

Molekulární genetik a zvířat 1

Antonín Stratil

**Česká zemědělská univerzita
v Praze**

V Prezentacích jsou doplněny stručné vysvětlující texty

Studijní literatura

- Hruban V., Majzlík I.: Obecná genetik. ČZU Praha, 2000.**
- Rosypal S.: Úvod do molekulární biologie. Díl první. Brno, 1996**
(případně pozdější vydání).
- Hruban V. a kolektiv: Principy a aplikace molekulární genetiky**
ve šlechtění. ČZU Praha, 1999 (vybrané kapitoly).
- Alberts B. a kolektiv: Základy buněčné biologie. Úvod do**
molekulární biologie buňky. Espero Publishing,
Ústí nad Labem, 1998.
- Knoll A., Vykoukalová Z.: Molekulární genetik. Metody**
detekce polymorfizmů. MZLU Brno, 2002.
- Stratil A.: Přednášky a prezentace.**

Doporučená:

- Ruvinski A., Graves J. A. M.: Mammalian Genomics. CABI Publishing,**
Oxfordshire, UK, 2005.
- Brown T. A.: Genomes. BIOS Scientific Publishers, Oxford, UK,**
1999.
- Primrose S. B., Twyman R. M.: Principles of Gene Manipulation**
and Genomics. Seventh Edition. Blackwell Publishing,
Oxford, UK, 2006.
- Lewin B.: Genes VI. Oxford University Press, Oxford, 1997.**
- Vědecké časopisy, databáze, internet.**

Úvod

Struktura nukleových kyselin

Ústřední dogma molekulární genetiky

Replikace

Gen-organizace

Transkripce

Post-transkripční úpravy

Translace

Post-translační úpravy

Typy genomů

Organizace genomu eukaryontů

Molekulární biologie

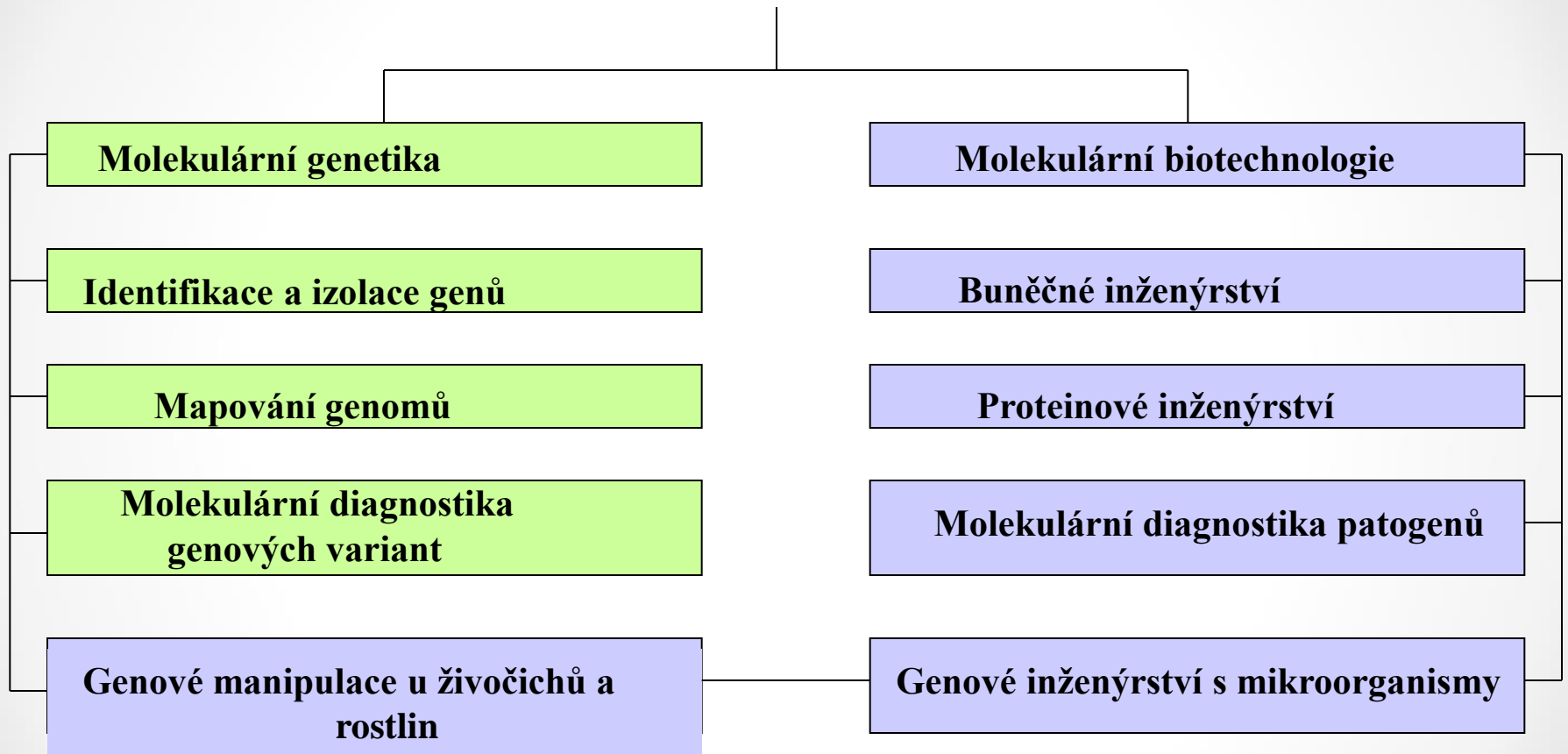
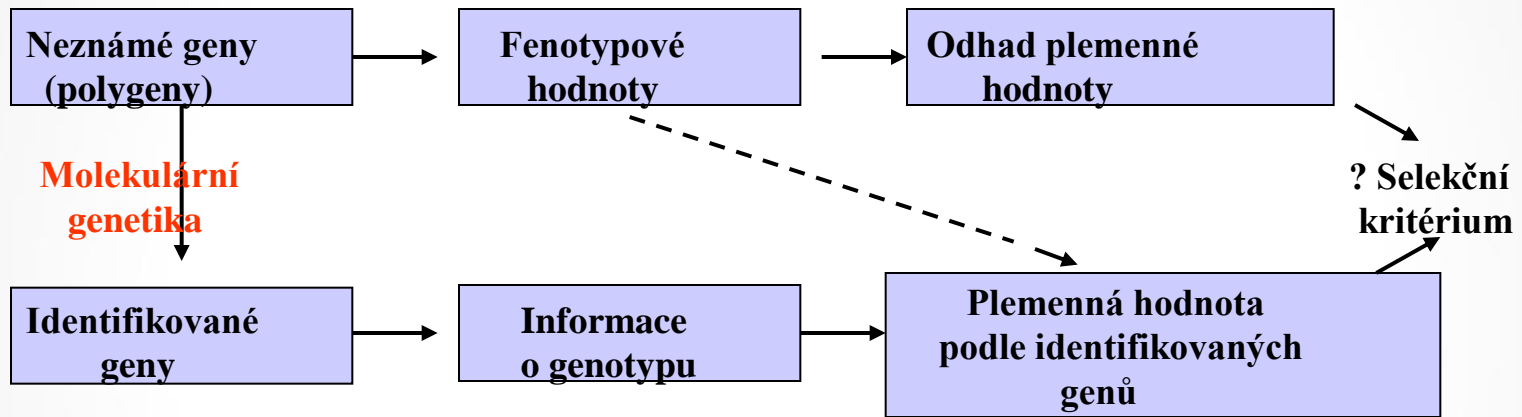
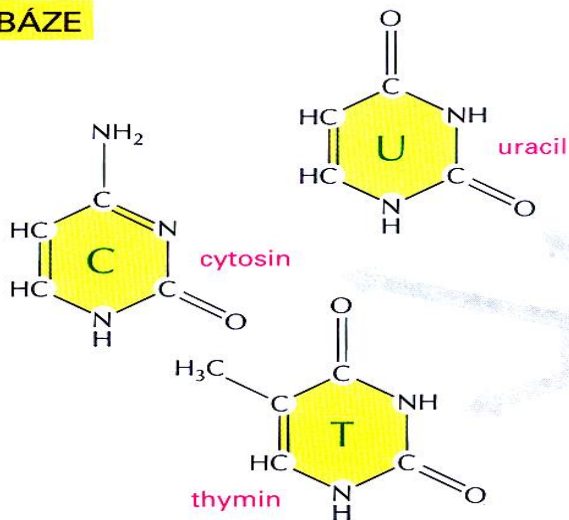


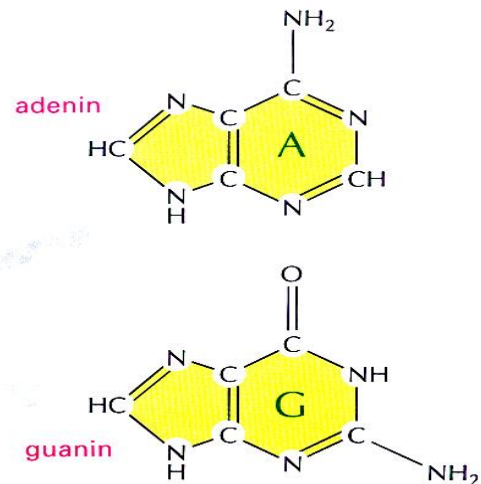
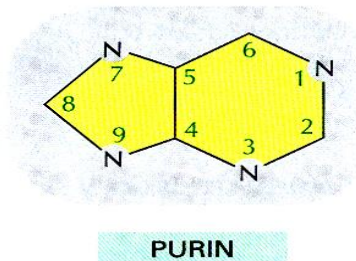
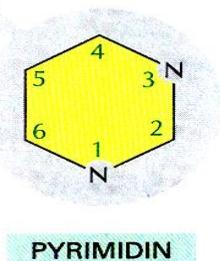
Schéma znázorňující začlenění poznatků molekulární genetiky do procesu šlechtění.



BÁZE

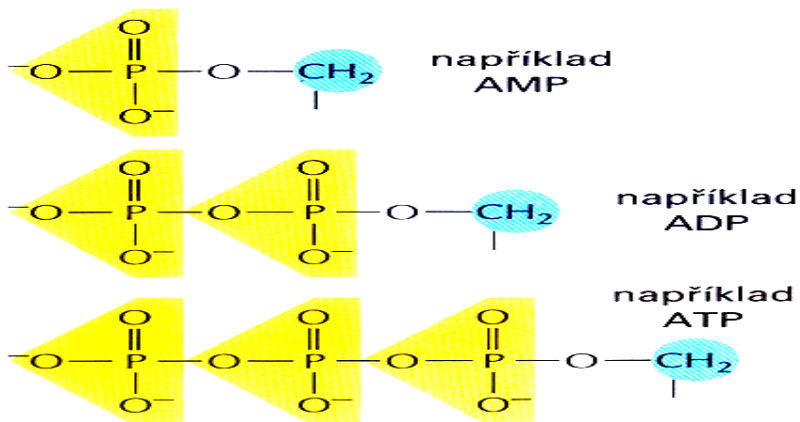


Báze jsou dusíkaté cyklické sloučeniny, buď pyrimidiny nebo puriny



FOSFÁTY

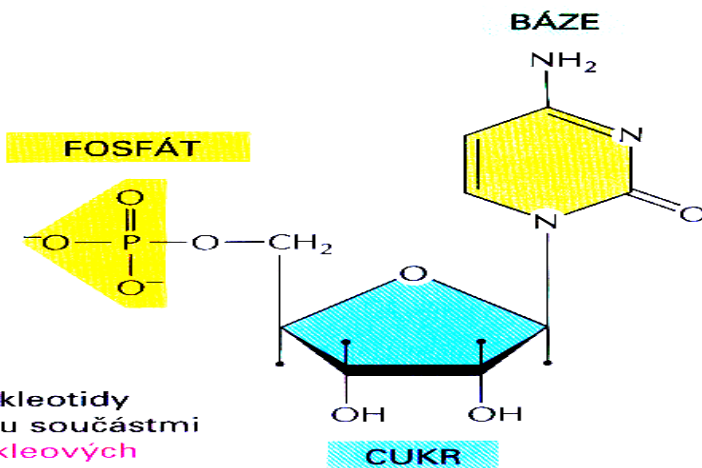
Fosfáty jsou obvykle připojeny k hydroxylové skupině na uhlíku C5 ribosy nebo deoxyribosy. Běžné jsou mono-, di- a trifosfáty.



vazba fosfátu vnaší do nukleotidu záporný náboj

NUKLEOTIDY

Nukleotid se skládá z dusíkaté báze, pětiuhlíkatého cukerného zbytku a jedné nebo více fosforylových skupin.

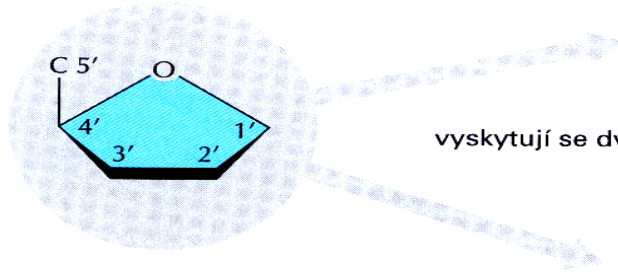


Nukleotidy jsou součástí nukleových kyselin

CUKRY

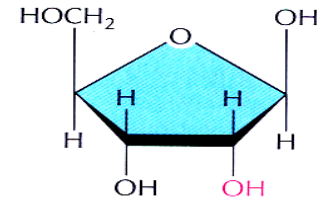
PENTOSA

pětiuhlíkatý cukr

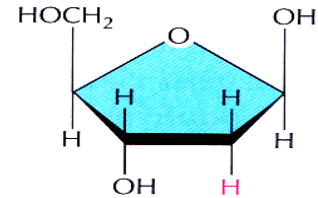


Každý číslovaný atom uhlíku v cukerném zbytku v nukleotidu se označuje znakem pro stopu ('), tedy 5' ; česky se říká "pět s čárkou".

vyskytují se dva druhy



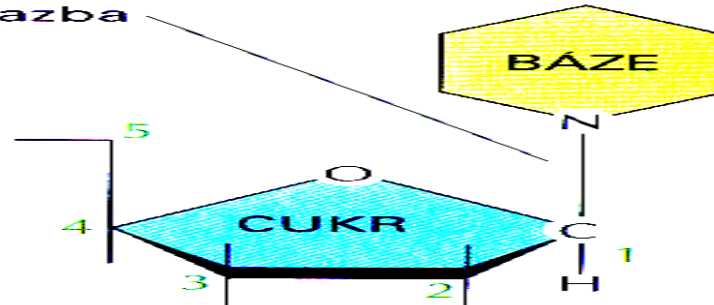
D-ribose
v ribonukleových kyselinách



2-deoxy-D-ribose
v deoxyribonukleových kyselinách

ZÁKLADNÍ VAZBY CUKRŮ

N-glykosidová
vazba



Báze je připojena k téměř uhlíkovému atomu (C1), jako při spojování jednotlivých cukerných zbytků

NÁZVOSLOVÍ

Názvy mohou být matoucí, ale zkratky jsou jasné.

BÁZE	NUKLEOSID	ZKRATKA
adenin	adenosin	A
guanin	guanosin	G
cytosin	cytidin	C
uracil	uridin	U
thymin	thymidin	T

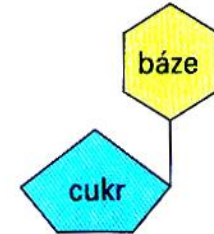
Nukleotidy se zkracují třemi velkými písmeny. Uvedme několik příkladů:

AMP = adenosinmonofosfát

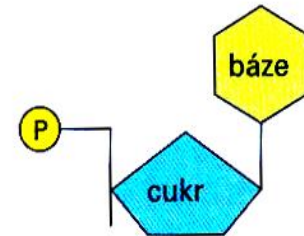
dAMP = deoxyadenosinmonofosfát

UDP = uridindifosfát

ATP = adenosintrifosfát



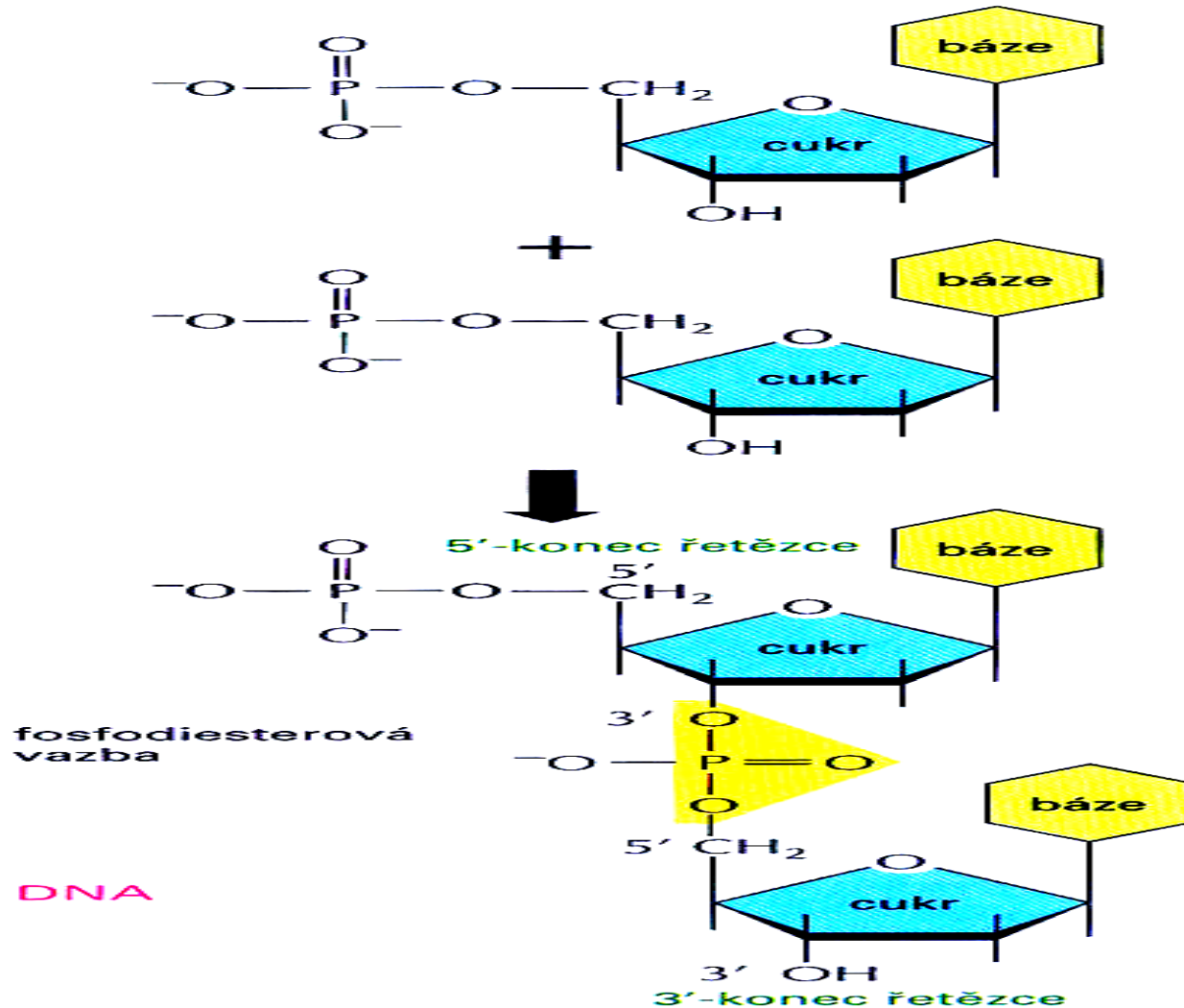
BÁZE + CUKR = NUKLEOSID



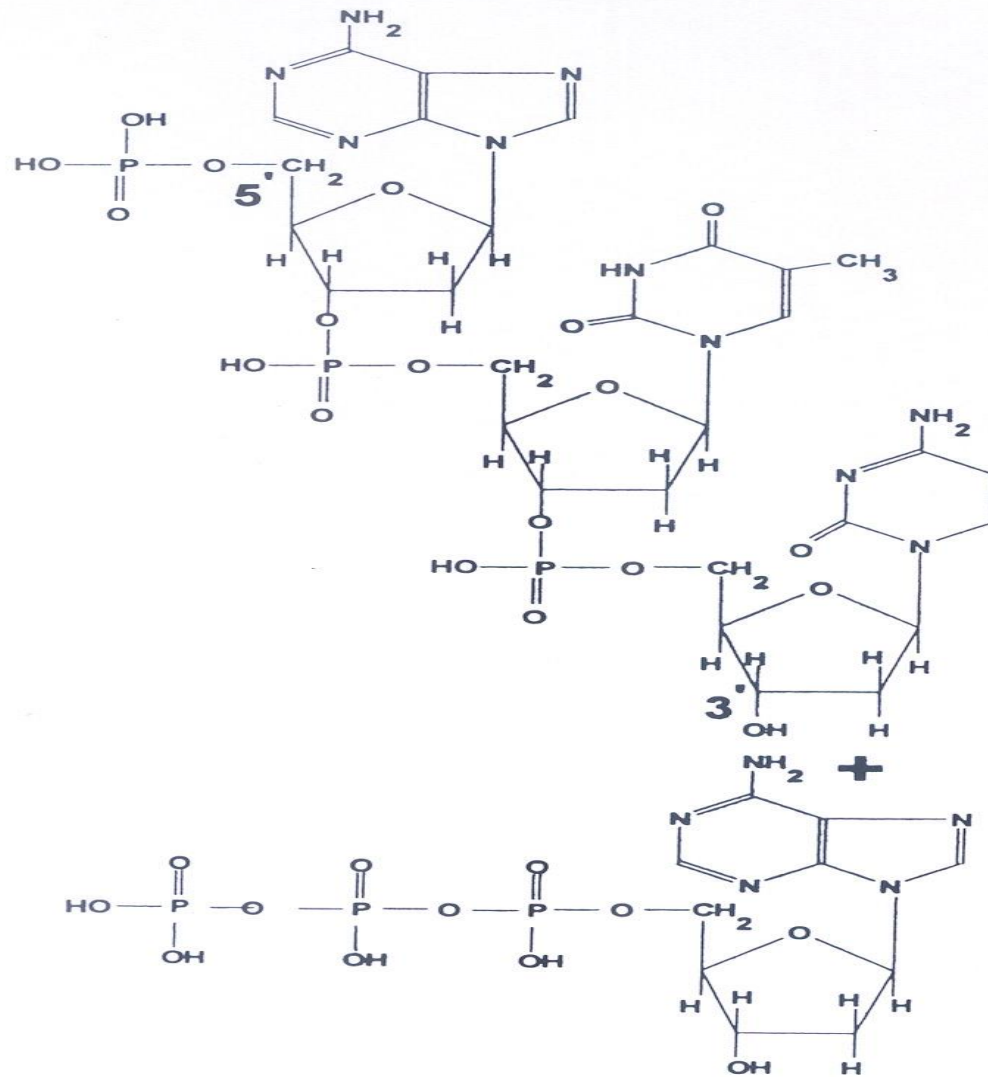
BÁZE + CUKR + FOSFÁT = NUKLEOTID

NUKLEOVÉ KYSELINY

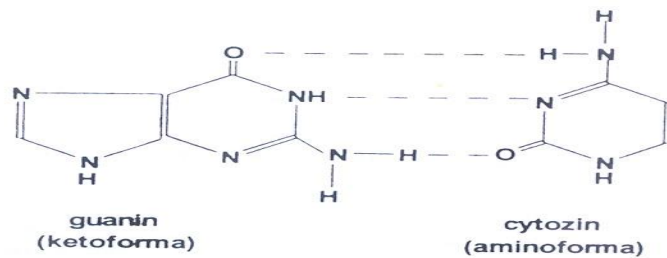
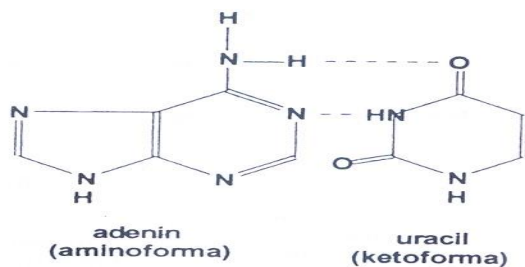
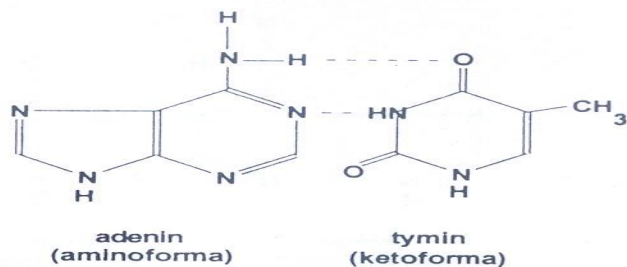
Nukleotidy jsou navzájem spojeny **fosfodiesterovou vazbou** mezi uhlíky 5' a 3', a tvoří tak nukleovou kyselinu. Lineární sled nukleotidů v řetězci nukleové kyseliny se obvykle nazývá **sekvence** a označuje se **jednopísmenným kódem**, například A-G-C-T-T-A-C-A, s 5'-koncem vlevo.



Syntéza polynukleotidového řetězce

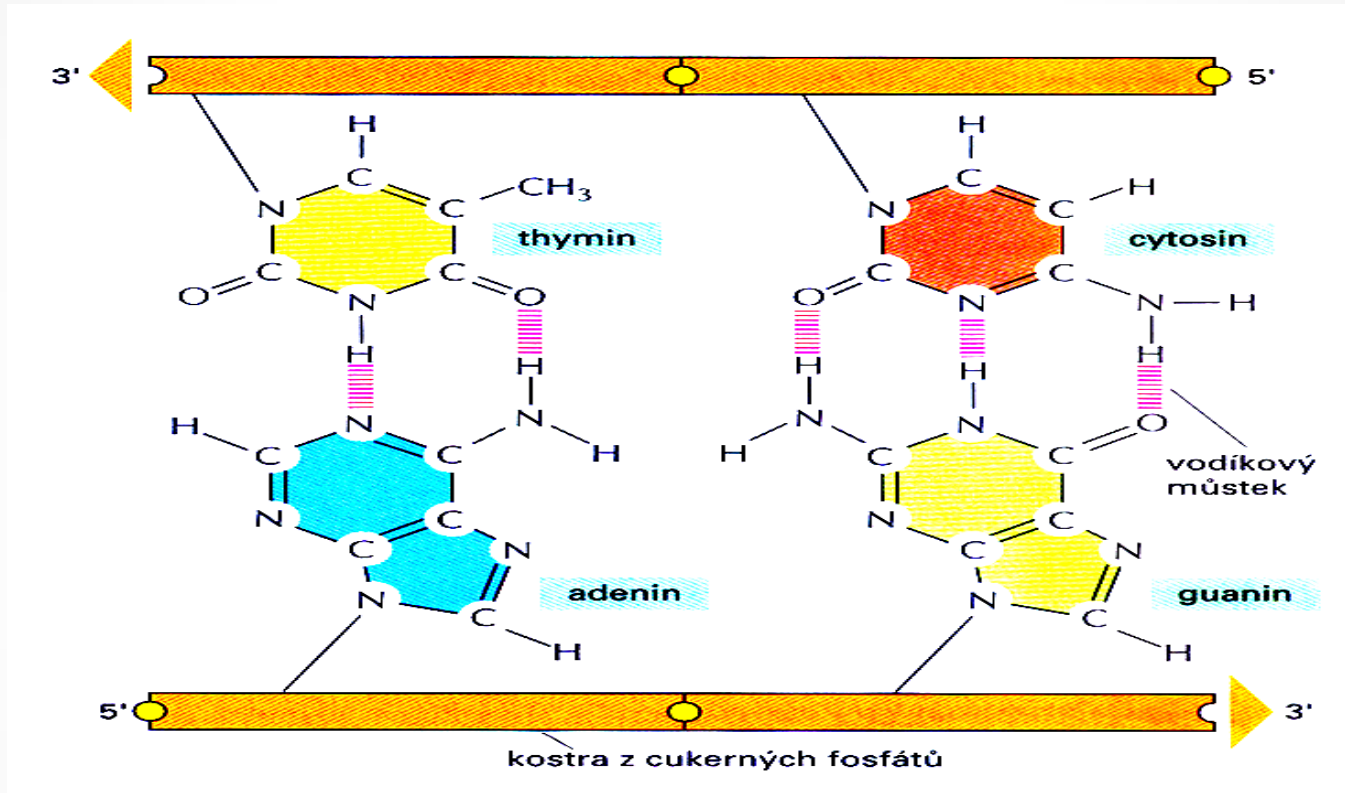


Watsonovo-Crickovo párování bází



Chargaffovo pravidlo: obsah A = obsah T
obsah G = obsah C
obsah purinů = obsahu pyrimidinů

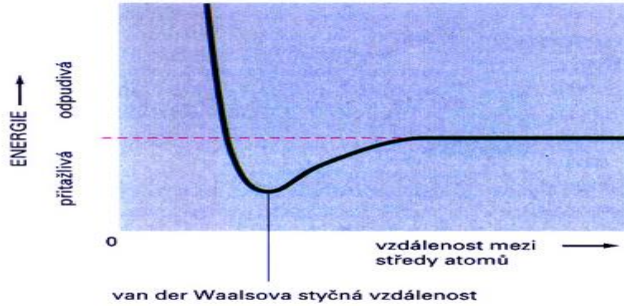
Komplementární páry bází ve dvoušroubovici DNA



Slabé chemické vazby

VAN DER WAALSOVY SÍLY

Na velmi krátké vzdálenosti vykazují kterékoli dva atomy slabou vazebnou interakci díky pohyblivým elektrickým nábojům. Pokud se však atomy dostanou do velmi těsné blízkosti, velmi silně se odpuzují.



Každý atom má svůj charakteristický „rozměr“, tzv. **van der Waalsův poloměr**; kontaktní vzdálenost mezi dvěma atomy je součtem jejich van der Waalsových poloměrů.



Dva atomy se přitahují van der Waalsovou silou až do vzdálenosti dané jejich van der Waalsovými poloměry. I když tyto interakce mohou být jednotlivě velmi slabé, **van der Waalsova přitažlivost** se může stát významnou silou, jestliže se dva povrchy makromolekuly dostanou do těsného kontaktu.

VODÍKOVÉ MŮSTKY

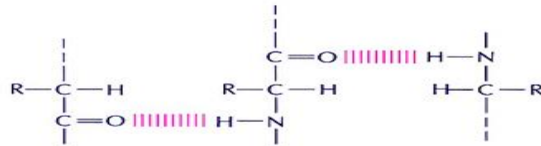
Jak už bylo popsáno v panelu 2-2, **vodíkové vazby** či **můstky** se tvoří, jestliže se atom vodíku dostane mezi dva elektronegativní atomy (obvykle kyslík nebo dusík).

elektronegativní

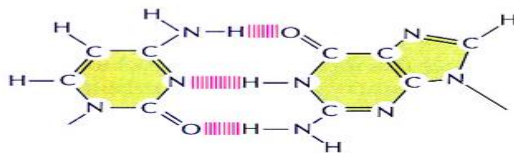
Vodíkové můstky jsou nejpevnější, jestliže jsou všechny tři atomy v jedné linii.



Příklady nacházíme v makromolekulách:
aminokyseliny v polypeptidových řetězcích



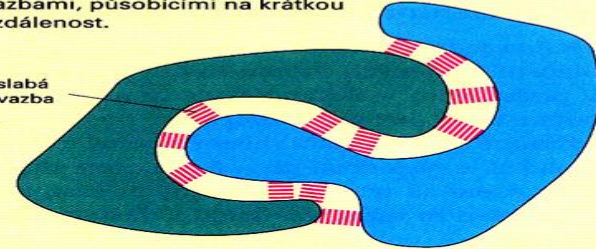
dvě báze, G a C, v DNA nebo v RNA



SLABÉ CHEMICKÉ VAZBY

Organické molekuly mohou interagovat s jinými molekulami nekovalentními vazbami, působícími na krátkou vzdálenost.

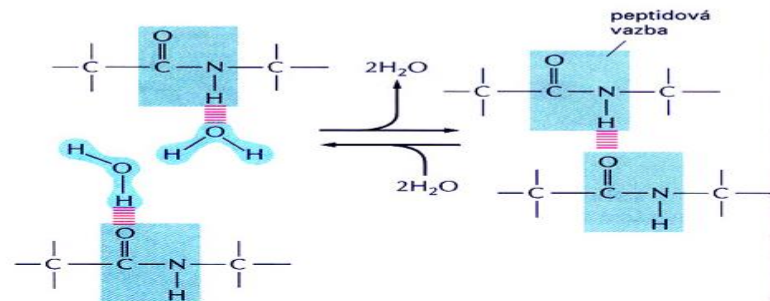
slabá vazba



Slabé chemické vazby reprezentují přibližně 5 % silné kovalentní vazby. Jsou však dostatečně silné, jestliže se jich účastní více najednou.

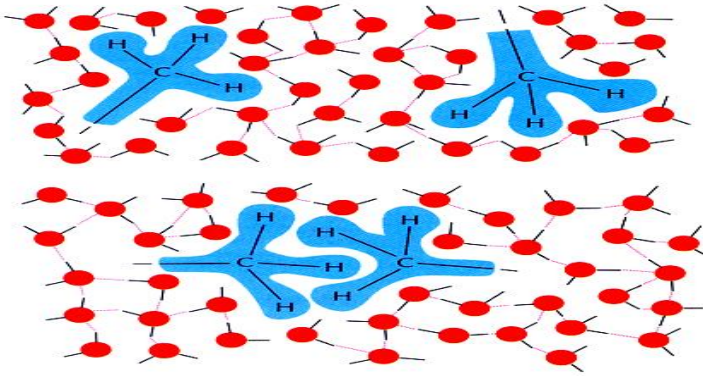
VODÍKOVÉ MŮSTKY VE VODĚ

Kterékoli molekuly, jež jsou schopny tvořit vodíkové můstky samy se sebou, to mohou činit i s molekulami vody. Díky této soutěži s molekulami vody jsou vodíkové můstky mezi dvěma molekulami rozpuštěnými ve vodě poměrně slabé.



Slabé vazby

HYDROFOBNI INTERAKKCE



Voda nutí hydrofobní skupiny, aby se navzájem spojovaly a minimalizovaly tak kontakt s vodou. Hydrofobní skupiny, které jsou tak k sobě tlačeny, jsou vlastně odpuzovány nemožností interakce s okolím; přesto se této „přitažlivosti“ často říká hydrofobní interakce.

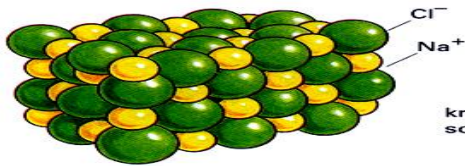
IONTOVÉ VAZBY

Iontové interakce se vyskytují mezi plně nabitými skupinami (iontová vazba) nebo skupinami s částečným nábojem.



Přitažlivá síla mezi dvěma náboji, δ^+ a δ^- , se prudce zmenšuje se vzdáleností mezi nimi.

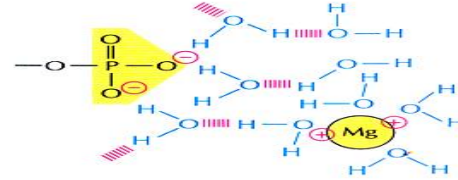
V nepřítomnosti vody jsou iontové interakce velmi silné. Zodpovídají za pevnost takových nerostů, jako je mramor nebo achát.



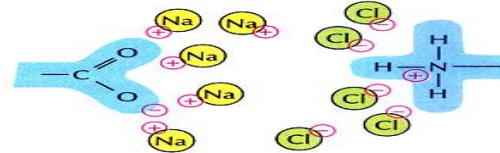
krystal kuchyňské soli NaCl

IONTOVÉ VAZBY VE VODNÝCH ROZTOCÍCH

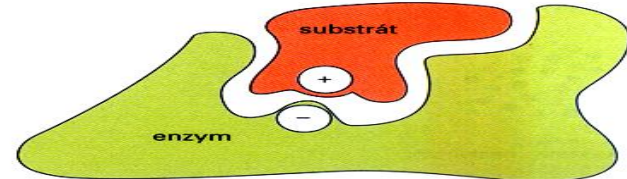
Nabitě skupiny jsou chráněny svými interakcemi s vodou. Proto jsou iontové vazby ve vodě velmi slabé.



Podobně se další ionty v roztoku shlukují kolem nabitých skupin a dále oslabují iontové vazby.

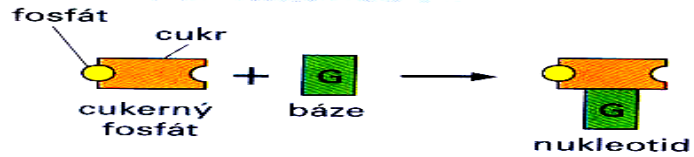


Přes oslabení vodou a solemi jsou iontové vazby velmi důležité v biologických systémech. Enzym, který váže kladně nabitý substrát, má často ve vazebném místě záporně nabitě aminokyselinové zbytky.

