

# Információtárolás az élőlényekben

Kozlovszky Miklós

Kozlovszky.miklos@nik.uni-obuda.hu

10. Előadás

# Sejt

## Ismétlés I.

Sejt: legkisebb élő egység

- Prokarióta (pl.: baktériumok, algák)
  - Törzsfejlődésileg ősibb „egyszerűbb”, nincsenek belső membránjaik, stb., (de sejtmembránjuk van)
  - DNS → RNS → fehérje szintézis, szinte egy helyen történik
- Eukarióta
  - „Fejlettebb” (gombáktól felfelé), Egysejtes/többsejtes élőlények
  - Körülhatárolt valódi sejtmag
  - „kompartmentáció” → munkamegosztás és ezáltal akár párhuzamos működés
  - A sejtmag burkolat: segít a szállításban, információt és anyagokat szűr, stabil/védett környezetet biztosít
  - DNS → RNS → transzport a membránon keresztül → fehérje szintézis

# Sejtalkotók, sejtciklus

## Ismétlés II.

- Eukarióta sejtalkotók (fontosabbak)

- Sejtmag, ER, Golgi-apparátus, mitokondrium, lizoszóma, magvacska, sejthártya, sejtbürok, (csillók)

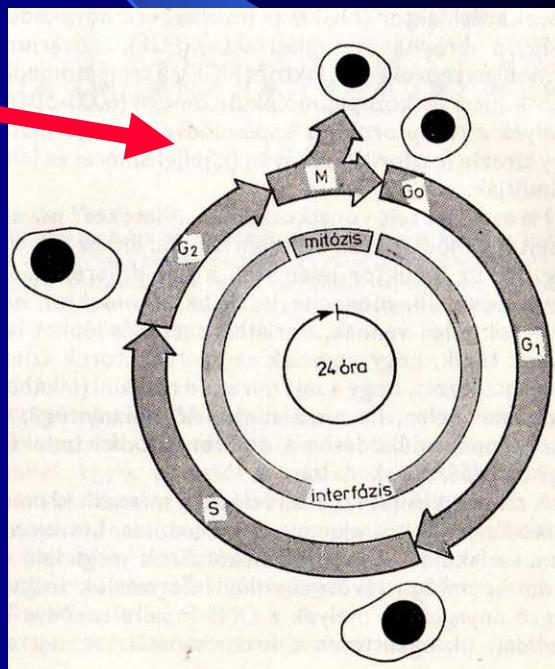
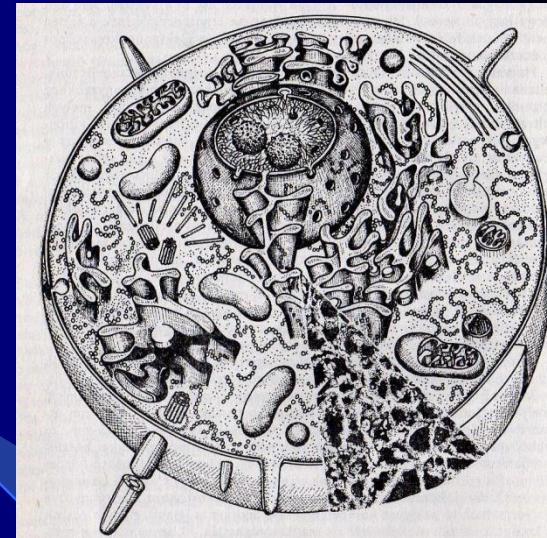
- Sejtciklus

- Sejtosztódás

- Túlélés alapja  $\leftrightarrow$  információ mentés
- Amitózis
  - Egysejtűeknél
- Mitózis
  - Testi sejtek, számtartó ( $2n \rightarrow 2*2n$ )
    - G0/1: működési fehérjék szintetizálódnak, működik
    - S: DNS + hisztonok + replikációs fehérjék szint.
    - G2: felkészülés az osztódásra  $\rightarrow$  M (Osztódik...)

Stimuláló kémiai jelek pl.: Növekedési faktorok

- Meiózis
  - Ivarsejtek, számfelező  $2n \rightarrow (2*2n) \rightarrow 4*(1n)$
  - (Megtermékenyítéskor gen. információ keveredik:  $1n(\text{♀}) + 1n(\text{♂}) \rightarrow 2n$ )



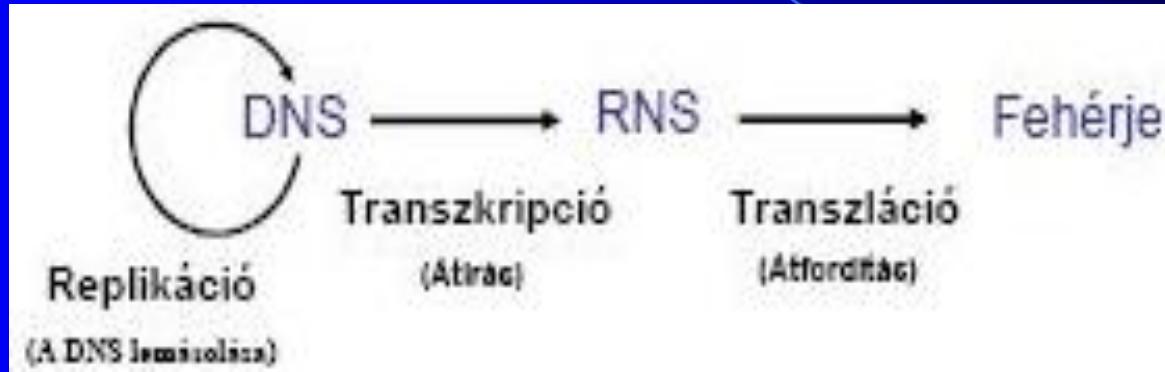
# Az információ rendezettsége

- Sejtmagban lévő bázikus anyagot (DNA-t) régóta ismerik (máshogy festődött)
- Aktívan működő sejtek sejtmagja világosabb
- Interfázisban (nyugalmi állapotban) :
  - Heterokromatin - sötéten festődik, tömött szerkezetű, nem íródik át
    - Nem megjelenő információk: konstitutív heterokromatin
    - Rendkívüli esetben előhívható információk: fakultatív heterokromatin
  - Eukromatin (igazi) – világosabban festődik. aktív kromatin, laza szerkezetű
    - Az átírásra hozzáférhető információt tartalmazza
- Osztódáskor a kromatin összecsomagolódik (kromoszómákba) → átkerül az utódokba → ismét felnyílik
- Elektronmikroszkópos vizsgálatok ezt a képet pontosították

# Extranukleáris DNS (Eukariótákban)

- Prokariótákra jellemző cirkuláris forma
- Mitokondrium
  - Teljes szekvencia ismert
  - Anyai úton öröklődik
    - Éva hipotézis (200.000 év, Kelet Afrika)
  - Sejtciklustól függetlenül replikálódik
  - Saját fehérjéket kódol: saját DNS + nukleáris DNS (önálló fehérjeszintézis, saját rRNS és tRNS-el)
- Növények kloroplasztisz-a

