

Bevezetés a Bioinformatikába

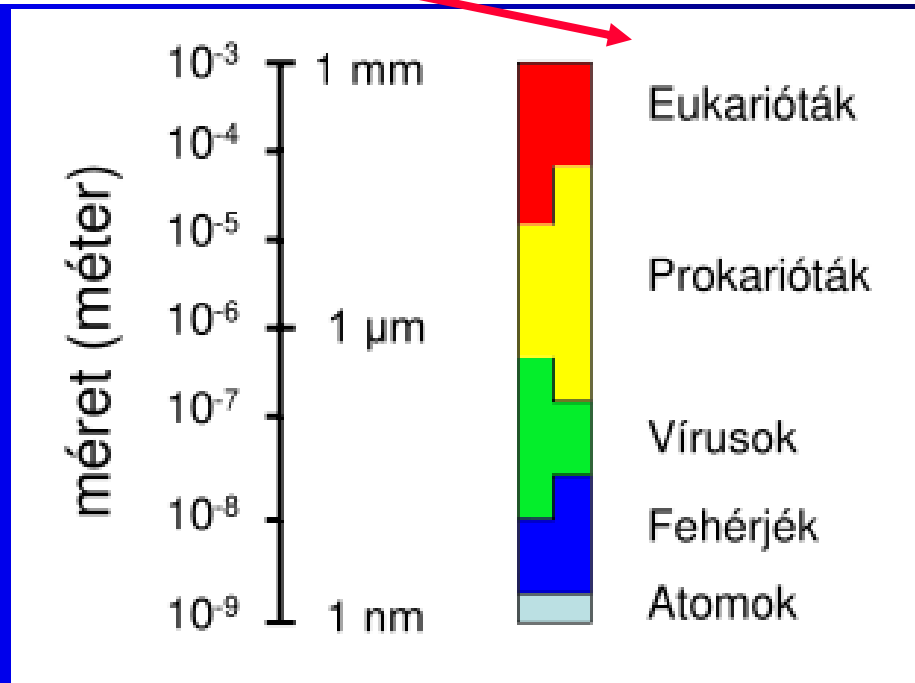
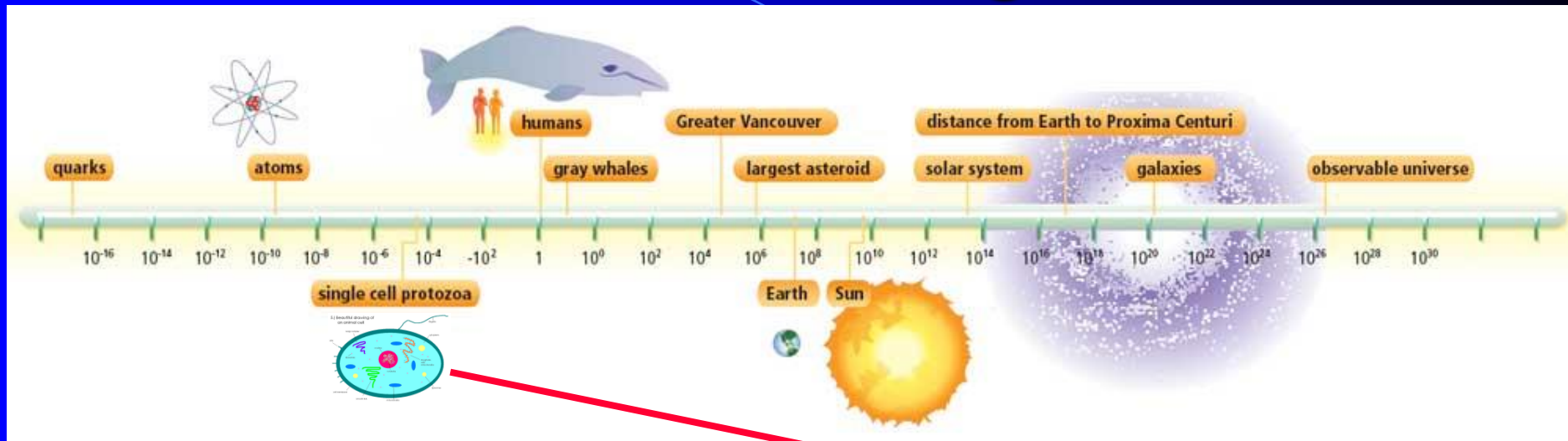
Fehérjék

Kozlovsky Miklós

kozlovsky.miklos@nik.uni-obuda.hu

4. Előadás

Méretetek a világban



Az „omics” világ

✓ Proteomics

- Vizsgálja, hogy hol keletkeznek, valamint milyen szerepet játszanak a fehérjék az élő szervezetben
- Structural proteomics
 - Szekvenciákból térszerkezetet vizsgál/jósol, fehérjéket osztályokba sorol, stb.