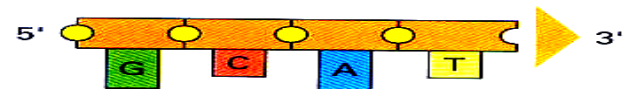


DNA a její stavební podjednotky

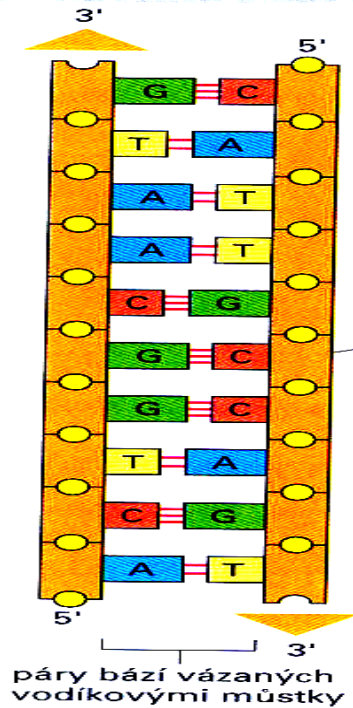
stavební kameny DNA



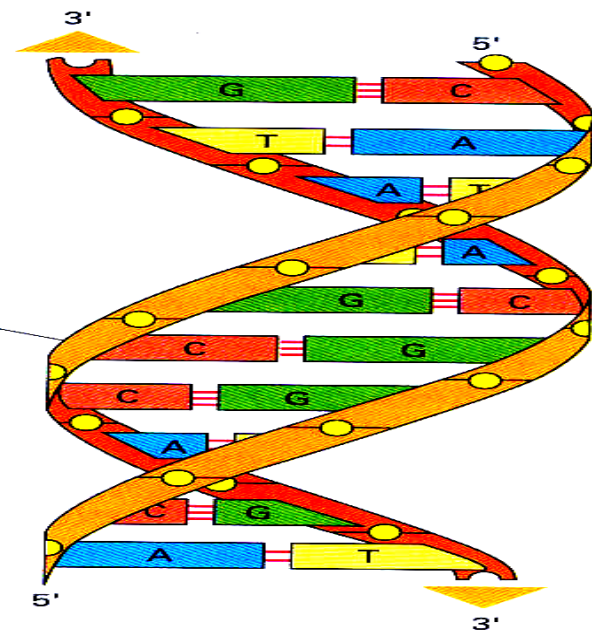
řetězec DNA



dvouřetězcová DNA

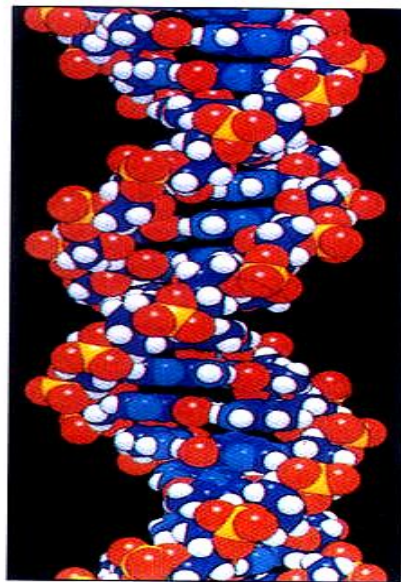


dvojšroubovice DNA



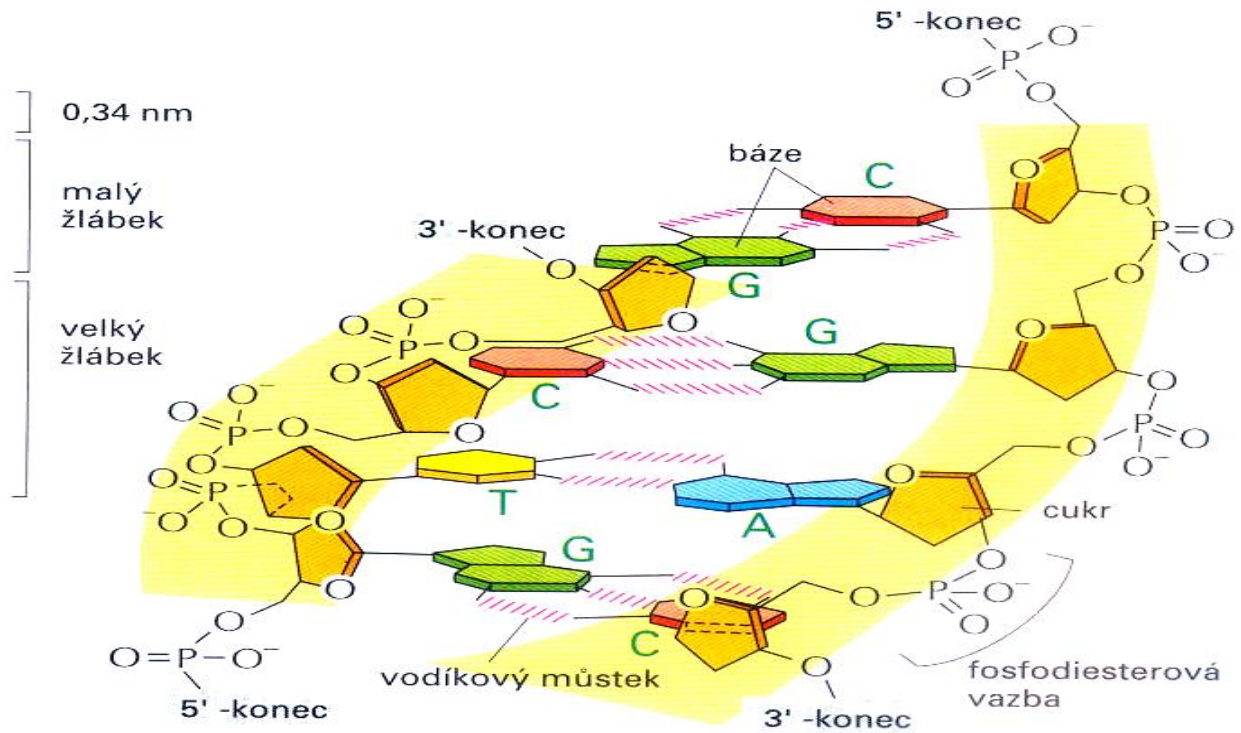
kostra cukerného fosfátu

Dvoušroubovice DNA



2 nm

(A)



(B)

Konformace B:

Vzdálenost mezi páry bází: 0,34 nm

Počet párů bází na jednu otáčku: 10,5

Délka jedné otáčky: 3,4 nm

Základní konformace DNA

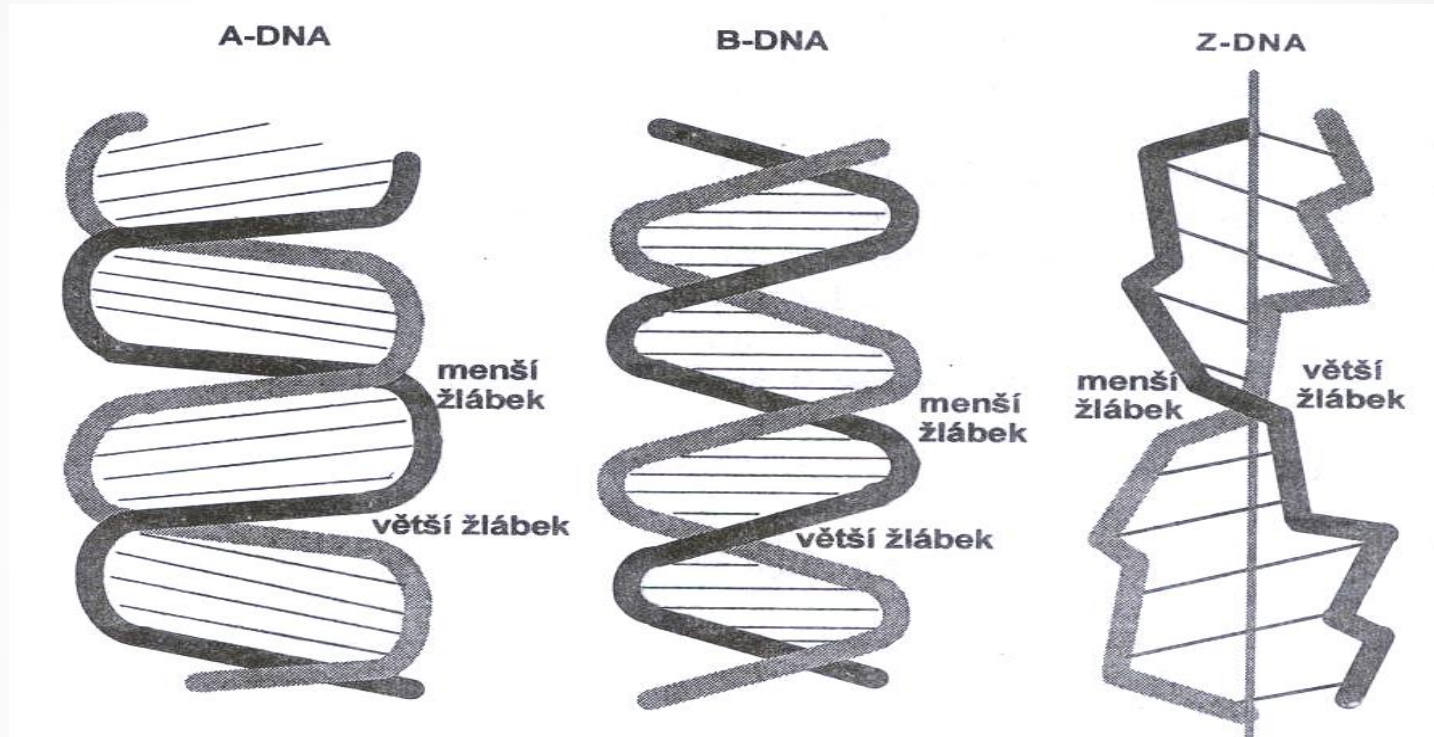
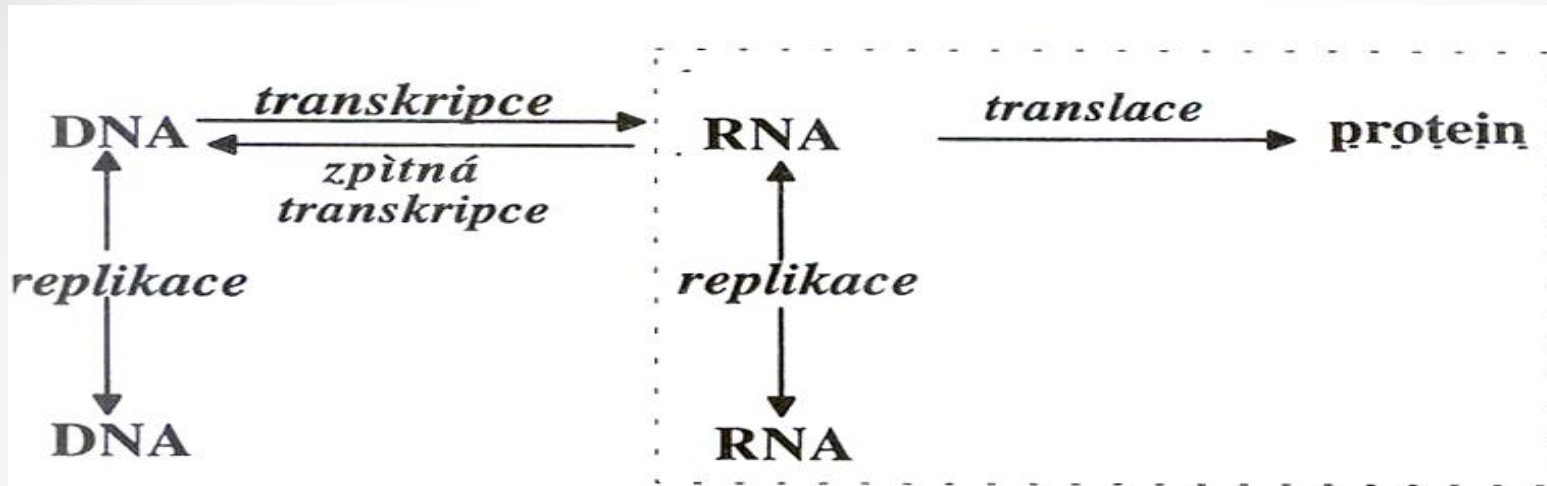


Schéma ústředního dogmatu molekulární biologie



Šipkou je znázorněn směr přenosu genetické informace.

U všech prokaryotických a eukaryotických organizmů, také u DNA-virů probíhá přenos genetické informace v plném rozsahu.

U RNA-virů probíhá jen v rozsahu vyznačeném obdélníkem,

Funkce DNA

Struktura DNA umožňuje uchování genetické informace, její přenos (replikace) a realizaci (transkripce a translace).

Replikace DNA

Protože oba řetězce DNA v dvoušroubovici obsahují sekvence nukleotidů, které jsou navzájem přesně komplementární, mohou obě vlákna sloužit jako **templát** (matrice) pro syntézu nového komplementárního řetězce. Pro tento proces se používá označení **replikace DNA**; vznikají při něm dvě identické molekuly z jedné původní. Tak je umožněn přenos genetické informace z mateřské buňky do dceřiné.

Přesto zde vzniká otázka, jak může buňka zkopírovat miliardy párů nukleotidů před každým rozdělením. Kopírování musí být velice rychlé a přesné: během 8 hodin dokáže živočišná buňka zkopírovat celý svůj genom průměrně s jednou či dvěma chybami. Tuto práci provádí soubor proteinů, který dohromady vytváří replikační aparát. Replikace DNA dává vznik dvěma novým dvoušroubovicím, které pocházejí z mateřského helixu, a každá dceřiná dvoušroubovice je tvořena jedním původním a jedním nově vytvořeným vláknem (**semikonzervativní replikace**).

Dvoušroubovice DNA je za normálních podmínek velmi stabilní, oba řetězce jsou spojeny mnoha vodíkovými můstky mezi bázemi. K oddělení vláken dochází při teplotě kolem 100 °C. Pokud má však řetězec DNA fungovat jako templát, musí být oba řetězce odděleny, aby se jejich nukleotidy mohly párovat s nukleotidy připojovanými do nového vlákna. Celý proces replikace začínají **iniciační proteiny**, které se vážou na DNA a rozvíjejí její dvoušroubovicovou strukturu přerušováním vodíkových můstků. Ačkoliv vodíkové můstky udílejí dvoušroubovici vysokou stabilitu, energie jednoho můstku je velice malá. Proto separace krátkého úseku DNA nevyžaduje velký přísun energie a může k ní docházet za pomoci těchto iniciačních proteinů i za normální teploty.

Místům, kde je struktura dvoušroubovice DNA nejdříve narušena, se říká **replikační počátky**, které jsou určeny speciální nukleotidovou sekvencí. V jednoduchých buňkách, jako jsou bakterie nebo kvasinky, jsou tyto replikační počátky určeny přibližně 100 páry bází. Tato místa jsou tvořena specifickými sekvencemi nukleotidů, které jsou jednak rozpoznávány iniciačními proteiny a navíc zde snadno dochází k oddělení obou řetězců. Z předchozího výkladu víme, že A se páruje s T dvěma vodíkovými můstky, zatímco v páru G-C jsou tři vodíkové můstky. Proto se v replikačních počátcích typicky vyskytují sekvence s vysokým obsahem A a T, které lze relativně snadno rozpojit.

Bakteriální genom, který je obvykle tvořen jednou kruhovou molekulou o délce několika milionů bází, má **jediný replikační počátek**. Na rozdíl od bakterií **lze v mnohem větším lidském genomu najít zhruba 10 000 takových počátků**. Jejich velký počet umožňuje lidské buňce zreplikovat veškerou DNA během relativně krátké doby.

Jakmile jsou iniciační proteiny navázány na DNA a otevřou dvoušroubovicovou strukturu, váže se na replikační počátek skupina proteinů, které spolupracují při syntéze nového vlákna DNA.

Syntéza DNA v replikačních vidličkách

Pro začátky replikace jsou typické útvary ve tvaru Y, které se nazývají **replikační vidličky**. V těchto replikačních vidličkách jsou navázány proteiny replikačního aparátu, které se pohybují ve směru replikace a rozvíjejí dvoušroubovicovou strukturu za současné syntézy nového řetězce. **V jednom replikačním počátku se vytvoří dvě vidličky**, které se pohybují směrem od sebe, a proto je replikace bakteriálního i eukaryontního chromosomu nazývána *obousměrná*. Vidlice se pohybuje velice rychle – zhruba 1000 nukleotidů za sekundu u bakterií a **100 nukleotidů za sekundu u člověka**.

Nejdůležitějším enzymem replikačního aparátu je **DNA-polymeráza**, která syntetizuje nové vlákno DNA podle původního řetězce. Tento enzym katalyzuje připojování nukleotidů na 3'-konec rostoucího řetězce DNA, kdy je vytvářena **fosfodiesterová vazba** mezi skupinou 3'-OH řetězce a 5' fosfátovou skupinou přidávaného nukleotidu. Nukleotidy vstupují do reakce jako energeticky bohaté **deoxynukleosidtrifosfáty** a dodávají energii polymerační reakci. Energie uvolněná hydrolýzou jedné fosfodiesterové vazby v nukleosidtrifosfátu je dostatečná pro kondenzační reakci, při které se váže **nukleotidový monomer** do nově syntetizovaného řetězce za současného uvolnění **difosfátu (PP_i)**. DNA-polymeráza využívá hydrolýzou uvolněnou energii k polymerační reakci. Difosfát je dále hydrolyzován na anorganické fosfáty (P_i), což zaručuje naprostou nevratnost polymerační reakce.

DNA-polymeráza se neodděluje od DNA po každém přidání nukleotidu, ale zůstává navázána na DNA a během polymerace se podél ní pohybuje. K DNA je připojena pomocí speciálního proteinu.

Asymetričnost replikační vidličky

Jelikož je DNA syntetizována pouze ve směru $5' \rightarrow 3'$, nastává v replikační vidličce určitý problém. Jeden nový řetězec je syntetizován podle templátu ve směru $3' \rightarrow 5'$ a druhý podle templátu ve směru $5' \rightarrow 3'$. Oba řetězce však zdánlivě rostou směrem, jakým se pohybuje replikační vidlička, a tak se zdá, že první vlákno je syntetizováno ve směru $5' \rightarrow 3'$ a druhé $3' \rightarrow 5'$.

DNA-polymeráza však může katalyzovat růst řetězce DNA pouze v jednom směru – přidává nové podjednotky na jeho 3'-konec. To řeší syntézu nového vlákna podle $3' \rightarrow 5'$ templátu, avšak zůstává problém kopírování podle $5' \rightarrow 3'$ templátu, protože neexistuje DNA-polymeráza, která by dokázala prodlužovat 5'-konec DNA. V tomto směru **roste DNA diskontinuálně**, to znamená, že jsou ve směru $5' \rightarrow 3'$ syntetizovány krátké úseky DNA, které jsou následně spojovány v kontinuální řetězec. Tyto krátké úseky se nazývají **Okazakiho fragmenty**. **Řetězec DNA, který je tvořen kontinuálně (ve směru $5' \rightarrow 3'$) se nazývá vedoucí řetězec; druhé vlákno, syntetizované diskontinuálně, se nazývá opožďující se nebo váznoucí řetězec.**

Dvě funkce DNA-polymerázy

DNA-polymeráza je velice přesně pracující enzym, který udělá průměrně jednu chybu na 10^7 zreplikovaných párů bází. Typicky se párují A-T a G-C, ale mohou se párovat i G-T a A-C, avšak tyto interakce jsou mnohem méně stabilní. K atypickému párování dochází relativně často, a pokud by nebylo opravováno, hromadily by se v buňce mutace, které by mohly vést i k její smrti. Hromadné fixaci mutací v genomu je zabráněno **opravnou funkcí samotné DNA-polymerázy**. DNA-polymeráza katalyzuje tedy **nejen syntézu DNA, ale má i opravnou funkci nazývanou korektura (proofreading)**. DNA-polymeráza má tedy dvě základní funkce: **$5' \rightarrow 3'$ polymerační a $3' \rightarrow 5'$ nukleázovou aktivitu (štěpící nukleové kyseliny)**.

Syntéza DNA v replikační vidličce

DNA-polymeráza může připojit nový nukleotid pouze ke spárovaným nukleotidům, ale nedokáže začít syntetizovat nové vlákno. Musí tedy existovat jiný enzym, který by dokázal spojit dva volné nukleotidy a začal tak syntetizovat zcela **nové vlákno podle jednořetězcové DNA**. Tento enzym však nevytváří DNA, ale **krátký úsek RNA**. Tyto úseky dlouhé přibližně 10 nukleotidů se párují s templátovým řetězcem a poskytují svůj 3'-konec jako začátek pro DNA-polymerázu. **Slouží tedy jako primer pro syntézu DNA a enzym, který ho tvoří, se nazývá primáza. RNA-primer je syntetizován podle templátu na základě komplementarity bází stejně jako DNA.**

Při syntéze vedoucího řetězce je třeba pouze jeden RNA-primer. Syntéza opoždujícího řetězce je však diskontinuitní a vyžaduje neustále tvorbu RNA-primerů. Jakmile se replikační vidlice posune a uvolní se nový úsek nespárovaných nukleotidů, je na něm vytvořen nový RNA-primer. DNA-polymeráza pak přidává deoxyribonukleotidy na 3-konec RNA-primeru a prodlužuje řetězec DNA, dokud nenarazí na další primer.

Pro vytvoření souvislého vlákna DNA z opoždujícího se řetězce tvořeného mnoha oddělenými úseky (Okazakiho fragmenty) jsou nutné tři enzymy: **RNA-primery jsou odstraněny nukleázou a jsou nahrazeny DNA, což provádí opravná DNA-polymeráza, a v konečné fázi jsou všechny úseky spojeny DNA-ligázou.** DNA-ligáza dokáže spojit 5'-fosfátový konec prvního úseku s 3'-hydroxylovým koncem dalšího úseku.

Replikační aparát

Replikace DNA vyžaduje spolupráci několika druhů enzymů. Jednou ze základních složek je **helikáza**, která rozvíjí dvoušroubovicovou strukturu. Dalšími bílkovinami nutnými pro replikaci jsou proteiny vázající se na jednovláknovou DNA, tzv. **SSB proteiny (z anglického single-strand binding proteins)**, které chrání jednořetězcovou DNA uvolněnou helikázou před znovuspárováním. **Svírací protein (anglicky sliding clamp)** je další součástí replikačního aparátu. Tento protein pevně váže DNA-polymerázu na templát, na opoždujícím se řetězci ji navíc uvolňuje po dokončení syntézy každého Okazakiho fragmentu. Svírací protein má tvar prstence, který obemyká DNA a pohybuje se podél ní i s navázanou DNA-polymerázou ve směru replikace.

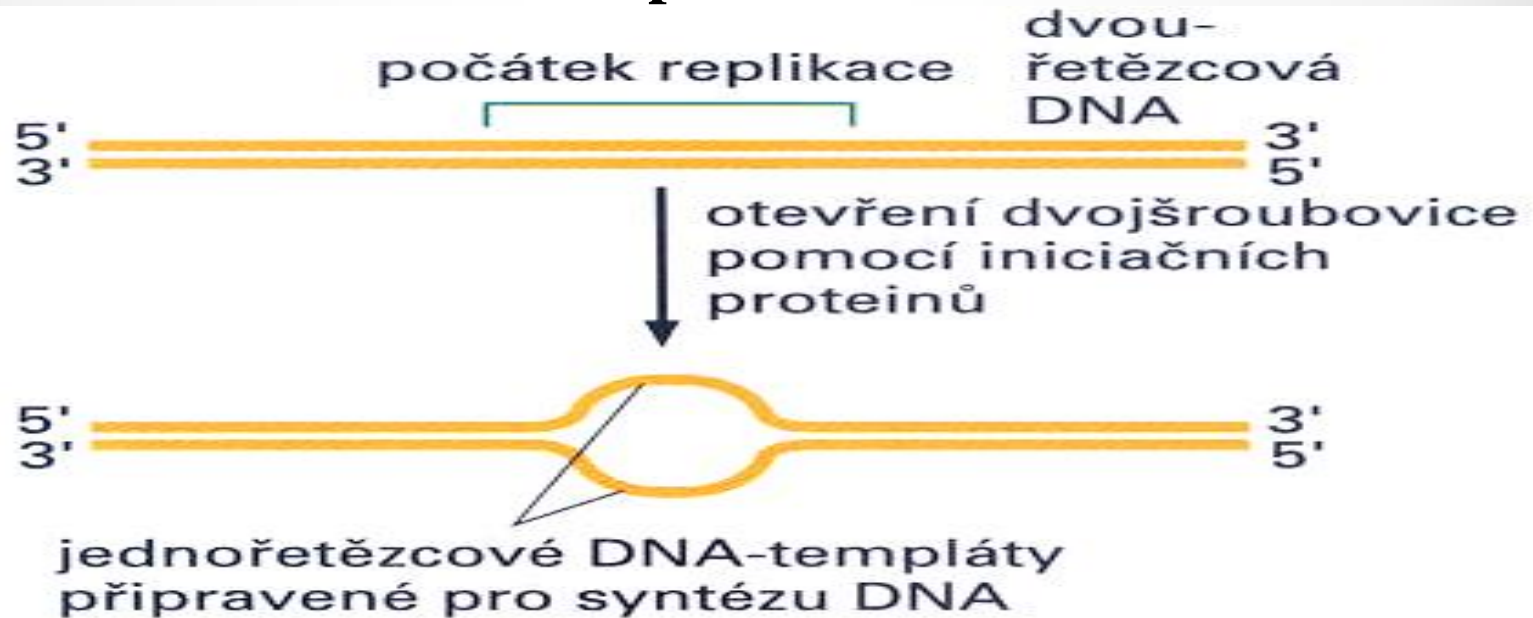
Tok genetické informace (DNA→RNA→protein)

Princip předávání genetické informace z DNA do RNA a následně z RNA do proteinu je tak obecný, že byl nazván **ústředním dogmatem molekulární genetiky**.

Mechanismus, kterým buňky přepisují DNA do RNA, se nazývá **transkripce** (přepis), a využití informace zapsané v RNA k syntéze proteinů se nazývá **translace** (překlad).

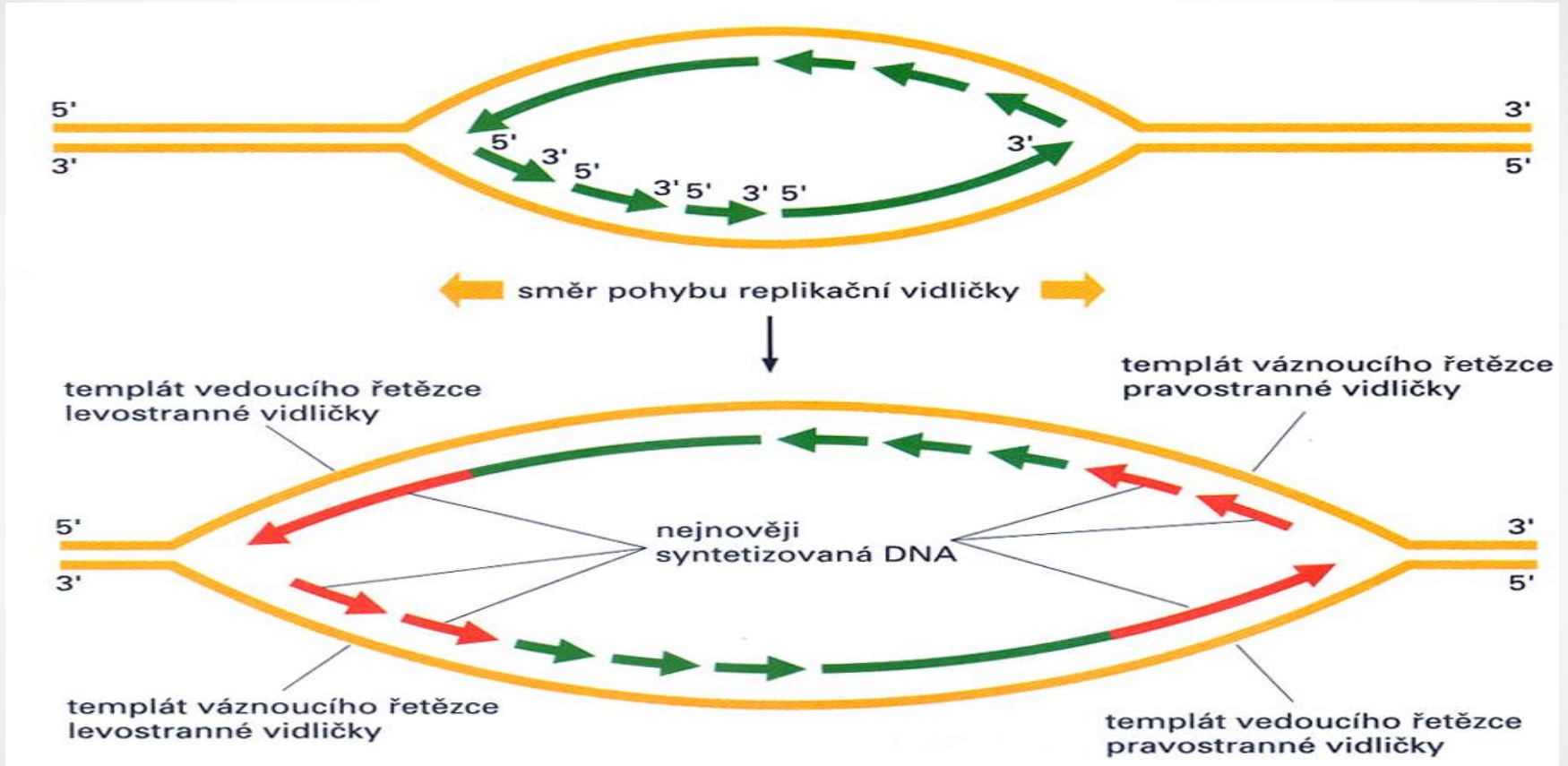
Transkripce a translace jsou procesy, kterými buňka realizuje své genetické instrukce – své geny. Podle jednoho genu může vzniknout mnoho kopií RNA a jedna RNA může dát vznik několika identickým molekulám proteinu. Protože v buňce bývá jen jedna kopie genu na haploidní genom, tato amplifikace umožňuje buňce nasyntetizovat potřebné množství proteinu mnohem rychleji, než kdyby DNA byla sama templátem pro syntézu proteinů. Každý gen může být přepsán a přeložen s různou účinností, což umožňuje buňce nasyntetizovat velmi odlišná množství různých proteinů. Navíc, buňka může regulovat expresi každého genu podle okamžité potřeby.

Replikace



Obrázek 6-13 Rozvolnění dvojšroubovicové struktury v počátku replikace. Iniciační proteiny replikace rozpoznávají sekvenci specifickou pro replikační počátky a lokálně oddálí oba řetězce. Vzniklé jednořetězcové úseky DNA mohou být využity jako templát pro kopírování DNA.

Asymetričnost replikační vidličky



(Okazakiho fragmenty)

Gen. Typy genů

Pojem gen se používá ve dvou základních významech: jako **synonymum pro vlohu** a jako **pojmenování pro konkrétní úsek DNA**.

Gen jako vloha je jednotkou informace, podle které se vytváří určitý znak – fenotyp. U diploidních organismů má gen dvě alely; buď jsou stejné (**homozygot**) nebo se liší (**heterozygot**). Na úrovni fenotypu se projevuje **dominance** (u heterozygota se projevuje funkce jedné alely) a **recesivita** (u heterozygota se funkce jedné alely neprojevuje). Dalšími pojmy jsou **neúplná dominance a kodominance** (projev obou alel současně). Geny jako vlohy můžeme rozdělit na **majorgeny** (geny velkého účinku) a **minorgeny** (geny malého účinku).

Majorgeny mají velký účinek na fenotyp. **Neovlivňuje je vliv prostředí. Kódují kvalitativní znaky** (barva hrachových semen, barva očí, zbarvení srsti, struktura hemoglobinu aj.).

Minorgeny - geny malého účinku mají obecně malý účinek na fenotyp a týkají se zpravidla vlastností kódovaných mnoha geny a alelami (celková velikost organismu, doživost...) - tzv. **kvantitativní znaky**. Jejich **mezialelové interakce** jsou velmi složité. Samotný jeden gen zpravidla nemá valný význam. Účinky jednotlivých alel různých genů (ale někdy i alel téhož genu) se s dominantním účinkem podobného typu (alely podporující růst například) často sčítají nebo násobí.

Gen jako úsek DNA je odpovědný za vytváření funkční RNA molekuly. Rozeznáváme několik typů genů:

Strukturní geny – jsou to úseky DNA, které jsou transkribovány do **mRNA**, které slouží jako matrice pro syntézu polypeptidových řetězců

Geny pro ribozomální RNA (rRNA)

Geny pro transferovou RNA (tRNA)

Geny pro malé RNA (small nuclear RNA, snRNA; small nucleolar RNA, snoRNA; mikroRNA, miRNA; small interfering RNA, siRNA; piwi interacting RNA, piRNA)

Organizace eukaryontního strukturního genu

Úsek DNA zahrnující všechny nukleotidy, které jsou transkribovány do mRNA (messenger; mediátorová RNA) se označuje jako **strukturní gen**. Jelikož je to právě mRNA, která je překládána do polypeptidu a jelikož translace začíná na 5' konci mRNA, je přijata konvence, že sekvence bází v genu je psána jako ekvivalent mRNA sekvence, ve směru od 5' ke 3' konci. Na rozdíl od replikace, **transkripce do RNA je nesymetrická**, tzn., že v určitém segmentu DNA je přepisováno jen jedno **vlákno (3' → 5')** označované jako **negativní (antikódující)**. Druhé **(nepřepisované) vlákno (pozitivní; kódující)** má potom stejnou sekvenci jako produkt transkripce, který je označován jako **primární transkript**:

5' ...TCAGTTACGTACG...3' DNA - pozitivní vlákno (kódující)

|||||

3' ...AGTCAATGCATGC...5' DNA - negativní vlákno (antikódující) – přepisované vlákno

|||||

5' ...UCAGUUACGUACG...3' RNA – primární transkript

Funkční strukturální gen sestává ze dvou částí: **transkripční jednotky**, která navazuje na **regulační oblast (promotor)** (viz schéma genu). Bez regulační oblasti je gen (transkripční jednotka) nefunkční.

Transkripční jednotka je úsek DNA, který je přepisován (transkribován) do pre-mRNA (heterogenní RNA, hnRNA). V této transkribované pre-mRNA je informace pro kódování polypeptidového řetězce. Ale nejenom to. Transkripce začíná i značný kus sekvence před start kodonem (začátkem kódování polypeptidu) a končí značný kus sekvence za stop kodonem (který ukončuje kódování polypeptidu). V pre-mRNA jsou úseky kódující polypeptid (**exony**) a úseky nekódující (**introny**).

Takže v transkripční jednotce (DNA) rozeznáváme: **iniciační místo transkripce, kde RNA-polymeráza II začíná přepisovat DNA; tato část je v exonu 1.** V určité vzdálenosti od iniciačního místa je v sekvenci DNA **sekvence ATG**, což je **start kodon (kodon pro aminokyselinu metionin).** **Tímto kodonem začíná kódování všech polypeptidů u eukaryontů.** Start kodon může být v exonu 1, 2, 3, nebo i dalších. Je to specifické pro každý gen. Sekvence od iniciačního místa transkripce po start kodon se označuje jako **5' UTR (5' nepřekládaná oblast; 5' untranslated region).** Nukleotid, od kterého začíná přepisování transkripční jednotky se označuje +1 (Cap site; Cap místo – viz dále). Nukleotidy číslované od startovacího nukleotidu doprava po směru transkripce se označují znaménkem +. Sekvence před iniciačním místem transkripce se označuje jako **5' upstream sequence (proti proudu) nebo 5' flanking sequence (5' přilehlá sekvence)** (tato sekvence není součástí transkripční jednotky). První nukleotid promotoru před začátkem transkripce se označuje -1, a další nukleotidy doleva proti směru transkripce se označují znaménkem - (např. -4 -3 -2 -1 +1 +2 +3 +4.....)

Určitá část kódující sekvence, označovaná jako signální sekvence, začínající ATG kodonem, kóduje signální peptid, který se odštěpuje při post-translačních modifikacích (viz dále).

Za stop kodonem (TAG nebo TGA nebo TAA) pokračuje nekódující sekvence, až do místa, kde končí transkripce. To znamená, že je to exonová nekódující sekvence, která se označuje 3' UTR (3' nepřekládaná oblast; 3' untranslated region). Délka 3' UTR může být stovky až několik tisíc bází. 10 až 30 bází před koncem terminace transkripce je lokalizován polyadenylační signál (sekvence AATAAA, případně ATTAAA), který je signálem pro ukončení transkripce. (Znovu zdůrazňuji, že transkripce DNA do pre-mRNA začíná exonem a končí exonem).

DNA sekvence, která následuje za 3' UTR, je označována jako 3' downstream sequence (po proudu) nebo 3' flanking sequence (3' přilehlá sekvence).

Úseky DNA, které po transkripci a post-transkripčních úpravách dávají vzniknout zralé mRNA, se označují exony. V DNA jsou jednotlivé exony od sebe odděleny nekódujícími úseky DNA (introny). Eukaryontní geny se výrazně liší v počtu exonů (některé geny mohou být bez intronů; tj. transkripční jednotka není fragmentovaná), jiné mají několik, několik desítek, nebo více než 70 exonů (např. gen pro dystrofin). Obecně, u živočichů je délka intronů mnohem větší než délka exonů. V sekvenci DNA, na začátku intronu je sekvence GT, a na konci intronu je sekvence AG. I když existují některé výjimky, tyto sekvence se vyskytují prakticky pravidelně, a proto se pro ně používá označení GT-AG pravidlo (GT-AG rule).

Regulační oblast genu

V regulační oblasti se nacházejí sekvence, na kterých dochází k **uspořádání obecných transkripčních faktorů a RNA polymerázy II do iniciačního komplexu a jejich navázání na DNA**. Tato sekvence se nazývá **promotor**. Nejvýznamnějším typem sekvence v promotoru je **TATA box**, který je prvním rozpoznávaným místem promotoru při uspořádávání transkripčních faktorů do iniciačního komplexu. Konsenzuální sekvence je TATAWAWR (W = A nebo T; R = A nebo G). Existují však i jiné motivy. Místo, ve kterém začíná RNA polymeráza přepisovat gen, leží obvykle 25-30 párů bází za TATA boxem. Značné procento eukaryontních genů (u člověka asi 76%) neobsahuje TATA box. Součástí promotoru jsou i další sekvence (**CAAT box, GC box aj.**). Polohy regulačních sekvencí se liší gen od genu.