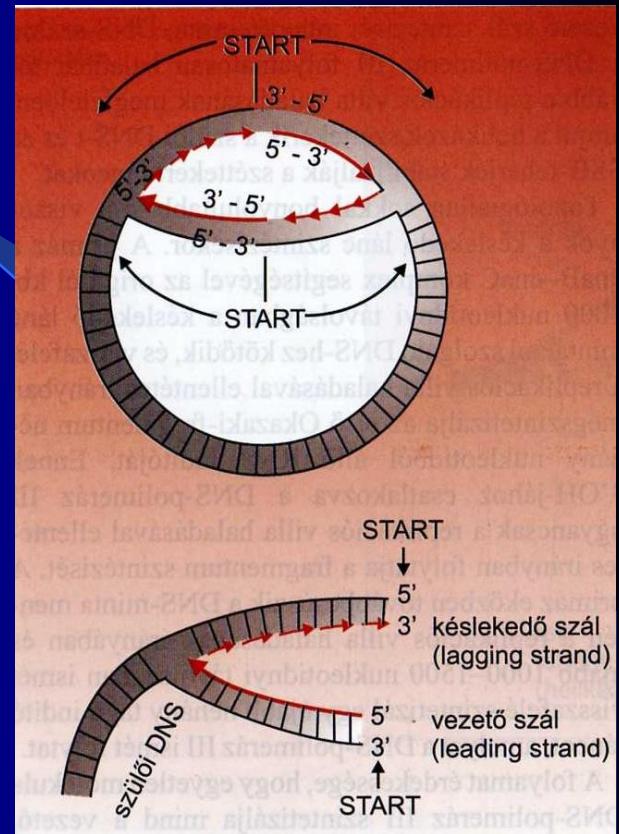
# A biológiai információ áramlás



- Három folyamat:
  - Replikáció – nem vesz részt a fehérje szintézis folyamatában
  - Transzkripció /átírás/
  - Transzláció /fordítás nukleinsav sorrend → aminosav sorrend

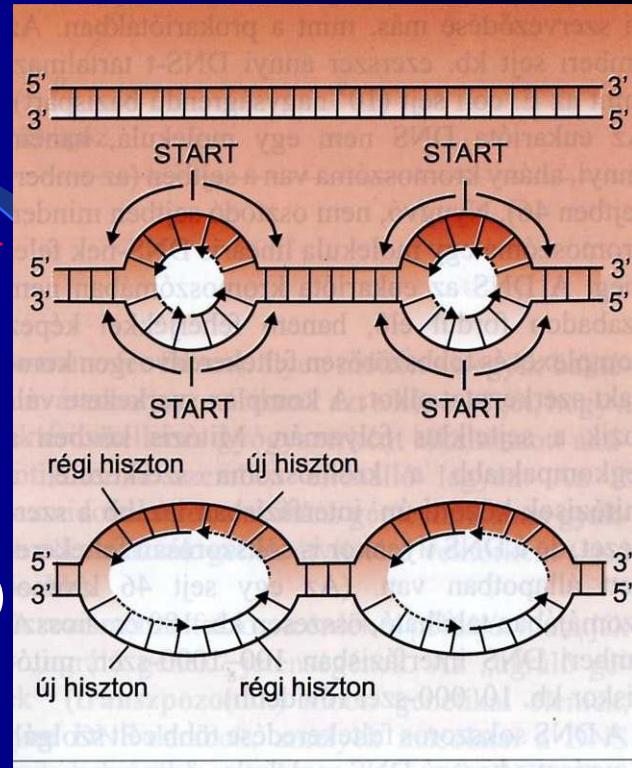
# Replikáció (DNS→DNS)

- Hibridizáció jelensége
  - Állandóan szeretnek összekapcsolódni a komplementer párok
- Hibák esetén: repair
- Szemikonzervatív replikáció
  - Az egyik mindig eredeti szál (template)
  - $3' \rightarrow 5'$  irányú olvasás,  $5' \rightarrow 3'$  irányú szintézis
  - A template és az új szál antiparallel lefutású
- Indítólánc (Primer) szükséges
  - Első néhány nukleotid szintetizálásához
  - Szabad 3'OH vég
- Okazaki fragmentumok
- Prokariótáknál
  - A genom zárt, cirkuláris a DNS, startpont meghatározott helyen
  - Replikációs villa (Y)
  - Leading strand: folyamatos
  - Lagging strand: 1000-1500 nukleotidnyi ugrások
  - DNS ligáz: összekötés (energiaigényes) (NAD<sup>+</sup>)



# DNS replikáció (DNS→DNS)

- Eukariótáknál
  - Bonyolultabb, nem cirkuláris és hosszabb a DNS
  - Replikációs buborékok
- Több startpont a lánc mentén
- „nick” megszüntetés: itt is DNS ligáz (ATP)
- Telomér szekvencia (3'-as véget védi)
  - A differenciált szomatikus sejtekben a telomér szintetizáció leáll
  - A korábban megszintetizálódott telomér szekvenciák hossza csökken



# Mutáció

A sejt örököltő anyagának (DNS-nek) megváltozása, és ezáltal a sejtmagban hordozott információ megváltozása

- Típusai:
  - Pontmutáció:minimális terület változik a DNS láncban
  - Kromoszóma aberráció: a DNS egy területe letörök, eltűnik, vagy más hova visszaépül, illetve megváltozik a sorrend
  - Genommutáció: amikor a kromoszómák száma változik meg
- Előfordulhat szomatikus (testi) és ivarsejtekben egyaránt
  - Szomatikus eset csak adott egyedet érinti, ivarsejteknél az utódot

# Mutáció

- Mutáció okai

- Spontán mutáció
  - Sejten belüli endogén folyamatok (pl.:hőmozgás)
- Mutagén tényezők
  - Ionizáló sugárzás (röntgen, UV, stb.)
  - Antibiotikumok
  - Kémiai anyagok (szabad gyökök fokozott képződése)

# Repair (javító mechanizmus)

## Karbantartó (repair) mechanizmus

- Enzimek felismerik a meghibásodott helyeket és kijavítják
- Mismatch repair – metiláció ellenőrzés
- Egyéb gyakori hibák
  - Dimerizáció - egymás melletti Timinek kovalens kötést alakítanak (UV)
  - Depurinizáció - lehasad a bázis (savak, hőhatás)  $10^4$  purinbázis naponta sejtenként
  - Dezaminálódás – „idegen” bázisok keletkeznek (spontán, vegyszerek, ionizáló sugarak) C → U
- A rendszer javítókapacitása nem végtelen (hatásfok <100%)
  - Hibafajtánként eltérő érzékenységű
  - Több hiba esetén több lesz a kijavítatlan hibák száma
  - Csírasejtes hibák öröklődnek
  - Szomatikus sejtek hibái
    - Sejt életfolyamatainak károsodása (életképtelen, csökkent funkció)
    - Kontrollálatlan szaporodás
  - Apoptózis programozott sejthalál
- Ames próba: mutagén hatás kimutatása
  - Genetikailag pontosan feltérképezett hisztidin hiányos mutáns *Salmonella* baktériumtörzs
  - Csak akkor növekszik ha a táptalajban hisztidin van

# Mutagén anyagok

Mutagén - Kémiai vagy egyéb anyagok, amelyek a genetikai anyag károsodását okozhatják.

- Ivóvíz klórozása
  - A felszíni vízből mesterséges tisztítás → klórozási melléktermékek
- Festékek / hígítók, háztartási szerek
- Bizonyos szépségápolási cikkek
  - Feketelistá
    - (de szürkelistán lévő szerek is /pl.:fogkrém- Trichlosan/)
- Azbeszt
- Növényvédő szerek, rovarölő szerek
  - dichlorvos 1955 óta ismert, foszforsav-észter típusú rovarölő hatású idegméreg (szúnyogirtásnál használjuk)  
Nincs szelektív hatása, tehát minden idegrendszerrel bíró élőlényt azonos hatásmechanizmus alapján mérgez.  
Ismert vízszenyező, erős mérgező hatással bír vízi ízeltlábúakra és a házimére.
  - Sugárzások (röntgen, stb.)

# Mutagén anyagok környezetünkben

## Dohányzás

- A dohányfüst mintegy 4000 különböző vegyületet tartalmaz, mely vegyületek részben irritálják a nyálkahártyákat, részben azokon átjutva mérgező, mutagén vagy karcinogén hatást fejtenek ki.
- A 4000 kémiai összetevőből hatvan komponens ismerten rákkeltő.
- A sokféle kémiai alkotórész kimutatható a cigaretta főfüstjében és többségük a mellékfüstben is.



