

# Bevezetés a Bioinformatikába

## Fehérjék

Kozlovsky Miklós

[Kozlovsky.miklos@nik.uni-obuda.hu](mailto:Kozlovsky.miklos@nik.uni-obuda.hu)

3. Előadás

# Ismétlés I.

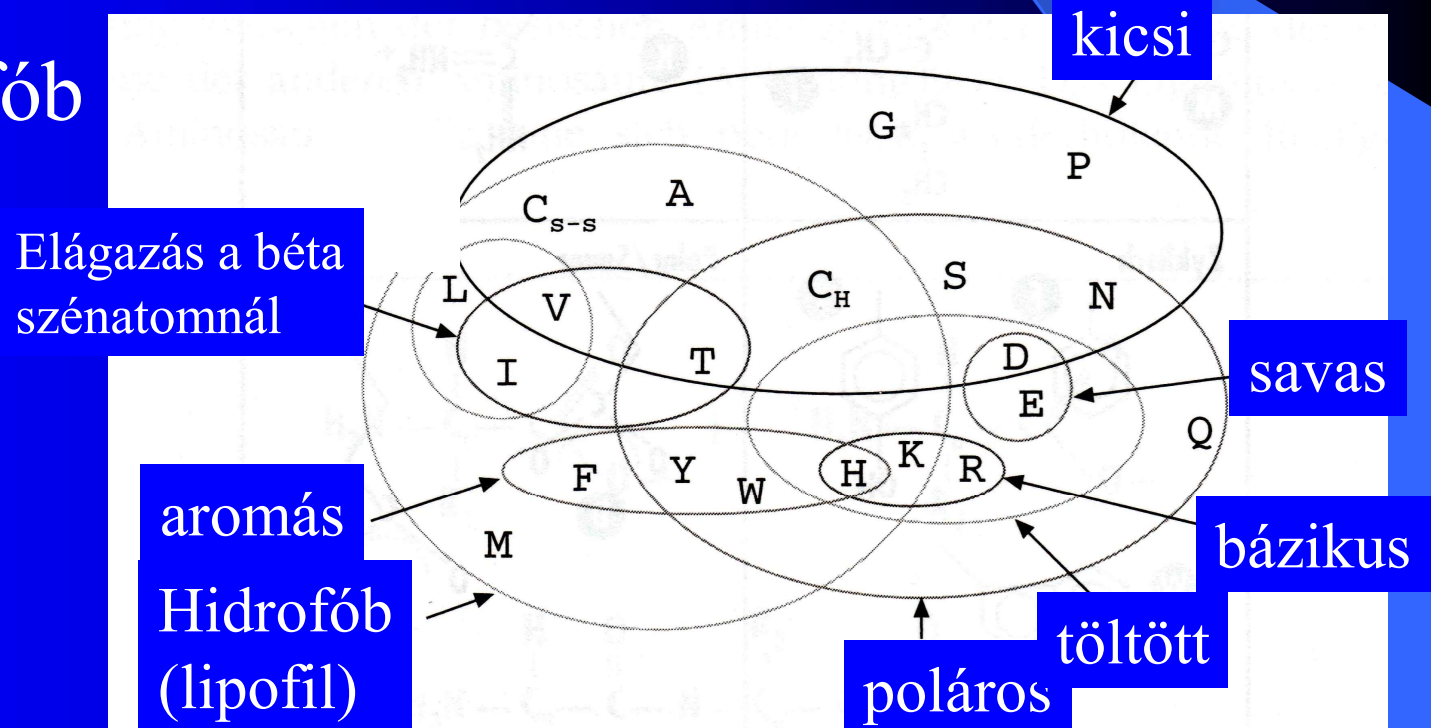
## Fehérjék szerepe az élőlényekben

- Enzimek (katalitikus folyamatok)
- Transzportfehérjék (pl. sejthártyáknál)
- Védőfehérjék
- Toxinok
- Hormonok
- Kontraktilis fehérjék
- Struktúrafehérjék
- Tartalékfehérjék (pl. tojás, növények magvai)
- stb.