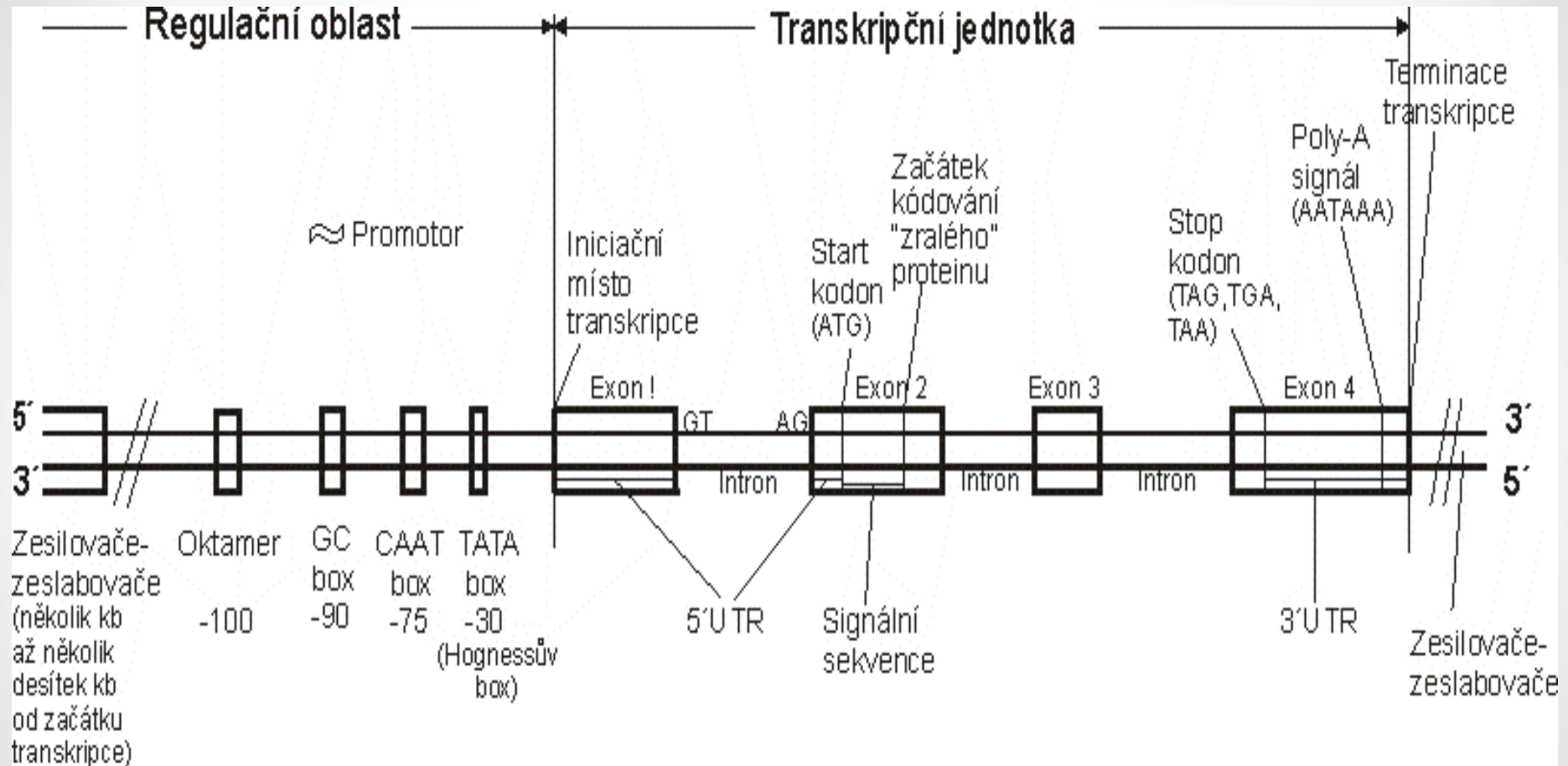
Transkripční faktory jsou proteiny, které se podílejí na iniciaci transkripce. Nejběžnější jsou TFIIA, TFIIB, TFIID (součástí je podjednotka zvaná TATA binding protein (TBP) – váže se specificky na sekvenci TATA boxu), TFIIE, TFIIF a TFH.

Zesilovače (enhancery) jsou místa, na která se vážou aktivátory (rozvolnění chromozómů).

Zeslabovače (silencery) jsou místa DNA, na která se vážou represory (spiralizace chromozómů).

Zesilovače a zeslabovače mohou být ve vzdálenosti až tisíců nukleotidových párů od promotoru a mohou být lokalizovány před nebo za genem. Regulační sekvence mohou být lokalizovány i v intronech.

Schéma eukaryontního strukturního genu



Transkripce je katalyzovaná RNA polymerázou II

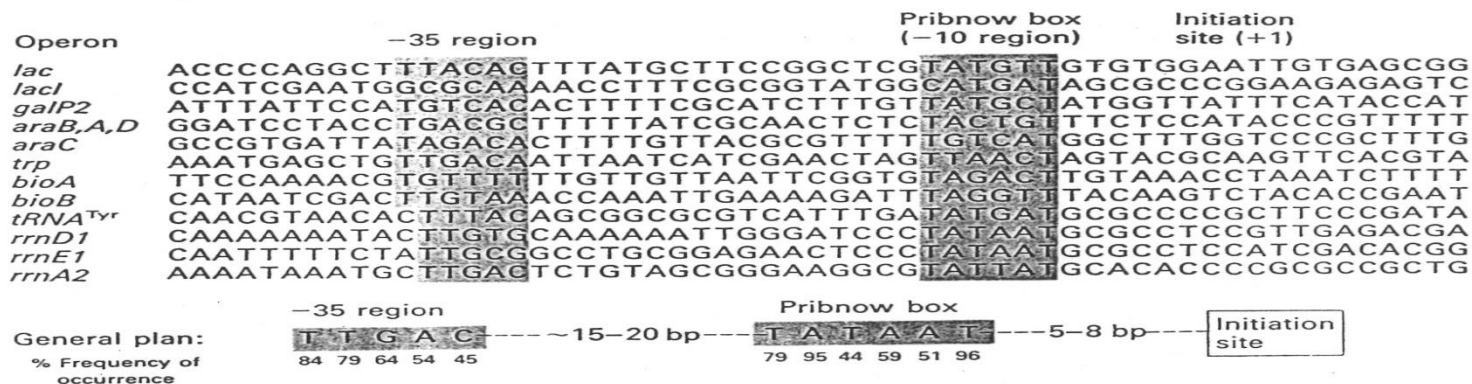
Jednoduché strukturní geny – nemají introny – přepisují se celé

Složené strukturní geny – skládají se z exonů a intronů

Introny: 5' GT.....AG 3' - GT/AG pravidlo

Genomová sekvence genu *EEF1A1* prasete

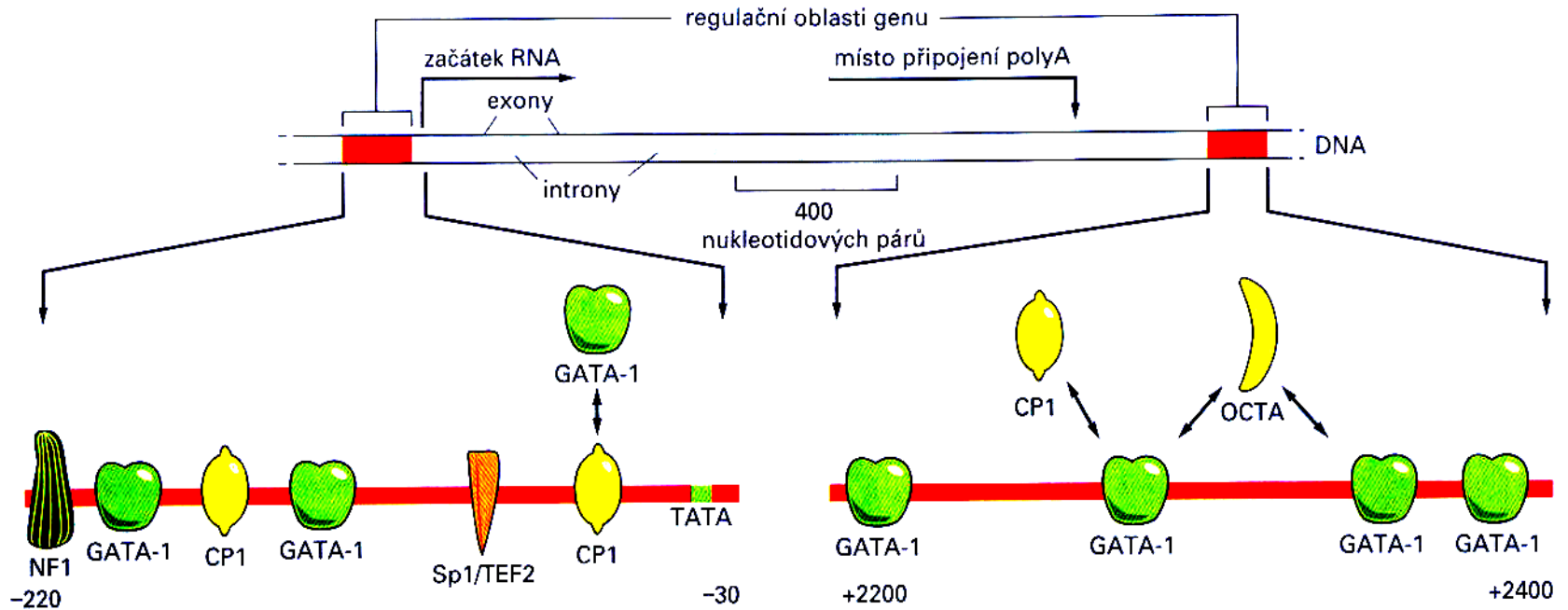
GTGGAGGTGG	CGCGGGGTAA	ACTGGGAAAAG	TGGTGTCTGTG	TGCTGGCTCC	GCCCTTTTCC	CCGAGGGTGG	GGGAGGACCA	5' upstream
<u>TATATAAG</u> CG	CCGTGGTCCC	CGCGAACGTT	CTTTTTTCGCA	ACGGGTTTGC	CGCCAGGACA	CAGGTGAGTA	CGGGTGTGGC	Exon 1
CTCCGTCCGC	ATGGCCTCCG	CCGGTGGCCA	CGGCCTTAGC	GTGCCTCCCG	GCCCCC CGC	CGTAGAGGGC	TCTGCGCCCT	(5' UTR)
GGTCTTGATT	CCGAGCTGCG	GCGGGGGGGA	GGTGGAGAAC	TCGAGGCCCT	CCGCTCTCGC	GGTTCCCTAC	CGCGTGCCCG	
GTGGCGGCCCT	GCTGGGGCGC	CGTGGCCGCC	GCGTGCATC	CGCGCCTTCG	CGCCCGCCCC	GTCTGTCTTC	GCGCTCGAGT	
TCGGCCGGCA	GCCGCGTGT	ACTGCCAGGG	CGGCCAGCC	GAGGCCGCC	AAGGTGGCCT	CCGTGCCCGG	CGGTCCCGGG	
GCAGACGGCC	GCGGCGGGG	GGGGCGGGCG	GTCGGAGACC	CAGCGCCCCG	CCCGGGCGGG	CTGCGAGCGC	GGAAAAGATGG	
CCGCTCCCGG	TCTCAAGATG	GAGGACGCGG	ACGGCGGGCG	GCGCACAAAAG	GGCCGGGCGG	TGCTGCCCGC	GCCGCTTCCC	
GGGACTCGCG	GCCC GGTTG	GGGAGGGGCT	GGGCGCCGGG	GCGGCCCCGT	TCGCTGGGTG	TAGGCCGGG	AGGGCCCGCA	
GGGCGCCGGT	ACTAATGCTC	CCTGGAGTTT	GCCCTTGGCT	GTTGGGGTTG	TGCCGTGAGTC	TCCGGTCGTC	GGGACAGTAG	
TTTAAAGAAG	GTTTTTGTG	TCTTAGGTGT	CGTGAAAGCC	ATCGCTAAAA	GCTAAGCAAA	<u>AATGGGAAAG</u>	GAGAAGACTC	(5' UTR)
ACATCAACAT	CGTCGTCATC	GGACACGTCG	ATTCGGGCAA	GTCACCACCT	ACTGGCCATC	TGATCTACAA	ATGCGGTGGG	Exon 2
ATCGACAAAA	GAACCATCGA	AAAGTTCGAA	AAAGAGGCTG	CGGAGTTCG	TAAAGTAACG	TCAGCAAAT	AAGTGGTAGT	
TCTCTAAAAAC	TAGTGTATAT	CTGACTTGAA	CTAGTACATT	GCAAAGTTTG	CTAAGTGGCT	TAACCTCAGC	CAAGTTTAGG	
AGATGGAAGT	GAAACCTAGC	ATTGCTTCCA	AGTGGGACTA	AGGAGTTTAG	TTACCGAGTA	ACCCTTCTTA	ATCCCTGCAG	
ATGGGAAAGG	GCTCCTTCAA	GTATGCCTGG	GTCTTGACA	AACTAAAGGC	TGAACGTGAG	CGTGGTATTA	CCATTGACAT	Exon 3
CTCCCTGTGG	AAATTGAGAG	CCAGCAAGTA	CTACGTGACC	ATCATTGATG	CTCCAGGACA	CAGAGACTTT	ATTAAGAACA	
TGATTACAGG	CACTTCCCAG	GTTTGTAGGA	TTAAGAAGTC	CTGAGAGTAA	CATTTTGAGT	TATCTATAAA	TTGGCGAAAC	
CATGGATCTG	TCCTTTTGG	AGAGGTTTTT	ACAGGGTGTG	GCCTTTTTTT	CTTTTTTAAA	<u>GGCTGACTGT</u>	<u>GCTGTCTTGA</u>	
TTGTTGTCTG	TGGTGTGGT	GAATTCGAAG	CAGGTATTTT	CAAGAACGGG	CAGACCCGTG	AGCATGCTCT	TCTGGCCTAC	Exon 4
ACCCTGGGTG	TGAAACAGCT	CATTGTTGGC	GTTAACAAAA	TGGATTCCAC	TGAGCCACCC	TACAGCCAGA	AGAGATATGA	
AGAGATTGTA	AAGGAAGTCA	GCACCTACAT	TAAGAAAAAT	GGCTACAACC	CTGATACAGT	AGCATTGTGT	CCAATTTCTG	
GCTGGAATGG	TGACAACATG	CTGGAGCCAA	GTGCTAATGT	AAGTGGTTTT	GATGAGTTCT	CAAAACTTAA	GTCTGGGAAG	
GGATAAGTTT	GAAAAACTA	ATTGATGATC	TTGTGTTTTA	GATGCCTTGG	TTCAAGGGAT	GGAAAGTCAC	CCGTAAGAT	Exon 5
GGCAATGCCA	GTGGAACCAC	ACTGCTTGAA	GCTCTGGATT	GCATTCTACC	ACCAACTCGT	CCAACCTGACA	AGCCCTTGGC	
CCTGCCCTTC	CAGGACGTCT	ACAAAAATTGG	TGGTAAGTTG	AACGTTACAC	TCAGCTGGTA	TGGGTTGCAG	TGATGTTTTGT	
CTTCATACGA	GATAATTATG	CTGTGCTGTG	AATACTTACG	GTATTGGTAC	AGTCCCTGTG	GGCCGAGTGG	AGACCGGTGT	
TCTCAAACCT	GGCATGGTGG	TCACCTTTGC	TCCAGTCAAT	GTTACAACCTG	AAGTCAAGTC	TGTTGAAATG	CACCATGAAG	Exon 6
CTTTGAGTGA	AGCCCTTCCT	GGGGACAACG	TGGGCTTTAA	TGTCAAAAAC	GTGCTGTGCA	AAGACGTTTCG	TCGTGGCAAT	
GTGGCTGGTG	ACAGCAAAAA	TGACCCACCT	ATGGAAGCAG	CTGGCTTTAC	AGCTCAGGTA	ACACTTCAGT	GACAGGTTTT	
TAAACTGCCA	GAACCTTAAC	TGCTAAAATG	TCTAAAATGT	GCTAAAATGCA	AATTTTCTCC	CTAAGGTGAT	TATCCTGAAC	
CATCCTGGCC	AGATCAGTGC	TGGTTCAGCA	CCTGTGCTGG	ATTGCCACAC	AGCTCACATT	GCTTGCAAGT	TTGTGAACT	Exon 7
GAAGGAGAAG	ATTGATCGTC	GTTCTGGGAA	AAAGCTGGAA	GATGGCCCCA	AGTTCCTGAA	ATCTGGTGAT	GCTGCCATTG	
TCGATATGGT	TCCTGGCAAG	CCCATGTGTG	TTGAGAGCTT	CTCTGACTAT	CCTCCTCTGG	GTAAGAGTGC	CTTTTATAGT	
TAATTAAGCT	TCAGTAAAAA	TGGA AAAAGCA	GTTATGTAGA	GAAAGTGTTT	CTCATGGTGA	TTTAATTTTT	GTTTTAATAG	
GCCGATTTGC	TGTTCTGTGAC	ATGAGACAGA	CAGTTGCTGT	GGGTGTCATC	AAAGCTGTGG	ACAAGAAGGC	AGCTGGAGCT	Exon 8
GGCAAGGTCA	CCAAGTCTGC	TCAGAAAAGCT	CAGAAGGCTA	<u>AA TGA</u> ATATT	ATCCCCAATA	CCTGCCACCC	CAGTCTTAAT	(3' UTR)
CAGTGGTGGG	AGAACGGTCT	CAGAACTGTT	TGTCTCAATT	GGCCATTTAA	GTTTAATAGT	AAAAGACTGG	TTAATGATAA	
CAATGCATCG	TAAAACCTTC	AGAAGGAAAAG	GAGAATGTTT	TGTGGACCAT	TTGTTTTTTT	TGTGTGGCAG	TTTTTAAGTT	
ATTAGTTTTT	AAAATCAGTA	CTTTTTAATG	GAAACAACCT	GACCAAAAAA	CTGTCACAGA	ATTTTGTAGAC	CC <u>ATTAAAAA</u>	
AAGTTTTAATG	AGAAACCTGT	GTGTTCTTTT	GGTCAACAGA	GTAACCTCCT	ATGAAGGTAA	AGGCTGGTCA	TAACCTATTT	3' downstream
CTAGTGTTAC	ATAGAAAACAA	GATTTGTGGA	TGTTTTTTTC	TGGGTATGAG	CTTCATGTTT	CTCAAGACTG	TCTACAGAGA	



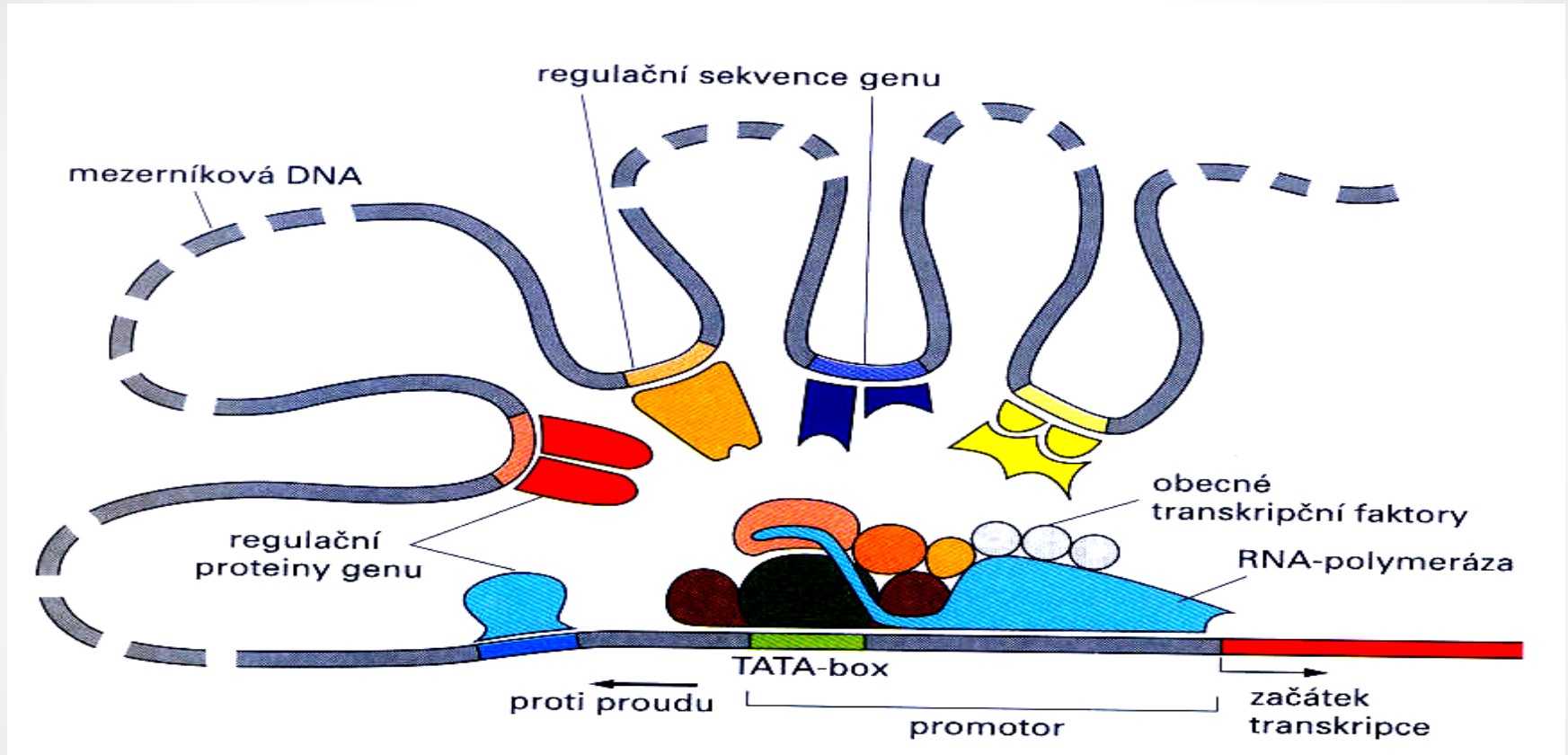
▲ **Figure 7-3** The nucleotide sequences in several promoter sites that direct transcription by *E. coli* RNA polymerase containing the major sigma factor, σ^{70} . Each sequence shown belongs to a specific operon, that is, a cluster of genes that are all controlled by the same promoter. The Pribnow box is a strongly conserved region of six bases; a sequence around the -35 position, which has a five- or six-base core, is also conserved. The bases in red type are initiation sites. Usually transcription starts at a single nucleotide; occasionally multiple, alternate sites are present (e.g., in the *lac* and *rrnA2* op-

erons). The general plan of the promoters is shown at the bottom of the diagram; the percentages below each base indicate the frequency of 112 different promoters. Note that the initiation site for *bioB* (a gene for the vitamin biotin) is at a thymine (which becomes a uracil in the mRNA) and that the site for *rrnA2* (a gene for rRNA) has three cytosines; all of the other initiation sites occur at an adenine or a guanine. [See D. Hawley and W. R. McClure, 1983, *Nucleic Acids Res.* 11:2237.]

Model kontroly exprese lidského β -globinového genu



Regulační sekvence typického eukaryontního genu



Transkripce eukaryotického strukturního genu a posttranskripční úpravy

Při **iniciaci transkripce** katalyzované RNA-polymerázou II se na Hognessově boxu (TATA) vytvoří aktivní iniciační komplex RNA-polymerázy II a transkripčních faktorů schopný za přítomnosti volných ribonukleotidů zahájit transkripci. Transkribovaný úsek musí být přechodně rozvinut, zpřístupněn regulačním proteinům a DNA dvoušroubovice rozpletena pro činnost RNA polymerázy.

Prodlužování (**elongace**) **transkriptu**: RNA polymeráza syntetizuje RNA řetězec podle DNA předlohy, přičemž asi 12 nukleotidů nového vlákna zůstává ještě připojeno v místě transkripce. Za tímto segmentem se řetězec RNA odděluje od matrice a přepsaný úsek DNA se opět svinuje do dvoušroubovice.

Terminace transkripce: U eukaryontů je konec transkripce signalizován sekvencí AATAAA, která se označuje jako **polyadenylační signál**, neboť signalizuje, že 10 až 30 nukleotidů za touto sekvencí se hnRNA bude štěpit a tím se uvolní z DNA. Tento konec je pak polyadenylován za katalytického účinku poly(A)-polymerázy.

Z tohoto pravidla jsou však i výjimky. Např. hnRNA vznikající transkripcí genů kódujících histony nemá polyadenylační signál. Taková hnRNA není polyadenylována, ale její 3' konec je určen tvorbou vlásenky a rozeznáván U7-snRNA vázající se k této hnRNA prostřednictvím specifických sekvencí.

Post-transkripční úpravy (RNA processing)

Transkripční jednotka se přepisuje do primárního transkriptu (pre-mRNA; hn-RNA) celá – od iniciačního místa transkripce až po konec transkripce. Primární transkript následně prochází úpravami:

- **Připojení čepičky:** 5' konec hnRNA se hned při jejím uvolňování z chromatinu pokrývá strukturou ve formě $m^7G^{5'}ppp^{5'}N_1mpN_2mp\dots$

(m = metylová skupina, G = guanozin, N_1 a N_2 = purinový nebo pyrimidinový nukleozid, p = fosfát) (7-metylguanozin vázaný 5',5'-trifosfodiesterovou vazbou k sousednímu metylovanému nukleotidu). Tato struktura se označuje jako čepička (5' cap) a vyjadřuje se zkratkou m^7G . K tomuto procesu obvykle dochází kotranskripčně, tedy ještě v průběhu transkripce genu.

- **Polyadenylace 3' konce:** hnRNA se štěpí 10 až 30 nukleotidů za poly-A signálem a následně se ke 3' konci hnRNA přidá 50 až 200 zbytků kyseliny adenylové, což je katalyzováno poly(A)-polymerázou.

Tyto dvě modifikace konců mRNA – čepička a poly(A) pravděpodobně stabilizují molekulu RNA a pomáhají při jejím transportu z jádra do cytoplasmy. Později jsou také využity aparátem syntetizujícím proteiny jako signál, že daná mRNA má oba konce a informace, kterou kóduje, je kompletní.

- **Sestřih (splicing):** Většina strukturních genů eukaryontů nejsou souvislé kódující sekvence, ale jsou přerušovány segmenty (introny), které nejsou součástí zralé funkční mRNA (ty jsou vystřiženy, a exonové sekvence jsou spojeny). Tak vznikne zralá mRNA, která přechází z jádra do cytoplasmy, na ni přisedají ribozómy a dochází k syntéze polypeptidového řetězce.

Podstatné rysy genu a jeho mRNA

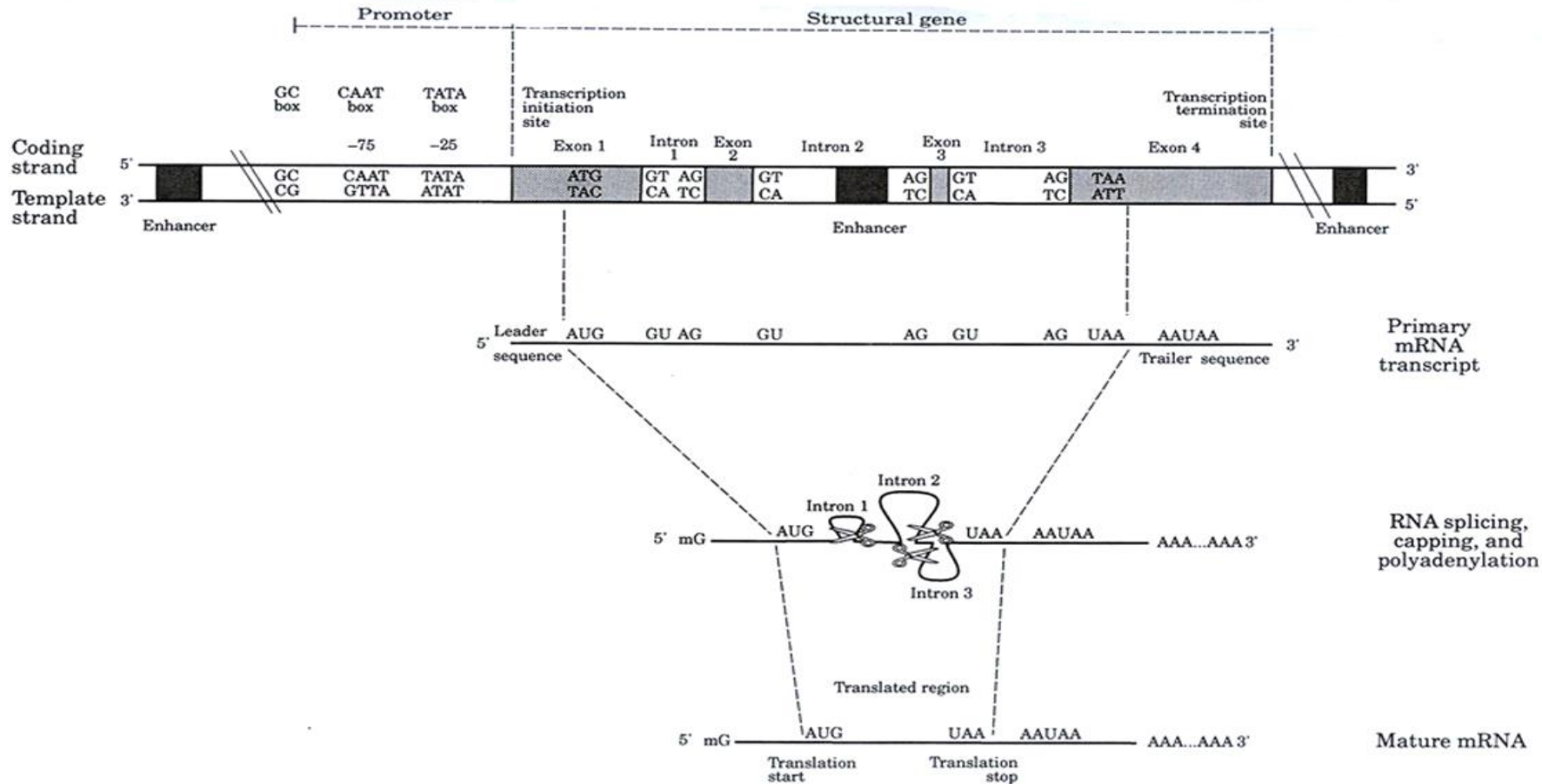


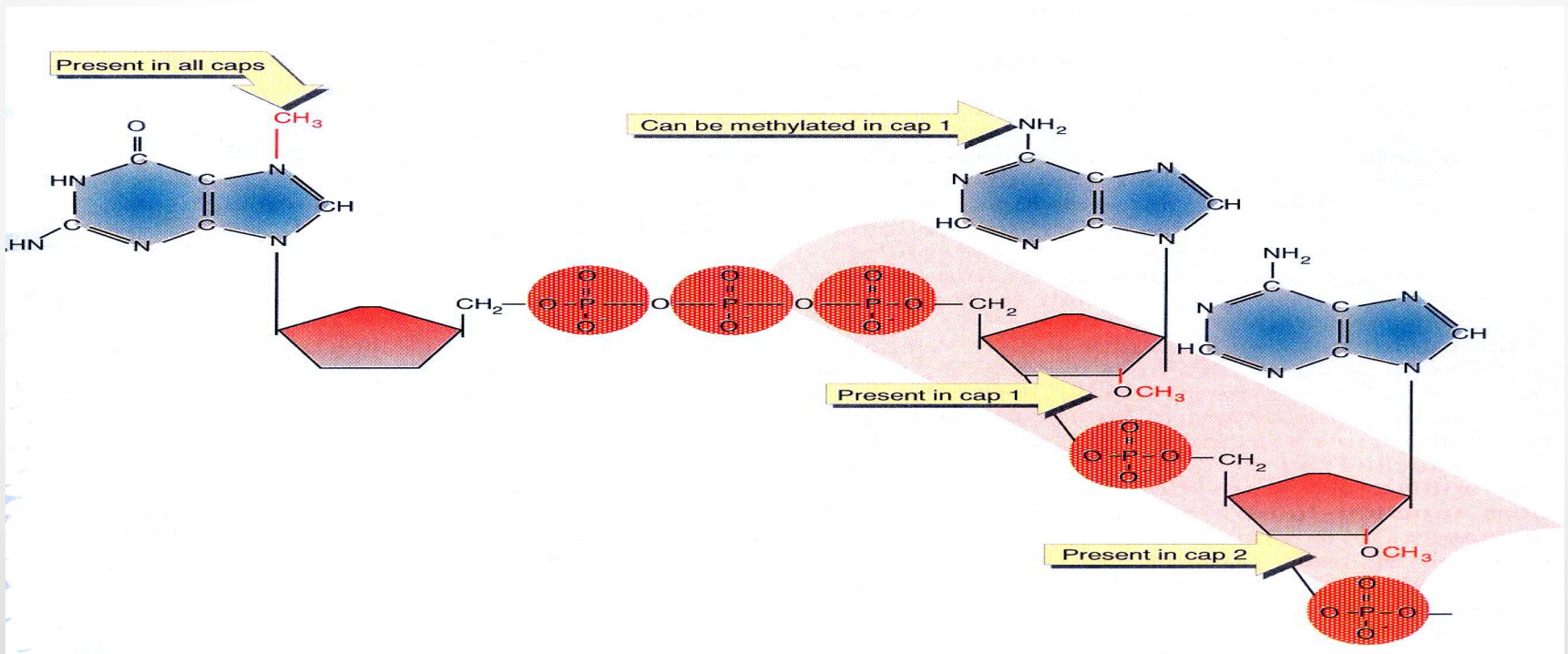
Fig. 1.10 The essential features of a gene and its mRNA.

Primary mRNA transcript = pre-mRNA = hnRNA (heterogenní jaderná RNA)

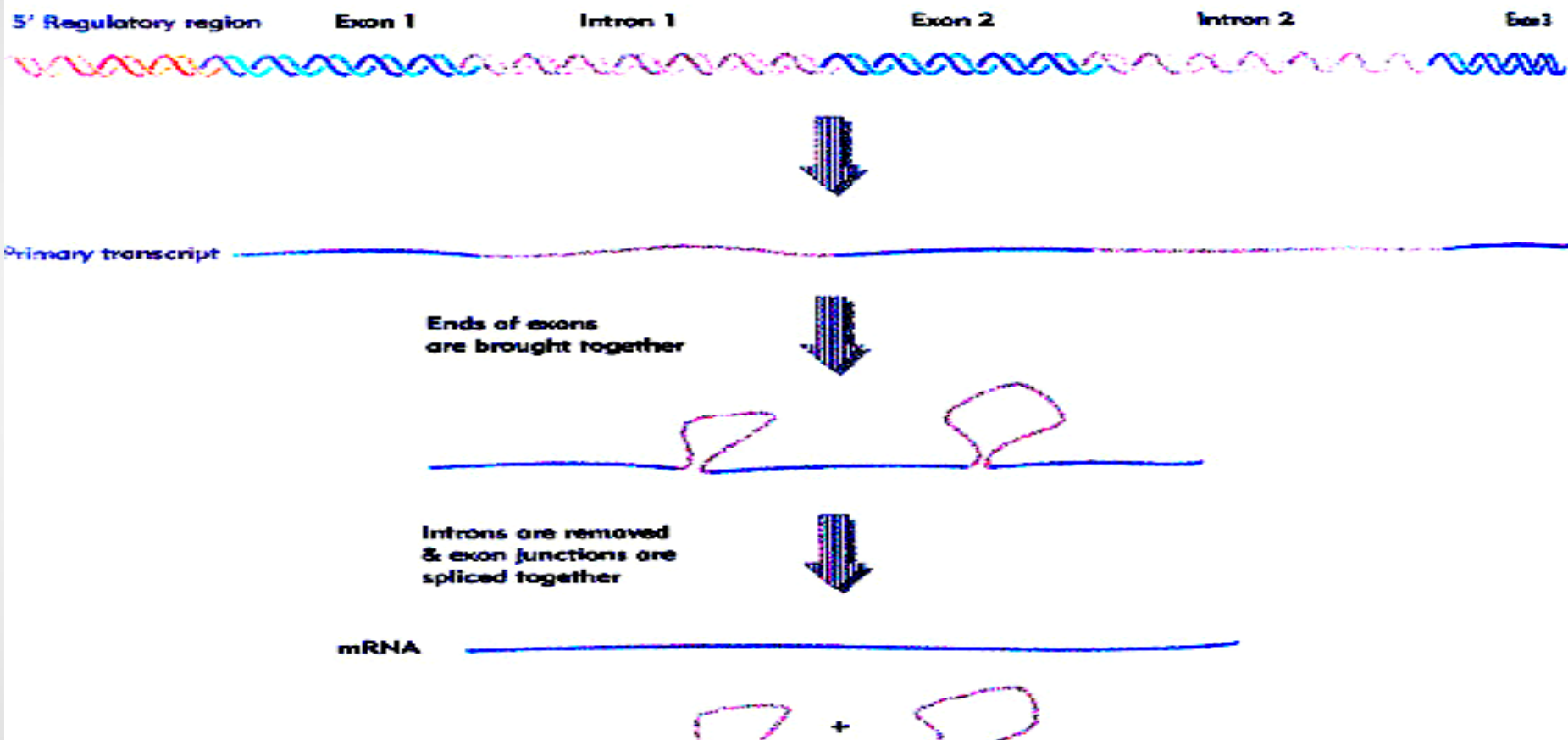
Úprava 5'-konce hnRNA čepičkou

5'-konec hnRNA se hned při jejím uvolňování z chromatinu pokrývá strukturou ve formě $m^7G^{5'}ppp^{5'}N_1mpN_2mp\dots$ (koncový metylguanozin se váže

5',5'-trifosfodiesterovou vazbou k sousednímu metylovanému nukleotidu). Tato struktura se označuje jako **čepička** a vyjadřuje se zkratkou **m^7G** .

