

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

Hegedűs Tamás

tamas@hegelab.org

MTA-SE Membránbiológiai Kutatócsoport
SE Biofizikai és Sugárbiológiai Intézet



BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Mai témák

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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Fehérjék dinamikájának jelentősége

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

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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Számítógépes modellezés jelentősége

Atomi szintű információt adnak mozgásokról.

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

PI. NMR és MD - igen

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

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](#)

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

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. PI. hidrofób aminosavak klasztereződése

Strukturálisan igen flexibilisek.

Nincs kompakt globuláris hajtogatódás, reziduális szerkezet.

Megdönt a paradigma,
mely szerint jól definiált 3D szerkezethez kapcsolható fehérje funkció.

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Rendezetlen fehérjék II.

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

Mire jó?	Entrópikus lánc: Effektor: Scavengers: Összeszerelődés: Bemutató felület:	K ⁺ csatorna inaktiválása peptid inhibitorok kazein calmodesmon, F-aktin foszforilációs és proteolitikus helyek
----------	---	--

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Rendezetlen fehérjék III.

A rendezetlenség jóslása

Tanuló algoritmusok
PDB-ben előforduló
rendezetlen fehérjék szekvenciája alapján

Disopred2

Kölcsönhatási energiák becslése

IUPred.enzim.hu

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

K. Dunker – Indiana University

Tompa Péter – MTA Enzimológiai Intézet

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

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 (Isd. 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 ☺

Enzimek osztályozása (EC)

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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

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
- összeillesztés jósága a meghatározó

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Szekvencia-illesztés

[illegible]

CLUSTAL W (1.83) multiple sequence alignment

Alignement – pl. ClustalW

```

2HYD      -----MIKRYLQFVK-----PYKYRIFATIIVGIIKFGIPMLIP
3B5X      -----WQTFKRLWTYIR-----LYKAGLVSTIALVINAAADTYMI
CFTR_HUMAN MQRSPLEKASVSVSKLFFSWTRPILRKGYQRLELSDSIQYIPSVDSADNLS
          *           *           *           *           *

```



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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Negyedleges szerkezet

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

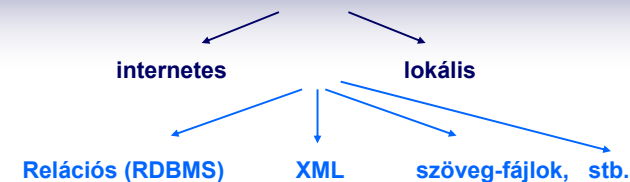
PISA - Protein Interfaces, Surfaces and Assemblies

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

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

Adatbázisok I.

```
graph TD; A[Adatbázisok I.] --> B[Információ]; A --> C[Iskolák]
```



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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Adatbázisok II.

Lokális adatbázisok:

- RDMBS
- fájlok

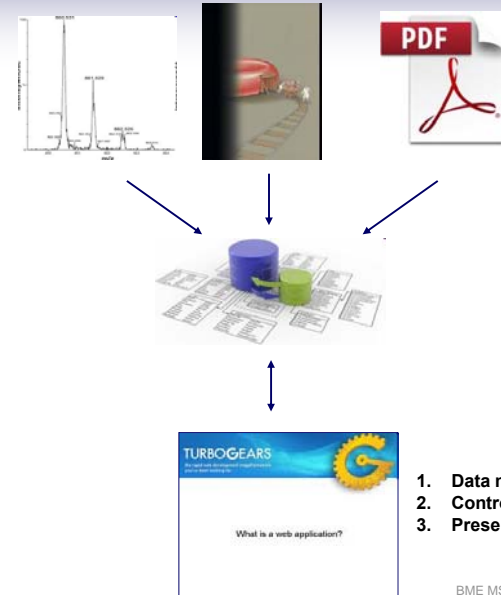
Előnyei:

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

Hátrányai:

- lokális
- erőforrás-igény
- hozzáértés-igény

Meta-adatbázisok



1. Data model
2. Controller (business logic)
3. Presentation (GUI)

Szekvencia fájl formátumok

FASTA

```
>CFTR_HUMAN | P13569 | Cystic fibrosis transmembrane conductance regulator...
MQRSPLKASVVS KLFFSWTRPILRKGYRQRLSDIYQIPSVDSADNLSEKLEREWDR
LASKKNPKLINALRRCCFFWRPFYGFILYLGVEVTKAVQPLLGRITIASYDPDNKEERSIA
IYLGIGLCLLFIVRTL LLLPAIFGLHHIGMQMRIAMFSLIYKTKLSSRVLDKISIGQL
VSLSSNNLNKFDGLALAHFVWIAPLQVALLMGLIWELLQASAFCGLGFLIVLALFQAGL
GRMMMKYRDQRAKGISERLVITSEMIENIQSVKAYCWEAEKMIENLRQTELEKLRKAA
... DTRL
```

PIR

```
>P1;CRAB_ANAPL
ALPHA CRYSTALLIN B CHAIN (ALPHA(B)-CRYSTALLIN).
MDITIHNP LI RRPLFSW LAP SRIFDQIFGE HLQSELLPA SP SLSPFPLMR
SPIFRMP SWL ETGLSEMRLE KDKFSVNLDV KHFSPEELKV KVLGDMVEIH
GKHEERQDEH GFIAREFNRK YRIPADVDFL 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

Mai témák

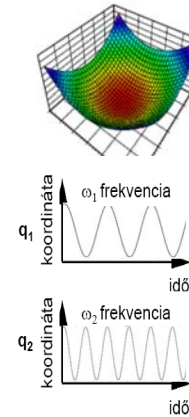
- Bevezetés – a fehérje dinamika és a szimulációk jelentősége
- Fehérjék jellemzése bioinformatikai eszközökkel
 - Másodlagos szerkezeti mintázatok jóslása
 - Rendezetlen fehérjék
 - Funkcionális régiók azonosítása
 - A harmadlagos és negyedleges szerkezet
- Informatikai eszközök – biológus szempontból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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Fehérje dinamika vizsgálata