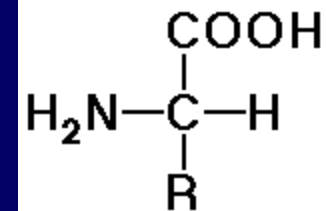
# Ismétlés II.

## Aminosavak

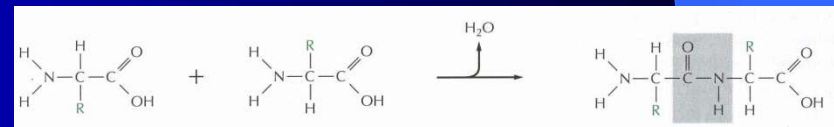
- Felépítésük
- Kapcsolódásuk
- Tulajdonságaik (Venn-diagrammal)
- N-terminális C-terminális
- Hidrofil/fób



# Aminosavak



- Szerkezetük az R csoportban különbözik
- 20 természetes aminosavból építkezünk (esszenciális, nem esszenciális)
- Az egyes R-el jelzett gyökök kiemelkednek mint oldalláncok, és ezek lesznek a biológiailag aktív hatócsoporthok
- Az aminosavak egymással való összekapcsolódása: peptidkötéssel



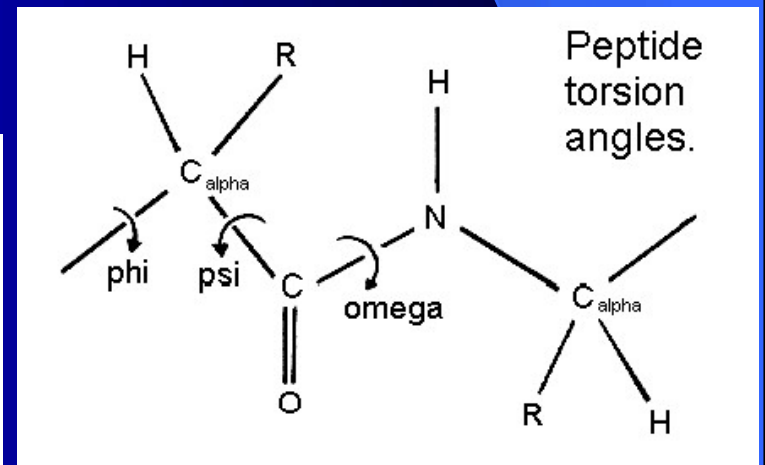
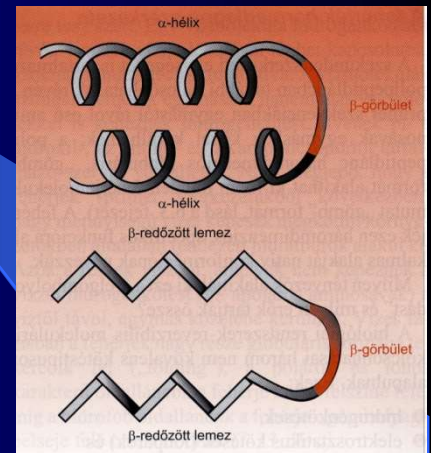
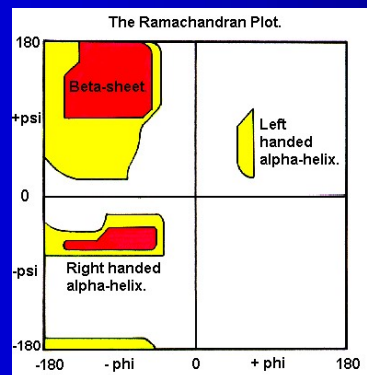
# Elsődleges szerkezet

- A peptid láncot alkotó aminosavak minősége és sorrendje határozza meg
- 100 aminosavból álló fehérje esetén  $20^{100}$  számú egymástól eltérő kombináció létezhetne!