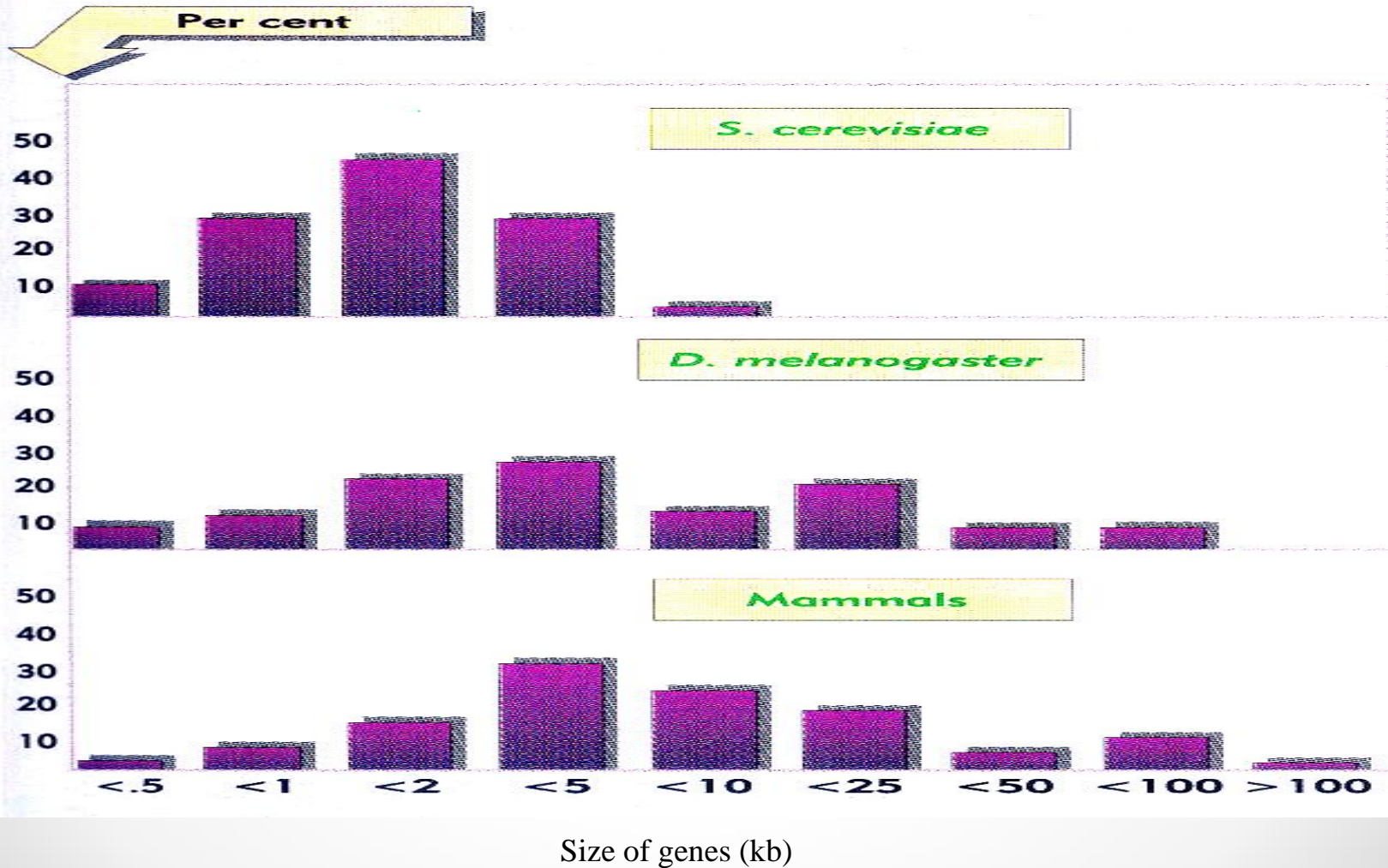


Overview: an interrupted gene consists of alternating exons and introns; the introns are removed by splicing of the RNA transcript, generating an mRNA consisting only of exon sequences. The regulatory region includes the promoter at the 5' end.



Velikosti genů

Yeast genes are fairly small, but genes in flies and mammals have a dispersed distribution extending to very large sizes.



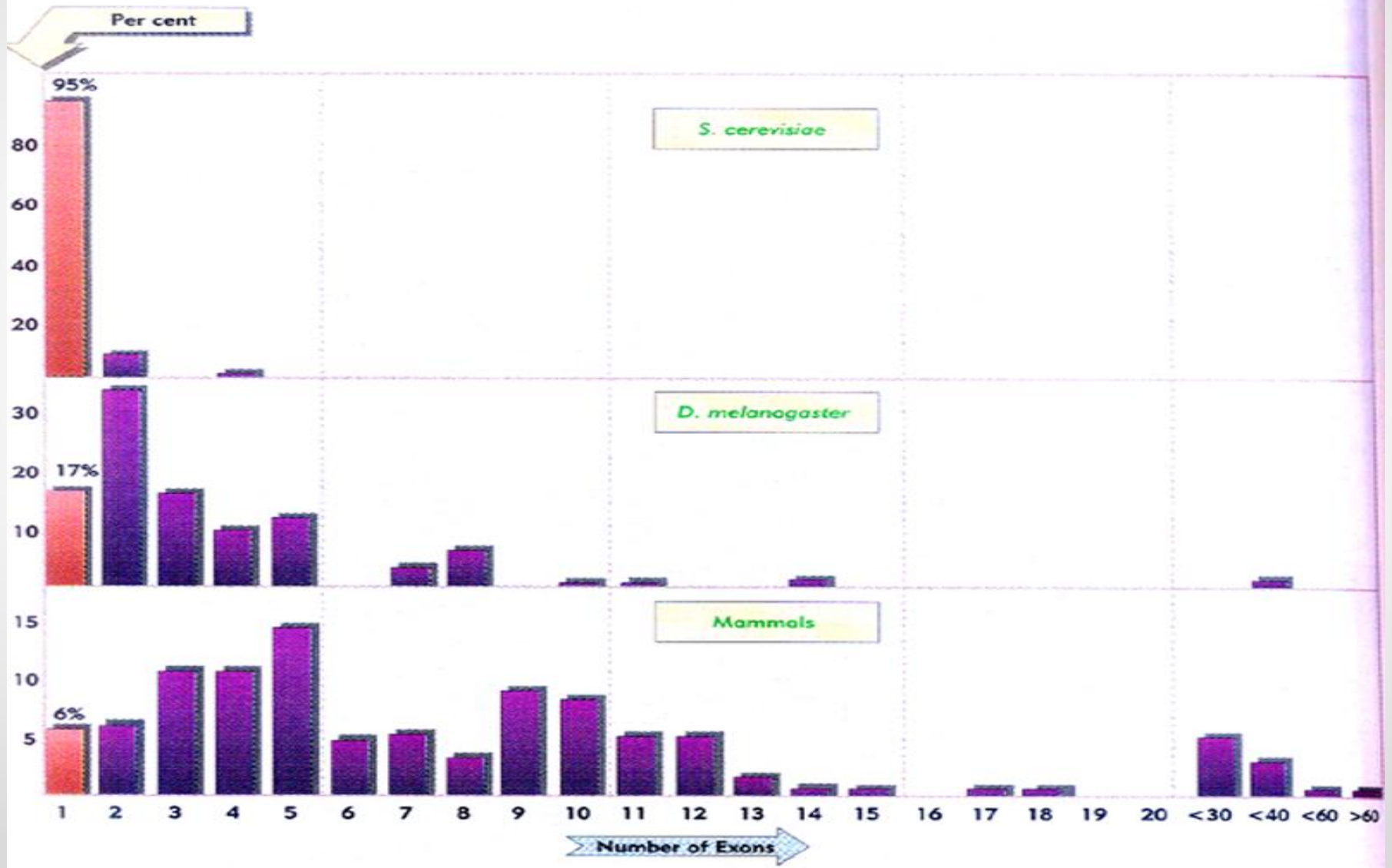
Velikosti genů

The average lower eukaryotic gene is similar in size to the average mRNA, but the average size of the higher eukaryotic gene is about 5 times the length of its mRNA product.

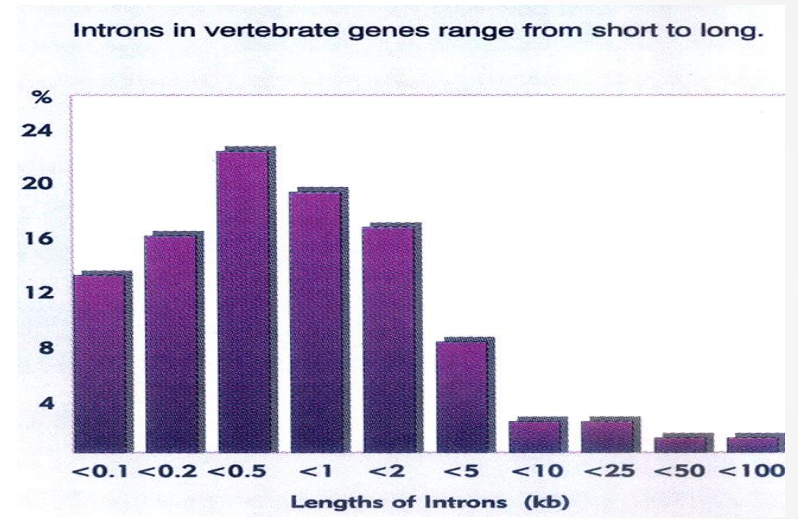
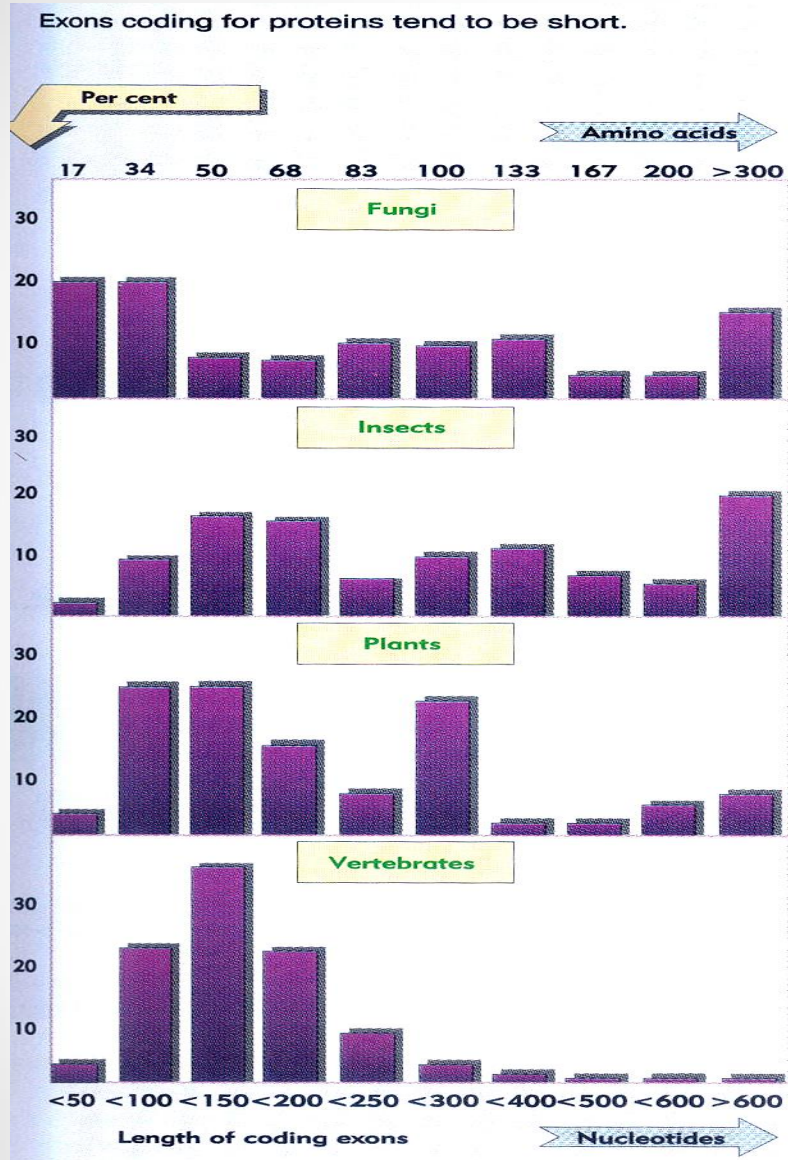
Species	Average exon number	Average gene length (kb)	Average mRNA length (kb)
<i>S. cerevisiae</i>	1	1.6	1.6
Fungi	3	1.5	1.5
<i>C. elegans</i>	4	4.0	3.0
<i>D. melanogaster</i>	4	11.3	2.7
Chicken	9	13.9	2.4
Mammals	7	16.6	2.2

Počty exonů v genech

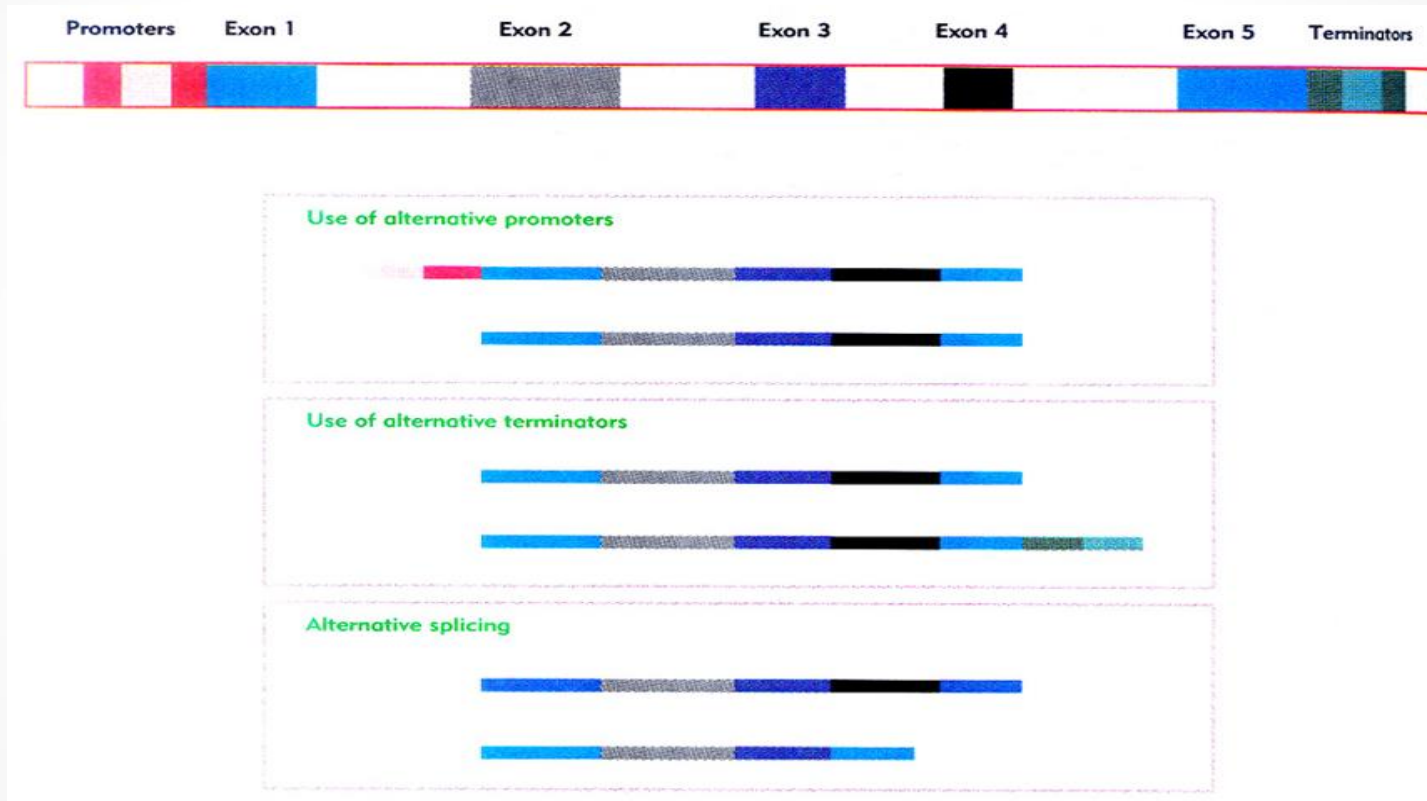
Most genes are uninterrupted in yeast, but most genes are interrupted in flies and mammals. (Uninterrupted genes have only 1 exon, and are totalled in the leftmost column.)



Velikosti exonů a intronů

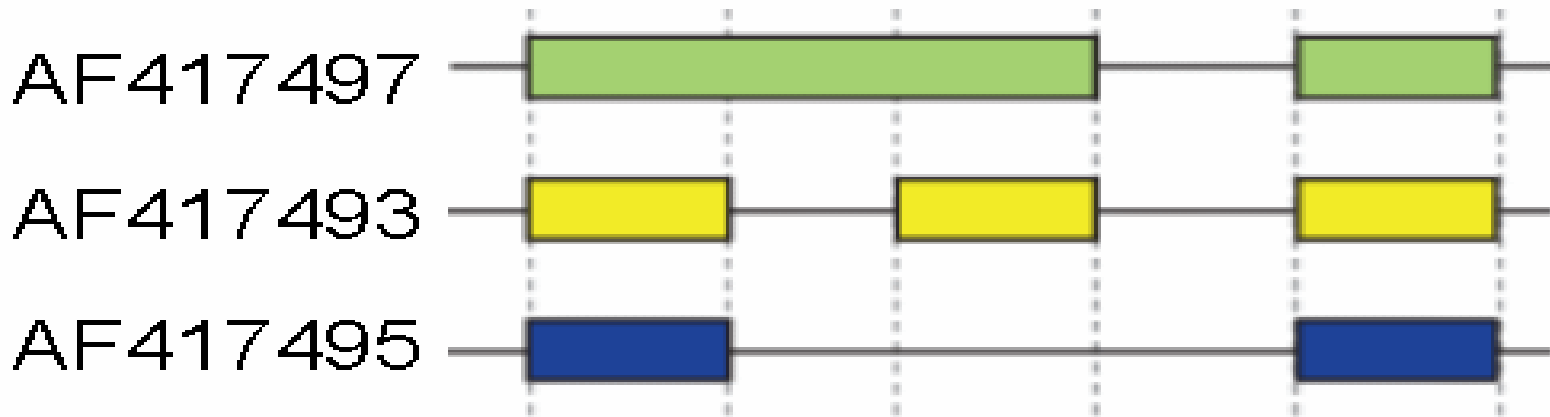


Gen může kódovat různé RNA produkty změnou iniciace, terminace, nebo sestřihu



Alternativní sestřih

Myší hyaluronidasa



Tři typy:

- **zachování intronu:** namísto vystřížení intronu, tento je zachován v transkriptu mRNA. Intron však musí správně kódovat aminokyseliny. Případný stop kodon nebo posun fáze translace může způsobit, že protein bude nefunkční.
- **vystřížení exonů:** v tomto případě některé exony jsou vystříženy a dojde ke změně aminokyselinové sekvence v exprimovaném proteinu.
- **Alternativní místa sestřihu** (alternative donor/acceptor site mode): v tomto případě jsou pro daný exon využívána alternativní 5' a 3' místa, a to má za následek posun fáze translace nebo inzerci/deleci aminokyselin v exprimovaném proteinu.

(Rekordmanem v alternativním sestřihu je gen *Dscam* Drosophily, který má 38 016 variant sestřihu.)

Transkripty genu *LMNA* u člověka

(Databáze Ensembl)

Name	Transcript ID	Protein ID	Description
LMNA-001	ENST00000368300	ENSP00000357283	protein_coding
LMNA-002	ENST00000368301	ENSP00000357284	protein_coding
LMNA-003	ENST00000368298	ENSP00000357281	protein_coding
LMNA-004	ENST00000368299	ENSP00000357282	protein_coding
LMNA-005	ENST00000496738	No protein product	processed_transcript
LMNA-006	ENST00000368294	ENSP00000357277	protein_coding
LMNA-007	ENST00000368297	ENSP00000357280	protein_coding
LMNA-008	ENST00000473598	No protein product	processed_transcript
LMNA-009	ENST00000469565	No protein product	processed_transcript
LMNA-010	ENST00000495341	No protein product	processed_transcript
LMNA-011	ENST00000478063	No protein product	processed_transcript
LMNA-012	ENST00000470199	No protein product	processed_transcript
LMNA-013	ENST00000470835	No protein product	processed_transcript
LMNA-014	ENST00000498722	No protein product	processed_transcript
LMNA-015	ENST00000459904	No protein product	processed_transcript
LMNA-201	ENST00000292302	ENSP00000292302	protein_coding
LMNA-202	ENST00000347559	ENSP00000292304	protein_coding
LMNA-203	ENST00000361308	ENSP00000355292	protein_coding
LMNA-204	ENST00000392353	ENSP00000376164	protein_coding
LMNA-205	ENST00000392355	ENSP00000376165	protein_coding
LMNA-206	ENST00000448611	ENSP00000395597	protein_coding

Příklad alternativní transkripce

114 Chapter 4

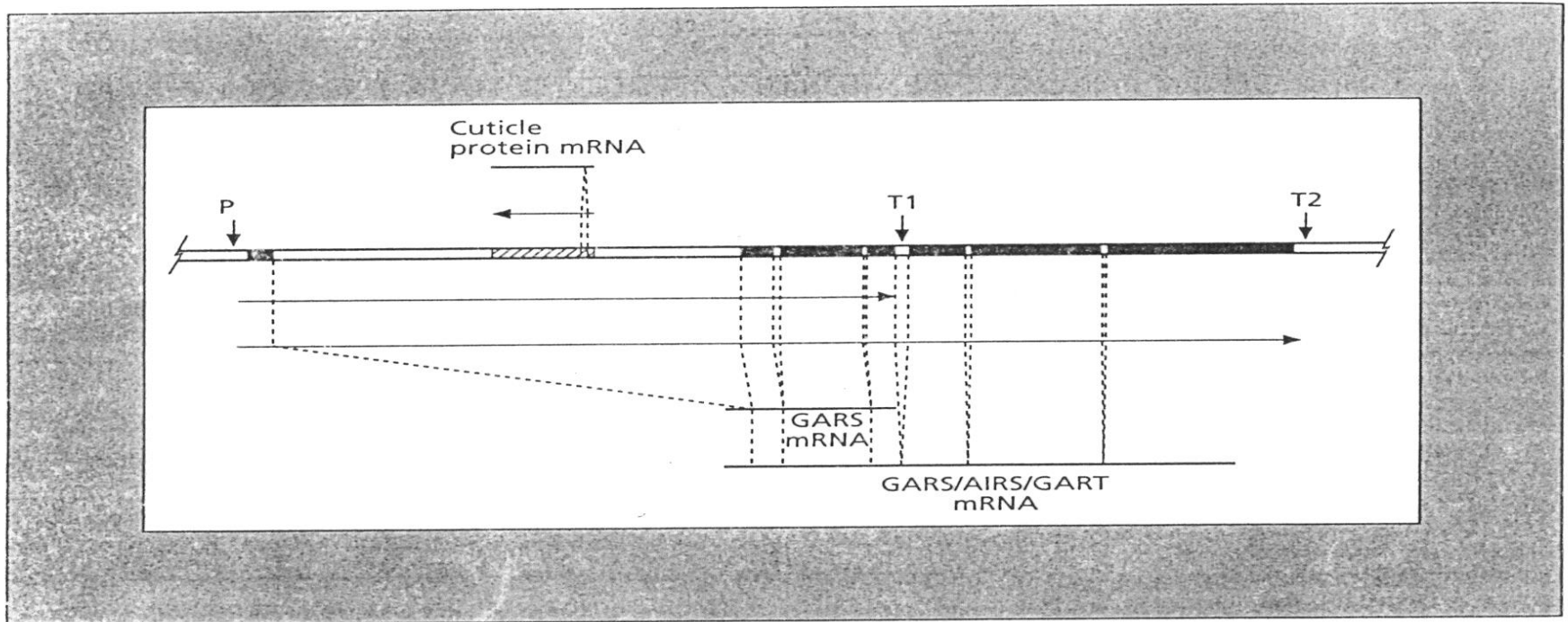


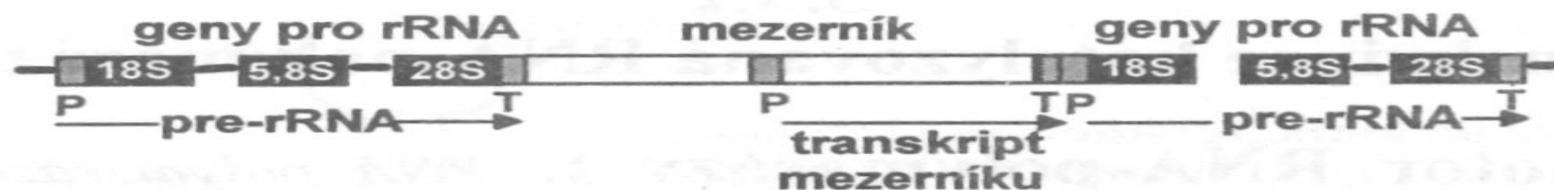
Fig. 4.10 An example of a gene 'nested' within the intron of another gene. The *Gars* gene of *Drosophila*, transcribed and translated from left to right, encodes a trifunctional enzyme responsible for three enzyme activities in the pathway of purine biosynthesis, abbreviated to GARS, AIRS and GART. *Gars* exons and introns are shown as stippled and empty bars, respectively. Because of a partially effective transcription termination sequence (poly-adenylation site, T1) in the fourth intron, two messenger RNAs (mRNAs) are produced, the smaller one encoding only GARS activity. The long first intron contains a gene for cuticle protein, shown as a hatched bar. It itself contains a small intron, and is transcribed and translated from right to left. Thus *Gars* uses one DNA strand as template and the cuticle protein gene uses the other. P = promoter; T1, T2 = transcription terminators. After Henikoff *et al.* (1986).

Geny pro rRNA

Části eukaryotické chromozomové DNA, na nichž jsou lokalizovány geny přepisované do rRNA, se označují zkratkou rDNA. Zvláštností rDNA je, že je tvořena mnohonásobnými repeticemi dvou transkripčních jednotek, z nichž každá je přepisována RNA-polymerázou I do 45S pre-rRNA. V transkripční jednotce jsou seřazeny geny pro 18S-rRNA, 5,8S-rRNA a 28S-rRNA. Mezi oběma jednotkami se nachází dlouhý mezerník obsahující regulační oblasti pro transkripci, t.j. mezerníkový promotor, zesilovače transkripce a terminátor. Základní část promotoru předchází startovacímu nukleotidu. Celý úsek mezerníku od promotoru až po terminátor se přepisuje RNA-polymerázou I (viz schéma transkripční jednotky RNA-polymerázy I a její primární transkript). pre-rRNA obsahuje exony a introny, které podléhají sestřihu. Introny jsou vyštěpovány, ale exony nejsou při sestřihu spojovány. V ribozomech působí jako nespojené RNA, které jsou navzájem drženy prostřednictvím ribozomových proteinů.

Rostoucí eukaryontní buňka musí syntetizovat miliony kopií každého typu molekul ribozomální RNA v každé buněčné generaci, aby mohla vytvořit miliony ribozomů. K tomu potřebuje mnohonásobné kopie rDNA genů, které kódují ribozomální RNA. Lidské buňky obsahují asi 200 rDNA genů na haploidní genom, které jsou rozloženy na pěti různých chromozomech. Geny jsou uspořádány v tandemových sériích.

Transkripce genů pro rRNA



P = *promotor*

T = *terminátor*

— = *mezerník mezi geny pro rRNA*

□ = *mezerník mezi skupinami genů (jednotkami repetice)*







Toto schema nepodává informace o velikostech genů a mezerníků. Zdůrazňuje se v něm toliko jejich pořadí.

Obr. 128

Transkripční jednotka RNA-polymerázy I a její primární transkript

pre-RNA obsahuje exony a introny, které podléhají sestřihu. Introny jsou vyštěpovány, ale exony nejsou při sestřihu spojovány. V ribozomech působí jako nespojené RNA, které jsou navzájem drženy prostřednictvím ribozomových proteinů.

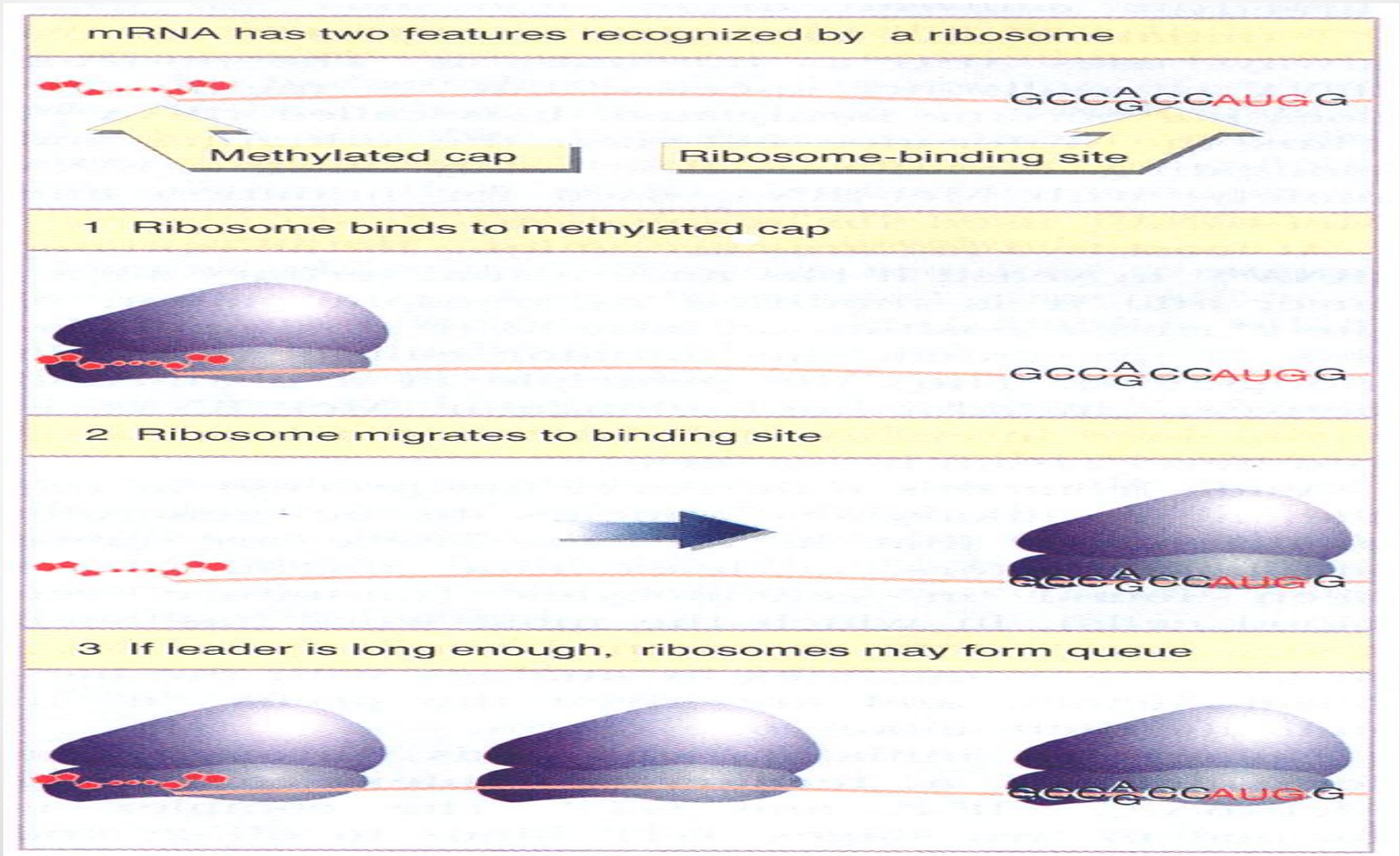
Ribozomy jsou velké ribonukleoproteinové částice, které obsahují více RNA než proteinu a disociují na velké a malé podjednotky.

Ribosomes	Subunits	rRNAs	Proteins
<p>Bacterial</p>  <p>70S mass: 2.5×10^6 D 66% RNA</p>	<p>50S</p>  <p>30S</p> 	<p>23S = 2904 bases 5S = 120 bases</p> <p>16S = 1542 bases</p>	<p>31</p> <p>21</p>
<p>Mammalian</p>  <p>80S mass: 4.2×10^6 D 60% RNA</p>	<p>60S</p>  <p>40S</p> 	<p>28S = 4718 bases 5.8S = 160 bases 5S = 120 bases</p> <p>18S = 1874 bases</p>	<p>49</p> <p>33</p>

28S – transkripce RNA polymeráza I
5,8S – transkripce RNA polymeráza I
18S – transkripce RNA polymeráza I

5S – transkripce RNA polymeráza III

Eukaryotické ribozomy migrují od 5' konce mRNA k vazebnému místu pro ribozom, které obsahuje iniciační kodon AUG



Genetický kód

Tripletový, nepřekrývající se, univerzální, degenerovaný
(více kodonů kóduje jednu aminokyselinu)

Crack the code with Trends in Genetics

trends in GENETICS

First position (5' end)	Second position				Third position (3' end)
	U	C	A	G	
U	Phe (F) Phe (F) Leu (L) Leu (L)	Ser (S) Ser (S) Ser (S) Ser (S)	Tyr (Y) Tyr (Y) Stop Stop	Cys(C) Cys (C) Stop Trp (W)	U C A G
C	Leu (L) Leu (L) Leu (L) Leu (L)	Pro (P) Pro (P) Pro (P) Pro (P)	His (H) His (H) Gln (Q) Gln (Q)	Arg (R) Arg (R) Arg (R) Arg (R)	U C A G
A	Ile (I) Ile (I) Ile (I) Met (M)	Thr (T) Thr (T) Thr (T) Thr (T)	Asn (N) Asn (N) Lys (K) Lys (K)	Ser (S) Ser (S) Arg (R) Arg (R)	U C A G
trends in GENETICS	Val (V) Val (V) Val (V) Val (V)	Ala (A) Ala (A) Ala (A) Ala (A)	Asp (D) Asp (D) Glu (E) Glu (E)	Gly (G) Gly (G) Gly (G) Gly (G)	U C A G

Kodon **UGA** (je i stop kodon) kóduje také **selenocystein** (u eukaryontů). Pro jeho rozlišení jako kodonu pro Sec je nutná vlásenkovitá struktura mRNA v 3' UTR oblasti. (Selenoproteiny: glutathion peroxidáza, selenoprotein P, selenoprotein W, deiodináza typ 1, 2 a 3).

Kodon **UAG** (je i stop kodon) kóduje také **pyrolysin** (Pyl). Nutná přítomnost vlásenkovité sekvence v 3' UTR oblasti. (Vyskytuje se u některých metanogenních archaebakterií).

Geny pro tRNA

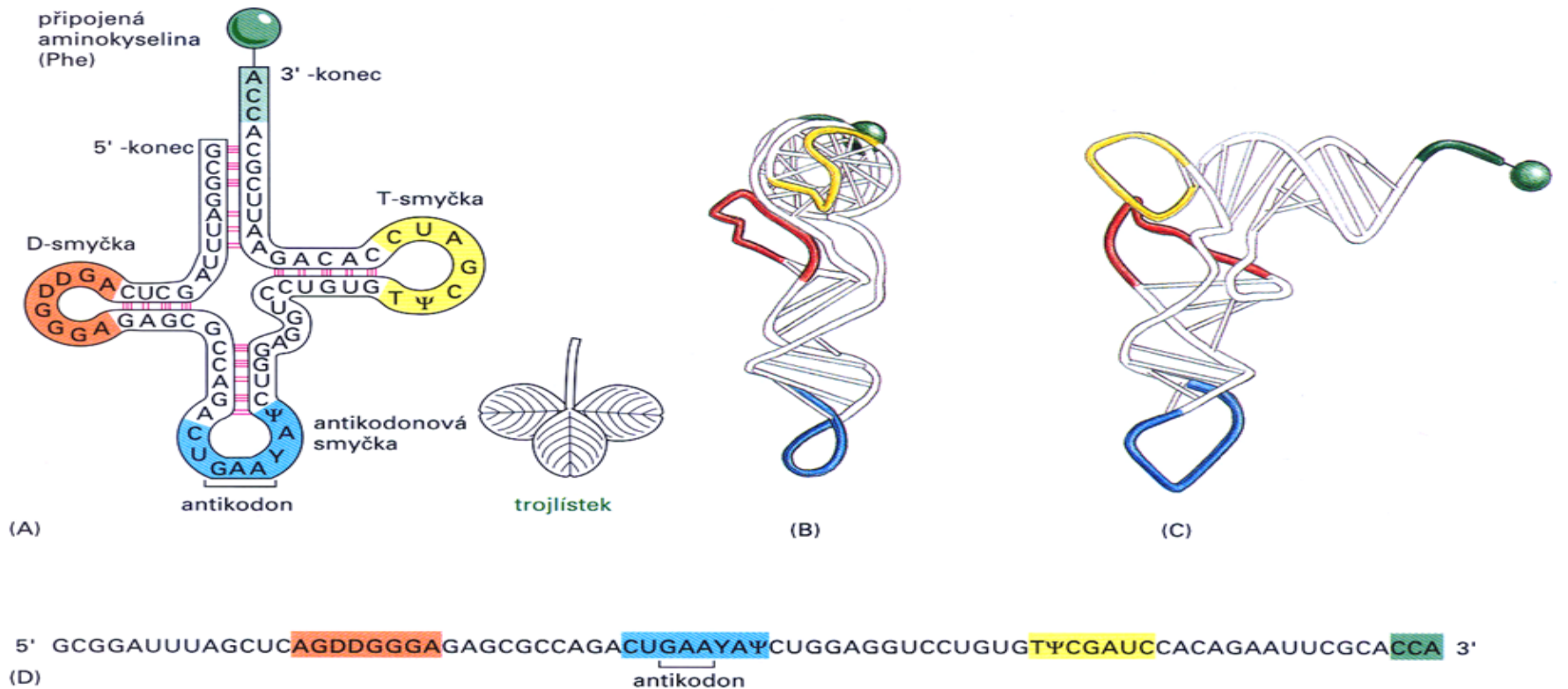
Geny pro tRNA (a 5S-rRNA) se přepisují RNA-polymerázou III do pre-tRNA (a pre-5S-rRNA), které mají **exon – intronovou strukturu**. V lidském genomu bylo identifikováno 497 genů pro tRNA.