# Mutagén anyagok környezetünkben

## UV sugárzás

- Az ultraibolya (UV) sugárzás : 100 nm - 420 nm
- Hullámhossz alapján:
  - *UV A* (320-420 nm) sugárzás
  - *UV B* (280-320 nm) sugárzás
  - *UV C* (< 280 nm) sugárzás
- Veszélyesség alapján:  $UV\ B > UV\ A > UV\ C$
- A napfény minden két (A és B) tartományt magába foglalja. Káros hatású a túlzott napozás.
- Az UV sugárzás hatásmechanizmus:
  - a bőr alapi rétegében elnyelődik, gyulladást, felégést okoz.
  - A bőrpír elmúlását követően a pigmentáció fokozódásával jár.
  - Festékes anyajegyek irritációja és fokozott UV hatás eredményeként nő a bőr rosszindulatú festékes daganatainak (melanoma) előfordulása is.



# Mutagén anyagok környezetünkben

## Ételek

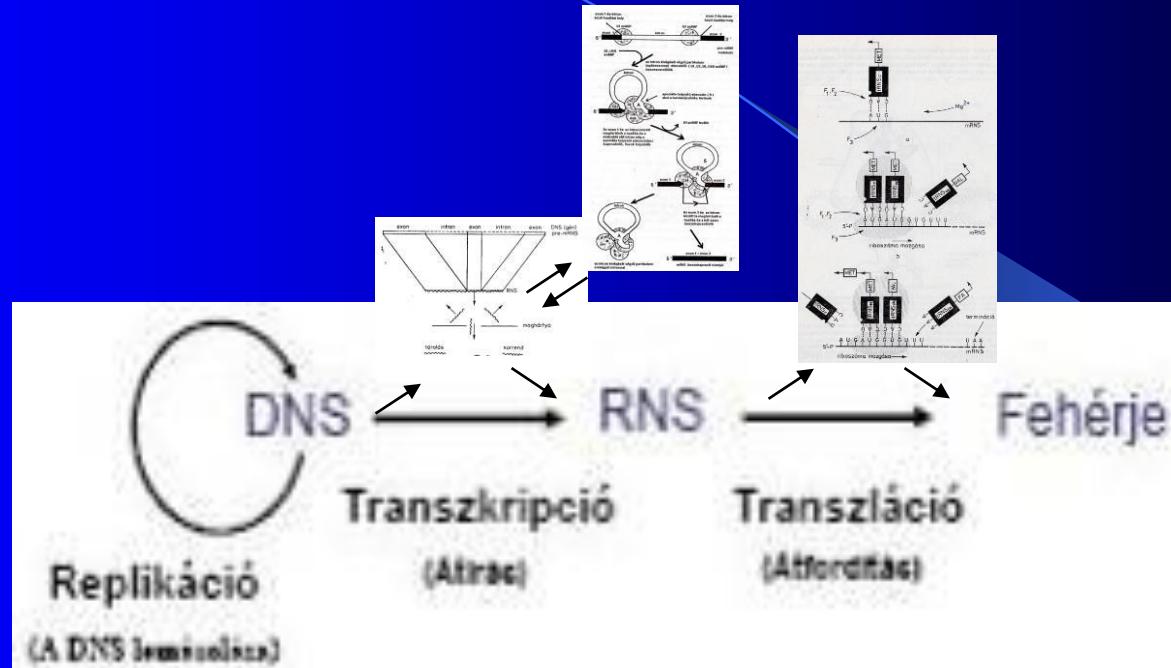
- Élvezeti cikkek és a fűszernövények mutagénjei:
  - a kávésav a kávéban (a fejes salátában is megtalálhatjuk).
  - sziningrin a mustárban (ezt a káposztafélék és a torma is termelik)
  - szafrol a borsban
  - d-limonén a narancsfélék héjában (ez utóbbi friss gyömbérben, de kandírozott narancs- és citromhéjat tartalmazó süteményekben, héjjal préselt narancslevékben is előfordul).
  - ...
- Bizonyos zöldségfélék nemesítése (mesterséges) során ezeknek a mutagén vegyületeknek a mennyisége jócskán megemelkedhet.

# Mutagén anyagok környezetünkben

## E-számok

- **E-123 - Amarant: Színezőanyag**
  - Hatás: allergén, karcinogén, mutagén. Szintetikus anyag, amely viszonylag gyakorta okoz allergiát. Az állatkísérleteknél karcinogén és mutagén hatást okozott. Az USA-ban 1976 óta tiltott.
- **E-240 - Formalin, formaldehid : Tartósítószer**
  - A formaldehid Magyarországon bejelentett veszélyes anyag, fokozottan tűzveszélyes, gáz formájában a levegővel keveredve robbanásveszélyes is lehet.
- **E-249 - Káliumnitrit: Térfogatnövelő:**
  - A nitritet a húskészítmények (töltelékáruk) színezésére használják, emellett a kolbász- és szalámifélék eltarthatóságának növelésére is. A nitrátokat (E251, E252) ezenkívül a kemény (szeletelhető) sajtok tartósításához és speciális halkészítmények érleléséhez is alkalmazzák.

# A biológiai információ áramlás

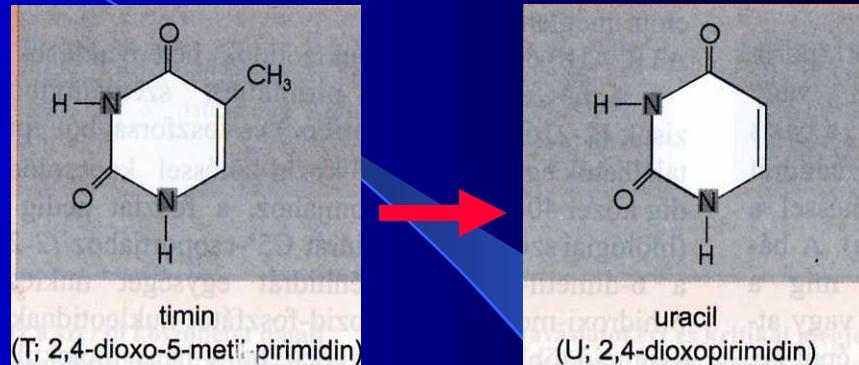


- Három folyamat:
  - Replikáció – nem vesz részt a fehérje szintézis folyamatában
  - Transzkripció /átírás/
  - Transzláció /fordítás nukleinsav sorrend → aminosav sorrend

# A ribonukleinsav (RNS)

- Eltérő felépítés:

- Hasonló a DNS-hez, mert
    - 2 purin, 2 pirimidin bázisos
    - Timin helyett Uracil



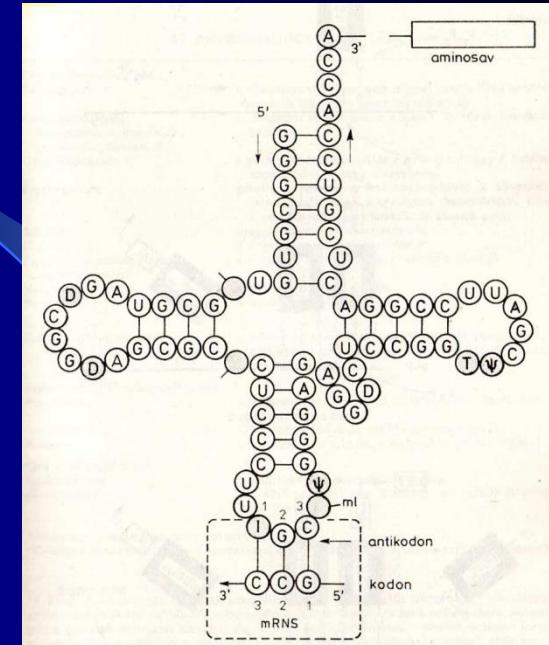
- De itt is négy típus, mint a DNS-nél

- Fajtái

- mRNS – messenger RNS (hírvivő, klasszikus)
  - tRNS – transzfer RNS (szállító, klasszikus)
  - rRNS – riboszomális RNS (riboszóma alkotó, klasszikus), legtöbb termelődik
  - hnRNS – heterogén nukleáris RNS (egy rész az mRNS átlakulási folyamatának köztes terméke)
  - snRNS – small nuclear/kis magi RNS (az átalakulási folyamatban fontos)