● Genomics

- Foglalkozik:
 - génstruktúrákkal
 - szabályozó szekvenciákkal
 - nem kódoló részekkel

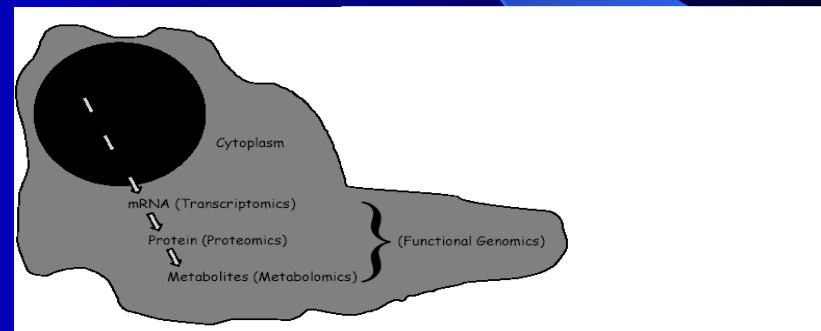
Megpróbálja leírni az élőlényt a genom szekvenciája alapján

● Transcriptomics

- Az átírás törvényszerűségeit vizsgálja

● Metabolomics

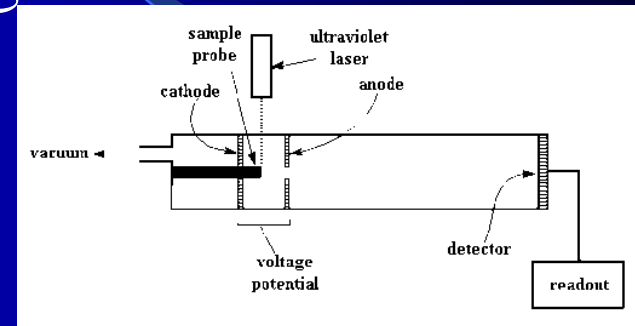
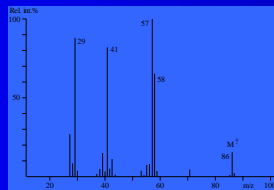
- Az élő sejt számára fontos kis molekulásúlyú molekulák vizsgálata



Fehérjeszerkezet megismerési metodikák

- Fehérje alkotók mennyiségi meghatározása

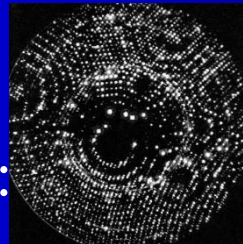
- MALDI - Matrix-Assisted Laser Desorption Ionization



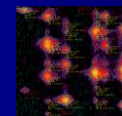
- 3D fehérjeszerkezet

- Röntgen kristallográfia:

- <https://youtu.be/QuCRBxjk3fg>
 - <https://youtu.be/fZOm8wustVk>

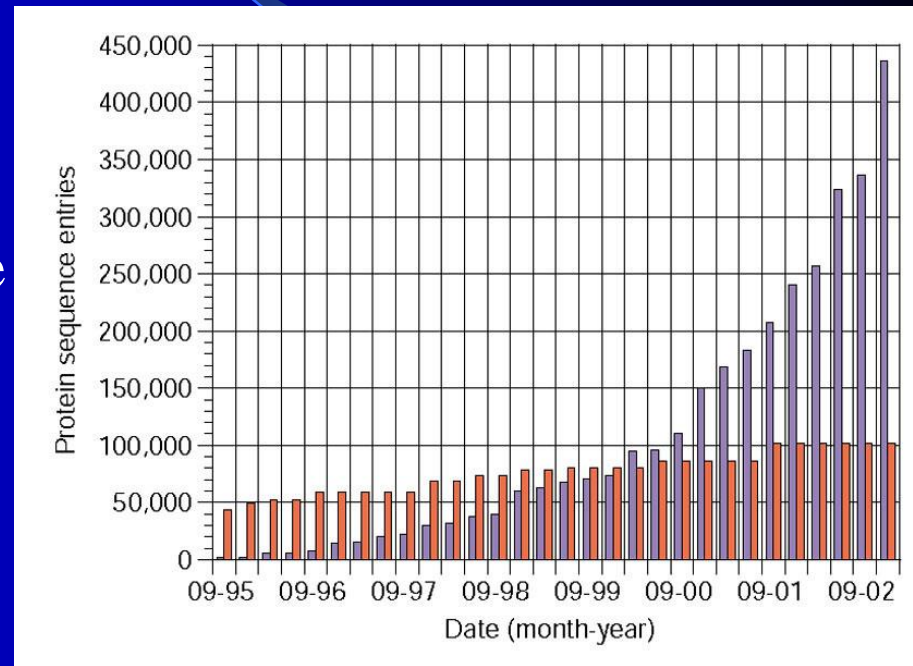


- Mágneses magrezonancia (NMR - Nuclear Magnetic Resonance)