Promotory tRNA a 5S-rRNA mají před startovacím nukleotidem vnější regulační část promotoru a za startovacím nukleotidem dva vnitřní boxy.

Post-transkripční úpravy zahrnují vystřížení intronů (za účasti enzymů endonukleázy a RNA ligázy). Posttranskripční úpravy jsou také zodpovědné za výskyt **minoritních bází** v sekvenci tRNA (nejčastěji pseudouridin, dihydrouracil a thiouracil). Celkem je známo několik desítek minoritních bází v tRNA.

Molekula tRNA

(transkripce transkripční jednotky pro tRNA je katalyzovaná RNA polymerázou III)



Geny pro malé RNA

Geny pro malé RNA (small nuclear RNA, snRNA; small nucleolar RNA, snoRNA; mikroRNA, miRNA; small interfering RNA, siRNA; piwi interacting RNA, piRNA)

Transkripty těchto genů jsou krátké úseky RNA (80 – 300 nukleotidů; miRNA asi 22 nukleotidů). Tyto malé RNA mají regulační funkci.

21 různých aminokyselin, které se nacházejí v proteinech

AMINOKYSELINA	POSTRANNÍ ŘETĚZEC
Asparagová kys.	Asp D negativní
Glutamová kys.	Glu E negativní
Arginin	Arg R pozitivní
Lysin	Lys K pozitivní
Histidin	His H pozitivní
Asparagin	Asn N polární bez náboje
Glutamin	Gln Q polární bez náboje
Serin	Ser S polární bez náboje
Threonin	Thr T polární bez náboje
Tyrosin	Tyr Y polární bez náboje

— POLÁRNÍ AMINOKYSELINY —

hydrofilní

AMINOKYSELINA	POSTRANNÍ ŘETĚZEC
Alanin	Ala A nepolární
Glycin	Gly G nepolární
Valin	Val V nepolární
Leucin	Leu L nepolární
Isoleucin	Ile I nepolární
Prolin	Pro P nepolární
Fenylalanin	Phe F nepolární
Methionin	Met M nepolární
Tryptofan	Trp W nepolární
Cystein	Cys C nepolární
Selenocystein	Sec U nepolární

— NEPOLÁRNÍ AMINOKYSELINY —

hydrofobní

Struktura proteinů

Všechny standardní aminokyseliny mají uhlík, na který se váže aminoskupina $-\text{NH}_2$ (jen prolin má NH), karboxylová skupina $-\text{COOH}$, **vodík** a zbytek značený symbolem $-\text{R}$, který představuje boční, postranní řetězec. Ten určuje většinu chemických vlastností aminokyselin. **Skupina $-\text{NH}_2$** na uhlíku α jedné aminokyseliny reaguje se **skupinou $-\text{COOH}$** na uhlíku α jiné aminokyseliny za vzniku **peptidové vazby** (kovalentní vazba) a vyloučení vody. NH_2 konec se považuje za začátek polypeptidového řetězce a označuje se jako **N-konec polypeptidu**. Zakončení karboxylovou skupinou se označuje jako **C-konec polypeptidu**.

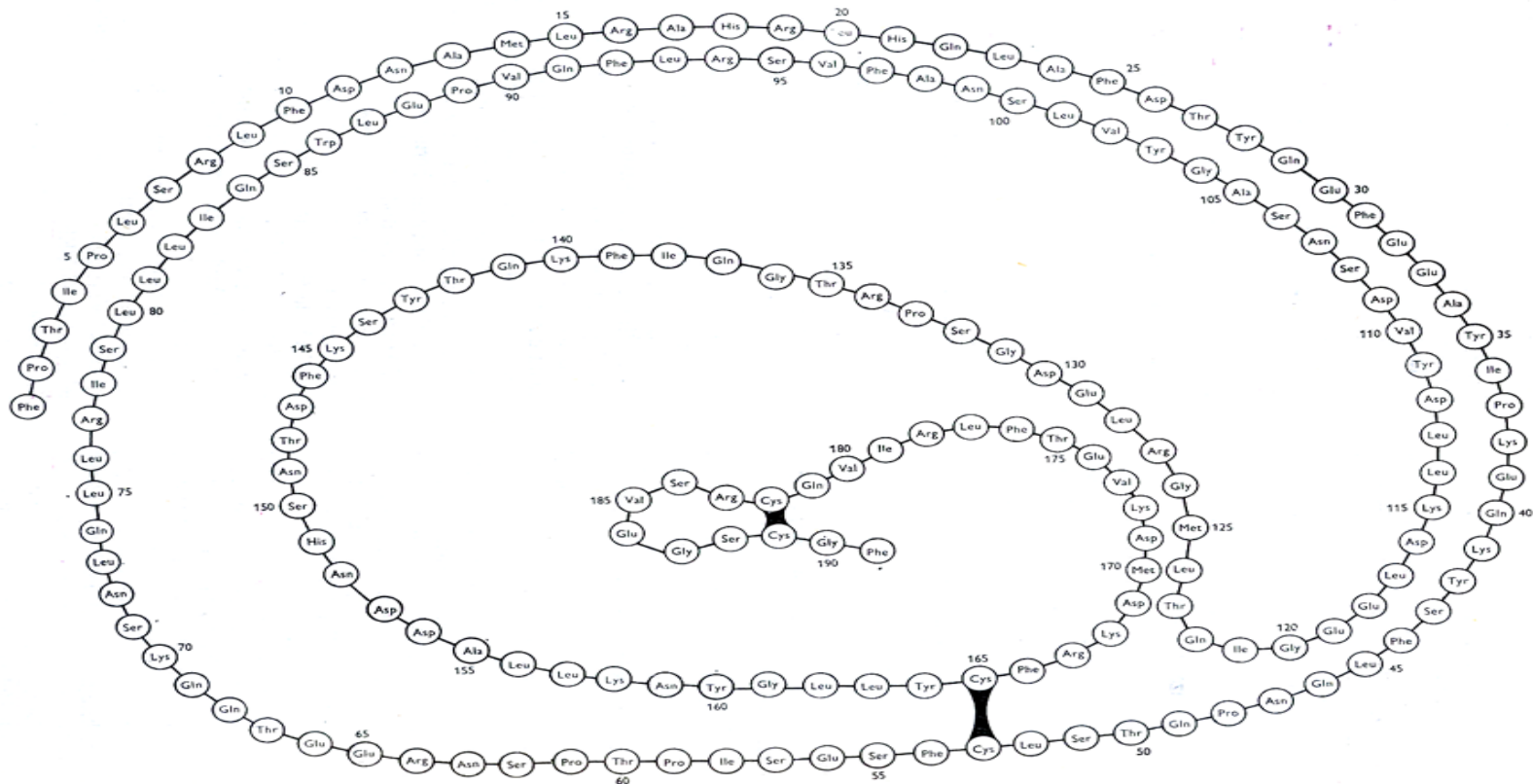
Primární struktura je polypeptidový řetězec sestávající z aminokyselin spojených peptidovou vazbou.

Sekundární struktura – je prostorové uspořádání polypeptidového řetězce (peptidové vazby, vodíkové můstky). Vyskytuje se jako α -šroubovice nebo β -list.

Terciární struktura – asymetrické trojrozměrné uspořádání. Zahrnuje peptidové vazby, vodíkové můstky, Van der Waalsovy interakce, hydrofobní interakce, iontové vazby, disulfidové můstky).

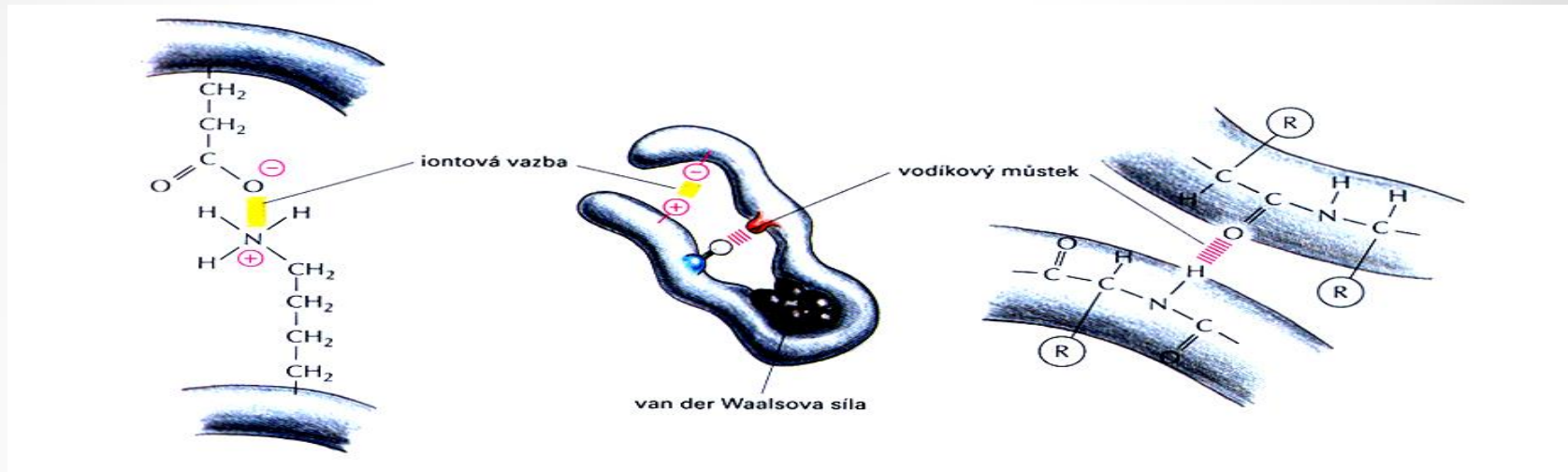
Kvarterní struktura – uspořádání jednotlivých polypeptidových řetězců v molekule proteinu. Zahrnuje všechny vazby, které se vyskytují v terciární struktuře. Podjednotky jsou spojeny vodíkovými můstky, iontovými vazbami, případně i disulfidovými můstky.

Struktura lidského růstového hormonu

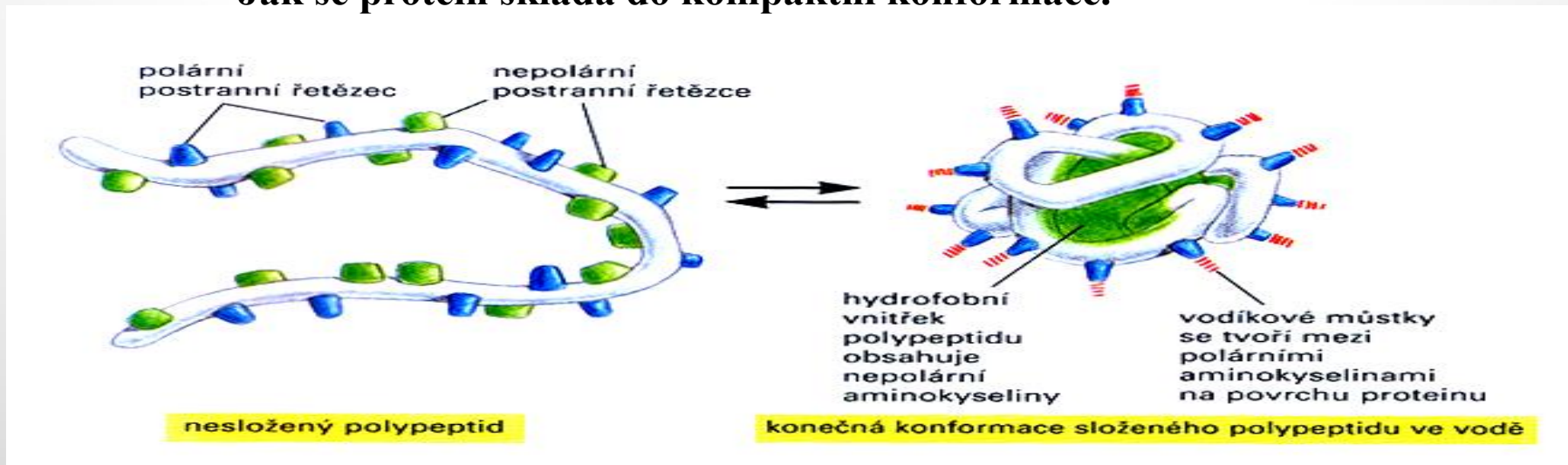


Obr. 20-26. Pravděpodobná struktura lidského růstového hormonu. Čísla označují residua aminokyselin počínaje od N-konce molekuly. (Strukturu zjišťoval C. H. Li.)

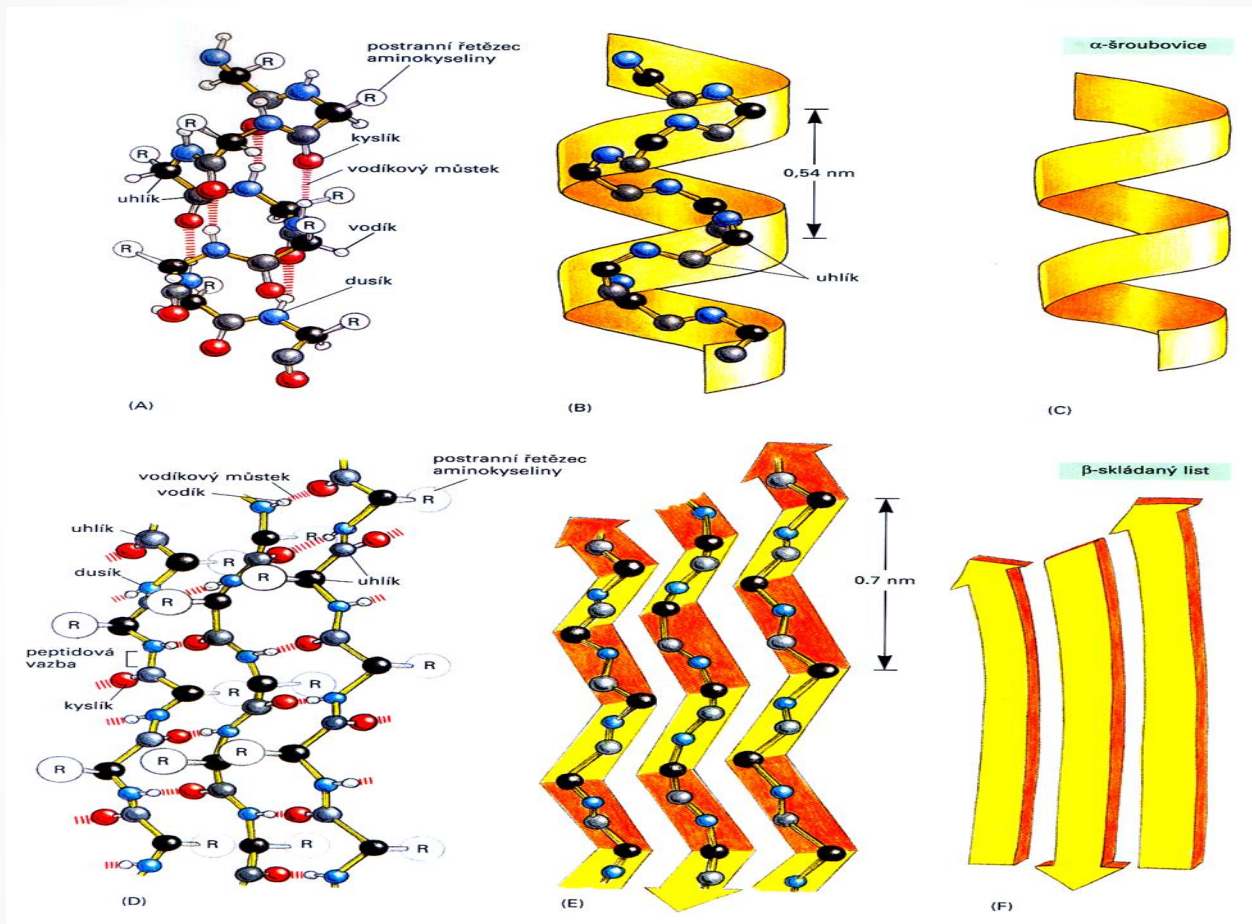
Tři typy nekovalentních vazeb, které napomáhají skládání proteinu



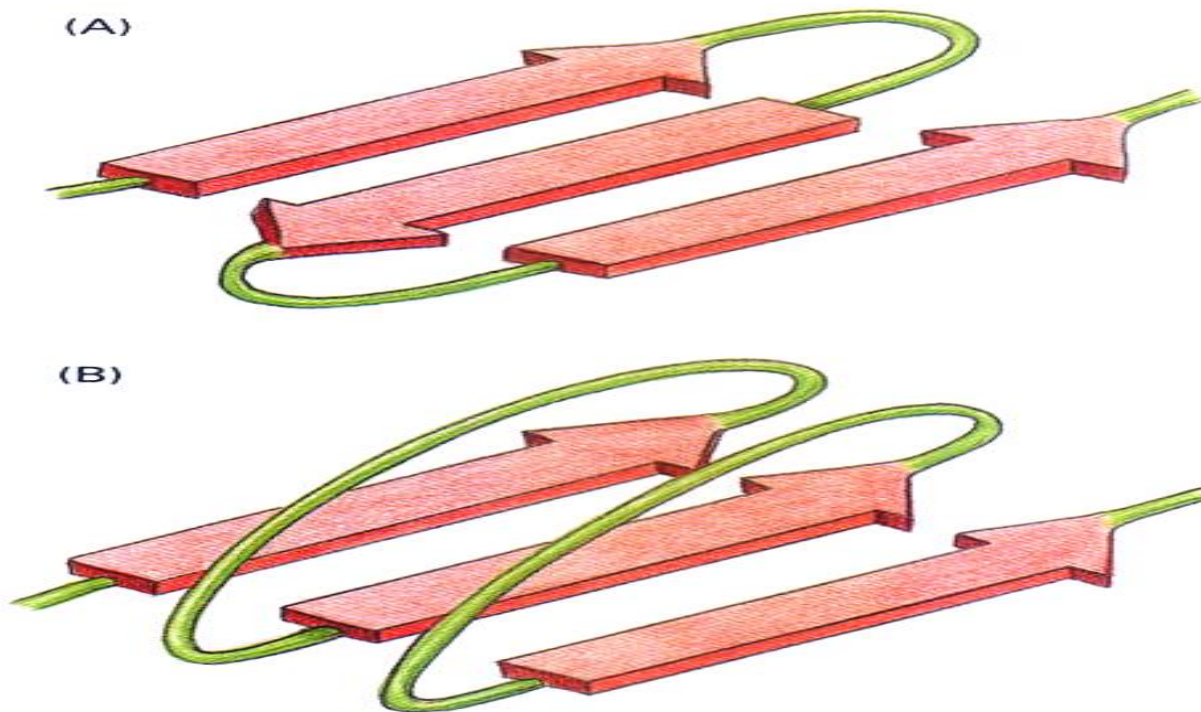
Jak se protein skládá do kompaktní konformace.



Pravidelné uspořádání polypeptidové kostry pozorované v α -šroubovici a v β -listu

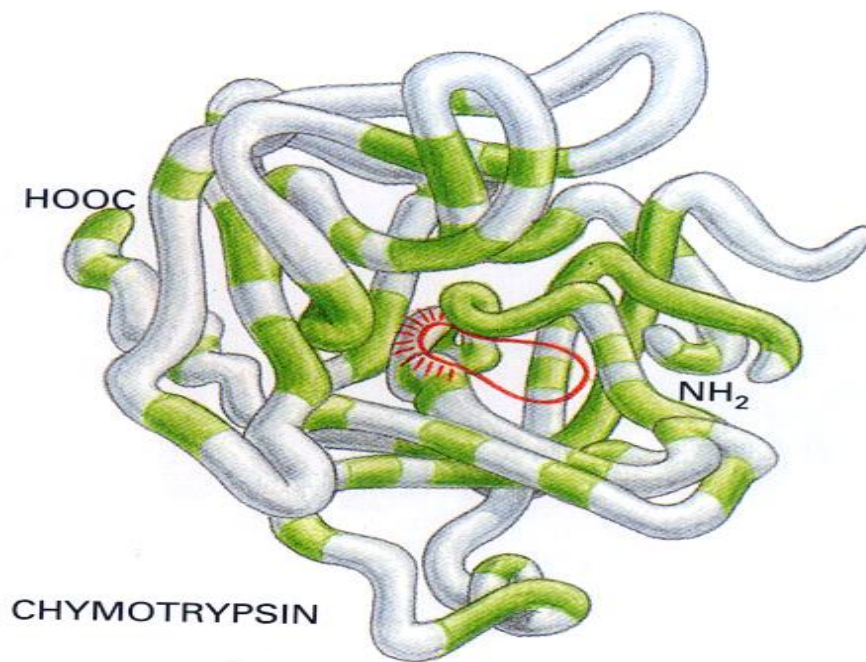
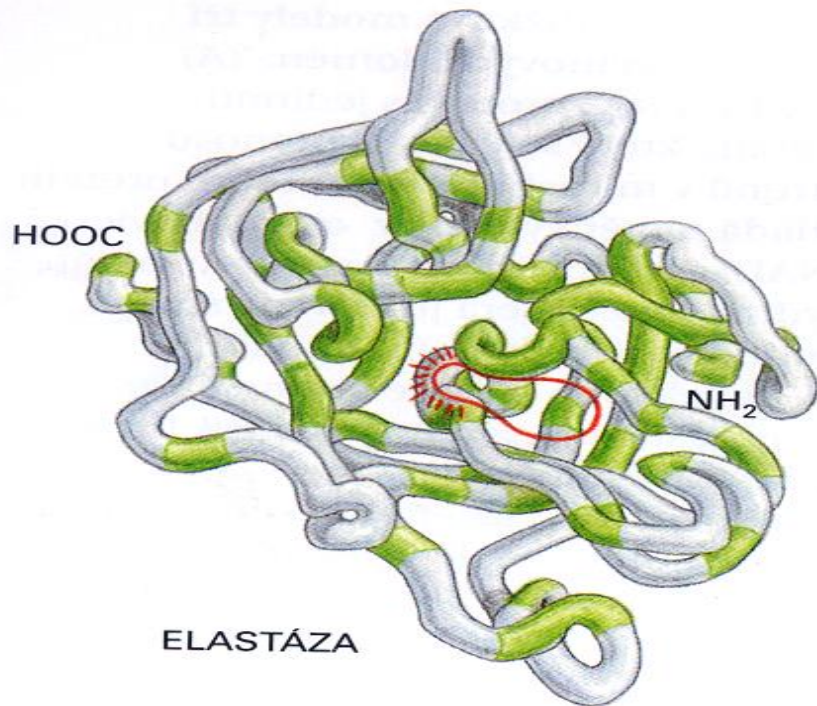


Dva typy struktury β -listu

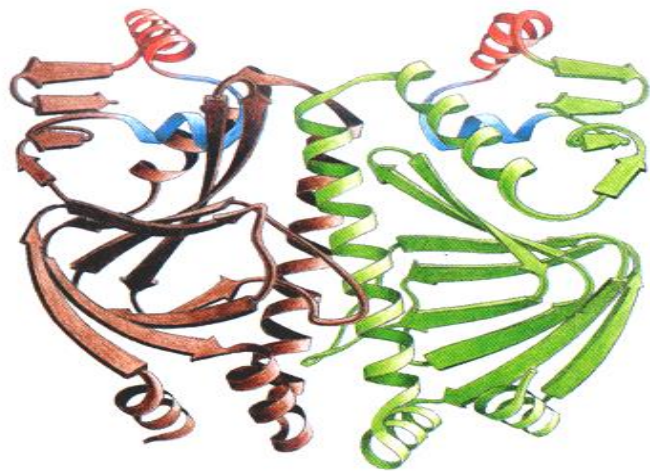


Obrázek 5-10 Dva typy struktury β -listu.
(A) Antiparalelní β -list (viz obrázek 5-9D).
(B) Paralelní β -list. Obě tyto struktury jsou v proteinech běžné.

Srovnání konformací dvou serinových proteáz

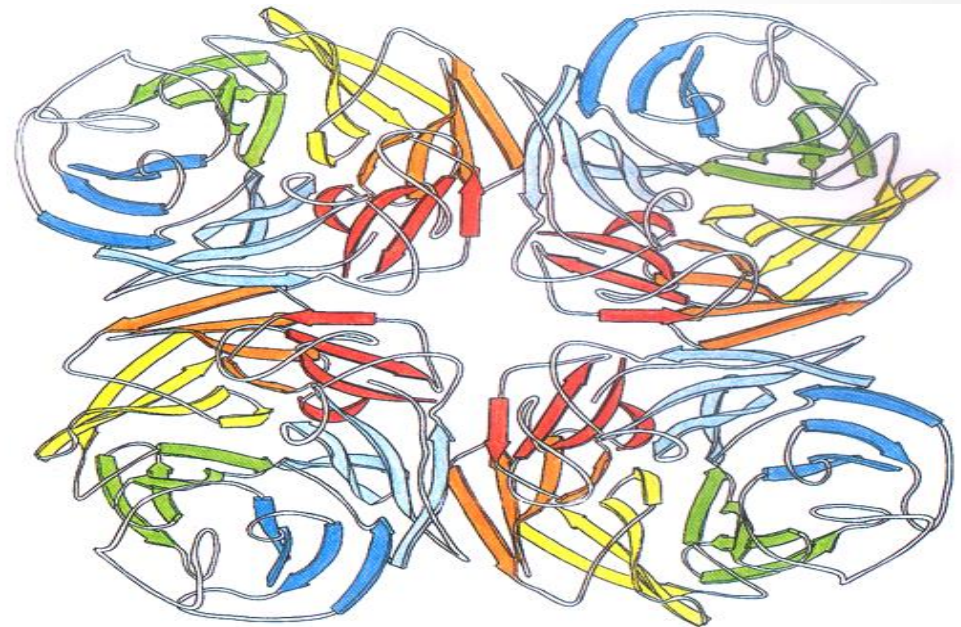


Řada proteinových molekul obsahuje násobné kopie
jediné proteinové podjednotky (**homodimer**, **homotetramer**, atd.)
Když se spojují odlišné podjednotky, označujeme je jako
heterodimer, **heterotrimer**, **heterotetramer**, atd.)



dimer proteinu CAP

(A)



tetramer neuraminidázy

(B)

Post-translační modifikace polypeptidového řetězce (1)

1. Zkrácení a proteolýza polypeptidového řetězce

Odštěpení signálního peptidu, proteolýza inzulinu, melitinu a mnoha dalších.

2. Chemické modifikace aminokyselin

- (deformylace – u prokaryot; formyl se váže na Met až po připojení k tRNA)
- hydroxylace prolinu a lyzinu (některé zbytky Pro a Lys jsou hydroxylovány; hydroxyprolin a hydroxylyzin v kolagenu)
- fosforylace serinu, treoninu a tyrozinu (některé zbytky; u některých proteinů významných pro přenos signálů)
- acetylace lyzinu (v histonech), N-Acetylace
- metylace lyzinu (v histonech)
- N-formylace N-terminálního glycinu v melitinu
- ubiquitinace
- S-nitrosylace

3. Glykosylace

- glykosylace s vazbou na N (vazba na asparagin)
- glykosylace s vazbou na O (vazba na serin nebo treonin)

4. Acylace a myristylace (lipidace)

- připojení postranních lipidových řetězců (mastných kyselin) (na serin, treonin, cystein)
- připojení kyseliny myristové (14-ti uhlíková nasycená mastná kyselina) na N-koncový glycin (některé proteinkinázy zahrnuté v přenosu signálu)

Post-translační modifikace polypeptidového řetězce (2)

5. Vytváření sekundární a terciární struktury

(spontánně, během syntézy nebo po uvolnění z ribozómů)

Většinou začíná ko-translačně (N-terminální část se začíná svinovat, zatímco C-terminální část proteinu je stále syntetizována na ribozómu. Na svinování některých proteinů se podílejí specializované proteiny – **chaperony**. U eukaryontů jsou známy jako „heat shock proteiny (HSP-70; HSP-90).

α -helix a „ β -sheet“ struktury se vytvářejí rychle, protože jsou stabilizovány intramolekulárními vodíkovými vazbami. Terciární struktura může zahrnovat kovalentní vazby ve formě disulfidových můstků (mezi dvěma cysteinovými zbytky), iontové vazby, vodíkové můstky, van der Waalsovy síly a hydrofobní interakce.

6. Vytváření kvarterní struktury

Četné polypeptidy se po translaci spojují vodíkovými můstky, iontovými vazbami, van der Waalsovými silami, hydrofobními interakcemi a disulfidovými můstky – vzniká kvarterní struktura (dimery, trimery, tetramery aj.).

7. Připojení některých dalších prostetických skupin

Prostetické skupiny enzymů a jiných proteinů se připojují k polypeptidovému řetězci až po uvolnění z ribozómů. Připojení může být spontánní nebo katalyzované enzymově (různé koenzymy, hem, kovové ionty apod.).

(A)



(B)

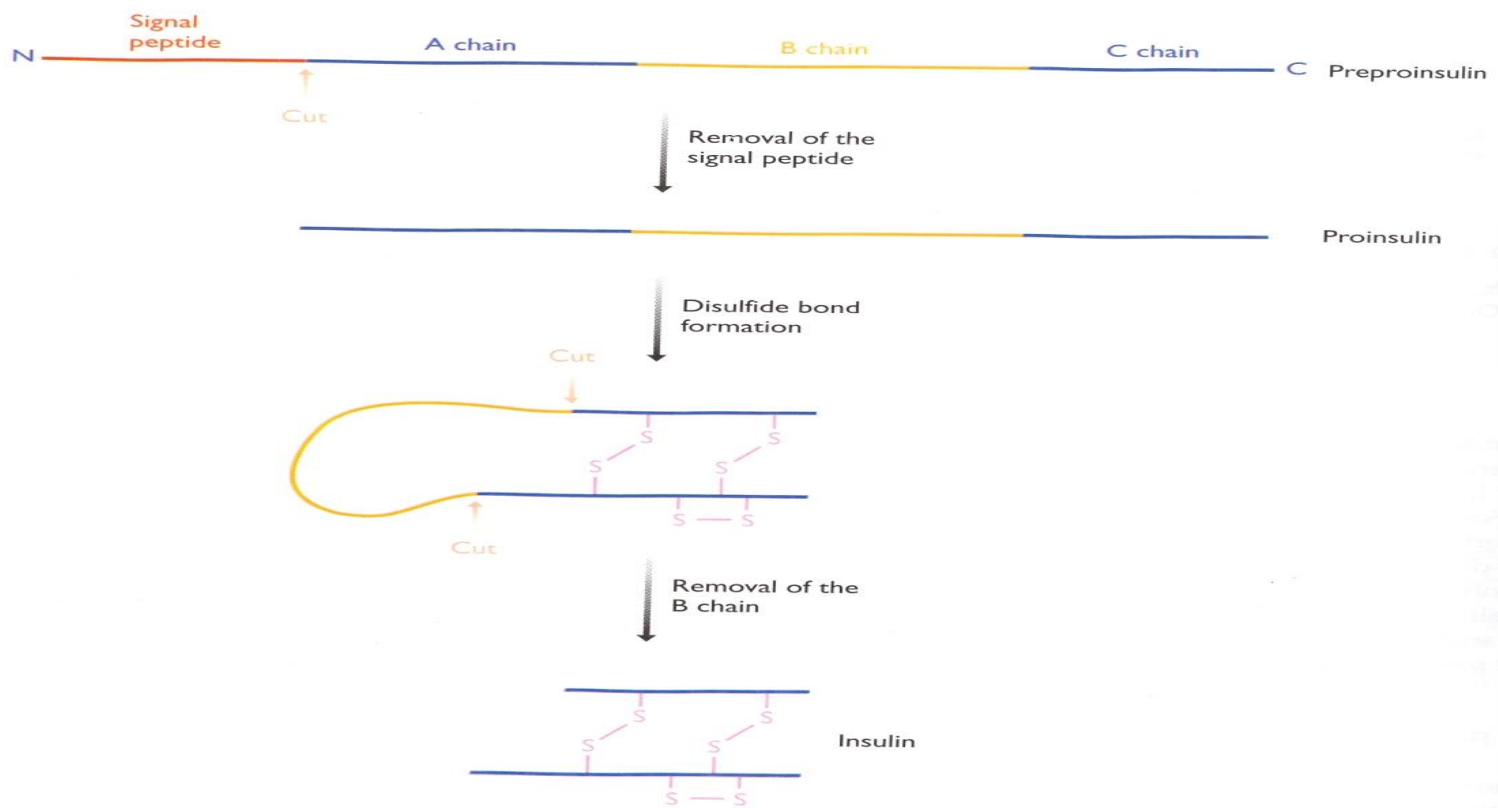


Figure 10.22 Post-translational processing by proteolytic cleavage.

(A) Processing of promelittin, the bee-sting venom. Arrows indicate the cut sites. For the one-letter abbreviations of the amino acids see *Table 7.3*, p. 155. (B) Processing of preproinsulin.

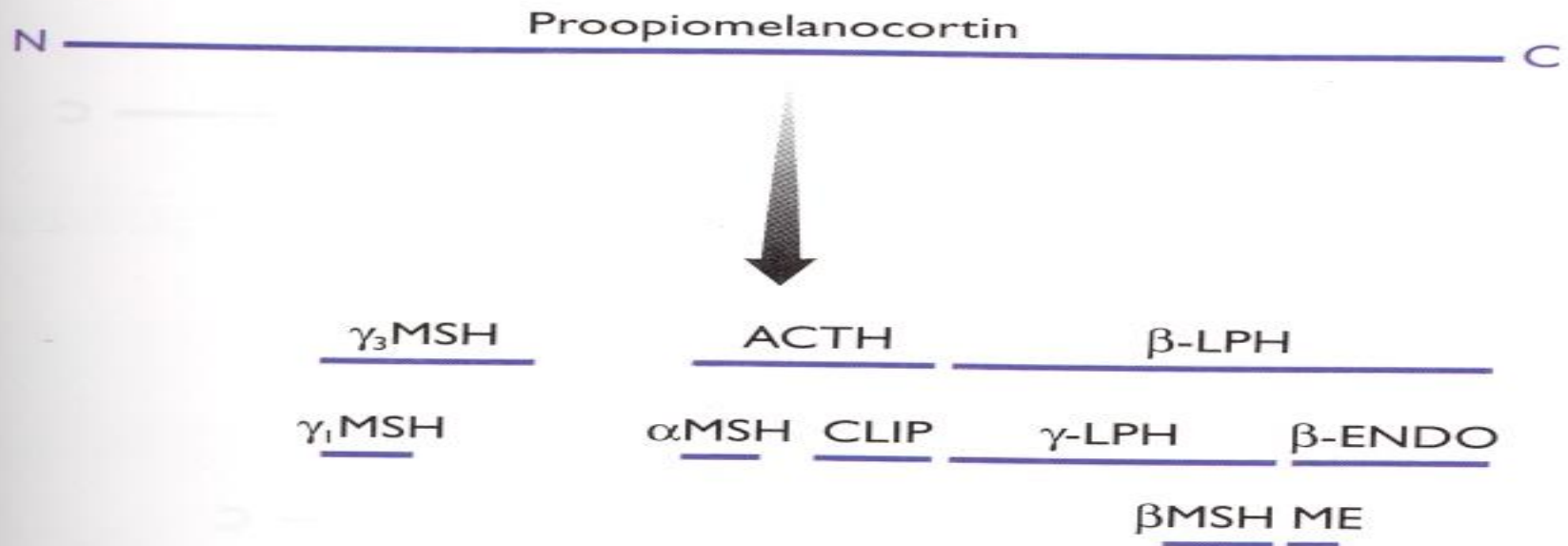
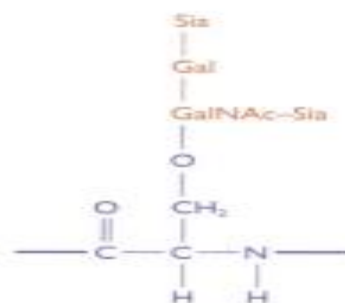


Figure 10.23 Processing of the proopiomelanocortin polyprotein.

Abbreviations: ACTH, adrenocorticotrophic hormone; CLIP, corticotropin-like intermediate lobe protein; ENDO, endorphin; LPH, lipotrophin; ME, met-enkephalin; MSH, melanotropin.