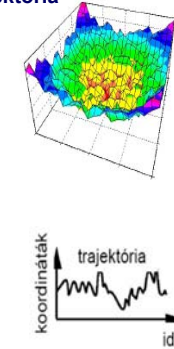
Normál-módus elemzés

- harmonikus potenciál
- analitikus mozgásegyenletek
- normál módusok



Molekuláris dinamika (MD)

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



BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

A „force field” - I.

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

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

TABLE I. Solvation Parameters^a

Atom types ^a	Volume	$\Delta G_{\text{rot}}^{\text{ref b}}$	$\Delta G_{\text{atr}}^{\text{ref c}}$	$\Delta G_{\text{rep}}^{\text{ref b}}$	$\Delta C_{\text{p}}^{\text{ref d}}$
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)

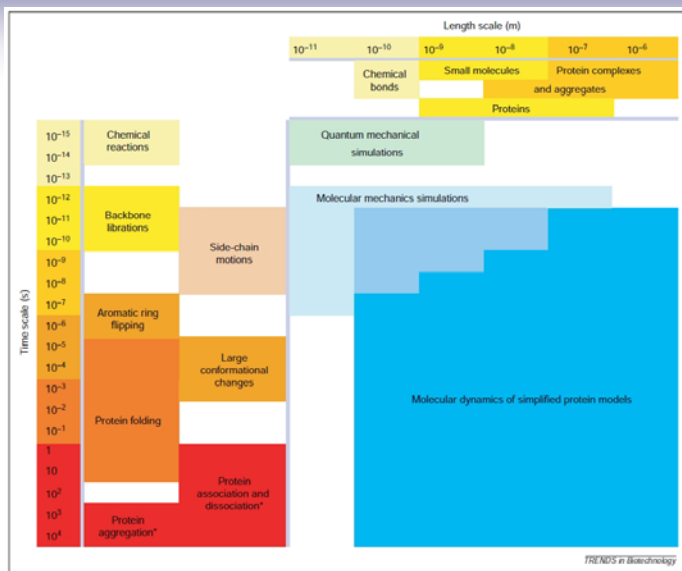
BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Az MD korlátjai

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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

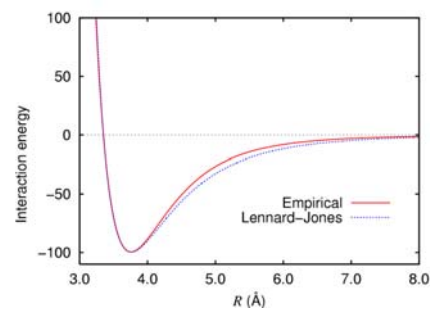
„Események” időskálája



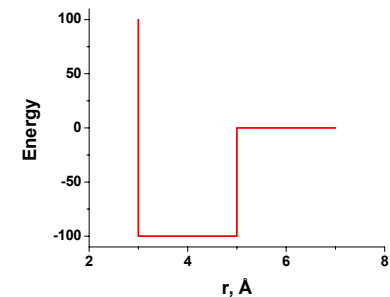
F. Ding and N.V. Dokholyan, TRENDS in Biotechnology, **23**:450 (2005)

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Diszkrét Molekuláris Dinamika (DMD)



wikipedia



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