Szekvencia-struktúra deficit

- Nagy projektek/programok szerepe
- Nagy információs deficit
- Jó lenne „jósolni”
 - Jellemzőket, struktúrát, stb.



Szekvencia<?>struktúra

Szent grál keresése

CASP

- Aminosav szekvenciából struktúra jóslás
- Két évente rendezik a versenyt: <http://predictioncenter.org/>
- Népszerű verseny (sok algoritmus)
- A verseny nyílt, az algoritmusokat többféle szempont alapján rangsorolják
- Jól lehet tanulni a konferencia előadásokból:
<http://www.predictioncenter.org/casp8/docs.cgi?view=presentations>
és a cikkekből :
<http://predictioncenter.org/index.cgi?page=proceedings>

Szekvencia<?>struktúra

Szent grál keresése

CASP 2020

- Covid-19 verseny zajlik jelenleg
- eredmények novemberben
- <https://predictioncenter.org/caspcommons/index.cgi>
- File minta:
- https://predictioncenter.org/caspcommons/models_QAresults.cgi

CASP14 (2020)

- Mennyire hasonlít a modell a valósághoz?
 - Domain orientációk, interakciók modellezése pontos?
 - Távolságok és kapcsolódások erőssége pontos?
 - Tudunk a modellel kérdéseket megválaszolni?

CASP14 (2020)

- Negyedleges szerkezetek jóslása (TS)
 - Template alapú modellezés
 - Template nélküli modellezés
 - Oldalláncok, hurkok és aktív kötőhelyek analízálása (ha a struktúra gerince ismert)

CASP (folyt.)

- Egyéb jóslási kategóriák
 - Reziduum-reziduum kapcsolat fehérjékben (RR)
 - Struktúrálatlan régiók meghatározása (disordered regions - DR).
 - Funkciók jóslása kötőhelyek jóslása (FN).
 - Modellek minőségének és megbízhatóságának meghatározása (QA)

CASP15 (2022)

- Most volt a jelentkezési határidő
- <https://predictioncenter.org/casp15/index.cgi>
- Decemberben lesz a verseny
 - Terv: megcsinálni mintha versenyeznénk
 - MI + kvantumszámítási infrastruktúra
 - Összehasonlítani a versenyeredményekkel

Másodlagos struktúrajóslás pontosság mérése

- Q_3 mérőszám
 - N-aminosavak száma a láncban
 - N_{α} -helyesen jószolt struktúra (alfa-hélix, stb.)
- $$Q_3 = \frac{N_{\alpha}}{N} * 100$$

$$Q3 = \frac{N_{r3}}{N} * 100$$

Interpretation	<i>Q3</i> classification
Right-handed helix	H
Strand	E
Left-handed helix	H
Random coil	C

- **Sov mérőszám (Segment overlap score)**
 - (Burkhard Rost et al.)
 - Erősebb mint a Q_3
 - Büntetőpontokat ad ha nem összefüggő a szekvencia (lukakért)
 - Megjegyzés
 - Legyünk mindig konzervatívak, érdemes átlagos értékekkel számolni,
- Egyszerű statisztikák, nincs bennük „fontossági” mérőszám

$$SOV = \frac{\sum \left(\frac{\text{minov}(s_1, s_2) + \delta}{\text{maxov}(s_1, s_2)} \right) \cdot \text{len}(s_1)}{n}$$

Sequence: KETAAKFERQHMDSS TSAASSSNYCQMKSRLTKDRCKPVNTFVHESLADVQAVCSQKNVACKNGQTNCYQSYSTM\$ITDCRETGSSKYPNCAYKTTQANKHIIIVACEGNPPYVPVHFDAS

(b)

[illegible]

(c)

Resulted: CCCCCCCCCDEECCECFCCGGDCCCCCCGDGECCCEFCEEEEEEFFCCCCCCCCDCEFFFCGDFDDEEFFCEFFEEEEEFFCCDFEFDEFEEEEEEEEEEEEEEEEEAEFEEDDEEC
Native: -FFCCCCCCCCDEECCECFCCGGDCCCCCCGDGDCEFCEEEEEEFFCCCCCCCCDCEFFFCGDFDDEEFFCEFFEEEEEFFCCDFEFDEFEEEEEEEEEEEEEEEEEAEFEEDDEECGE-

Szürke szín=egyezés

Megoldások, eszközök I.