(A) O-linked glycosylation



(B) N-linked glycosylation

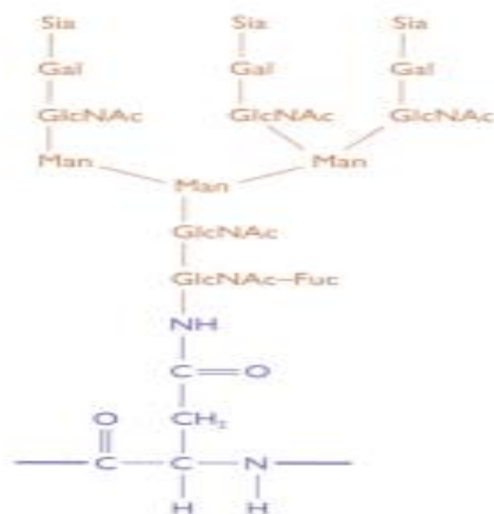


Figure 10.25 Glycosylation.

(A) O-linked glycosylation. The structure shown is found in a number of glycoproteins. It is drawn here attached to a serine amino acid but can also be linked to a threonine. (B) N-linked glycosylation usually results in larger sugar structures than seen with O-linked glycosylation. The drawing shows a typical example of a complex glycan attached to an asparagine amino acid. Abbreviations: Fuc, fucose; Gal, galactose; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; Man, mannose; Sia, sialic acid.

Komplementární analýzy a data – ucelený pohled

DNA - genom

kompletní popis genů v organismu

GENOMIKA

RNA - transkriptom

Kompletní popis mRNA v genomu

TRANSKRIPTOMIKA

PROTEINY - proteom

Kompletní popis proteinů
exprimovaných genomem

PROTEOMIKA

METABOLITY - metabolom

Kompletní popis metabolitů

METABOLOMIKA

INTERAKCE - interactom

protein – DNA, protein- RNA
protein-protein

SYSTÉMOVÁ BIOLOGIE

Organizace genomu u prokaryontů a eukaryontů

Termín **genom** zahrnuje úplný soubor genetického materiálu (DNA, RNA), kterým je daný organismus vybaven. Molekuly DNA a RNA jsou složeny z nukleotidů se specifickými sekvencemi bází. Termín genom používáme jak pro podbuněčné formy (viry, fágy), buňky, jedince, druh, tak také pro vyšší taxonomické jednotky (např. genom savců, krytosemenných rostlin aj.). Poprvé jej použil Winkler v roce 1920.

Genomy mohou být malé i velké.

Nejmenší genomy rostlinných viroidů mají několik set nukleotidů.

Největší genomy eukaryontní mají někteří obojživelníci (až 10^{10} bp) a některé vyšší rostliny (až 10^{11} bp).

Genomy mohou být jednoduché nebo složité. Každá z molekul informační soustavy může být vyjádřena jen v jednom exempláři, nebo v dvojité či vícenásobné dávce. Genom buňky může být soustředěn do jednoho prostoru, nebo rozdělen do několika oddílů.

Genom eukaryontní buňky je rozdělen do dvou oddělených buněčných oddílů – jaderného a cytoplasmatického prostoru.

V jádře se nachází větší část – proto ji nazýváme hlavní, centrální, nebo jaderný genom.

V cytoplasmě je cytoplasmatický, nebo vedlejší genom.

Jaderný genom člověka je rozdělen do 46 molekul DNA, každá v jednom chromozómu. Cytoplasmatický genom je tvořen větším množstvím identických kopií menších molekul DNA, lokalizovaných v mitochondriích v podobě “nahých“ kruhových molekul DNA.

Existují značné rozdíly v organizaci a složitosti genomů u prokaryontů a eukaryontů. Obecně lze říci, že se vzrůstající složitostí mnohobuněčných organismů vzrůstá složitost genomu. Množství DNA nebo počet nukleotidů v genomu však nijak nesouvisí s evoluční vyspělostí.

Genom virů a fágů

Viroidy a viry obsahují převážně jednovláknovou RNA. Některé viry obsahují dvouvláknovou RNA, uspořádanou do tvaru dvojité šroubovice.

Některé viry mají dvouvláknovou DNA nebo jednovláknovou DNA (např. některé fágy E. coli), ale jsou známy i fágy E. coli, které mají jednovláknovou RNA.

Některé viry mají DNA lineární, jiné kruhovou.

RNA viry – nejmenší RNA genom obsahuje 3 geny.

DNA viry – ssDNA geminiviry – 5 – 6 genů

dsDNA caulimoviry – 6 – 8 genů

badnaviry – nejméně 3 geny

chlorellaviry – více než 50 genů (genom 350 kb)

U virů probíhají s nejvyšší četností substituce, delece a inserce jednoho nebo několika nukleotidů.

Maximální využívání kapacity genomu virů je zajišťováno překrýváním čtecích rámců, mnohačetnou funkcí jednotlivých bílkovin nebo variabilitou délky a funkce jedné konkrétní bílkoviny. Větší nebo menší překrývání **čtecích rámců je u mnoha skupin virů běžné: v extrémních případech mohou být všechny čtecí rámce v určitém místě funkční a mohou kódovat různé bílkoviny.**

Některé nekódující oblasti na koncích, popřípadě uprostřed genomu, jsou regulačními místy řídícími replikaci a translaci. Struktury na 5' konci často řídí přepis jednotlivých čtecích rámců a obsahují zesilovače translace. Na 3' konci genomu bývají sekvence s funkcemi terminátorů.

Velikosti genomů

TABLE I

Error rates and genome sizes of RNA viruses as compared to autonomous organisms

Virus ^a	Genome size: v (number of nt or bp)	Error rate: $(1 - q)$ (per replication round and per nt) ^b
RNA		
Bacteriophage Q _B	4 200	3×10^{-4}
Polio-1 virus	7 400	3×10^{-5}
Vesicular stomatitis virus	11 000	1×10^{-4}
Foot and mouse disease virus	8 400	1×10^{-4}
Influenza-A virus	14 000	6×10^{-5}
Sendai virus	15 000	3×10^{-5}
HIV-1 (AIDS virus)	10 000	1×10^{-4}
Avian myeloblastosis virus	7 000	5×10^{-5}
DNA		
Bacteriophage M13	6 400	7×10^{-7}
Bacteriophage λ	48 500	8×10^{-8}
Bacteriophage T4	166 000	2×10^{-8}
<i>E. coli</i>	4.7 million	7×10^{-10}
Yeast (<i>Saccharomyces cerevisiae</i>)	13.8 million	3×10^{-10}
<i>Neurospora crassa</i>	41.9 million	1×10^{-10}
Human	3 billion	$\sim 10^{-12}$

Typy virových genoforů

Názvy hlavních skupin virů podle typu jejich genoforu

Typ virového genoforu	Označení skupiny virů s příslušným genoforem
dsDNA lineární nebo kružnicová	dsDNA-viry
ssDNA lineární nebo kružnicová	ssDNA-viry
dsRNA	dsRNA-viry
pozitivní ssRNA	+ssRNA-viry
negativní ssRNA	-ssRNA-viry
RNA určená k zpětné transkripci v hostitelské buňce do DNA, jejíž kopie po replikaci slouží jako matrice pro syntézu virového RNA-genoforu	RNA-viry se zpětnou transkriptázou
DNA určená k transkripci v hostitelské buňce do RNA jako matrici pro zpětnou transkripci do DNA opět s funkcí virového genoforu	DNA-viry se zpětnou transkriptázou
Vysvětlivky	
<p>dsDNA = dvouřetězcová DNA ssDNA = jednořetězcová DNA dsRNA = dvouřetězcová RNA ssRNA = jednořetězcová RNA Zkratky (symboly) pro názvy virů s příslušným genoforem uvedené ve druhém sloupci tabulky se čtou tímto způsobem: dsDNA-viry = viry s dvouřetězcovou DNA ssDNA-viry = viry s jednořetězcovou DNA dsRNA-viry = viry s dvouřetězcovou RNA +ssRNA-viry = viry s pozitivní RNA -ssRNA-viry = viry s negativní RNA Názvy posledních dvou skupin virů se čtou tak, jak je uvedeno ve druhém sloupci.</p>	

- ◆ dsDNA lineární nebo kružnicová;
- ◆ ssDNA lineární nebo kružnicová;
- ◆ dsRNA lineární;
- ◆ pozitivní ssRNA lineární;
- ◆ negativní ssRNA lineární;

Bakteriofág lambda

VECTORS

EMBL 3, EMBL4

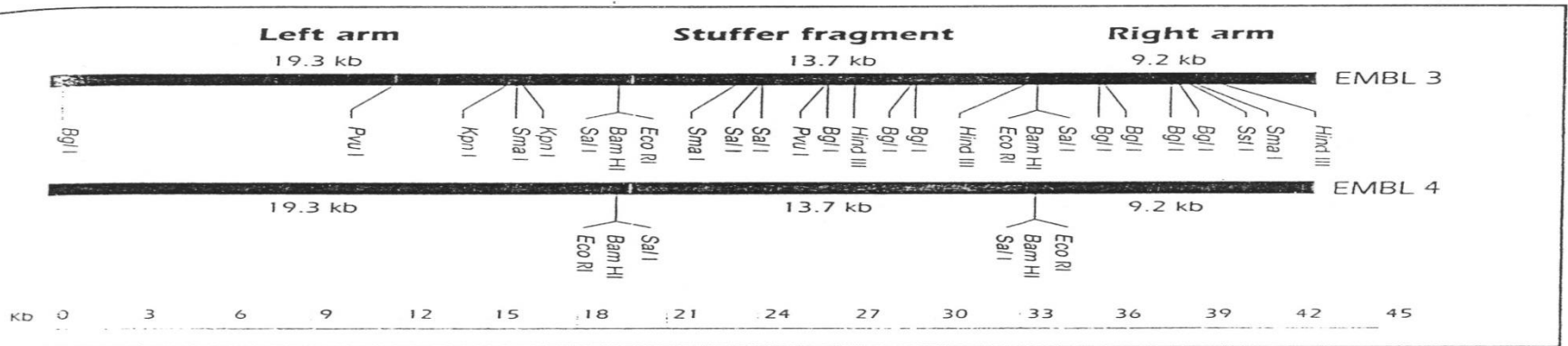


Figure 12.38 Genomic cloning vectors EMBL3 and EMBL4.

Genom prokaryontů

Prokaryontní (jednobuněčné) organismy mají dvouvláknovou kruhovou molekulu DNA (nukleoid) o velikosti 2 – 4 milionů bp. Genom prokaryontů je maximálně využit. Strukturální geny s regulačními oblastmi na sebe navazují a jen výjimečně jsou proloženy nekódujícími sekvencemi. Strukturální geny jsou většinou v jednom exempláři. Velikosti genomů jsou v rozmezí 6 – 8 tisíc genů. Introny (nekódující sekvence uvnitř genu) a mezerníky (nekódující sekvence mezi geny) jsou přítomny jen výjimečně. (Celkový obsah nekódující DNA u *E. coli* je asi 11%).

Buněčné organely

Plazmidy jsou malé kruhové molekuly extrachromozomální DNA bakterií. DNA je ve formě dvoušroubovice a velikosti jsou v rozmezí 1 000 až 200 000 párů bází. Mohou kódovat různé doplňující vlastnosti (rezistence proti antibiotikům; fixace vzdušného dusíku aj.). V jedné bakteriální buňce může být několik set plazmidů.

Některé typy plazmidů (s upravenými sekvencemi) se využívají v molekulární genetice, zejména pro klonování cizorodých fragmentů DNA (tj. zvolených úseků DNA). Využívá se při tom odolnosti k různým antibiotikům a přítomnosti mnohonásobných restričních míst. Tyto plazmidy jsou malé – např. pUC18 má 2690 bp.

Mitochondrie jsou cytoplasmatické organely eukaryontních organismů, obsahující jednu nebo několik molekul DNA ve formě kruhové dvoušroubovice. V buňce jich může být několik stovek, ale i sto tisíc. Mitochondriální DNA (mtDNA) je velmi malá; např. u savců má velikost asi 16 500 bp.

Genetický kód mitochondriální DNA má určité odlišnosti od univerzálního genetického kódu.

Chloroplasty jsou buněčné organely rostlin, ve kterých probíhá fotosyntéza. V jedné buňce je přítomno 10 až 100 chloroplastů a v každém chloroplastu je 10 až 50 molekul DNA (cpDNA). Chloroplastová DNA je ve formě cirkulární molekuly, ale také se uvádí, že je ve formě lineární. cpDNA obsahuje asi 100 genů, jednak pro fotosyntézu a asi 100 proteinů. Také geny pro rRNA (20 typů) a tRNA (37 typů). Chloroplastové geny obsahují introny.

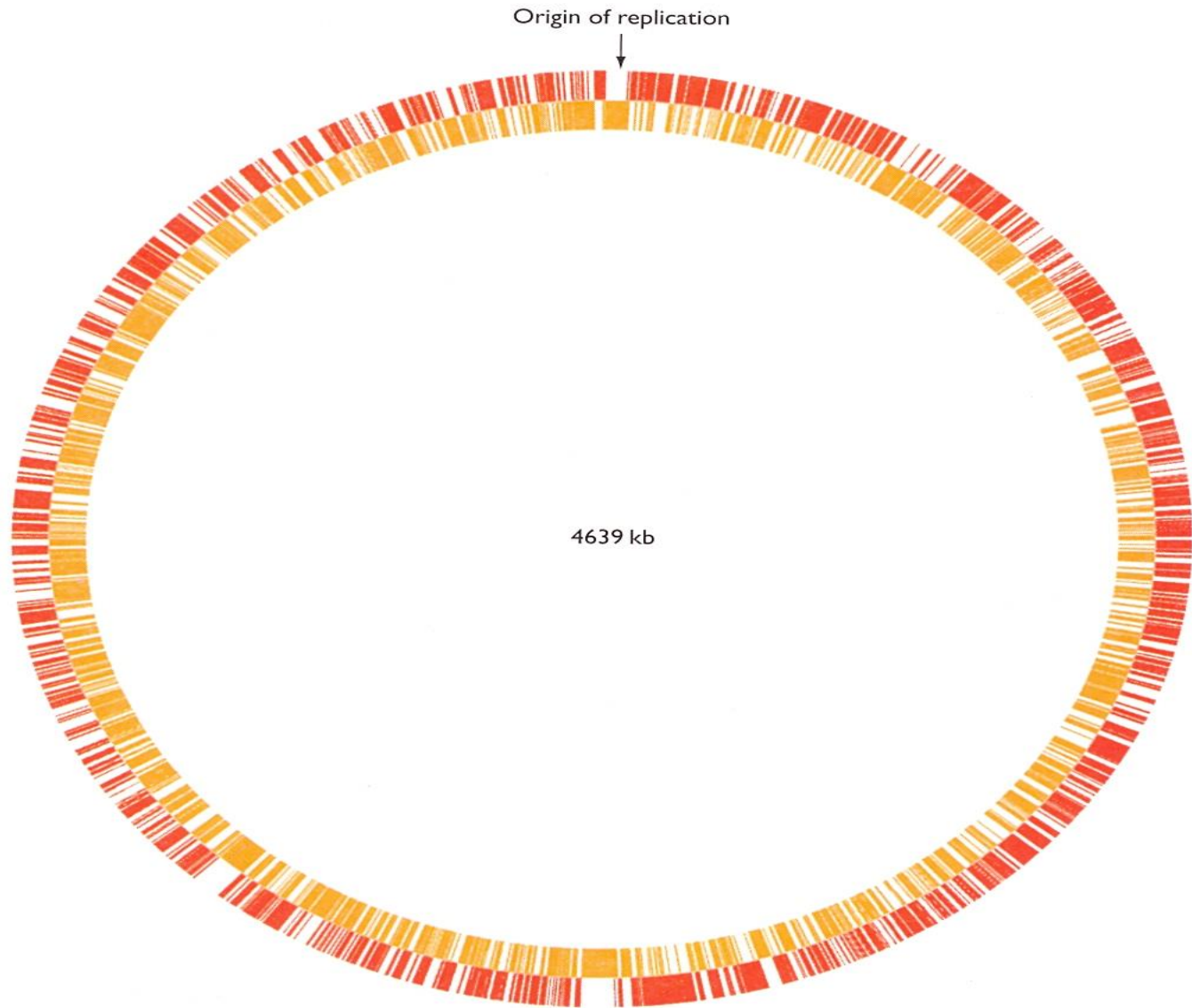


Figure 6.13 The *E. coli* genome.

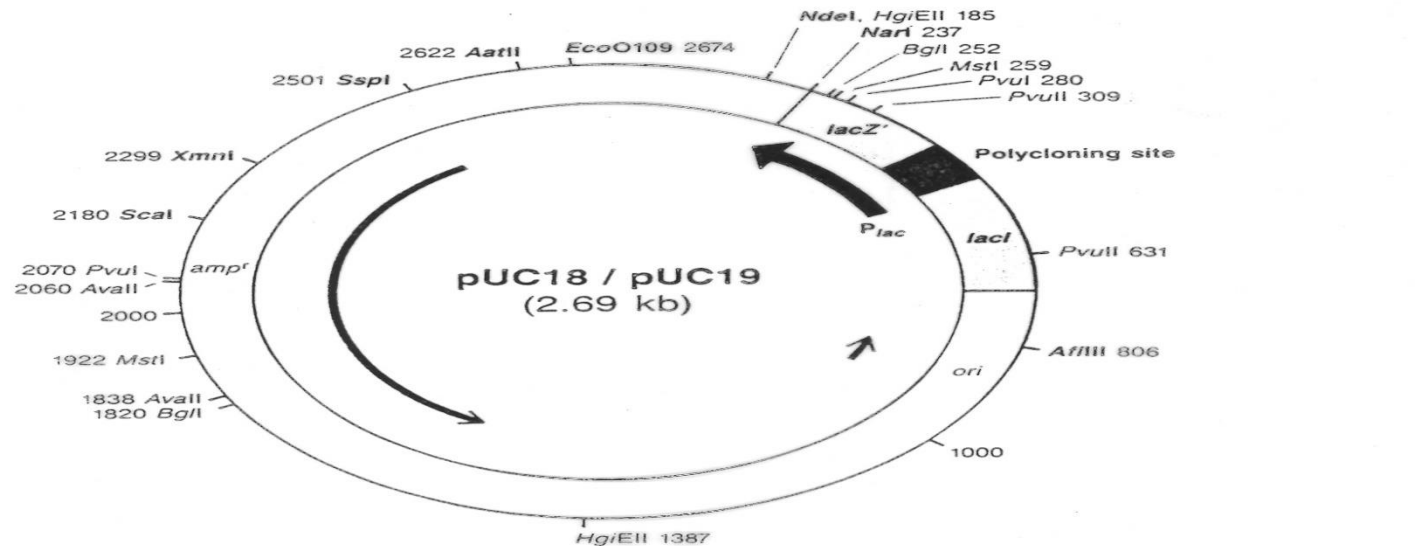
The map is shown with the origin of replication (Section 12.3.1) positioned at the top. Genes on the outside of the circle are transcribed in the clockwise direction and those on the inside are transcribed in the anticlockwise direction. Image supplied courtesy of Dr FR Blattner, Laboratory of Genetics, University of Wisconsin-Madison. Reproduced with permission.

Plazmid pUC18

pUC18, pUC19

These vectors are identical except that they contain polycloning sites arranged in opposite orientations. pUC plasmids lack the *rop* gene (see page 1.3), which normally is located close to the origin of DNA replication and is involved in the control of copy number. As a result, these plasmids replicate to a much higher copy number than do other plasmids that carry a pMB1 (or ColE1) origin. pUC vectors express the amino-terminal fragment of the *lacZ* gene product (β -galactosidase) and display α -complementation in appropriate hosts (see pages 1.8–1.9). Recombinants can therefore be identified by histochemical screening.

References: Messing (1983), Norrander et al. (1983), and Yanisch-Perron et al. (1985).



Polycloning Sites

pUC18

1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	7	8	
Thr	Met	Ile	Thr	Asn	Ser	Ser	Ser	Val	Pro	Gly	Asp	Pro	Leu	Glu	Ser	Thr	Cys	Arg	His	Ala	Ser	Leu	Ala	Leu	Ala	
ATG	ACC	ATG	ATT	ACG	AAT	TCG	AGC	TCG	GTA	CCC	GGG	GAT	CCT	CTA	GAG	TCG	ACC	TGC	AGG	CAT	GCA	AGC	TTG	GCA	CTG	GCC
				<i>Eco</i> RI		<i>Sac</i> I		<i>Kpn</i> I		<i>Sma</i> I <i>Xma</i> I		<i>Bam</i> HI		<i>Xba</i> I		<i>Sal</i> I <i>Acc</i> I <i>Hinc</i> II		<i>Pst</i> I		<i>Sph</i> I		<i>Hind</i> III				

pUC19

1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	5	6	7	8	
Thr	Met	Ile	Thr	Pro	Ser	Leu	His	Ala	Cys	Arg	Ser	Thr	Leu	Glu	Asp	Pro	Arg	Val	Pro	Ser	Ser	Asn	Ser	Leu	Ala	
ATG	ACC	ATG	ATT	ACG	CCA	AGC	TTG	CAT	GCC	TGC	AGG	TCG	ACT	CTA	GAG	GAT	CCC	CGG	GTA	CCG	AGC	TCG	AAT	TCA	CTG	GCC
				<i>Hind</i> III		<i>Sph</i> I		<i>Pst</i> I		<i>Sal</i> I <i>Acc</i> I <i>Hinc</i> II		<i>Xba</i> I		<i>Bam</i> HI		<i>Sma</i> I <i>Xma</i> I		<i>Kpn</i> I		<i>Sac</i> I		<i>Eco</i> RI				

In pUC18, the *Eco*RI site lies immediately downstream from *P_{lac}*.
In pUC19, the *Hind*III site lies immediately downstream from *P_{lac}*.

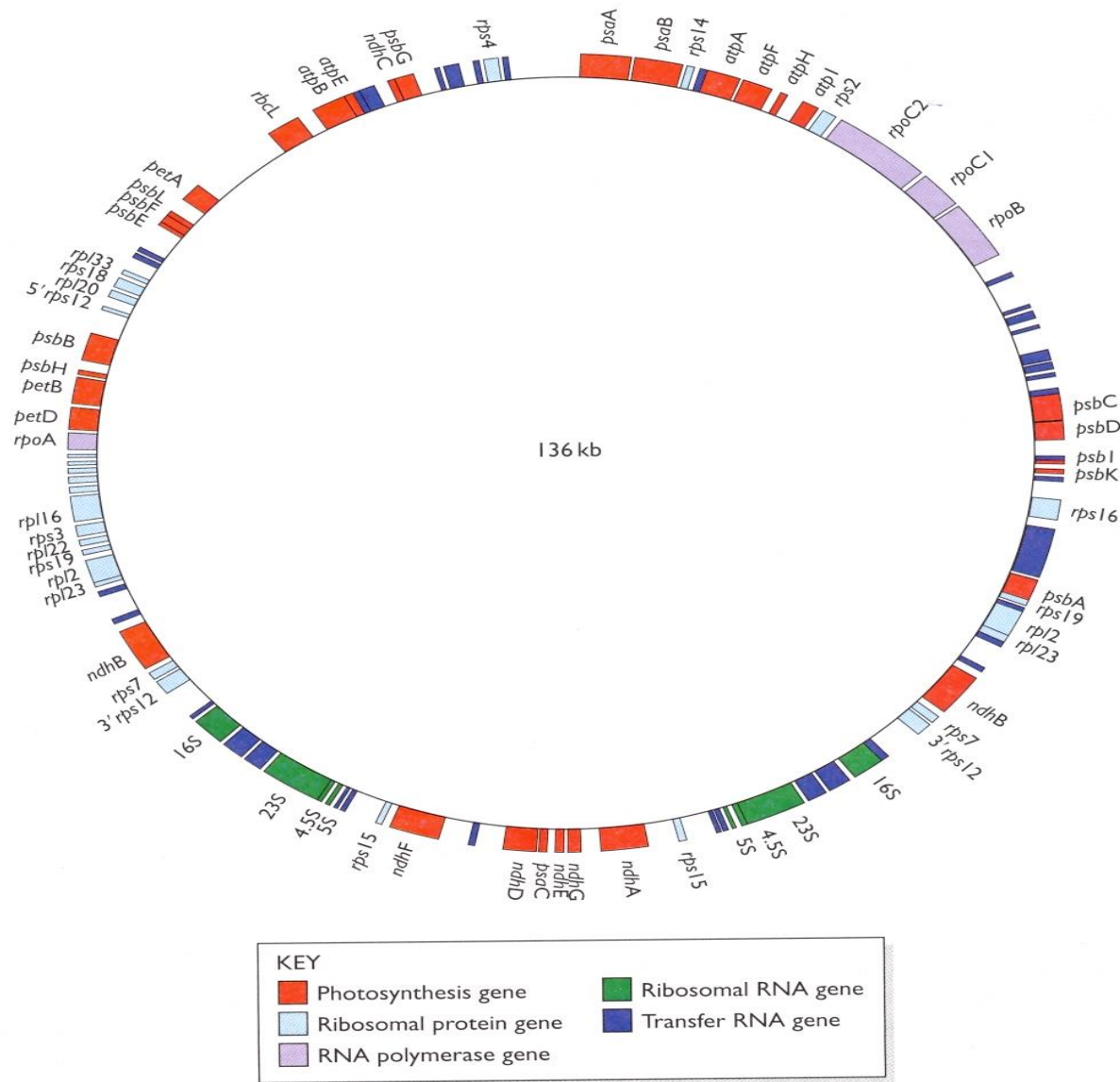


Figure 6.10 The rice chloroplast genome.

Only those genes with known functions are shown. A number of the genes contain introns which are not indicated on this map. These split genes include several of those for tRNAs, which is why the tRNA genes are of different lengths even though the tRNAs that they specify are all of similar size.

Mitochondriální DNA

Mitochondrie jsou orgány eukaryotických buněk. Slouží jako energetické centrum buňky. Obsahují vlastní genetickou informaci a dokonce i vlastní proteosyntetický aparát.

Mitochondrie jsou kulovité až podlouhlé orgány široké asi 1 μ m a dlouhé až 10 μ m. Jsou uzavřeny 2 membránami - vnější mitochondriální membránou a vnitřní mitochondriální membránou, která vybíhá v četné výběžky - tzv. kristy. Prostor mezi oběma membránami označujeme jako intramembranózní, prostor ohraničený vnitřní membránou označujeme jako interkristální. V prostoru mezi kristami najdeme matrix mitochondrie - amorfní hmotu, obsahující proteiny, koncentrovanou směs enzymů (včetně enzymů Krebsova cyklu a β -oxidace mastných kyselin), mitochondriální genom, speciální mitochondriální ribozomy a RNA. Většina buněk obsahuje alespoň 1000 mitochondrií.

Mitochondrie je energetické centrum buňky. Probíhá zde cyklus kyseliny citrónové (Krebsův cyklus, citrátový cyklus), kde je oxidován získaný acetylkoenzym A (acetyl-CoA). Cyklus vydává CO₂ a energii ve formě vysokoenergetických elektronů. Ty jsou z cyklu vyneseny nosiči (NADH, FADH₂), které v dýchacím řetězci (elektrontransportní řetězec) odevzdávají tyto elektrony za vzniku vody a syntézy ATP díky protonovému gradientu, získanému transportem elektronů vnitřní membránou. Bez mitochondrií by buněčné zásoby ATP byly odkázány jen na méně účinné metody získání energie, jako je např. glykolýza.

Díky částečné nezávislosti mitochondrií na zbytku buňky je označujeme jako tzv. semiautonomní orgány. Z evolučního hlediska je považujeme za bakteriální buňky, které pronikly do jiné buňky a tomuto prostředí se dokázaly přizpůsobit (tzv. endosymbióza). Pro bakteriální původ hovoří např. dvojitá membrána, cirkulární molekula DNA nebo

Nyní k vlastnímu genomu mitochondrie. Bylo prokázáno, že tento je jako u bakterií tvořen kruhovou molekulou DNA (mtDNA) o velikosti (u člověka) 16569 nukleotidových párů, které představují 37 genů, z toho 24 genů kóduje různé části proteosyntetického aparátu mitochondrie (2 typy rRNA a 22 tRNA) a zbytek se podílí na malé části enzymatické výbavy mitochondrie. Naprostá většina mitochondriálních proteinů je ale kódována v jádře buňky a tyto sem musí být přeneseny z místa vzniku v cytosolu. Stejně jako u bakterií, nejsou v mitochondriálním genomu nekódující sekvence (introny).

Proteosyntetický aparát mitochondrie se také výrazně liší od normálního aparátu eukaryotické buňky. Liší se např. velikostí ribozomálních podjednotek, menším počtem typů molekul tRNA (pouhých 22!) a odlišnosti najdeme i v genetickém kódu, zejména v kódování methioninu a stop-kodonů.

Přes zdánlivou nepodstatnost mohou být mutace v mtDNA zdrojem různých závažných chorob. Vzhledem k tomu, že **všechny mitochondrie v zygotě jsou z vajíčka** (všechny mitochondrie ze spermie jsou odbourány), mluvíme zde o tzv. **maternální dědičnosti** (tedy jeden z typů nemendelovské dědičnosti). Jinými slovy, pokud je matka přenašečkou mutace v mtDNA, pak tuto předá všem svým potomkům, zatímco pokud je přenašečem otec, pak tuto mutaci žádnému potomkovi předat nemůže. Další zvláštností je, že při dělení buňky nedochází k žádné kontrolované segregaci mitochondrií do dceřiných buněk. Proto je mtDNA mezi nové buňky distribuována zcela náhodně a dceřiné buňky tak mohou získat zcela různý počet normálních a mutací zatížených mitochondrií. Stav, kdy buňka obdrží pouze normální mitochondrie, či naopak pouze mutované mitochondrie se nazývá **homoplazmie**. Stav, kdy obdrží směs normálních a mutovaných mitochondrií, se nazývá **heteroplazmie**.

Případná nefunkčnost mitochondrií je neslučitelná se životem, ale vzhledem k tomu, že fenotypový projev závisí na podílu normálních a zmutovaných mitochondrií v buňce, je život takového jedince nejen možný, ale v případě malého počtu zmutovaných mitochondrií v buňce, se tato skutečnost nemusí projevit vůbec. Proto se u pacientů s těmito mutacemi mitochondrií setkáváme s variabilní expresivitou choroby a její neúplnou penetrací v rodokmenu (a aby to nebylo jednoduché, existuje mezi jadernou DNA a mtDNA poměrně těsný vztah a byly pozorovány takové interakce, které v rodokmenu následně vykazují AD či AR dědičnost). Existují různorodá onemocnění, podmíněná mutací v mtDNA. Uvedu zde několik příkladů:

Leberova hereditární optická neuropatie

Náhlá a rychlá nekróza optického nervu s pregresivní ztrátou zraku vedoucí k úplné slepotě. Projeví se zejména u mužů, a to mezi 16 - 25 lety věku.

Hluchota

Progresivní hluchota, jen málo případů dědičné hluchoty je způsobeno mutací v mtDNA.

NARP

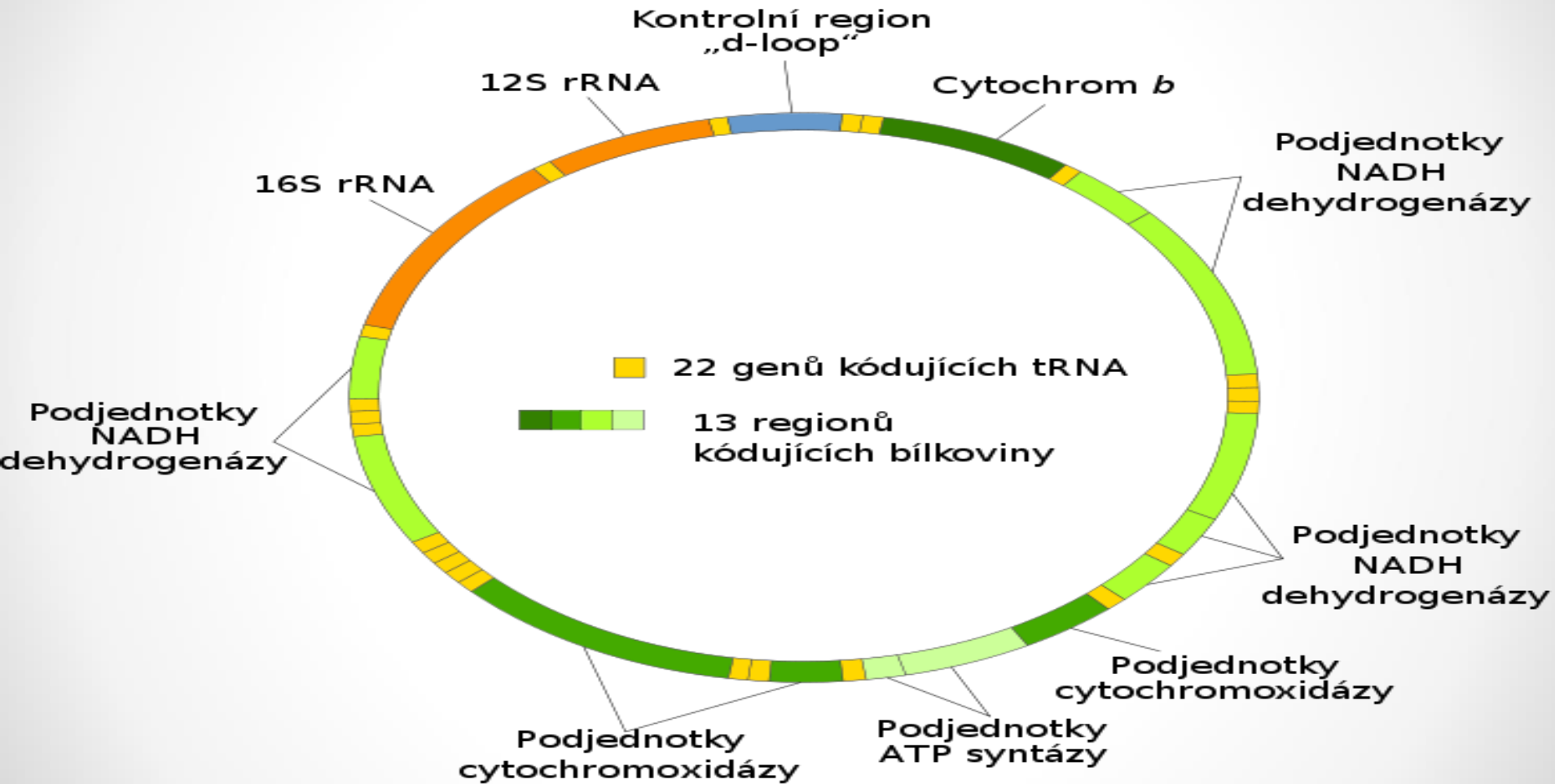
MELAS

Pearsonův syndrom

Vložil [Azrael](#) 26. 3. 2005

<http://genetika.wz.cz>

Mitochondriální DNA člověka



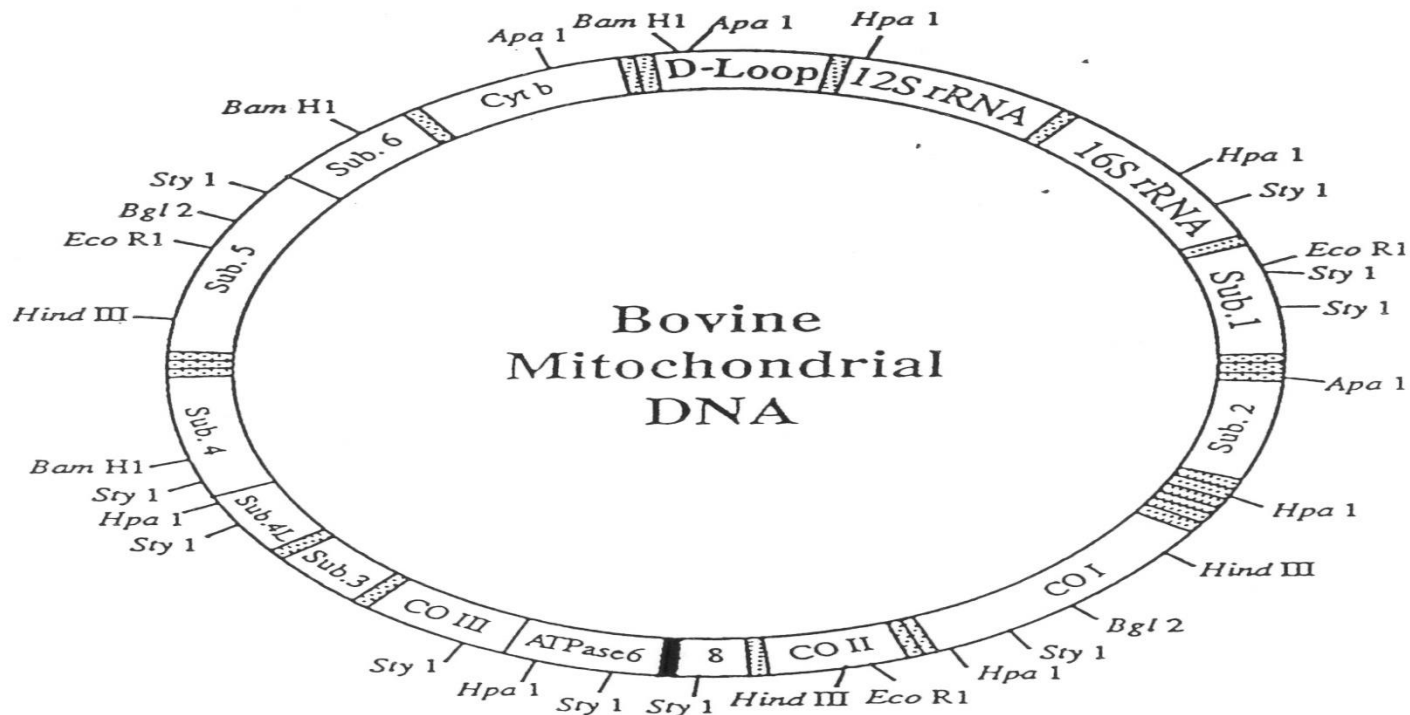


Fig. 1. Organization of the bovine mitochondrial genome. The shaded areas represent tRNA genes, designated by the single letter amino acid code. Protein genes are listed as follows: NADH dehydrogenase subunits 1-4, 4L, 5 and 6 (Sub. 1-4, 4L, 5 and 6), ATPase (8), Cytochrome oxidase subunits 1-3 (CO I, CO II and CO III) and cytochrome b (Cyt b). The relative locations of restriction sites surveyed are indicated.