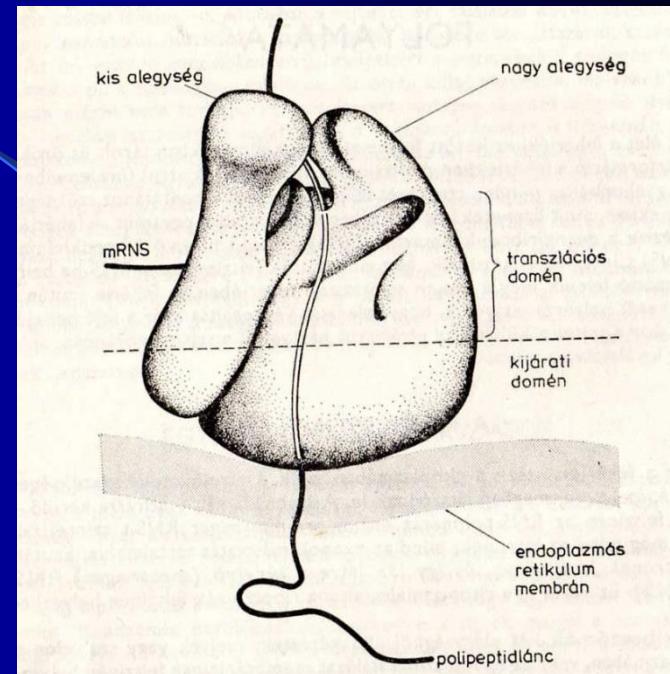
# tRNS – transzfer RNS

- Csonka lóhere alakú képlet
- Fő feladata a szállítás
- Első karján: egy antikodon (bázistriplet) van
- Második karja: a riboszómához rögzít
- Harmadik kar: aminosav aktiváló enzim felismerő hely
- Aktivált tRNS: Az aminosav kötő helyen megjelenik aminoacil-adenilát



# Riboszóma

- Két alegységből épül fel
  - Kis alegység
  - Nagy alegység
- Vagy a citoplazmában, vagy az endoplazmás hálózat membránjainak felszínén helyezkednek el
- Belsejében rRNS van, köpenyét fehérjék alkotják
- 3 lényeges pont van a fehérje köpenyen
  - A hely, P hely és R hely

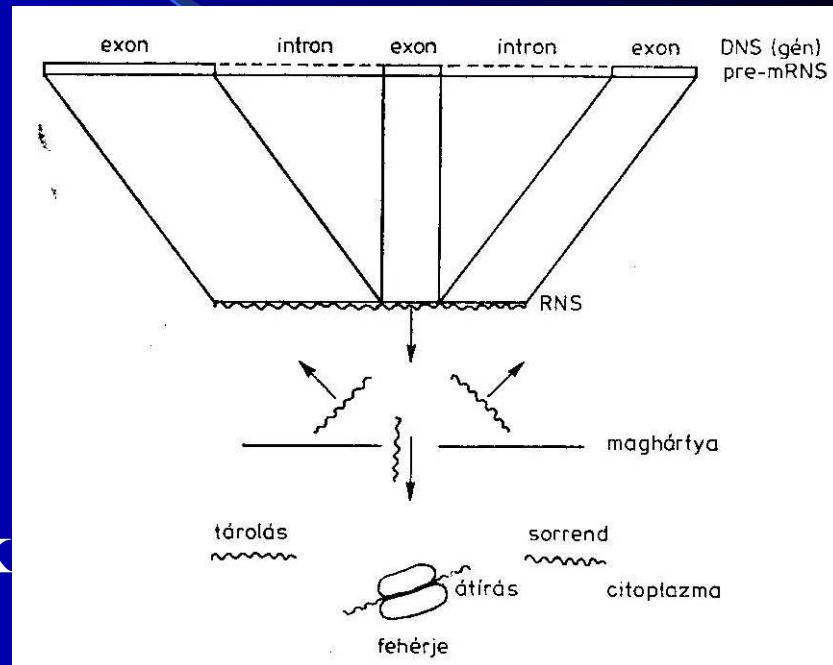


# Fehérjeszintézis

- A nukleinsavakban tárolt és örökített információ a fehérjékben realizálódik
- A fehérjeszintézis a citoplazmában zajlik
- Általában:
  - DNS → pre-messengerRNS → mRNS → fehérje
- Szótár
  - cDNS = kópia, vagy komplementer DNS, DNS kódoló rész Olyan DNS, amely az mRNS-ről, mint mintáról reverz transzkripcióval készül. Intront nem tartalmaz. Szekvenciából következtetni lehet a kódolt fehérje aminosav sorrendjére. (Jók vagyunk ha ez ismert)

# Fehérjeszintézis II.

- Exon-intron rendszer már a legprimitívebb eukariótákban is megfigyelhető
- Baktériumban minden megtermelődő mRNS átirásra kerül
- Eukariótáknál a kész mRNS-nek csak töredéke jut ki a citoplazmába (9/10 része heterogén nukleáris RNS-ként a sejtmagban marad)



# Pre-mRNS és mRNS

- Pre-messenger RNS

- RNS polimeráz enzim szintetizálja a DNS-en
  - Tartalmazza az intronokat és az exonokat egyaránt

- mRNS

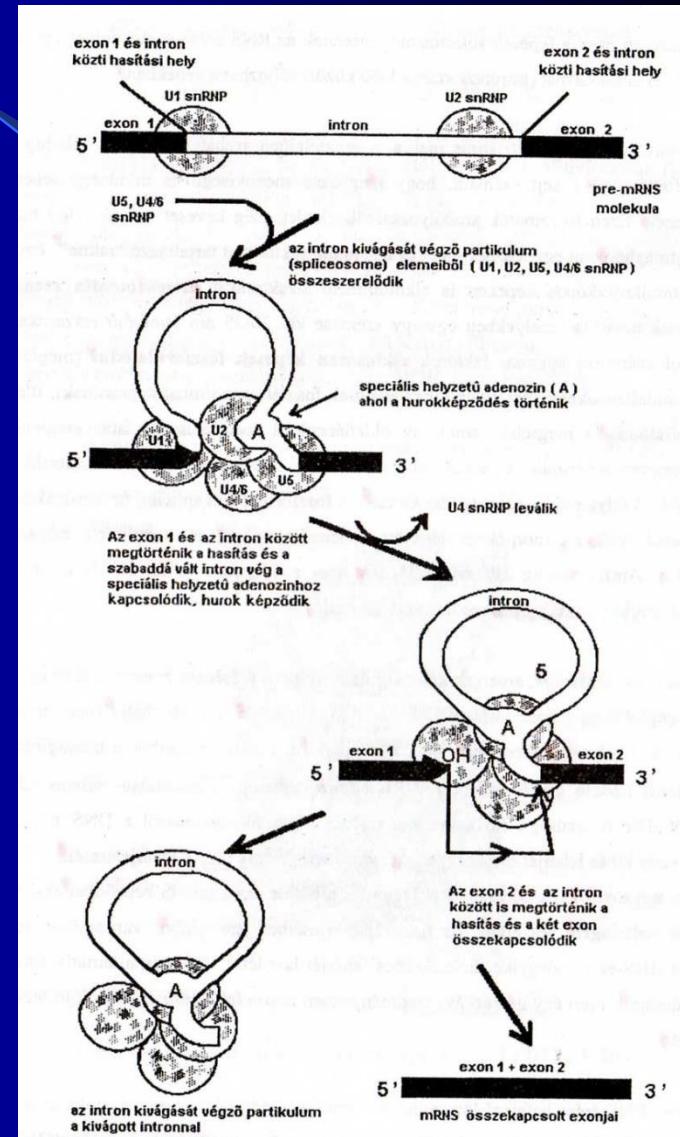
- Már nem tartalmazza az intronokat
  - Kikerül a citoplazmába
  - Eljut és a kapcsolódik a riboszómákhoz (több riboszóma dolgozik egyszerre=poliszóma)
  - mRNS átalakulása
    - 5` sapkaképződés (capping) az első 30 nukleotid elkészülése után guanin tartalmú nukleotid adódik hozzá az 5`-ös véghez
    - Poli-A farok (200-300 adenin tartalmú nukleotid) kapcsolódik (poly-A polimeráz segítségével) a 3`-as véghez (ha elkészült!)
    - Splicing – intronok eltávolítása pre-mRNS → mRNS (rövidebb)