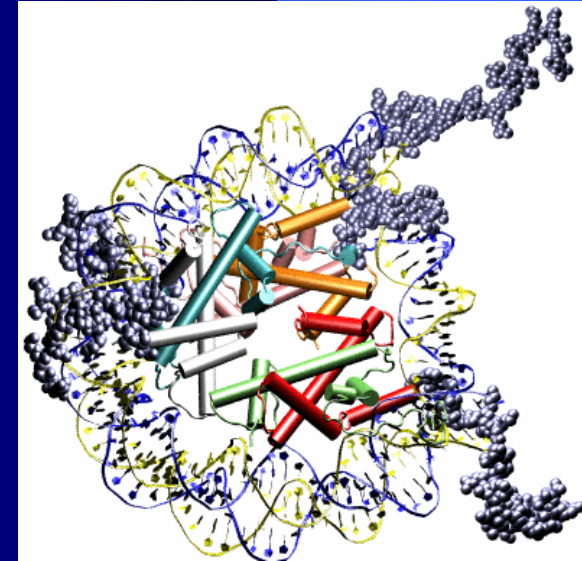
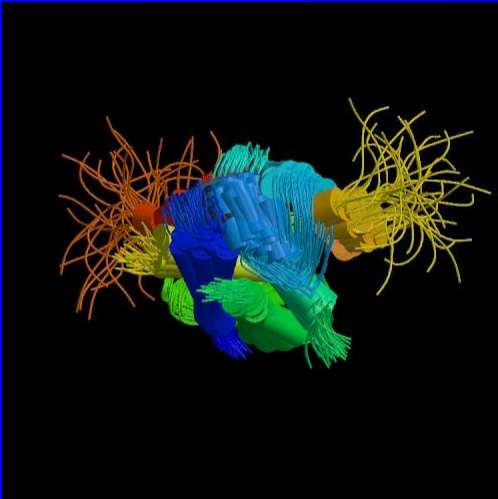
- Első generációs megoldások (70-es évek)
egyszerű aminosav sorrenden alapultak:
 - Chou & Fasman (1974) (Aminosav táblázatból)
 - Lim (1974)
 - Garnier, Osguthorpe & Robson (1978)
 - **Q3 pontosság kb: 50-55%**
- Második generációs megoldások (80-as évek)
Peptid szegmenseken alapulnak:
 - GOR III (1987) ismert aminosav sorrendből felállított struktúrák alapján gyakoriságot vizsgál
 - Neurális hálózatokkal: Qian & Sejnowski (1988)
 - Korábbi algoritmusokból kapott tudást MI-vel megtámogatták
 - **Q3 pontosság kb: 60-65%**

Megoldások, eszközök II.

- Harmadik generációs megoldások (90-es évek)
Evolúciós információk alapján, többszörös szekvencia illesztéssel
 - **PHD (Q3 ~ 70%)**
 - Rost B, Sander, C. (1993)
 - **PSIPRED (Q3 ~ 77%)** (PSIPRED 2.6 Q3 score ~ 80.7%)
 - <http://bioinf.cs.ucl.ac.uk/psipred/>
 - Jones, D. T. (1999)
 - Sok éven át vezette a CASP-ot
- Negyedik generációs megoldások?
 - MI alapú / heurisztikus megoldások
 - AphaFold
 - <https://onlinelibrary.wiley.com/doi/10.1002/prot.26257>
 - Max felső korlát?

Strukturálatlan részek predikciója

- Viszonylag új terület, kutatási fókuszban
- A genom kb 1/5-e ilyen területeket kódol
- DISOPRED és DISOPRED2 (2004, Ward et al.)
 - <http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1>
 - <https://github.com/psipred/disopred>



Domain jóslás

- Domain (1973, Wetlaufer):
 - Általában: fehérje szekvencia részlet, ami önállóan változhat/fejlődhet/struktúrázódhat a fehérje többi részétől függetlenül. (kompakt+funkcionális egység+fejlődik/változik+folding)
 - konzerválódott struktúra, hidrofób résszel
 - Régió, közös

<u>Group</u>	<u>Predictor</u>	<u>Responsible</u>
RR051 ==>	BAKER-RosettaDom	David Kim
RR333 ==>	BAKER-DP_HYBRID	David Kim
RR350 ==>	BAKER-GINZU	David Kim
RR069 ==>	MULTICOM-CMFR	Jianlin Cheng
RR443 ==>	MUProt	Jianlin Cheng
RR453 ==>	MULTICOM	Jianlin Cheng
RR118 ==>	Oka	Oxana Galzitskaya
RR229 ==>	CBRC-DP_DR	Tamostu Noguchi
RR293 ==>	LEE-server	Mina Oh
RR407 ==>	LEE	Jooyoung Lee

Key

Human

Server

Took part in CASP7
with this server

Took part in CASP7 but
with different server

Fizikai tulajdonságok és a szekvencia I.