# Másodlagos szerkezet

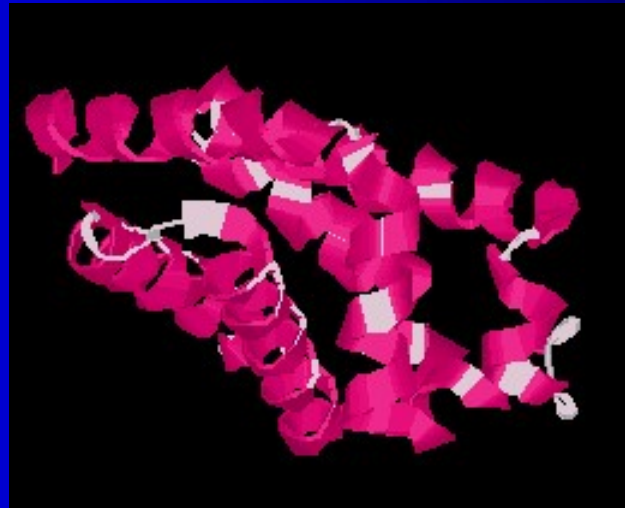
- Másodlagos szerkezet: a lánc gerincének rövid távú szerkezete
- Szakaszokat különböztethetünk meg:
  - Periodikus szakasz (pl. hélix, ill. béta-redő):
    - Homokonformációk: a ( $\phi$ ,  $\psi$ ) pár ismétlődik.
  - Aperiodikus szakasz (pl.:prolinban gazdag részek)
    - Heterokonformációk: a ( $\phi$ ,  $\psi$ ) változik
  - Kanyarok (angolul: turn)
    - Béta kanyar
    - Gamma kanyar

Ramachandran plot



# Harmadlagos szerkezet

- A teljes polipeptidlánc térbeli szerkezete, a másodlagos szerkezeti elemek térbeli elrendeződése
- A stabilitást jelentősen meghatározzák a feltekeredés miatt egymáshoz került oldalláncok között kialakuló kötések:
  - Diszulfidhidak
  - Ionkötések
  - Hidrogénkötések
  - Apoláris kötések