# Splicing

- Snurps (ejtsd:sznörpsz) - snRNS tartalmú (~250 bázisnyi) makromolekula komplex
- Spliceosome (ejtsd:szplájszoszóm) – különböző snRNS tartalmú snurps (5 db) egysége szerveződik

# Splicing lépései

- A spliceosome az exon közötti intron végekhez kapcsolódik (által ezek 5` végén GU, 3` végén AG bázisok vannak)
- A két intron véget közel hozza egymáshoz
- 5` végnél hasít, majd 3`-as végén hasít, a két exonvég összekapcsolódik
- Az intron hurokformában a spliceosome-hoz kötődik

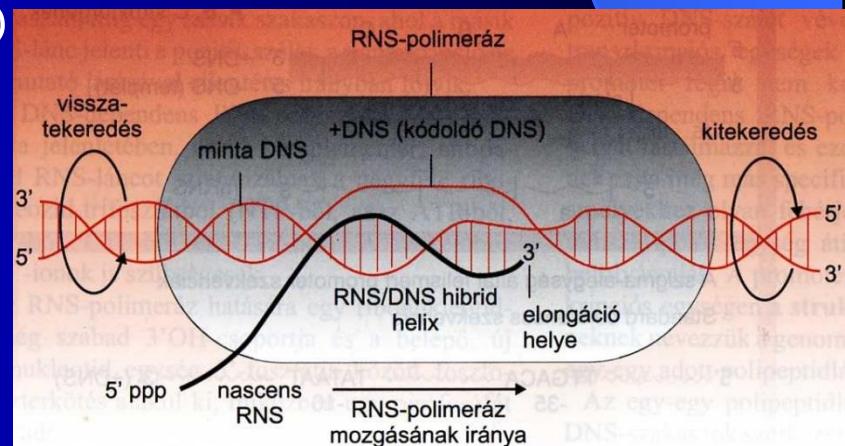


# Transzkripció – átírás (DNS→RNS)

- Az interfázisban (nem osztódó sejt) a DNS funkciója az információ átadása, és az átadási folyamat szabályozása
- Prokariótákban minden RNS-t ugyanaz az RNS polimeráz írja át
- Eukariótákban bonyolultabb a dolog
  - 3 féle polimeráz
    - RNS polimeráz I. → rRNS(nagy)
    - RNS polimeráz II. → mRNS
    - RNS polimeráz III → tRNS,rRNS(kicsi)

<http://www.youtube.com/watch?v=983lhh20rGY>

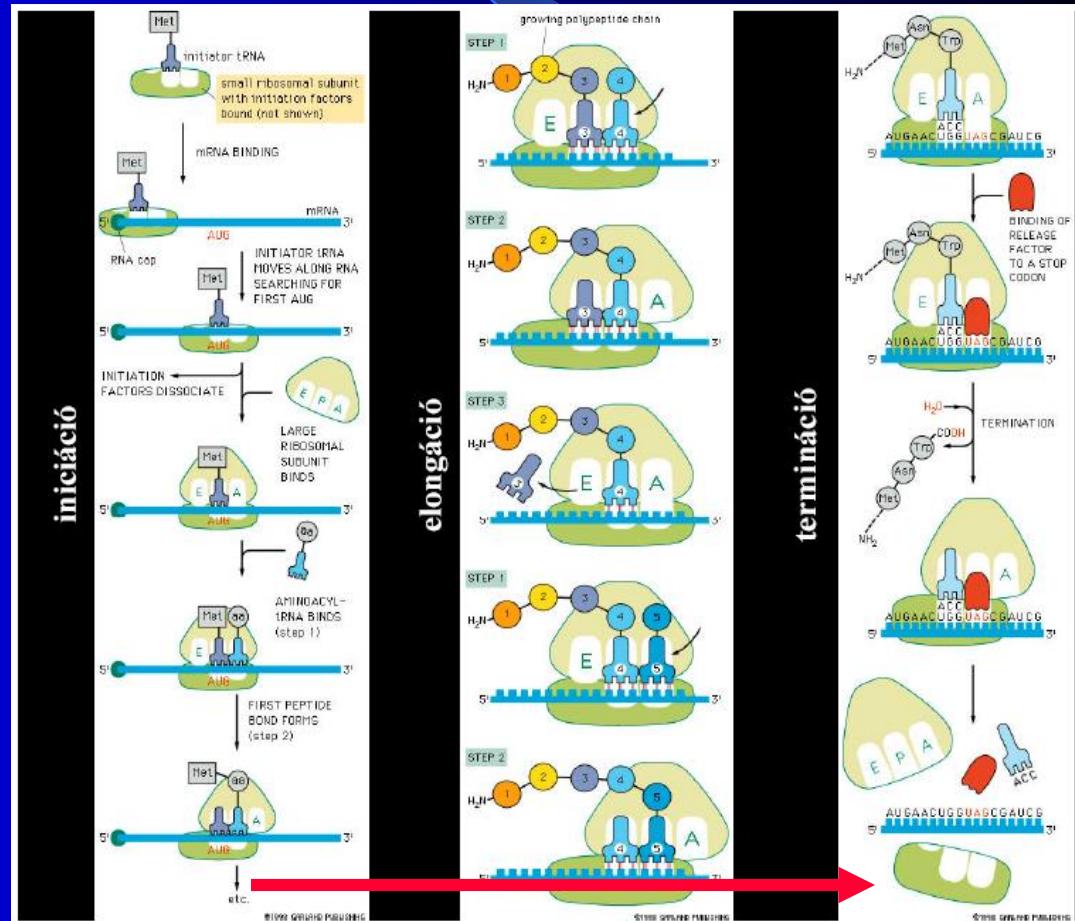
- Az elsődleges átiratok:
  - pre-mRNS
  - pre-rRNS
  - pre-tRNS mindig módosul