- Aktív területek jóslása
 - <http://www.biochem.ucl.ac.uk/bsm/ligplot/ligplot.html>
- Naccess (Oldékonyság/oldhatóság)
 - A hozzáférhető molekulákat számolja PDB file-ból
 - <http://wolf.bms.umist.ac.uk/naccess/>
- Elektrosztatikus potenciál számolás + molekula dinamika+ brown mozgás szimulációk

Struktúra osztályozás

- Adatbankok

- SCOP (Structural Classification of Proteins)

- <http://scop.mrc-lmb.cam.ac.uk/scop/>

- CATH

- hasonló a SCOP-hoz, csak más szempontok alapján osztályoz
 - <http://www.cathdb.info/>

Transzmembrán-hélixek

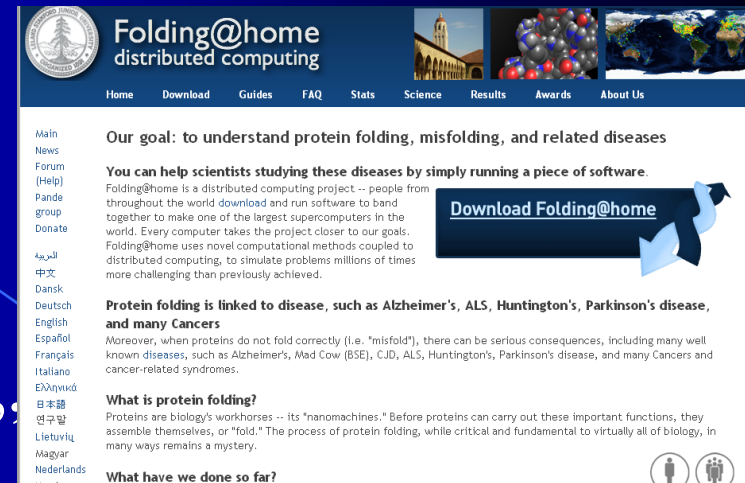
- Nagyon fontos terület (80-as évektől)
 - (Sejt)membránokhoz/ba kapcsolódás
 - Membránfehérjék szükségesek alapvető dolgokhoz pl.: fotoszintézis, idegi aktiváció, légzés, immunválasz, sejtek közötti jelátvitel, stb.
- Fehérje struktúrákat keresünk, melyek:
 - Adott helyeken hélix szerkezetűek (17-25 aminosav hossz)
 - Meghatározott helyeken hidrofób struktúrájuk van
 - Használható szoftverek
 - TMHMM
 - www.cbs.dtu.dk/services/TMHMM-2.0/
 - TOPPRED2
- Kb. 90-95%-os pontosság

3D Jóslási módszerek általánosságban

- Egyszerűnek tűnik, de nem az...
- Két út
 - Homológia modellezés
 - (Összehasonlítunk ismert dolgokkal)
 - <http://salilab.org/modeller/>
 - Ab-initio-jóslás
 - (Kezdeti szekvenciából jósolunk)
 - Már a másodlagos szerkezeteknél is gondban vagyunk, hát még ha bonyolítjuk a dolgot! (Samudrala- RAMP)
- Harmadik út...
 - MI és hibrid módszereket alkalmazunk

Protein Folding

- A fehérjék „hajtógatódása”
- Folding@Home
 - Desktop grid alapú/Boinc
 - <https://foldingathome.org/>
- Lényeges (3D szerkezet):
 - Alzheimer, Parkinson, sok fajta rákos megbetegedés, stb.



Modell ellenőrzés

- Fontos, mert messze vagyunk a 100%-tól
 - PROCHECK
 - <https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>
 - WHAT IF
 - <https://swift.cmbi.umcn.nl/servers/html/index.html>

Elterjedtebb bioinformatikai szoftverek

- Open Bioinformatics Foundation (OBF)

- Nyílt

- Főbb projektjei

- BioJava

- BioPerl

- Biopython

- BioRuby

- BioSQL

- DAS és Global Sequence Identifiers lista

- EMBOSS

- Bioconductor - R