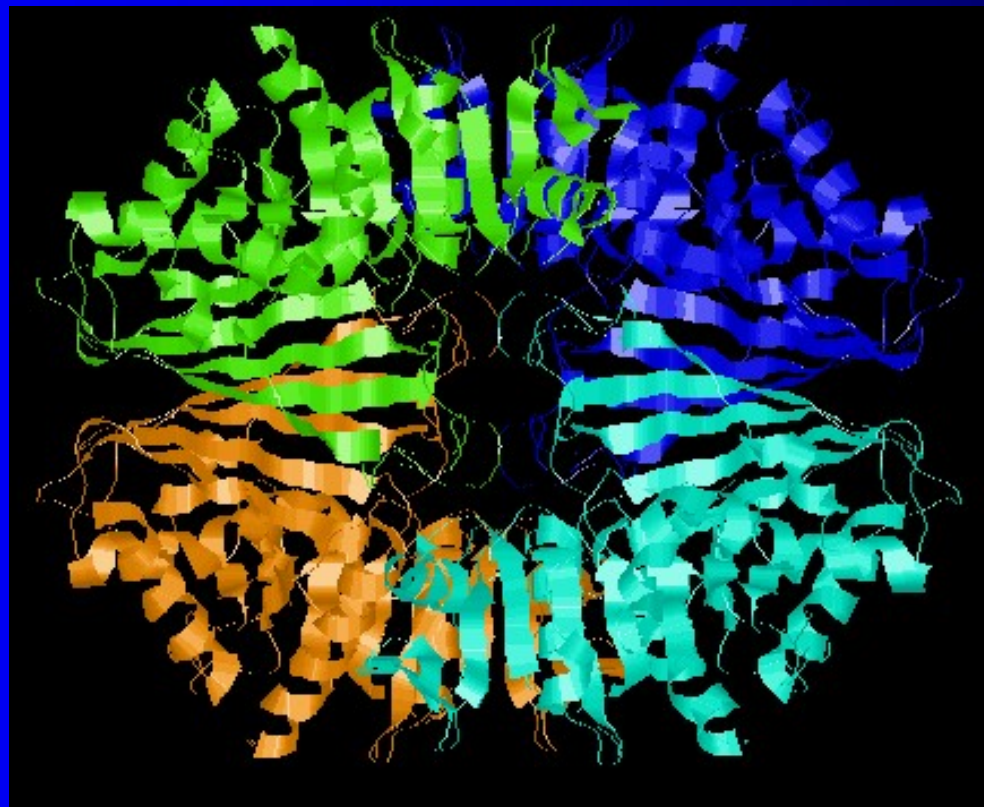


Az egymáshoz nagyjából hasonló térszerkezetű fehérjék általában egy szerkezeti családba tartoznak.



# Negyedleges szerkezet

A több polipeptid láncból álló fehérjék  
alegységszerkezete





# Felhasználási területek I.

- Molekulastruktúra megjelenítés
  - Számítógépes grafikával
  - Aktív centrumok meghatározása
  - Másodlagos szerkezet analízis
  - Domén identifikáció
  - Molekulageometria analízis
  - Molekulák közötti kapcsolati területek
  - Molekulafelszín vizualizáció
  - Elektrostatikus potenciál számolás

# Felhasználási területek II.

- Fehérjeazonosítás
  - Aminosav összetétel/tömeg/izoelektromos pont stb. alapján
  - Struktúra klasszifikáció, fehérje azonosításhoz
- Struktúra illesztés, összehasonlítás
  - Rokonsági (pl. evolúciós) kapcsolatok kutatása
- Fizikai tulajdonságok predikciója (jóslása) szekvenciából
  - Térszerkezet predikció
  - Oldószer általi hozzáférhetőség predikciója
  - Transzmembrán hélixek predikciója
  - Kölcsönhatások predikciója
  - Töltésviszonyok predikciója
  - Felszíni struktúrák, üregek meghatározása (hozzáférési pontok!)

# Biológiai információ kódolás (jelkészlet)

- IUPAC – International Union of Pure and Applied Chemistry
  - Nemzetközi “nomenklatura gyártó” szervezet
  - Leegyszerűsíti a műveleteket, kisebb redundancia
    - Gyorsabb, egyszerűbb algoritmusok



# Biológiai információ kódolás (file formátumok)

- GenBank (2 név)
  - Nukleotid szekvencia leíró file formátum
  - Génszekvencia adatbázis (később bővebben)
- PDB - Protein Data Bank
  - Biológiai makromolekulák 3 dimenziós struktúra leíró formátuma
- FASTA (2 név)
  - Általános szekvencia leíró file formátum
  - Szekvencia összehasonlító algoritmus (később bővebben)
- ASN.1
  - Általános file formátum

