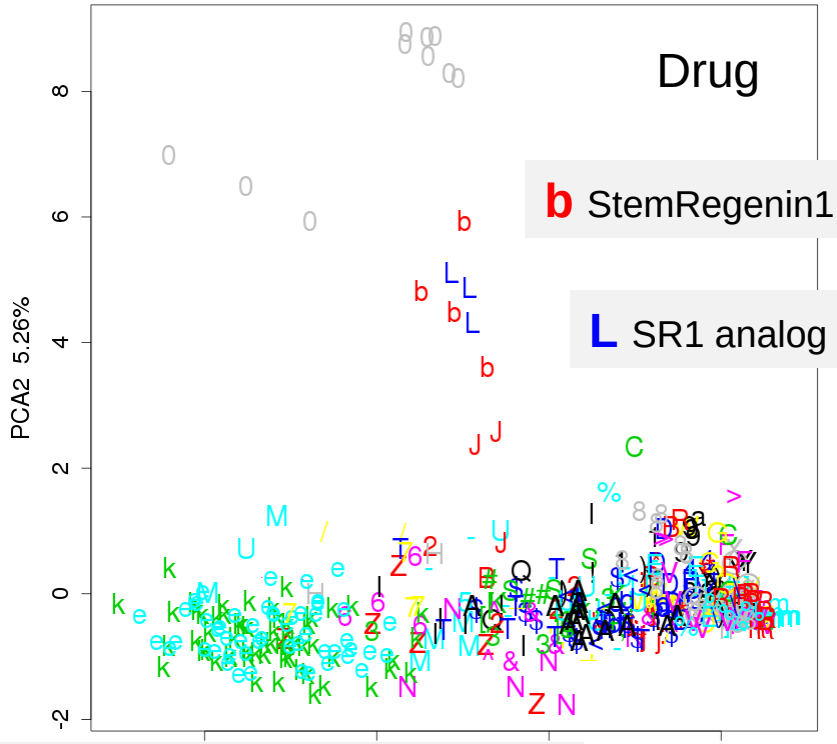
Heatmap (result of clustering)



PCA analysis

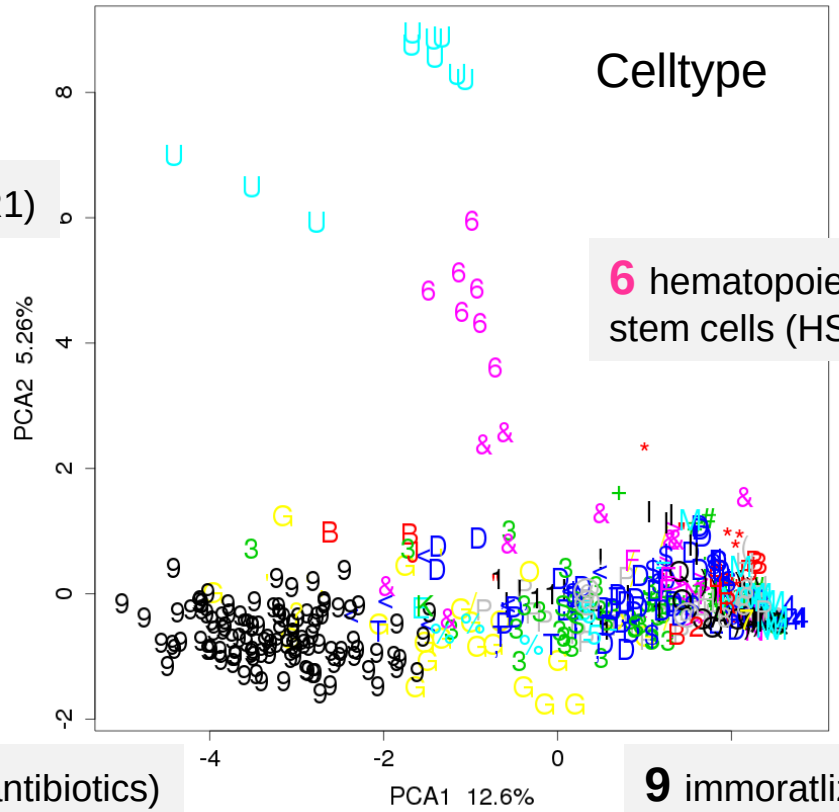
0 BPDE (carcinogen)

U normal lung WI-38 fibroblasts



b StemRegenin1 (SR1)

L SR1 analog



6 hematopoietic stem cells (HSC)

e thapsigargin (SERCA inhibitor)

k tunicamycin (antibiotics)

9 immoratlized B cells

Összefoglalás

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

Köszönetnyilvánítás

www.hegelab.org

Tordai Hedvig
Sarankó Hajnalka

Tóth Attila
Jakab Kristóf
Szöllősi Dániel

Erdei Áron
Erdős Gábor
Harmat Zita

Sarkadi Balázs

MTA-SE Membránbiológiai Kutatócsoport

Kellermayer Miklós

SE Biofizikai és Sugárbiológiai Intézet