# A transzláció főbb lépései

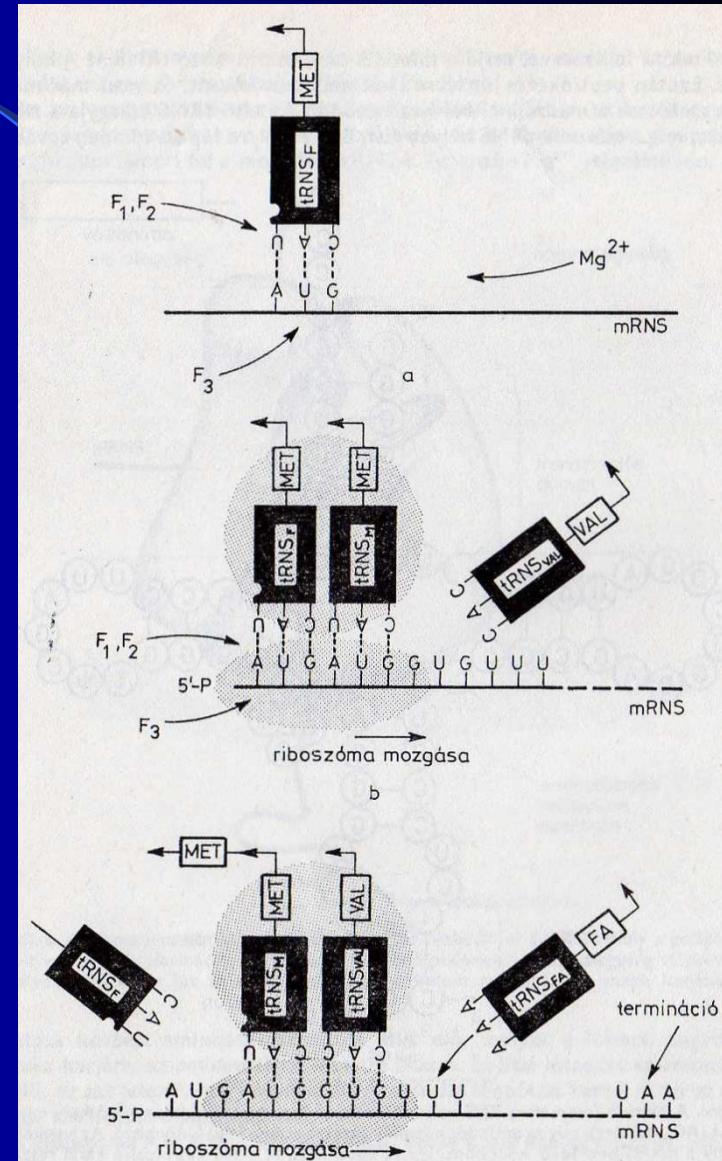
## (nukleinsav → aminosav)

- Iniciáció = Elindítás
  - aminosavak aktiválása, tRNS-hez kötése
  - mRNS-hez való kötődés – start kodon
- Elongáció = fehérjelánc hosszabbodás
  - polipeptid lánc hosszabbítás
- Termináció = befejezés
  - stop kodon
  - polipeptid lánc leszakad



# A transzláció főbb lépései (folyt)

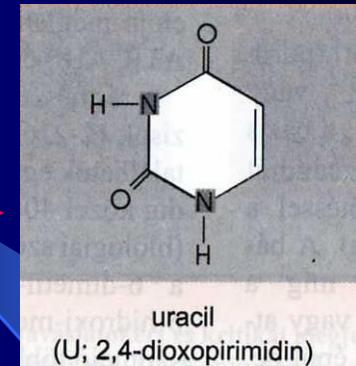
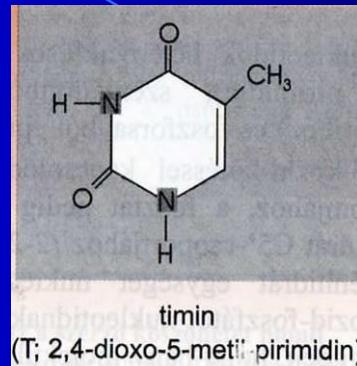
- Riboszómán P helyre ül az első tRNS, a második az A helyre → peptidkötés a két aminosav között → P-s tRNS elmege → A-ról P-re átmegy a másik tRNS → következő tRNS érkezik az A-helyre
- Eukariótákban a lánc minden MET (metionin)-al indul
- Lánc indításhoz szükségesek iniciációs faktorok ( $IF_{1,2,3}$ ) + energia
- Termináció STOP jel miatt következik be (a lánc elvágódik + a riboszómák szétesnek alegységeikre)



# A ribonukleinsav (RNS)

- Eltérő felépítés:

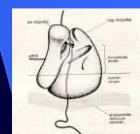
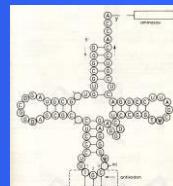
- Hasonló a DNS-hez, mert
  - 2 purin, 2 pirimidin bázisos
  - Timin helyett Uracil



- De itt is négy típus, mint a DNS-nél

- Fajtái

- mRNS – messenger RNS (hírvivő, klasszikus)
- tRNS – transzfer RNS (szállító, klasszikus)
- rRNS – riboszomális RNS (riboszóma alkotó, klasszikus), legtöbb termelődik
- hnRNS – heterogén nukleáris RNS (egy rész az mRNS átlakulási folyamatának köztes terméke)
- snRNS – small nuclear/kis magi RNS (az átalakulási folyamatban fontos)



# Szekvencia illesztés

- A bioinformatika egyik alapvető feladata DNS szekvenciák, valamint fehérje szekvenciák összehasonlítása.
- Kiindulási pont: A természet csak ritkán talál fel újat, többnyire meglévő dolgokat módosít.
- Ha ismerjük a szekvenciához hasonló szekvenciák tulajdonságait, akkor tudunk következtetni.
- Lehet: egyszeres és többszörös
  - mi most csak az egyszerest vizsgáljuk

# Szekvencia illesztés I.

- A szekvenciák illesztését régen „szemmel” végezték
  - Ma már nagyságrendekkel több az adat
  - Az illesztés automatizálható
  - Érdemes kihagyni az emberi tényezőt
- Legegyszerűbb megoldás:
  - +1 ha adott helyen egyezés van, -1 ha nincs

- HIGHLY RELATED:

HBA_HUMAN	GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKL G+ +VK+HGKKV A++++AH+D++ +++++LS+LH KL
HBB_HUMAN	GNPKVKAHGKKVLGAFSDGLAHLNDNLKGTFATLSELHCDKL

- RELATED:

HBA_HUMAN	GSAQVKGHGKKVADALTNAVAHV---D--DMPNALSALSDLHAHKL ++ +++++H+ KV + +A ++ +L L++++H+ K
LGB2_LUPLU	NNPELQAHAGKVFKLVYEAIIQLQVTGVVVTDATLKNLGSVHVSKG

- SPURIOUS ALIGNMENT:

HBA_HUMAN	GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSD---LHAHKL GS+ + G + +D L ++ H+ D+ A +AL D ++AH+
F11G11.2	GSGYLVGDSLTFVDLL--VAQHTADLLAANAALLDEFPQFKAHQE

# Fogalmak

- Szekvencia egyezés – ha a vizsgált szekvenciák minden pozíójában azonosak a betűk
- Szekvencia hasonlóság – a szekvenciák közötti eltérések valamelyen szempontból meghatározott valószínűségi szinten belül vannak (valószínűségeket kell számolni, hogy eldönthessük)
- Homológia – az evolúció során egymással rokon kapcsolatban állnak a szekvenciák (visszavezethetők hipotetikus közös ősre)  
A hasonló szekvenciák, nem feltétlenül homológok!
- Általános jelölések algoritmusoknál:
  - Egyezés = :
  - Kémiai sajátosságok egyeznek = .
  - Rés = -

# Manuális illesztés

- Illesztett párok:  
illesztés után az egymás alá írt szimbólumok.
- Optimális illesztés:  
Az az illesztés, melyhez a hozzárendelt mutáció sorozat súlya minimális.

1. szekvencia:	AGGVLI <span style="color:red">I</span> QVG 
2. szekvencia:	AGGV <span style="color:red">L</span> I <span style="color:red">I</span> QVG 
1. szekvencia:	AGGV <span style="color:red">L</span> I <span style="color:red">I</span> QVG 
2. szekvencia:	AGGV <span style="color:red">L</span> -I <span style="color:red">Q</span> VG     -

# Manuális szekvencia illesztés II.

	M	T	F	R	D	L	L	S	V	S	F	E	G	P	R	P	D	S	S	A	G	G
M	X																					
T		X																				
F			X									X										
R				X																X		
D					X																X	
L						X	X															
L						X	X															
S								X	X										X	X		
V									X													
S									X	X									X	X		
F		X									X											
E												X										
G													X								X	X
P														X	X							
R					X															X		
P																X	X					
O																						
S								X	X										X	X		
S								X	X										X	X		
A																					X	
G																	X				X	X
G																	X				X	X

# Felmerülő kérdések

- Mivel és hogyan büntessek, ha réseket/lukakat kell beraknom az illesztéshez?
- Melyik a legoptimálisabb algoritmus amivel illesztést végezhetek (erőforrásigények:idő, kapacitás, memória)?
- A kapott eredmény mennyire „igaz”, milyet kapnék ha egy véletlen szekvenciával hasonlítanék össze?