# Genbank formátum

- Egy bejegyzés 3 részből épül fel
  - Header - Fejléc
    - LOCUS/ DEFINITION/ ACCESSION/ VERSION/ KEYWORDS/ SOURCE/ ORGANISM/ REFERENCE/ AUTHORS/ TITLE/JOURNAL/ MEDLINE/ PUBMED/ REMARK/ COMMENT
  - Features - jellegzetességek
    - SOURCE/CDS/GENE Feature/RNA Feature
  - Sequence - Szekvencia
    - A teljes nukleotid sorrend

```
LOCUS       AF067844      218336 bp      DNA           PRI           08-FEB-1999
DEFINITION  Homo sapiens chromosome 10 clone PTEN, complete sequence.
ACCESSION   AF067844
VERSION     AF067844.1   GI:4240386
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 218336)
AUTHORS     Jensen,K., de la Bastide,M., Parsons,R., Parnell,L.D.,
             Dedhia,N.,
             Gottesman,T., Gnoj,L., Kaplan,M., Lodhi,M., Johnson,A.F.,
             Shohdy,M., Hasegawa,A., Haberman,K., Huang,E.M., Schutz,K.,
             Calma,C., Granat,S., Wigler,M. and McConbie,W.R.
TITLE       Genomic sequence of PTEN/MMAC1
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 218336)
AUTHORS     Jensen,K., de la Bastide,M., Parsons,R., Parnell,L.D.,
             Dedhia,N.,
             Gottesman,T., Gnoj,L., Kaplan,M., Lodhi,M., Johnson,A.F.,
             Shohdy,M., Hasegawa,A., Haberman,K., Huang,E.M., Schutz,K.,
             Calma,C., Granat,S., Wigler,M. and McConbie,W.R.
TITLE       Direct Submission
JOURNAL     Submitted (18-MAY-1998) Lita Annenberg Hazen Genome Sequencing
             Center, Cold Spring Harbor Laboratory, 1 Bungtown Rd., Cold
             Spring
             Harbor, NY 11724, USA
FEATURES    Location/Qualifiers
             source                1..218336
```

# PDB formátum

- Research Collaboratory for Structural Bioinformatics (RCSB)
- Formátum
  - A molekulán belüli atomi koordinátákat tartalmazza

```
HEADER      B-DNA
COMPND      G-C B-DNA BASE PAIR
AUTHOR      GENERATED BY GLACTONE
SEQRES      1 A      1      G
SEQRES      1 B      1      C
ATOM        1 P      G A      1      -6.620      6.196      2.089
ATOM        2 OXT     G A      1      -6.904      7.627      1.869
ATOM        3 O2P     G A      1      -7.438      5.244      1.299
ATOM        4 O5'     G A      1      -5.074      5.900      1.839
ATOM        5 C5'     G A      1      -4.102      6.424      2.779
ATOM        6 C4'     G A      1      -2.830      6.792      2.049
ATOM        7 O4'     G A      1      -2.044      5.576      1.839
ATOM        8 C3'     G A      1      -2.997      7.378      0.649
```

# FASTA formátum

- A legegyszerűbb formátum
- Nagyon elterjedt
- Formátum
  - „>” (kacsacsőr) és utána sortörés nélkül rövid megjegyzés
  - Majd a teljes szekvencia

```
>gi|4240386|gb|AF067844.1|AF067844 Homo sapiens chromosome 10 clone PTEN, complete sequence  
CAAGCTTTACACTAGAGCCTATATGAAGTTTTGATTCTAAGTGTTAATGTACCTTCTGACAACTGTGAAA  
TGAACCTTGTTTCCTGGGGAGCGCGTTCTGGTTTTCTCTTTGCACAGTTAAGCTGAGACTAGCATCATTCT  
AGTTTGCAGGTGACATTCTCTGGGAAGCTAGTCTATGGGGGAGATGACATCTTCTGAACCTAGTCCCCAC  
AGAGAACTTTGAATGAGTGGAATCAAGAGGTTGCCTGCATTCTTGCTCATGTCACAATGCTGGACATGTG
```



# Abstract Syntax Notation One

## ASN.1

- Jellemzői

- Formális leíró nyelv, abstract adattípusokhoz
- Alkalmazott nyelvtől, és fizikai adatrepresentációtól független
- Rendszerek közötti kommunikációhoz fejlesztették (VoIP, IN, mobil rendszerek...)
- Emberi fogyasztásra kevésbé alkalmas, automatikus feldolgozáshoz annál inkább
- Az adatokat csak hordozza, nem végez velük műveleteket (nem programnyelv)