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Egyszerűsített (Coarse Grain) modellek

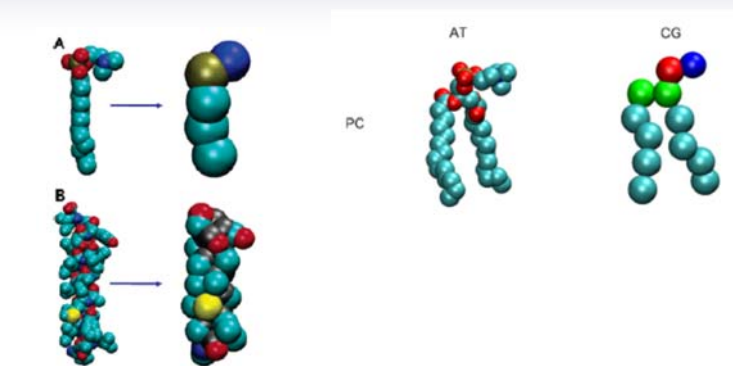


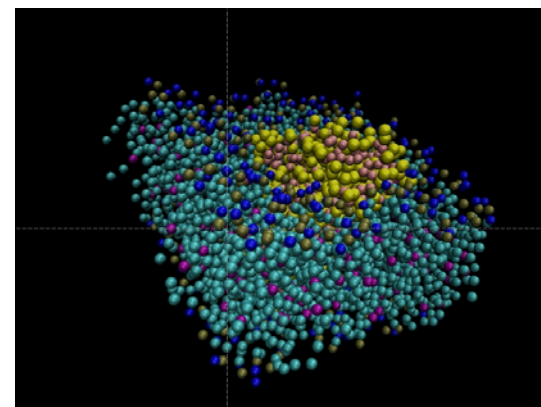
Figure 1. Atomistic (left-hand) and coarse-grained (right-hand) models compared for (A) a DPC molecule and (B) a GpA helix. Colors for atoms:

Bond, Sansom: MARTINI

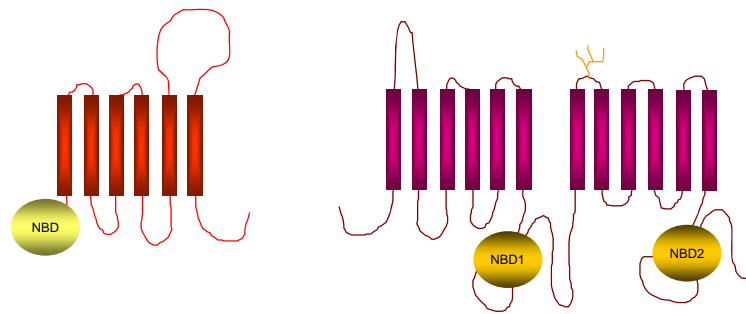
Fehérjére pl. 2 bead vagy 4+ bead modellek

BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Kettősréteg felépülése a fehérje köré

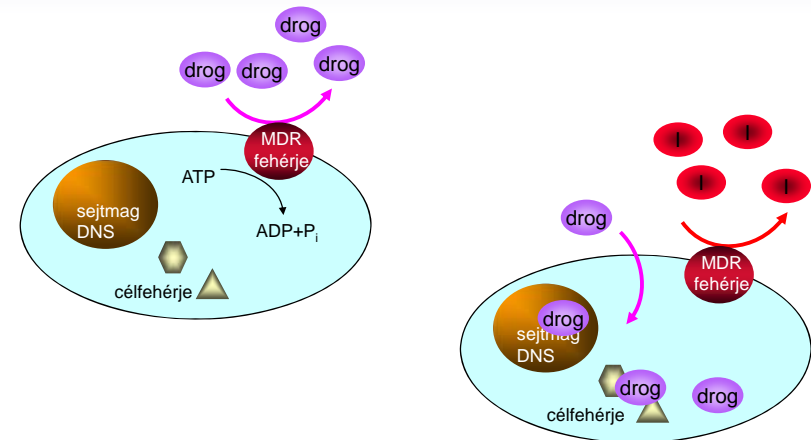


ATP Binding Cassette (ABC) fehérjék



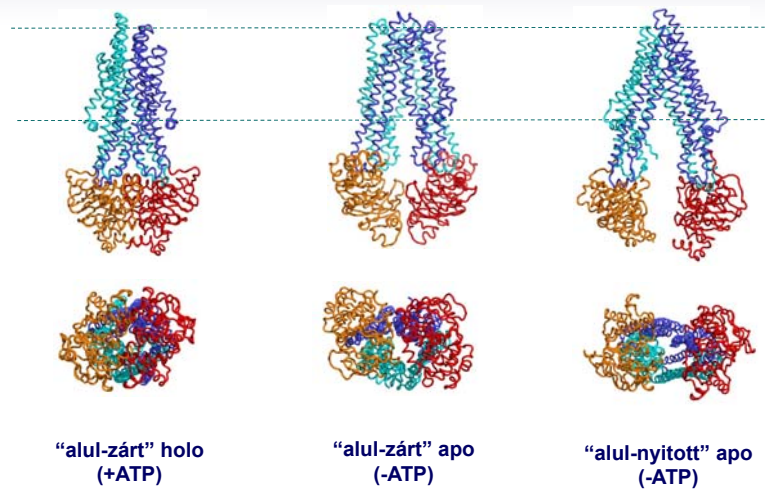
BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

A multidrog-rezisztencia és felfüggesztése



BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

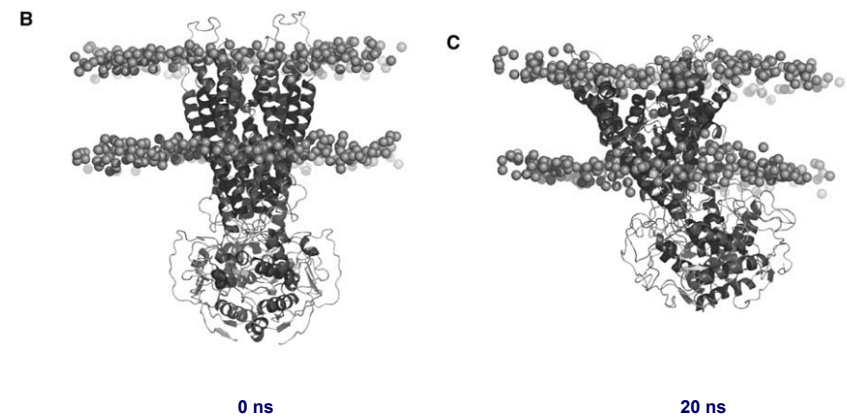
ABC fehérjék konformációi



BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Fehérjék konformációinak stabilitása

Eur Biophys J (2008) 37:403–409



BME MSc-Biofizika kurzus, 2014. - www.hegelab.org

Köszönetnyilvánítás

www.hegelab.org

Tordai Hedvig
Sarankó Hajnalka
Harmat Zita
Tóth Attila
Jakab Kristóf
Szöllősi Dániel
Erdei Áron
Erdős Gábor

Sarkadi Balázs

MTA-SE Membránbiológiai Kutatócsoport

Kellermayer Miklós

SE Biofizikai és Sugárbiológiai Intézet



Támogatások:

KTIA-AIK-12-2012-0025