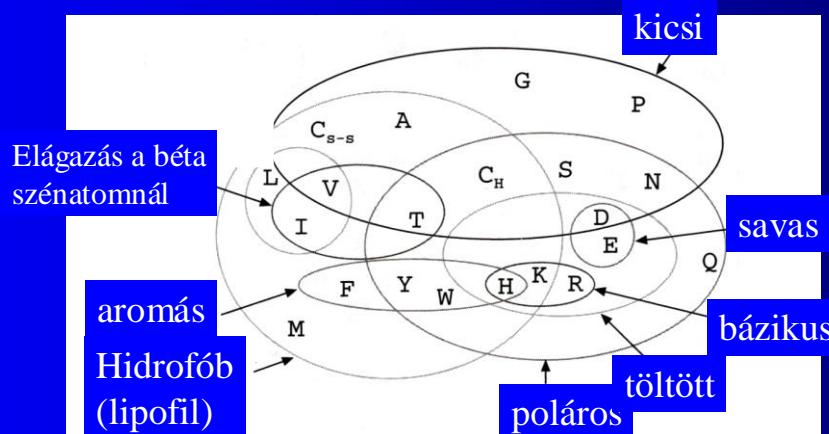
# Szekvencia illesztés (folyt.)

- Összerendeléskor nem mindegy mi mire cserélődik (eltérő aminosavak azonos tulajdonságok) a hasonlóságot próbáljuk meg pontozni!
  - Ha szigorúan csak azonosságot engedélyezünk:  
egységmátrix (ritka mátrix)
  - Engedékenység bevezetése:
    - Romlik a jel-zaj viszony!
    - Nagyobb a biológiai relevancia
    - Mátrix készítés -> külön kutatási ág:
      - fizikai tulajdonságok, mutációs gyakoriság, stb.
      - Mutációs gyakoriság alapján definiált mátrixok

	A	C	G	T
A	1	0	0	0
C	0	1	0	0
G	0	0	1	0
T	0	0	0	1

# Mutációk súlyozása

- Tegyük fel, hogy a mutációk egymástól függetlenek:
  - Egy mutáció sorozat az egyes mutációk valószínűségének szorzata
  - A különböző mutációk nem azonos valószínűsggel történnek. Ezeket súlyokkal megadhatjuk (nagyobb valószínűségű kisebbet, kisebb valószínűségű nagyobb súlyt kapnak)
    - Valószínűség logaritmusának mínusz egyszerese.
  - A mutáció súlya, az egyes mutációk súlyainak összege.
  - Legyen a mutáció és megfordítottja azonos valószínűségű.



# Percent Accepted Mutation Matrix PAM mátrixok

- Dayhoff –1970
- Szoros rokoni kapcsolatban álló szekvenciák illesztéséből állították elő őket
- A rövid evolúciós idő alatt történő változások képezik az alapját
- Evolúciós időre vannak extrapolálva a számolások
- PAM egység: átlag 1% változás a szekvenciában

PAM250

A mutációk számát adja meg adott mennyiségű generáció után (itt: 250 generáció).

Legelterjedtebben ezt használják a PAM-ok közül.

# PAM -Mátrixok (folyt.)

- Számolása:
- Rokonsági esély mátrix (relatedness odds mátrix)

a két aminósav egymás közötti cseréjének valószínű sége PAM egységnyi időben mérve  
a két aminósav gyakoriságának szorzatával

Két szekvencia esetén ezeket pozícióinként kellene összeszorozni  
(illetve adjuk össze a logaritmusukat)->log odds mátrix

PAM 250: ~20% azonosság/~80% eltérés (1 PAM= $\sim$ 1% eltérés)

De szekvencia azonosságot csak illesztés után lehet számolni

# BLOSUM mátrixok

- Henikoff & Henikoff 1992
- Blocks- adatbázisból származik: rés nélküli illesztéssel fehérjecsaládokból szekvenciákat különítenek el csoportokba  
Típusai: BLOSUM45, BLOSUM50, BLOSUM62, stb.
- A klaszterezés során a szekvenciák csoportosítását hasonlósági értékek alapján végzik  
pl: BLOSUM62 - akkor kerül bele a csoportba a szekvencia, ha legalább 62%-a megegyezik

Minél nagyobb divergenciát engedünk meg, annál nagyobb evolúciós idővel/különbséggel dolgozunk

A csoportban lévő szekvenciákat megvizsgálják és gyakoriságot számolnak az egyes szekvencia elemekre.

- Ismert fehérjéknél ellenőrizték, és általánosságban biológiaileg helyesebb eredményt ad mint a PAM
- Más aminosav cseréket részesít előnyben mint a PAM

# BLOSUM50 Mátrix

(a kerekített értékekkel, a könnyebb számolás miatt)

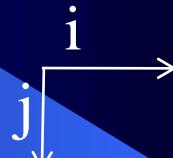
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	<b>5</b>	-2	-1	-2	-1	-1	-1	0	-2	-1	-2	-1	-1	-3	-1	1	0	-3	-2	0
R	-2	<b>7</b>	-1	-2	-4	1	0	-3	0	-4	-3	3	-2	-3	-3	-1	-1	-3	-1	-3
N	-1	-1	<b>7</b>	2	-2	0	0	0	1	-3	-4	0	-2	-4	-2	1	0	-4	-2	-3
D	-2	-2	2	<b>8</b>	-4	0	2	-1	-1	-4	-4	-1	-4	-5	-1	0	-1	-5	-3	-4
C	-1	-4	-2	-4	<b>13</b>	-3	-3	-3	-3	-2	-2	-3	-2	-2	-4	-1	-1	-5	-3	-1
Q	-1	1	0	0	-3	<b>7</b>	2	-2	1	-3	-2	2	0	-4	-1	0	-1	-1	-1	-3
E	-1	0	0	2	-3	2	<b>6</b>	-3	0	-4	-3	1	-2	-3	-1	-1	-1	-3	-2	-3
G	0	-3	0	-1	-3	-2	-3	<b>8</b>	-2	-4	-4	-2	-3	-4	-2	0	-2	-3	-3	-4
H	-2	0	1	-1	-3	1	0	-2	<b>10</b>	-4	-3	0	-1	-1	-2	-1	-2	-3	2	-4
I	-1	-4	-3	-4	-2	-3	-4	-4	-4	<b>5</b>	2	-3	2	0	-3	-3	-1	-3	-1	4
L	-2	-3	-4	-4	-2	-2	-3	-4	-3	2	<b>5</b>	-3	3	1	-4	-3	-1	-2	-1	1
K	-1	3	0	-1	-3	2	1	-2	0	-3	-3	<b>6</b>	-2	-4	-1	0	-1	-3	-2	-3
M	-1	-2	-2	-4	-2	0	-2	-3	-1	2	3	-2	<b>7</b>	0	-3	-2	-1	-1	0	1
F	-3	-3	-4	-5	-2	-4	-3	-4	-1	0	1	-4	0	<b>8</b>	-4	-3	-2	1	4	-1
P	-1	-3	-2	-1	-4	-1	-1	-2	-2	-3	-4	-1	-3	-4	<b>10</b>	-1	-1	-4	-3	-3
S	1	-1	1	0	-1	0	-1	0	-1	-3	-3	0	-2	-3	-1	<b>5</b>	2	-4	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	2	<b>5</b>	-3	-2	0
W	-3	-3	-4	-5	-5	-1	-3	-3	-3	-2	-3	-1	1	-4	-4	-3	<b>15</b>	2	-3	
Y	-2	-1	-2	-3	-3	-1	-2	-3	2	-1	-1	-2	0	4	-3	-2	-2	2	<b>8</b>	-1
V	0	-3	-3	-4	-1	-3	-3	-4	-4	4	1	-3	1	-1	-3	-2	0	-3	-1	<b>5</b>

Azonos reziduumok félkövérrel szedve

# Rés és résbüntetés

- A beszúrást ugyanakkora súllyal értékeljük, mint a törlést. Ezeket összefoglaló néven réseknek (angolul:gap) szokás nevezni, a hozzá tartozó súlyt pedig résbüntetésnek (angolul:gap penalty) .
- Általában egy súlyt adnak bármilyen karakter beszúrására és törlésére.
- A hosszú rések büntetése nem jó, ha a rés hosszából adódó büntetések összege. (nagyobb rész is mutálódhat egyetlen mutációval, különböző résbüntető függvényeket alkalmaznak emiatt.)