- Szabványosítása

- 1984 (CCITT) -> ISO 8824+8825->IEC *JTC 1*
- Jelenleg az ASN.1:2002 az aktuális

# Fehérje megismerési metodikák

- Különválasztás legjellemzőbb tulajdonságok alapján
  - Méret
  - Oldékonyság
  - Töltés
  - Kötődési affinitás

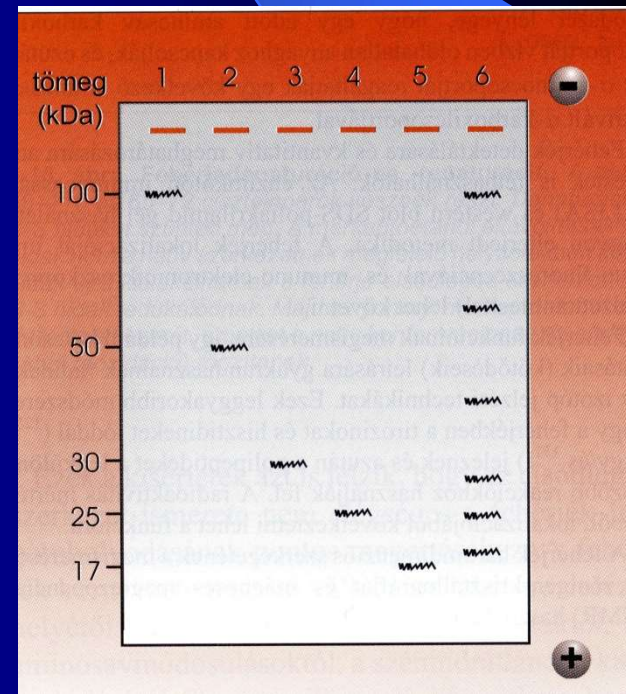
# Méret szerinti elválasztás I.

- SDS-poliakrilamid gélelektroforézis
  - Markerek használata, akár 0,1 µg protein detektálás is lehetséges

<http://www.youtube.com/watch?v=5CMVqRdh0Yk&feature=related>

<http://www.youtube.com/watch?v=toPpdoBYPWo&feature=related>

## Ultracentrifugálás

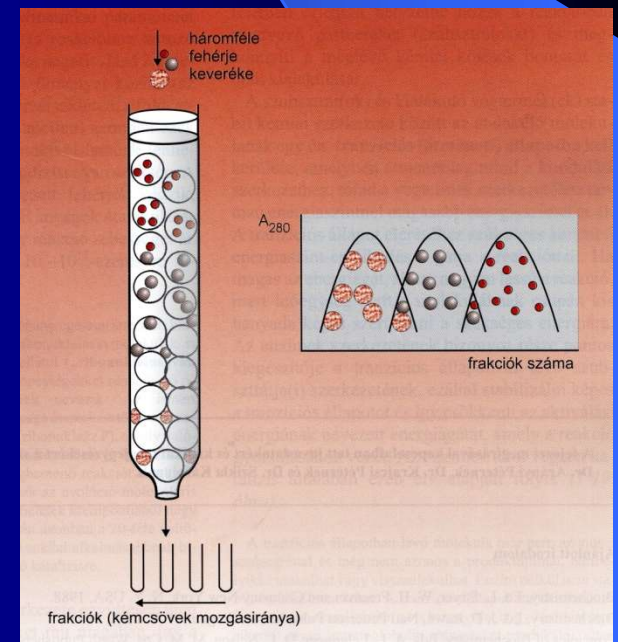


# Méret szerinti elválasztás II.

- Gélszűrési kromatográfia

- Oszlop tetejéről átfolynak az alja felé
- Kis molekula tömegűek bejutnak az üregekbe, gyorsabban leérnek
- az átfolyt rész (frakció) az alján gyűlik
- Fotométerrel vizsgálják az egyes frakciók fényelnyelését

<http://www.youtube.com/watch?v=P8Mmpb4OShw&feature=related>



# Oldékonyság

- Különböző oldékonyság jellemzi az egyes fehérjéket
  - Vizes fázisban
  - Poláros/apoláros oldószerekben
  - Detergensek jelenlétében
  - Különböző pH érték/ionerősségtől függően
- Elkülönítés alapja: egyes oldószerek oldják, mások nem...