(A) Human

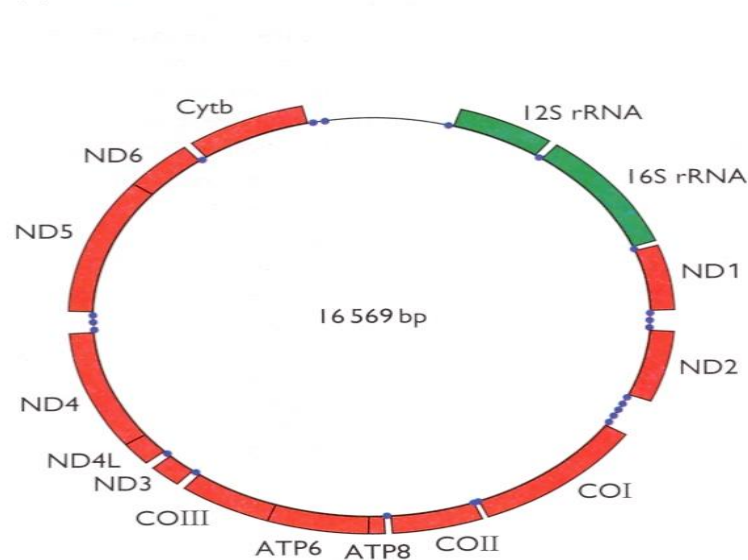
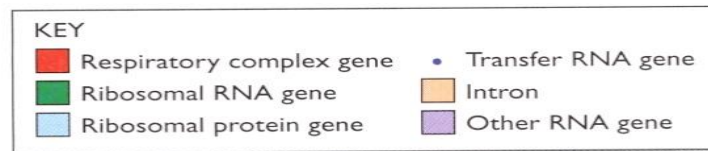
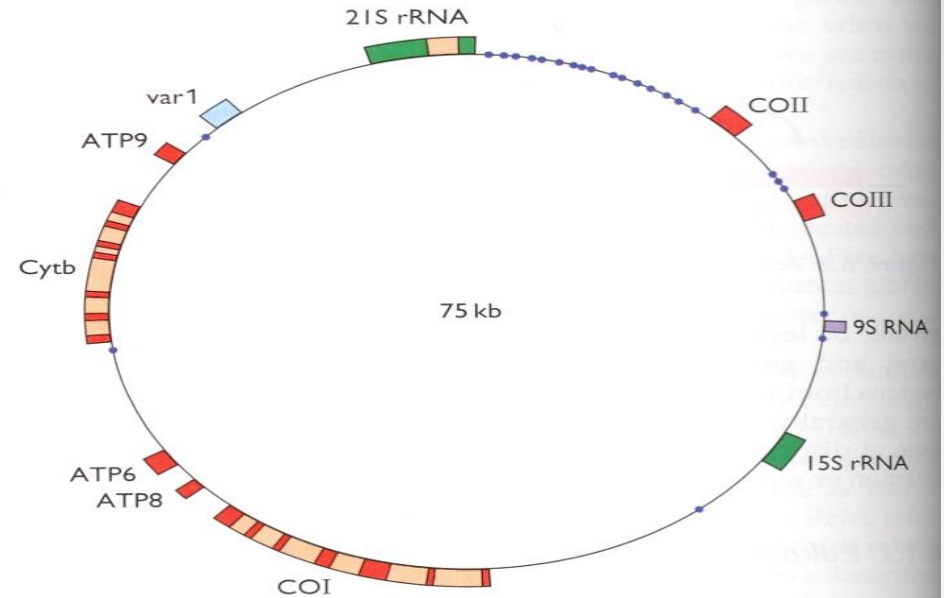
(B) *Saccharomyces cerevisiae*

Figure 6.9 The (A) human and (B) *Saccharomyces cerevisiae* mitochondrial genomes.

Because of their relatively small sizes, many mitochondrial genomes have been completely sequenced. The human and yeast genomes illustrate two types of organization, the human genome being small and compact with little wasted space, typical of multicellular animals, and the yeast mitochondrial genome having longer spaces between genes and possessing some introns, the typical organization for lower eukaryotes and plants. In the human mitochondrial genome the ATP6 and ATP8 genes overlap. The yeast genome contains five additional open reading frames, not shown on this map, that have not yet been shown to code for functional gene products, and there are also several genes located within the introns of the split genes. Most of the latter code for maturase proteins that are involved in splicing the introns from the transcripts of these genes (Section 9.3.3). The two genomes are drawn at different scales. Abbreviations: ATP6, ATP8, ATP9, genes for ATPase subunits 6, 8 and 9; COI, COII, COIII, genes for cytochrome *c* oxidase subunits I, II and III; Cytb, gene for apocytochrome *b*; ND1–ND6, genes for NADH hydrogenase subunits 1–6; var1, yeast mitochondrial gene for a ribosome-associated protein. The 9S RNA gene in the yeast genome specifies the RNA component of the enzyme ribonuclease P (Section 9.3.2).

Genetický kód, rozdílné kodony lidské mitochondrie

Kodon	Standardní kód	Kód mitochondrie
UGA	STOP	Trp
UGG	Trp	Trp
AUA	Ile	Met
AUG	Met	Met
AGA	Arg	STOP
AGG	Arg	STOP

Velikosti genomů mitochondrií a chloroplastů

Table 6.3 Sizes of mitochondrial and chloroplast genomes

Species	Type of organism	Genome size (kb)
A. Mitochondrial genomes		
<i>Chlamydomonas reinhardtii</i>	Green alga	16
<i>Mus musculus</i>	Vertebrate (mouse)	16
<i>Homo sapiens</i>	Vertebrate (human)	17
<i>Metridium senile</i>	Invertebrate (sea anemone)	17
<i>Drosophila melanogaster</i>	Invertebrate (fruit fly)	19
<i>Chondrus crispus</i>	Red alga	26
<i>Aspergillus nidulans</i>	Ascomycete fungus	33
<i>Reclinomonas americana</i>	Protozoa	69
<i>Saccharomyces cerevisiae</i>	Yeast	75
<i>Suillus grisellus</i>	Basidiomycete fungus	121
<i>Brassica oleracea</i>	Flowering plant (cabbage)	160
<i>Arabidopsis thaliana</i>	Flowering plant (vetch)	367
<i>Zea mays</i>	Flowering plant (maize)	570
<i>Cucumis melo</i>	Flowering plant (melon)	2500
B. Chloroplast genomes		
<i>Pisum sativum</i>	Flowering plant (pea)	120
<i>Marchantia polymorpha</i>	Liverwort	121
<i>Oryza sativa</i>	Flowering plant (rice)	136
<i>Nicotiana tabacum</i>	Flowering plant (tobacco)	156
<i>Chlamydomonas reinhardtii</i>	Green alga	195

Data taken from Brown (1998).

Genom eukaryontů

Eukaryonti mají diferencované jádro a jaderná DNA je lokalizována v **chromozómech**. V každém chromozómu je jedna molekula dvoušroubovicové DNA. Člověk, který má 46 chromozómů (v diploidní buňce) má v jádru každé buňky 46 molekul DNA; prase, s 38 chromozómy, má 38 molekul DNA (diploidní genomy). Obecně, velikosti genomů u savců jsou v rozmezí 2,5 až 3,5 Gb v haploidním genomu; u ptáků jsou podstatně menší – kolem 1,2 miliardy bází.

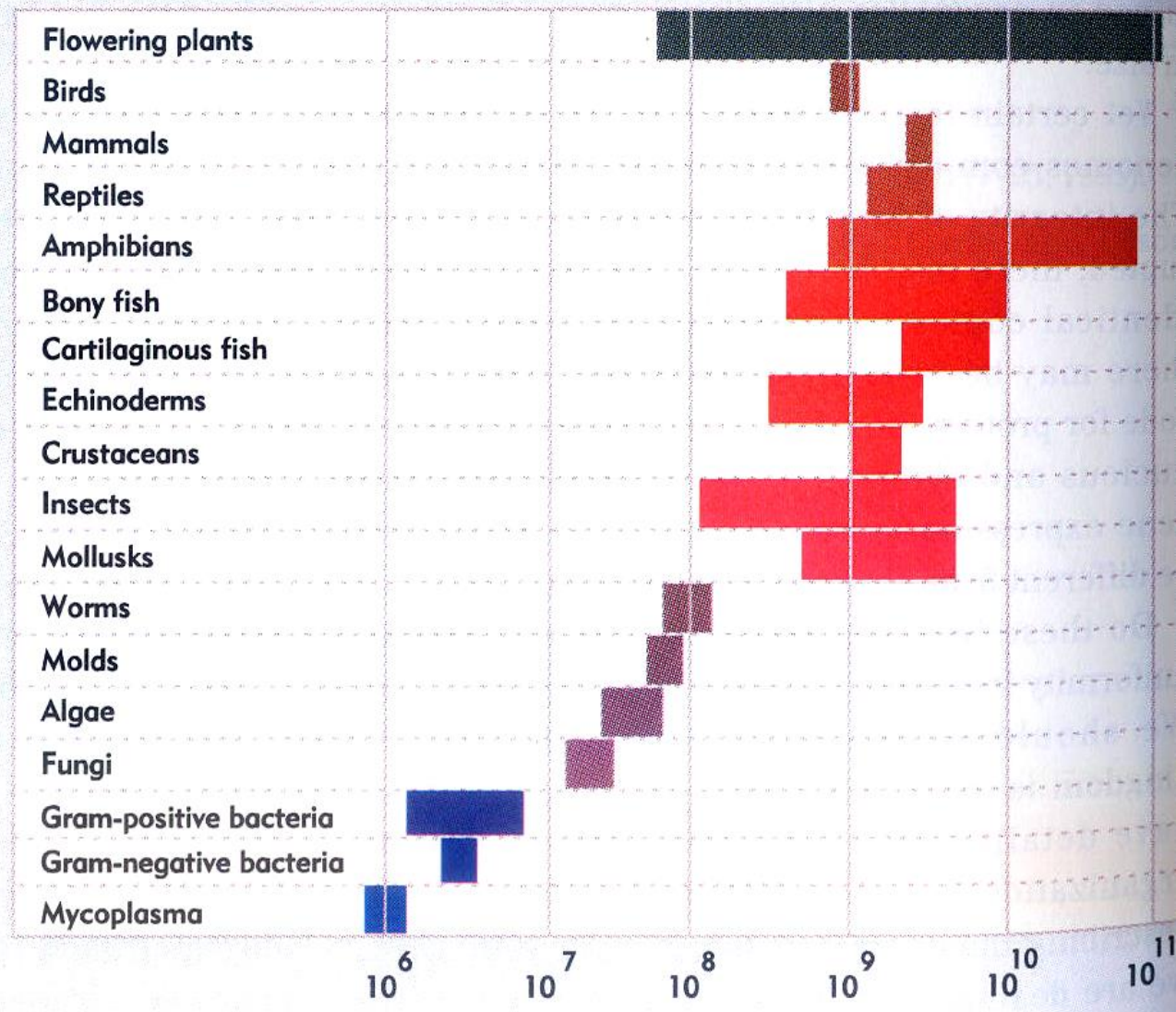
Rozdíly mezi genomem rostlin a živočichů:

- Rostlinné genomy mají větší frakci opakujících se sekvencí DNA ve srovnání s živočichy.
- Rozdíly v obsahu DNA v jádrech různých rostlinných druhů jsou v rozmezí téměř tří řádů, od přibližně 0,1 pg u *Arabidopsis thaliana* až po více než 100 pg v haploidním genomu jmelí.
- Druhy s vysokým celkovým obsahem DNA v jádrech se vyznačují vysokou proporcí repetitivních sekvencí.
- U struktury genů je zásadní rozdíl mezi velikostí a počtem intronů u rostlin a živočichů, speciálně savců, ptáků a obojživelníků. Zatímco u živočichů bývá velký počet intronů a jejich délka mnohonásobně převyšuje délku exonů, typický rostlinný gen má jeden až tři introny a jejich délka obvykle nepřevyšuje délku exonů.
- Transkripční jednotky (strukturní geny) živočišných genomů mají regulační sekvence před genem i za genem, kdežto rostliny je mají jen před genem.

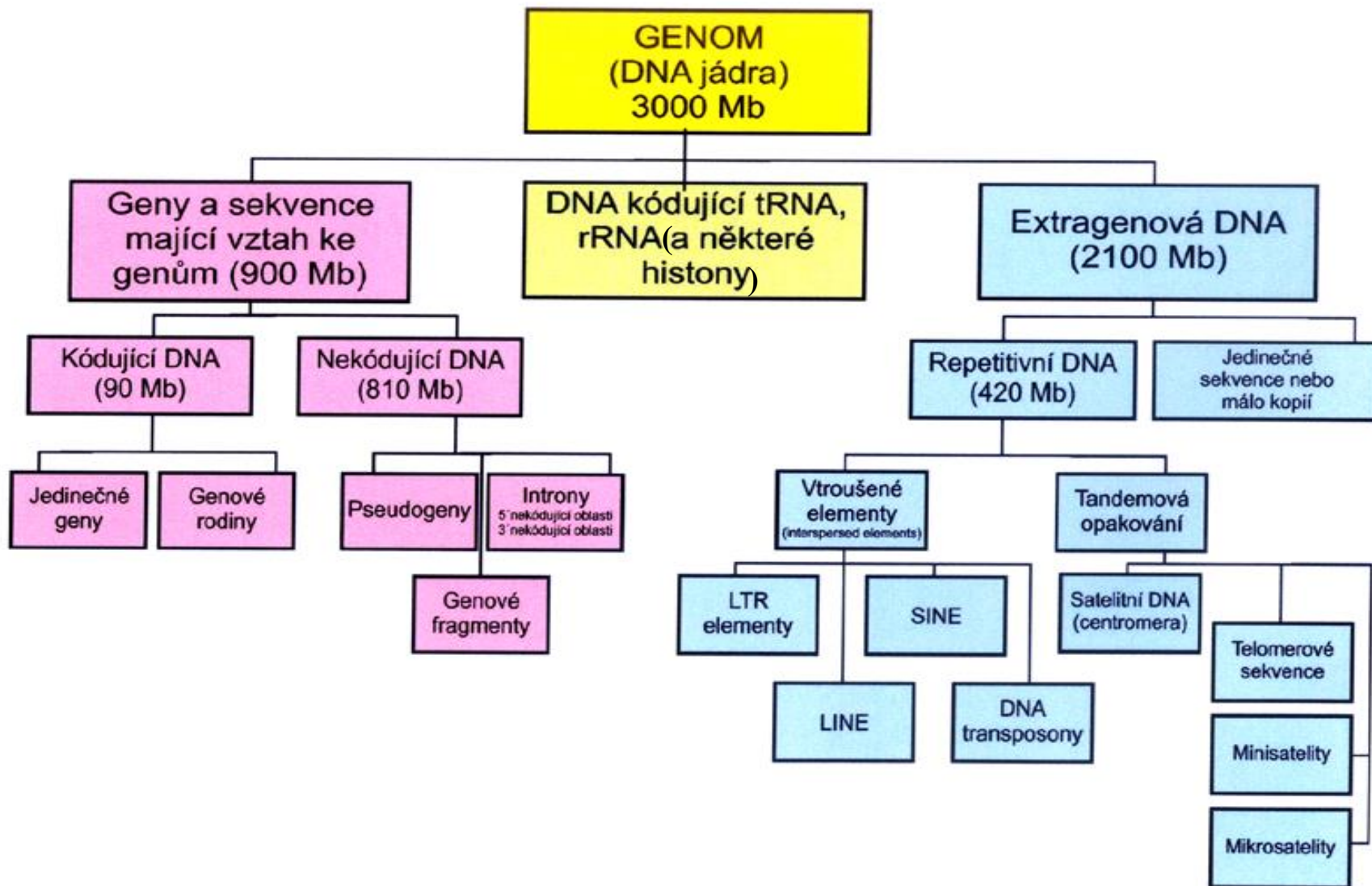
(To je jednou z hlavních příčin, proč se přenesené živočišné geny neprojevují v rostlinném genomu. Zatímco u živočichů sekvence zesilovačů (enhancers) působí do několika kilobází, a to nezávisle na své orientaci a nezávisle na tom, zda se nacházejí na 5' nebo 3' konci od kódující sekvence, u rostlin působí před promotorem nebo v oblasti promotoru a na menší vzdálenosti. Vlastní typické krátké sekvenční úseky promotorů se u rostlinných a živočišných genů liší jen nepatrně a neliší se ani signál pro polyadenylaci na 3' konci genu.)

Velikosti genomů u různých organismů

DNA content of the haploid genome is related to the morphological complexity of lower eukaryotes, but varies extensively among the higher eukaryotes. The range of DNA values within a phylum is indicated by the shaded area.

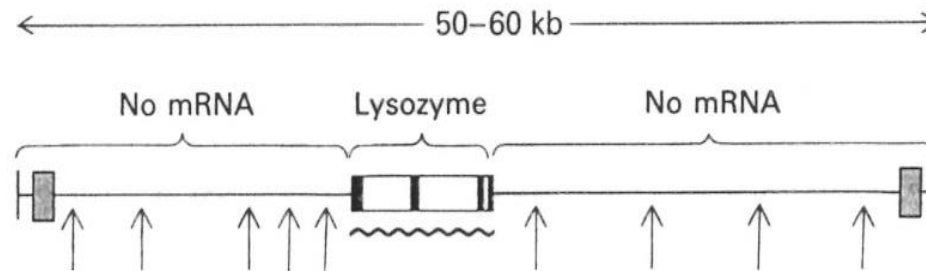


Organizace genomu - typy DNA sekvencí



Jedinečné geny (solitary genes)

Jsou přítomny jako jedinečné sekvence v genomu – žádná jiná stejná nebo podobná sekvence se v genomu nevyskytuje. Příkladem takového genu je gen pro lysozym slepičího vejce. Transkripční jednotka je 15 kb sekvence, která představuje jeden gen a obsahuje 4 exony a 3 introny.



▲ **Figure 10-1** The chicken lysozyme gene and its surrounding regions. The 15-kb transcription unit contains four exons (red) and three introns (white). The positions indicated by arrows are repetitious sequences found elsewhere in the genome. The solid squares (orange) indicate positions of scaffold-attachment regions, or SARs, that bind to the nuclear matrix. No mRNA other than lysozyme could be found in embryonic or adult tissues that hybridize to this region. [See P. Balducci et al., 1981, *Nucleic Acids Res.* 9:3575.]

Genové rodiny (duplikované geny)

Všechny geny kódující proteiny nejsou jedinečné jako gen pro lysozym. Velmi často ve vzdálenosti 5 – 10 kb od známého genu jsou sekvence, které jsou podobnými kopiemi tohoto genu. Takové geny představují téměř polovinu genů kódujících proteiny v genomu obratlovců.

Takové duplikované oblasti DNA se předpokládaly ještě dříve než byla známa jejich sekvence DNA – na základě sekvenční podobnosti proteinů. Tyto proteinové rodiny zahrnují od několika do více než 20 členů s podobnými, ale ne identickými sekvencemi (např. beta-globiny, inhibitory proteáz).

Pseudogeny jsou nefunkční sekvence DNA podobné sekvencím funkčních genů. Mohou mít stejnou exon – intronovou organizaci, ale mutace v sekvencích vedou k akumulaci sekvencí, které buď ukončují translaci nebo blokují úpravu pre-mRNA, čímž činí takové úseky nefunkčními.

Procesované pseudogeny vznikají reverzní transkripcí RNA (do DNA) a zabudováním takto vzniklého transkriptu (DNA) do genomu. (I kdyby měly identickou sekvenci, odpovídající otevřenému čtecímu rámci, budou nefunkční, protože budou postrádat regulační (promotorovou) oblast).

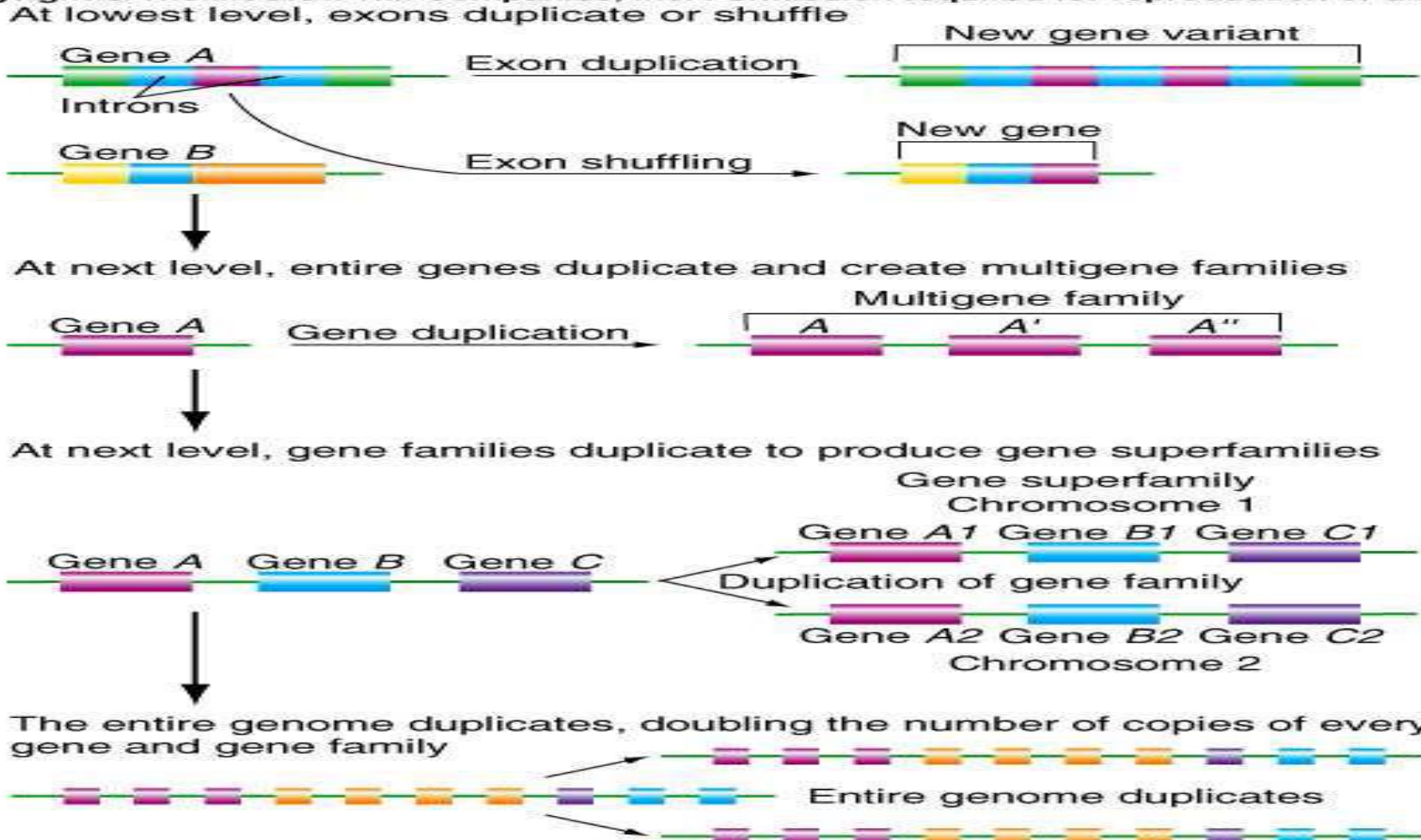
Genové rodiny

Table 10-2 Protein families in vertebrates and invertebrates: common and specific

Family	Number of proteins in family
COMMON PROTEINS	
Actins	5–30
Heat-shock proteins	3
Histones	100–1000
Keratins	>20
Myosin, heavy chain	5–10
Protein kinases	10–100s
Transcription factors	10–100s (?)
Tubulins, α and β	3–15
INSECT PROTEINS	
Eggshell proteins (silk moth and fruit fly)	50
VERTEBRATE PROTEINS	
Globins (many species)	
α -globin	1–3
β -like globins	5
Immunoglobins, variable regions (many species)	500
Ovalbumin (chicken)	3
Transplantation antigens (mouse and human)	50–100
Visual pigment protein (human)	4
Vitellogenin (frog, chicken)	5

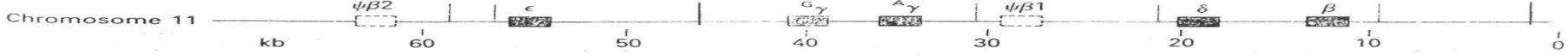
Základní mechanismy zvětšující velikost genomu

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



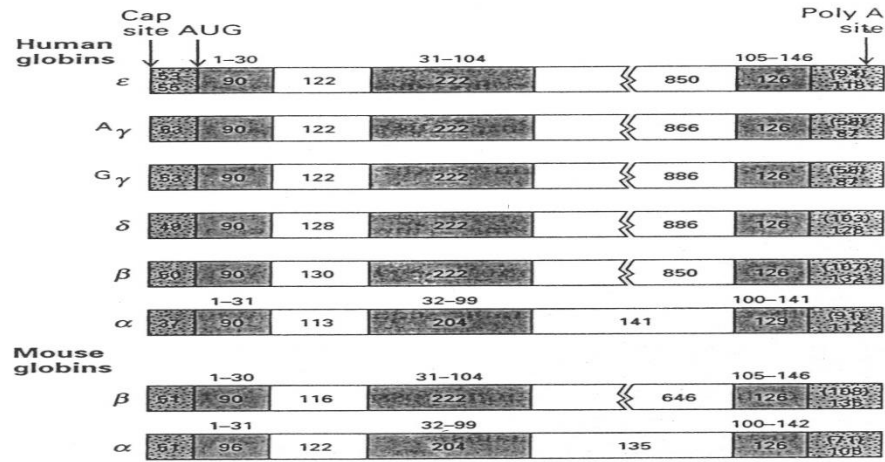
Exon shuffling – exony různých genů se vyměňují a vytvářejí tak geny, které jsou „mozaikou“ jiných genů; nové geny tedy mohou kódovat proteiny, které mají funkce „smíchané“ z původních genů.

Genová duplikace – duplikací a následnou divergencí vznikají multigenové rodiny, tj. sady genů, které jsou velice podobné svou sekvencí ale jejich produkty mohou mít jiné funkce.



4. Figure 10-4 Map of the β -like globin gene cluster on human chromosome 11. The positions of the five genes known to encode β -like globins are indicated by green boxes. The direction of transcription is from left to right for all these genes. Two nonfunctional pseudogenes, $\psi\beta 1$ and $\psi\beta 2$,

also lie in this region (yellow boxes). Vertical lines (red) mark the sites of repetitive DNA sequences, which occur at many other sites in the human genome. [See E. F. Fritsch, R. M. Lawn, and T. Maniatis, 1980, *Cell* 19:959.]



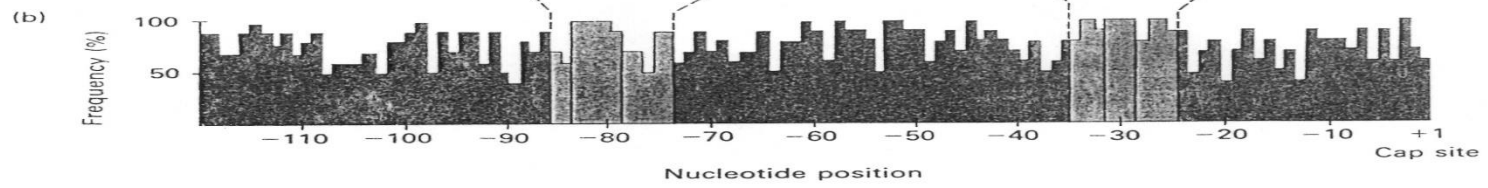
▲ Figure 10-6 General design of transcription units for α - and β -like globins, beginning with 5' cap site and extending to 3' poly A site, in humans and mice. All of these globin genes have two introns (yellow), which are located at similar positions within each gene. The lengths of the introns also are similar, although the second intron in the β -like genes is much longer than that in the α genes. All of these genes contain three exons (orange), some portions of which are untranslated (stippled areas). The numbers within the boxes indicate the number of nucleotides present in each region of the primary transcript; the numbers above the boxes designate the corresponding amino acid positions in the resulting polypeptides. In the right-hand boxes, the numbers in parentheses indicate the length of the 3' untranslated region up to the first A of the AAUAAA sequence; the other numbers indicate the length of (a) the site of poly A

Cell 21:653.]

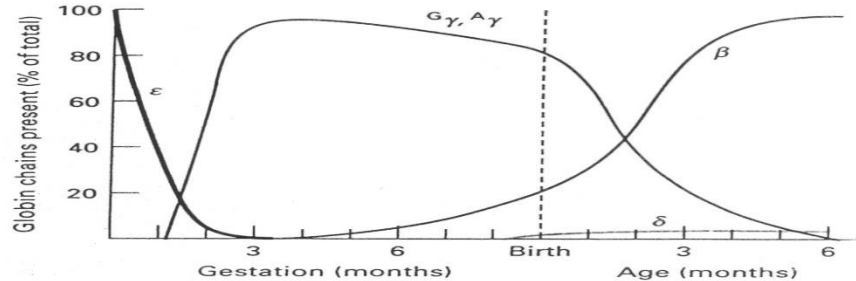
Mouse β^{min}
 Mouse β^{maj}
 Rabbit β
 Goat β^A
 Goat β^C
 Human β
 Human δ
 Human γ
 Human α

GGCCAATCTGCTC
 GGCCAATCTGCTC
 GGCCAATCTACT
 AGCCAATCTGCTC
 AGCCAATCTGCTC
 GGCCAATCTACT
 AACCAACCTGCTC
 GACCAATAGCCTT
 GACCAATAGCCTT
 GACCAATGACTTT

GGTATATAAAA
 AGCATATAAAG
 GGCATAAAAAG
 GGCATAAAAAG
 GGCATAAAAAG
 GGCATAAAAAG
 GCATAAAAAG
 GGCATAAAAAG
 GGCATAAAAAG
 GGCATAAAAAG



▲ Figure 10-7 Similarity of sequences just upstream from 5' cap site in globin genes. The sequences of about 150 nucleotides were aligned to give maximum similarity. (a) Sequences of two highly conserved regions—the CCAAT and the ATA boxes—are shown in red type. These regions are



▲ Figure 10-5 Proportion of total β -like globin protein in human red blood cells accounted for by the individual β -like globins (ϵ , γ , δ , and β) during gestation and the first 6 postnatal months. Comparison of these curves with map positions shown in Figure 10-4 indicates that the β -like genes are sequentially expressed during development. Only the β - and δ -globins are produced after 6 months. [See R. M. Winslow and W. F. Anderson, 1983, in *The Metabolic Basis of Inherited Disease*, 5th ed., J. G. Stanbury et al., eds., McGraw-Hill, p. 1666.]

present in all the globin genes sequenced to date. (b) The frequency of occurrence (%) of the predominant nucleotide at each position in the upstream regions of the ten genes listed in part (a) is shown in a histogram. [See A. Efstratiadis et al., 1980, *Cell* 21:653.]

Struktura lokusu *Serpina3* na myším chromosomu 12

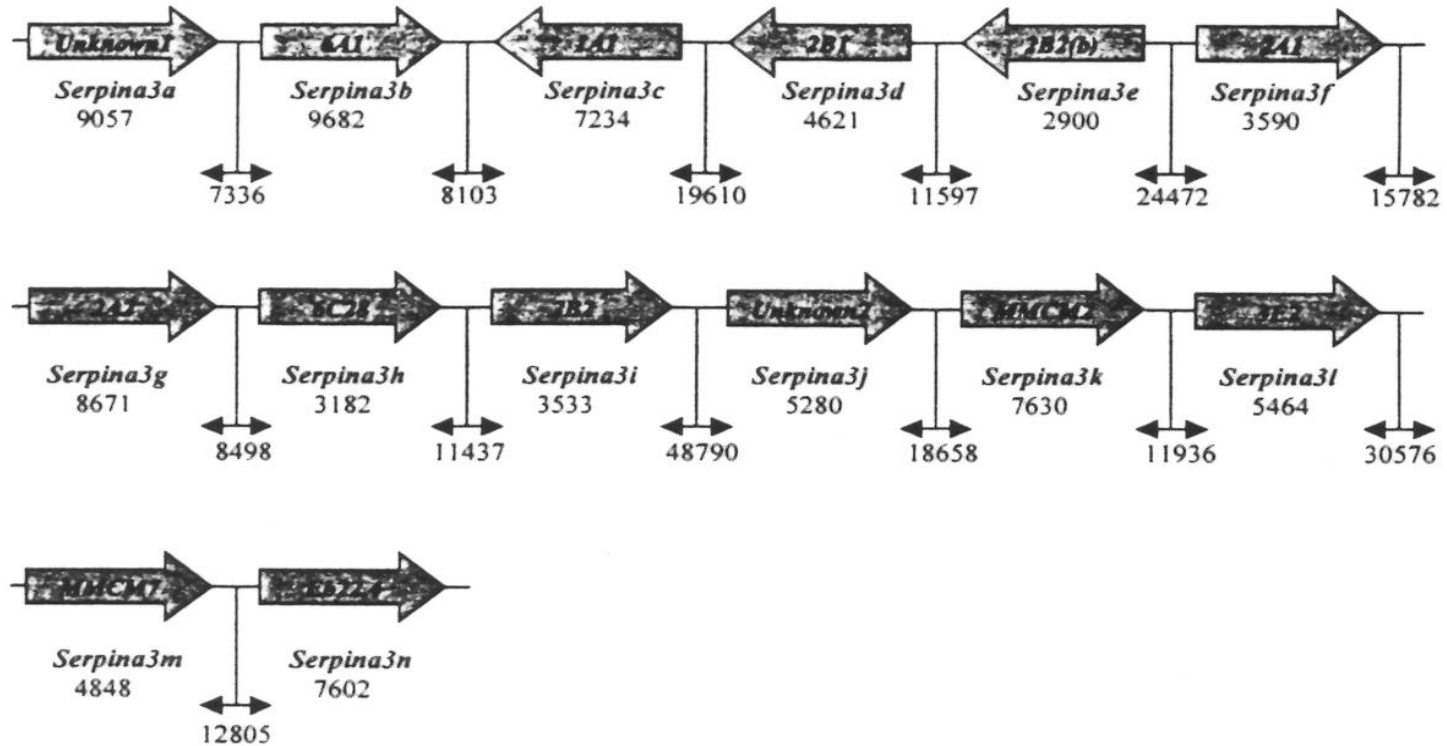


Fig. 2. The structure of the *Serpina3* (*Spi2*) locus on murine chromosome 12. All 14 genes are represented and the direction of transcription is indicated by the block arrow. The genes are labelled according to the classification in Table 2. The 'common' names are indicated within the arrow and the revised name is indicated below and in bold. The length of each gene as well as the intergenic sequence is shown. Information for this figure was acquired from the publicly available mouse genome sequence database (http://www.ensembl.org/Mus_musculus/).