# Gyakran használt szekvencia illesztési algoritmusok

- Egyszerű illesztés dinamikus programozással
    - Needleman & Wunsch (1970) – globális illesztés algoritmus
    - Smith & Waterman (1981) – lokális illesztési algoritmus
  - Egyszerű illesztés heurisztikával
    - Blast (Altschul et al. - 1990)
    - FASTA (1985)
  - Ismétlődő szekvenciák illesztése
    - Waterman & Eggert (1987)
- Adott: (x, y szekvenciák)  
Szabályok:  
•  $F(i,j) = \max \begin{cases} F(i-1,j-1)+s(x_i,y_i), \\ F(i-1,j)-d, \\ F(i,j-1)-d \end{cases}$   
•  $F(i,0) = -id$   
•  $F(0,j) = -jd$   
Ahol d = résbüntetés (PL:d=8)  
• Ha kész az út, utána leírjuk az illesztést:  
 $x_i, y_i$  –t leírni ha az előző lépés:  $(i-1,j-1)$   
 $x_i$  és „–” –t leírni, ha  $(i-1,j)$   
„–” és  $y_i$  –t leírni, ha  $(i,j-1)$
- 

# Algoritmus erőforrásigényének jelzése

- $O$  =ordó (big-O) nagyságrend „order of x”, ami lehet:
  - számolási idő
  - felhasznált memória méret
  - stb.
- Biológiai szekvenciák és hagyományos számítógépek esetén, egy algoritmus eredménye:
- Jól számolható, ha a bonyolultság  $O(n)$
- Számítható, csak lassú lesz, ha a bonyolultság  $O(n^2)$
- Csak nagyon rövid szekvenciánál használható, ha  $O(n^3)$

# Egyszerű illesztési algoritmusok

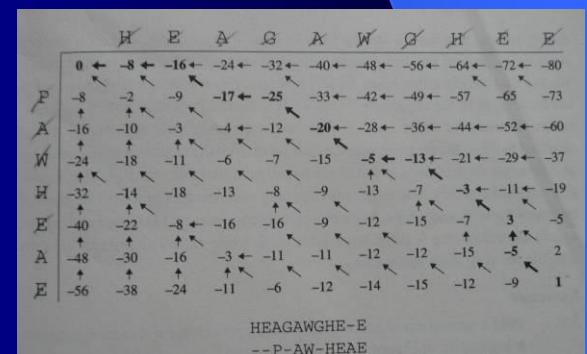
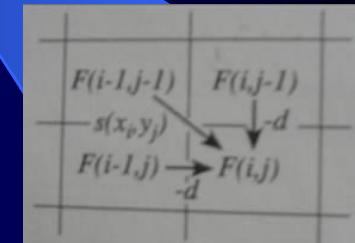
## Összefoglalás

algorithm	value calculated	scoring matrix	gap penalty	time required	
Needleman-Wunsch	global similarity	arbitrary	penalty/gap $q$	$O(n^2)$	Needleman and Wunsch, 1970
Sellers	(global) distance	unity	penalty/residue $rk$	$O(n^2)$	Sellers, 1974
Smith-Waterman	local similarity	$\hat{S}_{ij} < 0.0$	affine $q + rk$	$O(n^2)$	Smith and Waterman, 1981 Gotoh, 1982
FASTA	approx. local similarity	$\hat{S}_{ij} < 0.0$	limited gap size $q + rk$	$O(n^2)/K$	Lipman and Pearson, 1985 Pearson and Lipman, 1988
BLASTP	maximum segment score	$\hat{S}_{ij} < 0.0$	multiple segments	$O(n^2)/K$	Altshul et al., 1990

# GLOBÁLIS illesztés: Needleman – Wunsch algoritmus

- Lépései:
  - Szubsztitúciós mátrixot kiválasztani (használj pl.: BLOSUM 50-et, lásd korábban)
  - Táblázat keretét elkészíteni
    - egyik szekvenciát vízszintesen, másikat függőlegesen
    - 0. sor és oszlopot feltölteni ( $i^*d$  és  $j^*d$ -kel)
  - Kiszámolni a mátrix értékeit a szabályok alapján (nyilakat is írni, honnan jön az eredmény)
- Backtrace
  - jobb alsó sarokból indulva egészen a bal felső részig (hátról kezdve leírni az illesztett szekvenciákat a leírási szabály alapján)

## • Példa [1]:



Durbin, Eddy, Krogh; Biological sequence analysis, Cambridge, 2003

# LOKÁLIS illesztés

## Smith & Waterman algoritmus

- Új szabály: lehet 0 is.
- A 0. sor és oszlop most 0-kkal van feltöltve,
- Nem kell a jobb alsó sarokból indítani a backtrace-t (bárhonnan kezdhető)
- Példa[1]:

$$F(i, j) = \max \begin{cases} 0, \\ F(i - 1, j - 1) + s(x_i, y_j), \\ F(i - 1, j) - d, \\ F(i, j - 1) - d. \end{cases}$$

[1] Durbin, Eddy, Krogh; Biological sequence analysis, Cambridge, 2003

	H	E	A	G	A	W	G	H	E	E
P	0	0	0	0	0	0	0	0	0	0
A	0	0	0	5	0	5	0	0	0	0
W	0	0	0	0	2	0	20 ←	12 ←	4	0
H	0	10 ←	2	0	0	0	12	18	22 ←	14 ←
E	0	2	16 ←	8	0	0	4	10	18	20
E	0	0	8	21 ←	13	5	0	4	10	20
A	0	0	6	13	18	12 ←	4	0	4	16 ←
E	0	0	13	18	12	4	0	4	16	26

AWGHE  
AW-HE

# További fóliák az illesztési algoritmusokhoz

- Ismétléses illesztés
- Waterman & Eggert algoritmus
- Szabályok
  - Újabb bővítések:
- Példa[1]:

$$F(i,0) = \max \begin{cases} F(i-1,0), \\ F(i-1,j)-T, \end{cases} \quad j=1,\dots,m;$$
$$F(i,j) = \max \begin{cases} F(i,0), \\ F(i-1,j-1)+s(x_i, y_j), \\ F(i-1,j)-d, \\ F(i,j-1)-d. \end{cases}$$

[1] Durbin, Eddy,Krogh;Biological sequence analysis, Cambridge,2003

	H	E	A	G	A	W	G	H	E	E	
P	0	0	0	0	1	1	1	1	1	3	9
A	0	0	0	5	1	6	1	1	1	3	9
W	0	0	0	0	2	1	21	13	5	3	9
H	0	10	2	0	1	1	13	19	23	15	9
E	0	2	16	8	1	1	5	11	19	29	21
A	0	0	8	21	13	6	1	5	11	21	28
E	0	0	6	13	18	12	4	1	5	17	27

HEAGAWGHEE  
HEA.AW-HE.

# Féléves feladatok

- Prezentációk
  - 10 perces prezentáció (beszámolós órán elő is kell adni)
  - Doksi (5-6 oldalas összefoglaló)
- Féléves fejlesztési/programozási feladat beadásánál
  - Dokumentálás
  - Működő alkalmazás + bemutatás