# Elválasztás IV.

- Töltés alapján
  - Elektroforézis
  - Elektroforézis pH-grádiensben
- Ligand (kötődési) affinitás
  - Sok fehérje nagy affinitással kötődik bizonyos kémiai csoportokhoz (szintetikus vegyületek, enzimgátlók, szubsztrátok)

## Aminosavak menyiségi meghatározása

- Izolálni kell a cél fehérjét
- Savas hidrolízis után (6 NHCL, 110°C)
- Az aminosavak ioncserélő kromatográfiával egymástól elválaszthatók és ninhidrinnel, vagy fluoreszkaminnal mennyiségileg meghatározhatók
- Tehát itt a sorrendet nem kapjuk meg...



# Aminosavsorrend meghatározás

Az N-terminális aminosavát reagáltatni kell, egy olyan vegyülettel, mely az aminosavval együtt hidrolízissel eltávolítható a láncból

A leválasztott aminosav nagy nyomású folyadékkromatográfiával (HPLC) azonosítható

A következő ciklusban egy újabb aminosavat választanak le (until end...)

A ciklus ismételhető automatikus szekvenátorral

Meghatározható DNS-ből is, de (bár DNS-t meghatározni gyorsabb és pontosabb), a poszttranszlációs módosulások miatt kevésbé használható

# További metodikák

- Fehérjék detektálása és kvantitatív meghatározása
  - Antitestek segítségével
- Fehérjék lokalizációja
  - Immuno-fluoreszcencia
  - Immuno-elektronmikroszkóp (jelzett antitesttel)
- Funkció kutatás
  - Radioaktív izotóp jelzési technikák (a tirozint és a hisztidint  $I^{125}/I^{131}$  jódiszotópokkal jelzik) A radioaktivitás mértékéből és lokalizációjából következtetni lehet a funkcióra.
- Háromdimenziós fehérje szerkezet
  - Röntgen krisztallográfia
  - Mágneses magrezonancia (NMR)

# Enzimek I.

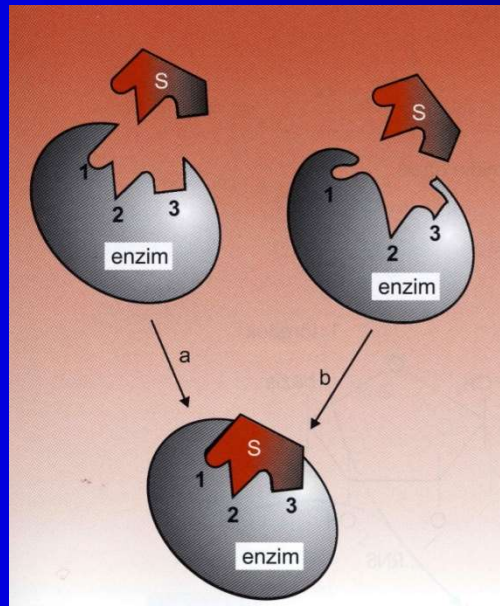
- A reakciók szabályozott lefolyását segítik elő (pl. gyorsítják akár  $10^9$ - $10^{12}$  x!)
- Szubsztrátok = reakció résztvevői
  - Az átalakulás ideje alatt tranzíciós (átmeneti) állapotba kerülnek
  - A fehérjék stabilizálják az állapotot és irányítják a lefolyását (szubsztrát specifikus kötő helyek)
- Produktum = végtermék
- A reakció tényleges irányát a kiindulási és végállapot közötti szabadenergia-változás határozza meg