Genový "cluster" pro Serpiny člověka

..... a srovnání s geny Serpina myši (na chromosomu 12F1)

(Forsyth et al., 2003)

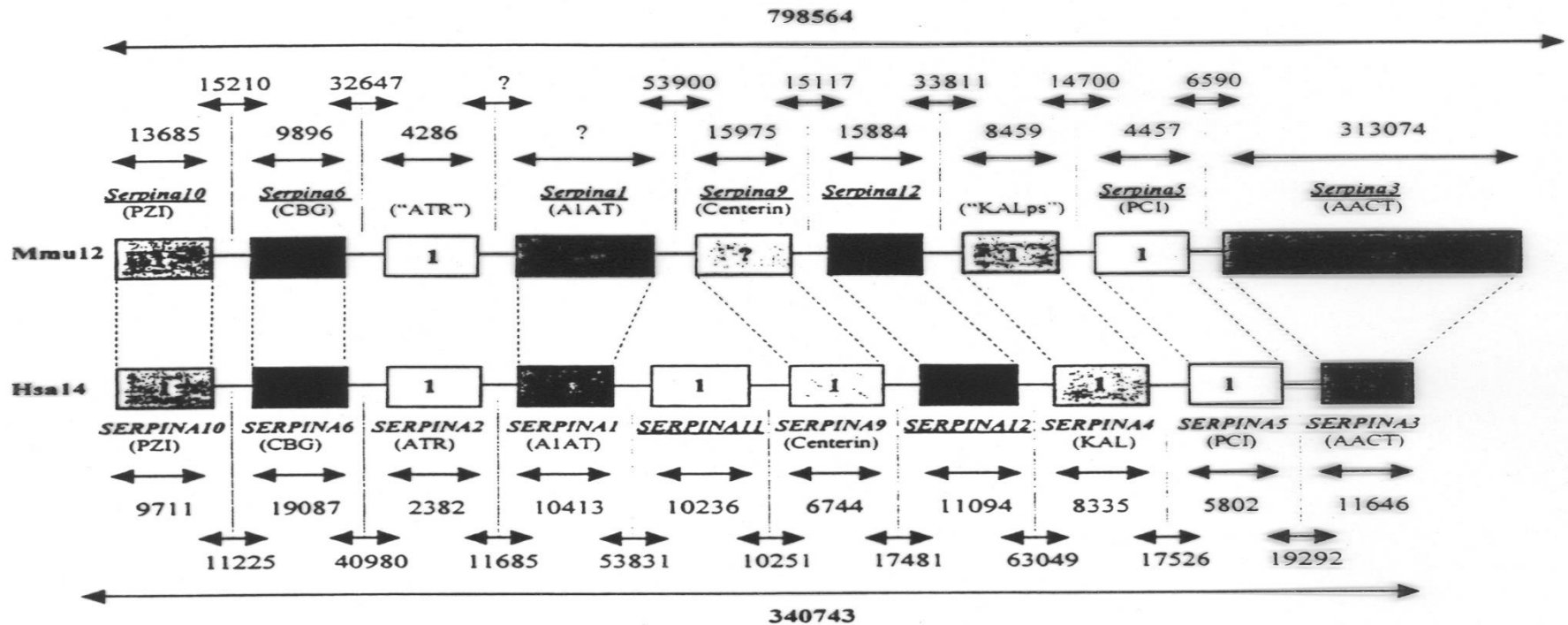


Fig. 3. Organization of mouse chromosome 12F1 containing the clade 'a' serpins versus the syntenic region on human chromosome 14q32.1. The number within each box indicates the number of genes within the locus. A question mark within a box indicates a gene fragment. The hatched boxes (*SERPINA11* and *SERPINA9*) are represented in the ENSEMBL database as one gene (ENSG00000170054) with two alternative transcripts but closer examination of the genomic sequence using the gene finding NIX tool (<http://www.hgmp.mrc.ac.uk/NIX/>) indicates two separate genes (as shown here). The size of the gene/s (base pairs) as well as the intergenic sequence is indicated below (Hsa) and above (Mmu). The length of the entire cluster of genes is also shown. The question mark for the *Serpina1* locus in the mouse indicates an area with poor sequence and only Sp1-5 (*Serpina1e*) has been confirmed here. However, the remaining four Sp1 genes (see text) have been identified within an unassigned chromosome fragment which is likely to belong in this locus. Murine "ATR" is in inverted commas as, although it has been identified in the databases as the homolog, it shows little identity with the human gene (Table 3). The murine gene in the corresponding position to human *SERPINA4* (Kallikrein inhibitor) has 57% amino acid identity with *SERPINA4* (Table 3) but appears to be a pseudogene and is described as "KALps". PZI = protein-Z dependent protease inhibitor, CBG = corticosteroid binding globulin, ATR = α_1 -antitrypsin related protein, AIAT = α_1 -antitrypsin, KAL = Kallikrein inhibitor, PCI = protein C inhibitor and AACT = α_1 -antichymotrypsin. New names, as approved by the MGNC and the HGNC, are underlined. Data for this figure was derived from the publicly available mouse and human genomes (<http://www.ensembl.org/>).

Evoluční vztahy mezi serpiny různých druhů

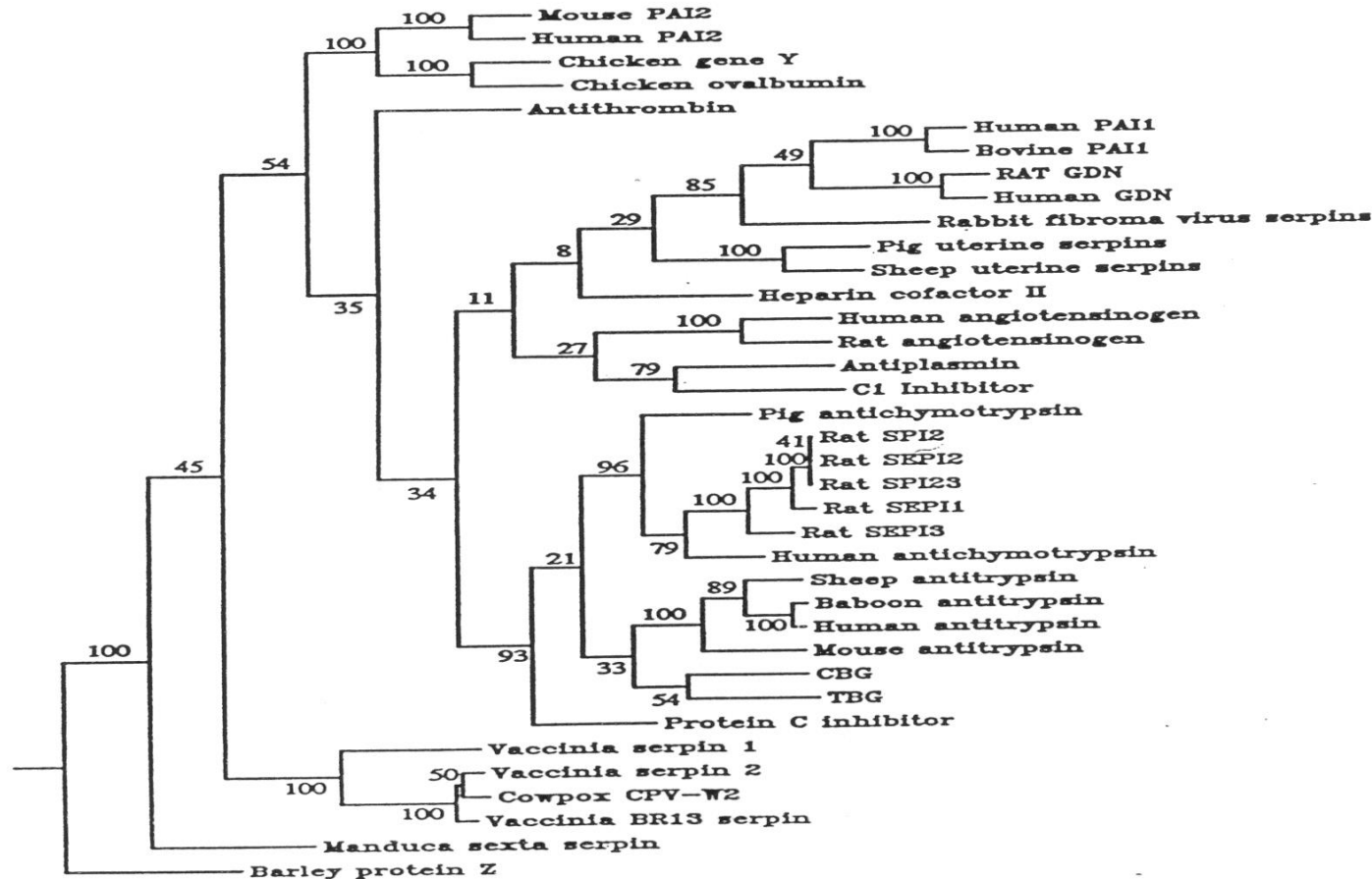


Figure 4. Gene tree of the serpins calculated as indicated in the text. This representation shows the relationship amongst branches of the tree. Note that the sequence HVPROTZG (barley protein Z) was defined as the outgroup as *a priori* it was considered the most evolutionary distant sequence. Branch lengths are proportional to evolutionary distance. All branch tips, unless stated otherwise, are human.

Homologní – podobný, stejného typu, odvozené od původní sekvence

Orthologní – homologní (sekvence aj.) u různých druhů

Paralogní – homologní (sekvence aj.) u stejného druhu

DNA kódující rRNA, tRNA a některé histony

Mnohonásobně se opakující kopie genu (v tandemu).

Geny pro rRNA - zahrnují 100 a více kopií (až 20 000 u obojživelníků)

Geny pro tRNA (v organismu je asi 50 různých genů pro tRNA) – každý gen je v 10 až 100 kopiích

Geny pro histony – zahrnují 50 – 100 kopií

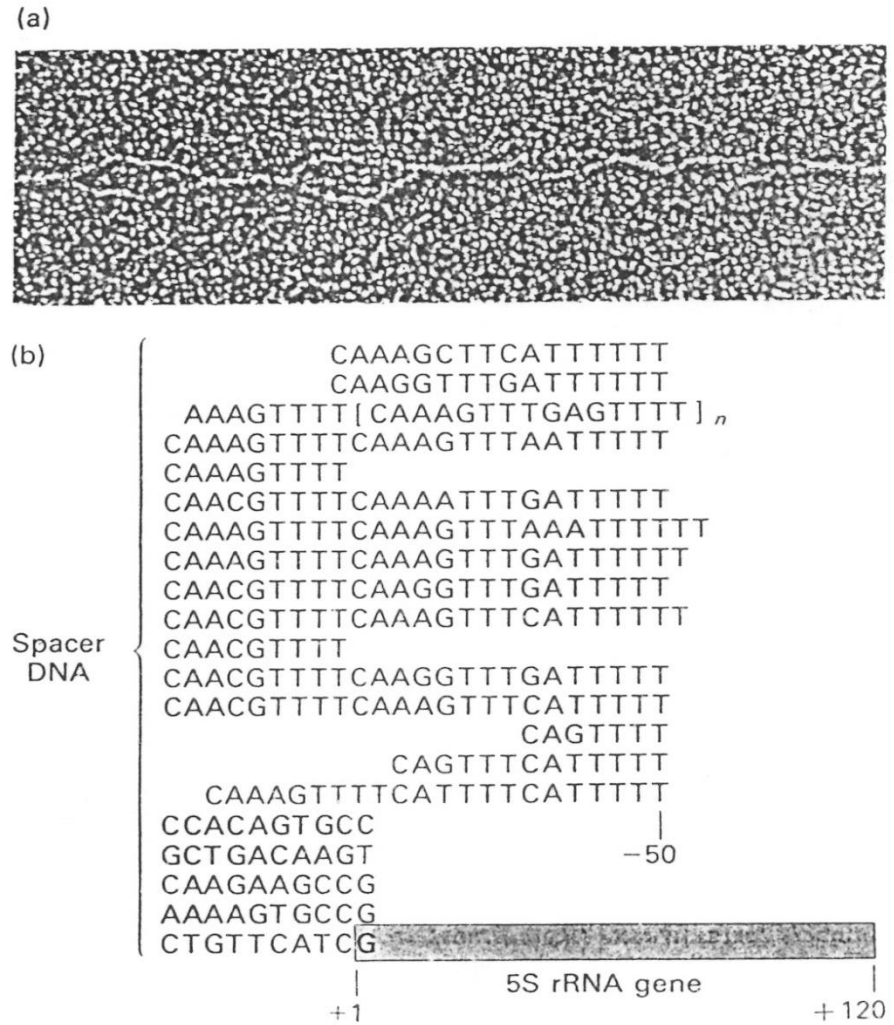
V **jadérku** dochází ke spojení částí různých chromozómů, ve kterých se nacházejí **geny pro ribozomální RNA**. V jadérku jsou transkribovány geny pro ribozomální RNA a jsou zde také z proteinů importovaných z cytoplazmy sestavovány ribozomální podjednotky. Ty jsou pak transportovány do cytoplazmy jadernými póry.

Geny pro ribozomální RNA jsou středně repetitivní sekvence organizované v tandemech, které obsahují jak transkribované, tak netranskribované segmenty. Dlouhé intergenové mezerníky (spacers) v lidských rDNA jednotkách obsahují mnoho Alu repeticí (SINE).

Histonové geny savců jsou lokalizovány na různých chromozómech, podobně jako jiné geny, jejichž produkty jsou nepostradatelné pro strukturu a funkci buněk (např. aktin a tubulin). U drosofilů se geny pro histony nacházejí v klastrech (shlucích).

Sekvence mezerníku mezi 5S rDNA geny u *X. laevis*

► **Figure 10-11** (a) Electron micrograph of partially denatured 5S rDNA from the frog *Xenopus laevis*. The “loops” correspond to spacer regions, which are rich in A and T residues and hence melt more easily than the transcribed regions. The spacer loops vary in length, whereas the transcribed coding regions are all the same length. (b) Sequence of spacer region between 5S-rRNA genes in *X. laevis*, starting from the 5' G that begins the transcribed portions of 5S rDNA. The end of the partially repetitious AT-rich portions of the spacer DNA (blue) begins about 50 bases (–50) upstream from the transcription start site. This sequence is shown in a stacked arrangement (the last nucleotide in each row is actually adjacent to the first nucleotide in the row below) to emphasize the repeated portions. About 200 bases upstream of the coding portion of the gene is a sequence (red) that appears a variable number of times ($n = 2$ to 12) in each spacer. This short, variably repeated sequence is responsible for all of the length variation among the spacers of a tandem 5S-gene array. [See N. V. Federoff, 1979, *Cell* 16:697.] Part (a) from D. Brown, P. C. Wensink, and E. Jordan, 1971, *Proc. Nat'l Acad. Sci. USA* 68:3175; courtesy of D. D. Brown.



Long terminal repeat elements (LTRs)

Long terminal repeats (dlouhé terminální repetice) jsou identické sekvence DNA ve stovkách nebo tisících opakování, které se nacházejí na obou koncích retrotranspozónů nebo provirové DNA vytvořené reverzní transkripcí retrovirové RNA. Mezi LTR elementy patří **retroviry**, **endogenní retroviry** a **retrotranspozóny**.

Retroviry: RNA genom je přepsán do DNA reverzní transkriptázou, která je produktem virového genu *pol*, a DNA kopie se zabuduje do hostitelského genomu. Nové viry mohou být vytvářeny kopírováním integrované DNA do RNA a obaleny membránovými proteiny kódovanými geny *env* virového genomu.

Endogenní retroviry (ERVs): Jsou to retrovirové genomy integrované do chromozómů obratlovců. Některé jsou stále aktivní a v určitém stadiu života buňky mohou řídit syntézu exogenních virů, ale většina je pozměněných, přičemž jeden z genů nezbytných pro sestavení infekčních částic (gen *env*) je zmutovaný nebo chybí, takže nemají možnost narušovat buněčnou stěnu. U savců představují asi 8% genomu.

Retrotranspozóny mají sekvence podobné endogenním retrovirům. U obratlovců se vyskytují v menším měřítku. Replikační a včleňovací proces je stejný jako u retrovirů, ale nemají gen *env*, nebo jej mají nefunkční, takže nejsou infekční.

Transpozóny jsou pohyblivé sekvence, které nemají virovou charakteristiku. Přesun z místa na místo se uskutečňuje pomocí enzymu transposázy. Jsou v genomech jak prokaryontů, tak eukaryontů. Vysoký je výskyt u kukuřice – způsobují změny viditelných znaků (zbarvení perikarpu, aleuronové vrstvy, endospermu, tvar zrna, zbarvení listů aj.)

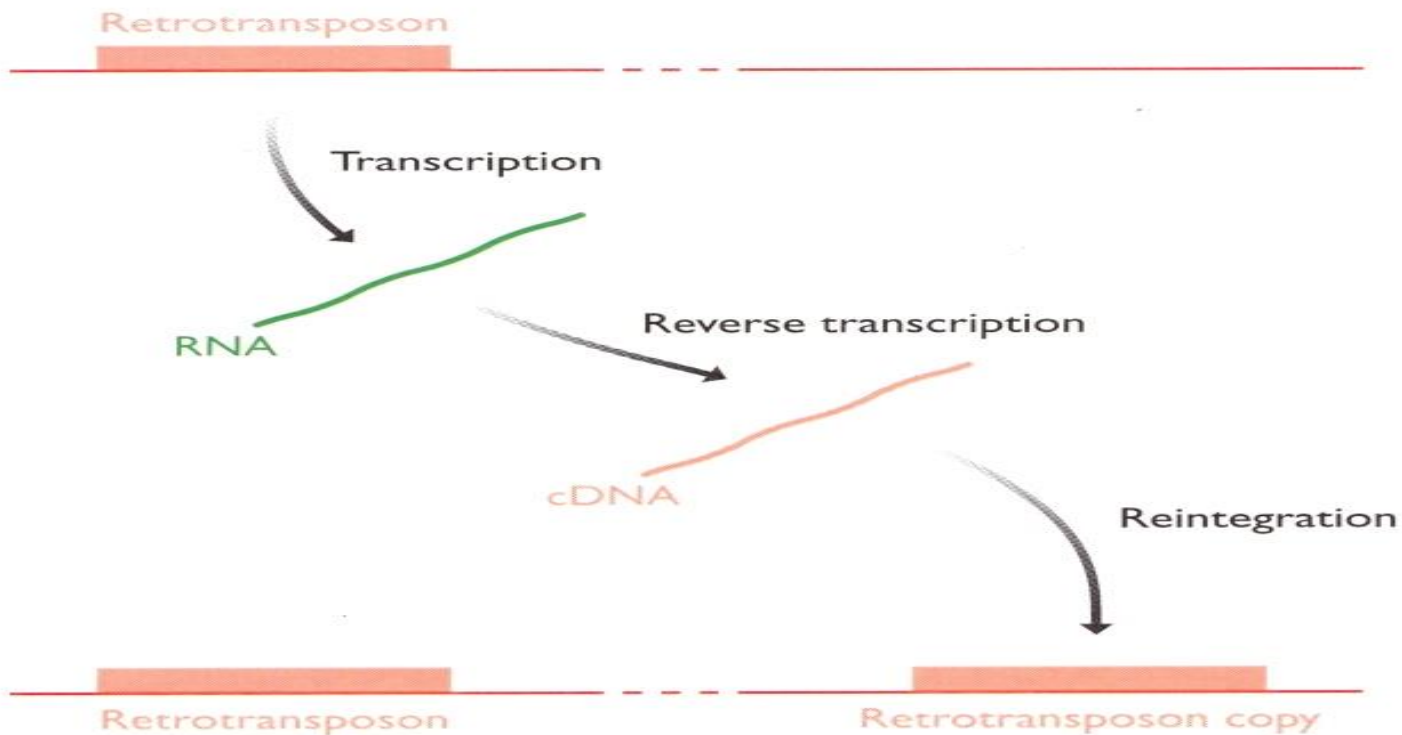


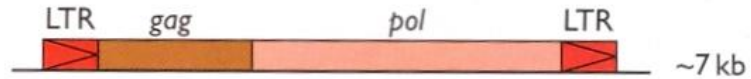
Figure 6.18 Retrotransposition.

Compare with *Figure 6.7*, p. 124, and note that the events are essentially the same as those that result in a processed pseudogene.

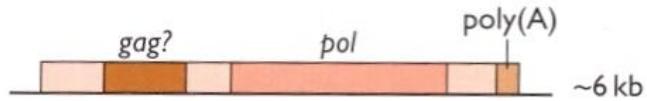
(A) Retrovirus



(B) *Ty1/copia* retrotransposon



(C) LINE



(D) SINE

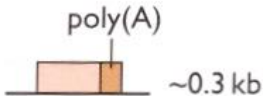


Figure 6.19 Retroelements.

A comparison of the structures of four types of retroelement. Retroviruses and retrotransposons are LTR elements that possess long terminal repeats at each end. The *gag* gene codes for a series of proteins located in the virus core, *pol* codes for the reverse transcriptase and other enzymes involved in replication of the element, and *env* codes for coat proteins. LINEs and SINEs are non-LTR retroelements or retrotransposons. Both have a poly(A) region (a long series of A nucleotides) at one end.

LINE a SINE

(Long interspersed nuclear element; short interspersed nuclear element)

Oba elementy patří mezi **retropozony**; tj. původně byly (a mohou být) transkribovány z genomové DNA (na RNA), reverzní transkripcí přepsány do DNA a zabudovány do náhodného místa na některém chromozómu. V podstatě mají charakter **procesovaných pseudogenů**.

LINE vznikly z transkriptů RNA polymerázy II, a SINE z transkriptů RNA polymerázy III.

Lidský SINE (Alu family) je odvozen od 7SL RNA, zatímco u většiny savců jsou **SINE odvozené od tRNA**.

LINE mají velikost přibližně 6 kb, a SINE asi 300 bp. Oba elementy končí poly-A sekvencí a jsou ohraničeny krátkými přímými repeticemi (7 – 14 bp)

V genomu savců je přítomno **20 000 – 40 000 kopií LINE a až 1 milion SINE**, které jsou roztroušeny na různých místech genomu.

Tandemová opakování (1)

Satelitní DNA

Satelitní DNA je charakterizovaná velkým množstvím krátkých nebo dlouhých sekvencí, které mohou zahrnovat několik milionů nukleotidů. Preferenčně jsou asociovány s heterochromatinem a jsou typickými **centromerickými sekvencemi** u většiny savců. Rozsah a počet různých satelitních DNA tříd u určitého druhu jsou vysoce variabilní. Satelitní DNA je typicky uspořádána hierarchickým způsobem. Např. převažující myší satelit je vytvářen nejméně čtyřmi úrovněmi hierarchií opakování. Devítinukleotidový motiv GAAAAATGA nebo jeho modifikace vytvářejí trimer s délkou 28 – 30 bp. Tento element je duplikovaný na 60 bp fragment a ten se dále opakuje čtyřikrát. Výsledný 240 bp element vytváří konečnou jednotku v dlouhých úsecích.

Lokalizace satelitní DNA v centromerách naznačuje, že má určitou strukturní funkci v chromozomu, zřejmě **ve spojitosti s procesem segregace chromozomů**.

Telomerové sekvence

Telomery stabilizují konce chromozómů. Téměř všechny známé telomery sestávají z **mnohanásobných repetic**; repetiční jednotka u člověka je CCCTAA a vytváří elementy nejméně 10 – 15 kb. Délka telomerové repetitivní oblasti se zkracuje věkem, a zkracování telomery může být přímo korelováno s buněčným stárnutím.

Tandemová opakování (2)

(Tandem repeats)

- málo repetitivní (kolem 10 kopií opakujícího se motivu)
- středně repetitivní (10 – 10 000 kopií)
- vysoce repetitivní (10^5 – 10^6 kopií)

Minisatelity. Tyto elementy jsou vytvářeny 10 – 60 bp jednotkami, které mohou být opakovány až mnohotisíckrát. Minisatelity v savčích genomech existují v několika rodinách, přičemž členové každé rodiny vykazují malou až výraznou sekvenční homologii. Minisatelity jsou hypervariabilní – skutečnost, která byla využita pro “DNA fingerprinting“.

Mikrosatelity. Čtvrtá třída tandemových repeticí je označována jako **mikrosatelity** (nebo krátké tandemové repetice – **short tandem repeats, STR**). Hranice pro tuto třídu jsou spíše záležitostí dohody – mono-, di-, tri- a tetranukleotidové repetice mohou být jednoznačně klasifikovány jako jednoduché repetice. Je však třeba zdůraznit, že genom zahrnuje kontinuitu tandemových repeticí vzhledem k délce repetiční jednotky. Např. nedávno byly charakterizována lidská hexanukleotidová repetice.

Výskyt: V eukaryontních genomech se hojně vyskytuje repetice $(CA)_n$. V savčích genomech je přítomna v 50 000 – 100 000 kopií (lokusech). Tyto elementy jsou rozptýleny v celém genomu. V některých oblastech mají nižší hustotu (centromery, telomery, organizátor jadérka). V genomech jsou přítomny ve velkých počtech také jiné motivy; podrobněji o mikrosatelitech pojednáme v souvislosti s genetickým polymorfismem.

Genetický polymorfismus repetitivní DNA

Minisatelity: vhodné jsou takové minisatelity, které mají repetiční jednotku 10 – 20 bází. Minisatelity jsou přítomny na několika místech různých chromozómů. V minisatelitech se vyskytují četné mutace, jak v počtu repeticí, tak v sekvencích. Použitím restriktivního enzymu, kterým naštěpíme genomovou DNA, následnou elektroforetickou separací v agarózovém gelu a hybridizací se značenou minisatelitovou sondou získáme unikátní vzory (DNA fingerprints) pro každé zvíře. Tato metoda je využitelná pro ověřování rodičovství, v ochraně přírodních zdrojů, evolučních studiích aj. Přispěla také ke genovému mapování.

Telomerový a satelitní (centromerový) DNA polymorfismus: Polymorfismus délky jako důsledek různého počtu opakování jednotek byl také pozorován pro centromerovou satelitní DNA. Podobně byl prokázán polymorfismus také v telomerových repeticích. Oba tyto typy polymorfismu mohou být užitečné v genovém mapování, jelikož mohou definovat začátek a konec chromozómových ramének.

Polymorfismus mikrosatelitů: Polymorfismus mikrosatelitů je dán opakováním různého počtu jednotek (např. CA). Studium tohoto polymorfismu bylo zjednodušeno zavedením a použitím PCR. Princip spočívá v syntéze primerů – úseků DNA, které přiléhají k mikrosatelitové oblasti, a tyto primery jsou využity k amplifikaci těchto fragmentů. Následnou elektroforézou můžeme identifikovat fragmenty rozdílné délky.

Studium polymorfismu mikrosatelitů se významně uplatňuje v projektech studia genomů hospodářských zvířat, zejména pro genové mapování, pro ověřování rodičovství aj. Výhodná aplikovatelnost je umožněna vysokým výskytem mikrosatelitů, hypervariabilitou, a ve většině případů stabilní mendelistickou dědičností.

Nemoci způsobené expanzí trinukleotidových repetic

*Pokud jsou uvnitř nebo v blízkosti genů, mohou mít mikrosatelity, resp. jejich různá délka, závažné důsledky, např. v heterogenní skupině monogenních nemocí podmíněných expanzí trinukleotidových repetic. Nejznámějším příkladem je **Huntingtonova chorea**, fatální neurologické onemocnění s nástupem v dospělosti, projevující se jako demence s extrapyramidovou poruchou motoriky. V genu pro huntingtin je **repetitivní sekvence (CAG)_n**, která kóduje úsek bílkoviny tvořený **zbytky glutaminu** (polyglutaminový úsek, polyglutamine tract). Za normálních okolností mají lidé méně než 20 trinukleotidů CAG a tedy i glutaminů v huntingtinu, kde tyto tvoří důležitou doménu pro interakce s jinými proteiny. Pokud se však mutací toto množství zvětší nad 30 glutaminů, protein nepracuje správně (jak přesně je předmětem rozsáhlého výzkumu a je mimo rozsah tohoto textu) s výsledným progresivním odumíráním neuronů v nucleus caudatus. U jiného onemocnění, **myotonické dystrofie** (svalová dystrofie se svalovou slabostí provázenou paradoxně zvýšeným svalovým tonem) se nachází patologická **expanze trinukleotidu CTG v 3' nepřekládané oblasti genu DMPK** (dystrophia myotonica protein kinase). Mutantní mRNA má sama o sobě patogenní potenciál, škodí pravděpodobně sekvestrací různých transkripčních faktorů.*

Mechanismy expanze/zmenšení tandemových repetic

Prvním mechanismem, který přispívá k polymorfismu tandemových repetic je **nerovnoměrný crossing-over**. To je typické zvláště pro větší repetice. Malé mikrosatelitní repetice často mění svoji délku díky **chybám při syntéze DNA**, např. mechanismem "klouzáni" polymerázy. Na čele replikace není dvojitá šroubovice DNA ještě extrémně stabilní a podléhá významným termálním fluktuacím. Pokud právě DNA polymeráza replikuje mikrosatelit, DNA řetězce během fluktuace nemusí reasociovat exaktně, ale s posunem o několik jednotek repetice. Tento mechanismus je zesílen u některých typech repetic, které mohou stabilizovat metastabilní stavy tvorbou dvouřetězcové kličky, např. CAG/CTG trinukleotid.

Minisatelitová sonda – použitá pro test paternity u psů (Southernův bloting, po štěpení DNA enzymem *Hae*III)

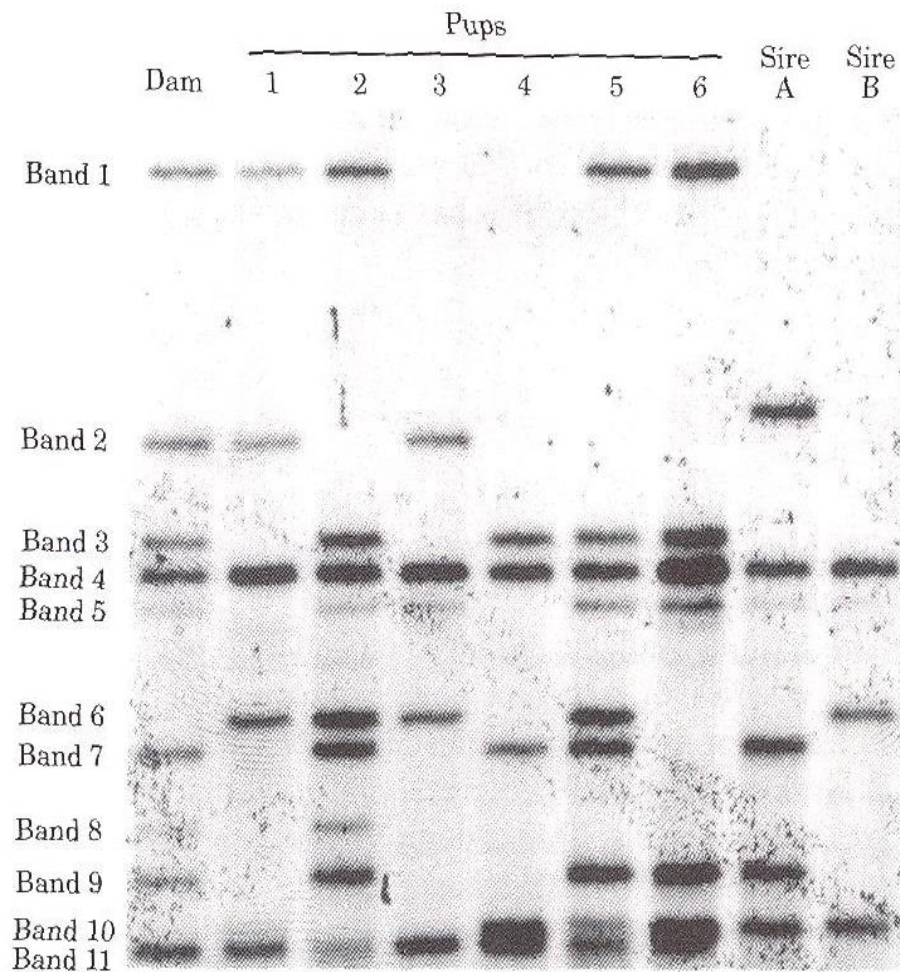


Fig. 12.2 A paternity test in dogs using Southern analysis with the enzyme *Hae*III and a multi-locus human minisatellite probe consisting of 29 tandem repeats of a 16 bp core sequence.

**Proužek 6 chybí u feny a u psa A, ale je přítomen u psa B a 4 štěňat.
Pes A tedy nemůže být otcem štěňat.**

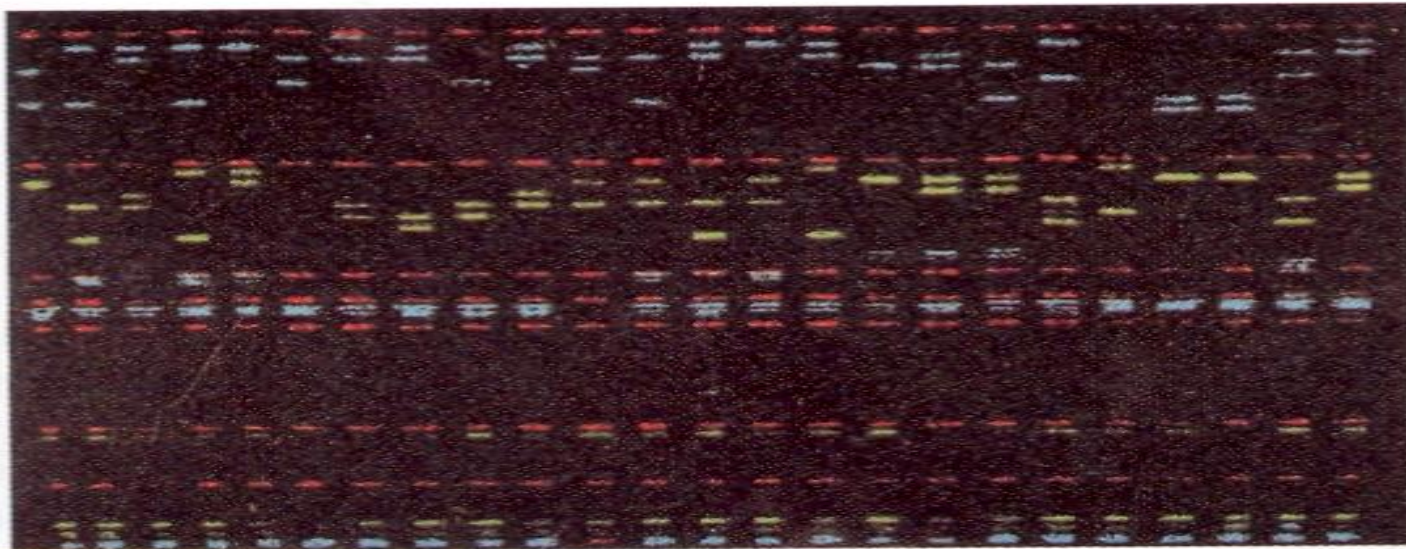
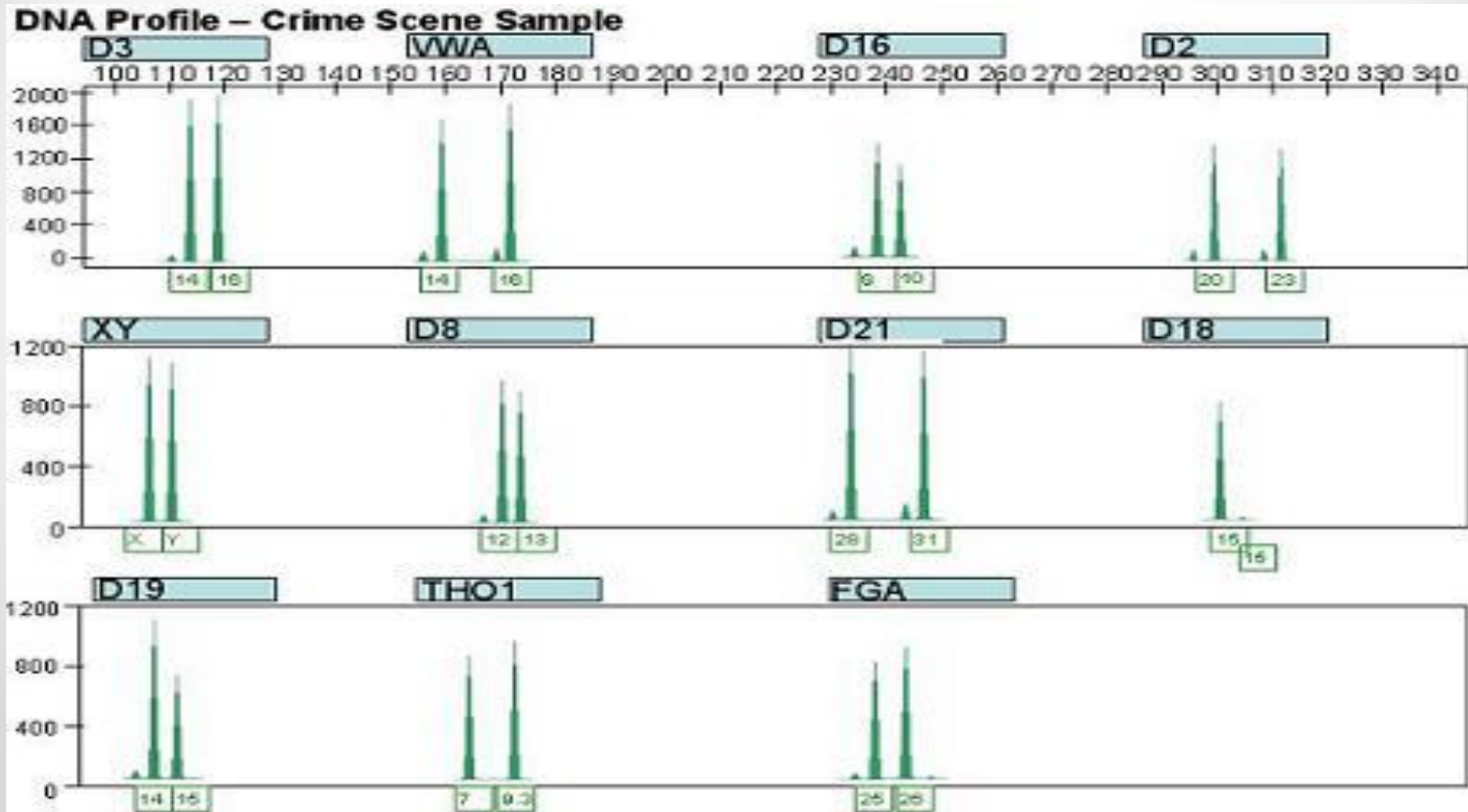


Figure 6.17 The use of microsatellite analysis in genetic profiling.

In this example, microsatellites located on the short arm of chromosome 6 have been amplified by PCR. The PCR products are labeled with a blue or green fluorescent marker and run in a polyacrylamide gel, each lane showing the genetic profile of a different individual. No two individuals have the same genetic profile because each person has a different set of microsatellite alleles, the alleles giving rise to bands of different sizes after PCR. The red bands are DNA size markers. Image supplied courtesy of PE Biosystems, Warrington, UK, and reproduced with permission.

Mikrosatelite – fragmentová analýza (multiplex PCR)

- Three chromatograms are shown, one for each of three distinct fluorophores.
- For each fluorophore, three or four loci are examined.
- The PCR primers are designed to give products with non-overlapping sizes for loci using the same fluorophore.



Analýza genomu

Odhalování variability DNA a celková analýza genomů zvířat zahrnuje:

- a) identifikaci genetických markerů
- b) konstrukci genetických map
- c) konstrukci map QTL
- d) lokalizaci a popis genů podílejících se na projevu užitkových vlastností
- e) určení úplné sekvence DNA genomů
- f) funkční analýzu genomu

Cíl: Využití v selekci podporované markery (MAS) a genomické selekci