# Enzimatikus folyamatok energiahatékonysága

|  | Aktiválási Energia (kJ/mol) |                    |
|--|-----------------------------|--------------------|
|  | Enzim nélkül                | Enzim jelenlétében |
| $\text{H}_2\text{O}_2 \rightarrow \text{H}_2\text{O} + \frac{1}{2} \text{O}_2$ | 75,6                        | 8,4                |
| $\text{Urea} \rightarrow \text{NH}_3 + \text{CO}_2$                            | 103,3                       | 52,1               |
| $\text{Szaharóz} \rightarrow \text{Glu} + \text{Fru}$                          | 109,2                       | 48,3               |
| $\text{Kazein} \rightarrow \text{peptidek}$                                    | 86,5                        | 50,4               |

# Szubsztrát-enzim kötődési modellek

- Kulcs zár modell (régi)
  - A szubsztrát és az enzim pontos kiegészítői egymásnak
- Indukált illesztés
  - A szubsztrát és a szubsztrát kötő hely egyes részeinek kölcsönhatása hozza létre a pontos illeszkedést



# Enzimek II.

- **Aktív hely**
  - Ezen belül található a **szubsztrát kötő hely**
  - Általában az enzimfehérjéhez képest kis terület
- **Szubsztrát kötő hely**
  - Jól konzerválódik az evolúció során
  - Kismértékű módosulása is komoly változást eredményezhet a működésben
  - A szubsztrát szerkezetének térbeli kiegészítője

# Fehérjeadatbázisok

- PDB Protein Data Bank
  - Az ismert térszerkezetek adatbázisa
- SWISS-PROT
  - Az ismert fehérjeszekvenciák adatbázisa (elég pontos)
- Egyéb adatbázisok
  - Fehérjecsaládok, domének, negyedleges szerkezetek, fehérjék génjei, szekvencia-összerendezések, szerkezetek minőségellenőrzéseinek eredményei, stb.



# Néhány információ a hemoglobinnról I.

- Aerob illetve anaerob folyamatok  
(aerob működés ~18x hatékonyabb)
- Az oxigén oldékonysága vízben kicsi
  - $O_2$  tartalom normál (Hb mentes) 5 ml/l és ~250 ml/l Hb-tartalmú vér esetén
- A megfelelő oxigénkoncentrációt speciális oxigén kötő és oxigén szállító molekulák biztosítják
  - Vérben: Hemoglobin (Hb)
  - Izomban: Mioglobin (Mb)
- Nobel-díjat ért (Max Perutz & John Kendrew)

## Néhány információ a hemoglobinról II.

- Hemoprotein, prosztetikus csoportként hem-et tartalmaz
- Hem = vas /Fe(II)/ és protoporfirin IX
- Az oxigént O<sub>2</sub>-molekula formájában köti (1db)
- Az oxigénkötése reverzibilis (különben probléma! - CO mérgezés)
- Globin részét négy peptidlánc alkotja, melyekhez egy-egy hem tartozik
- Felnőttben: két alfa és két béta lánc alkotja, mindkettőben több alfa-helix szakasz található

# Feladat I.

- Hemoglobin alfa lánc megkeresése
  - Ember, egér, csimpánz
  - Mentsük le FASTA-formátumba!
- Hasonlítsuk össze őket „kézzel”!
- Hol és milyen különbségek vannak?
- Keresési helyek:

<http://www.expasy.org/>

- SWISS-2DPAGE (egér és ember)
- SWISS-PROT/TrEMBL
- Uniprot: <https://www.uniprot.org>

Info: HBA\_MOUSE, HBA\_PANPA, HBA\_PANTR,  
HBA\_HUMAN

- Hemoglobin lánc megtekintése

# Kérdések

- Mekkora az egyes szekvenciák hossza?
- Van hosszbeli különbség?
- Hol találhatók aminosav különbségek és milyenek?