

Princíp kódovania informácie v DNA v skratke

Gubáš Marcel · Informačné technológie, Medicína

06.10.2010



Článok sa zaoberá základmi problematiky genetického kódu zapísaného v makromolekule DNA. Opisuje organizáciu zápisu dát na molekulárnej úrovni, uchovávanie a prenos informácie z generácie na generáciu a stručne opisuje spôsob aplikácie genetického kódu do reálneho fyzického sveta. V budúcnosti bude s veľkou pravdepodobnosťou genetické inžinierstvo silne závislé na odbore aplikovaná informatika.

Preto je potrebné, aby informatici pôsobiaci v tejto oblasti mali aspoň základné poznatky problematiky na biochemickej úrovni a funkčne porozumeli princípom uloženia, prenosu a aplikácie genetickej informácie. Informatika je neodmysliteľnou súčasťou mapovania genómu živých organizmov a ponúka možnosť simulácie pre budúce genetické manipulácie.

1. Úvod

Ľudstvo využíva metódy šľachtienia a kríženia už od dávnych prvopočiatkov civilizácie. Prví poľnohospodári intuitívne odpozorovali, že ak zasadia tie najlepšie semienka zo svojej úrody, ďalší rok bude nová úroda bohatšia. Takto sa po tisícky rokov vyšľachtilo napríklad obilie z obyčajných a na živiny málo bohatých tráv. Obdobne sa šľachtienie aplikovalo aj na plemená koňov, alebo psov.

V 19. storočí rozmach vedy a racionálneho myslenia neobišiel ani problematiku genetiky a dedičnosti. Poznatky sa preniesli z intuitívnej a empirickej úrovne na úroveň funkčných závislostí. Charles Darwin popísal všeobecný systém evolúcie, v ktorej predpokladal nosiče medzigeneračných informácií. Český prírodovedec Gregor Mendel odpozoroval a sformuloval základné pravidlá dedičnosti na konkrétnej elementárnej úrovni organizmu. Zo zistených skutočností vyplývalo, že „najzákladnejšou vlastnosťou všetkých živých organizmov je ich genetická kontinuita“[2].

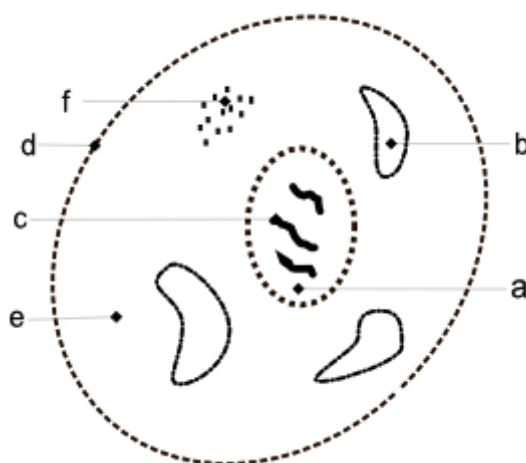
Vedecko-technická revolúcia v povojnovom období preniesla poznanie z funkčnej úrovne na materiálnu atomárnu úroveň. Pomocou precíznej techniky a pokročilých laboratórnych výskumov opísala organizáciu informácie, deje súvisiace s genetickým kódom a prenáša poznanie na novú úroveň – genetické inžinierstvo.

2. Lokalizácia genetickej informácie v tele organizmov

„Bunka je základná štruktúrna a funkčná jednotka rastlinných a živočíšnych tiel“[4]. Základnou stavebnou jednotkou mnohobunkových organizmov je eukaryotická bunka, ktorá na rozdiel od prokaryotickej obsahuje bunkové jadro. Polyméry kyseliny deoxyribonukleovej nesúce genetickú informáciu sú organizované v chromozómoch a nachádzajú sa v bunkovom jadre. Vďaka jadru sú oddelené od metabolických pochodov bunky, čo zvyšuje spoľahlivosť uchovania zápisu.

Na metabolizme eukaryotickej bunky sa podieľa organela zvaná mitochondria. Predpokladá sa, že mitochondrie sa vyvinuli z baktérií, ktoré boli pohltené predkami eukaryotických buniek. Nasvedčuje tomu aj fakt, že mitochondrie obsahujú vlastnú mDNA (mitochondriálna DNA). Z celkovej DNA obsiahnutej v bunke tvorí mDNA približne stotinový podiel a neopisuje na rozdiel od jadrovej DNA kompletný genóm. Uchováva napríklad informácie potrebné pri replikácii mitochondrií. Pre potreby spoznávania genómu organizmov vrátane človeka je zaujímavá hlavne jadrová DNA.

Ľudské telo obsahuje okolo 210 typov buniek v celkovej počte rádovo 10^{14} kusov. V každej somatickej (telovej) bunke sa nachádza kompletná genetická informácia, teda každá bunka má úplnú kópiu genómu jedinca. Výnimku tvoria bezjadrové bunky červených krviniek.



Obr. 1. Bunka.

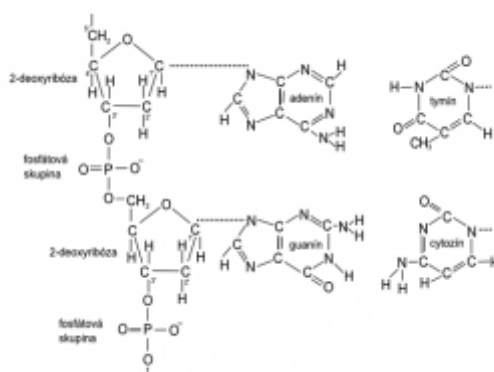
Na obrázku 1. je nakreslený silne zjednodušený model bunky obsahujúci len časti nevyhnutné pre základné pochopenie problematiky genetickej informácie. Bunku obaľuje d) bunková membrána. Vnútro bunky je vyplnené e) cytoplazmou, v ktorej sa nachádzajú organely ako b) mitochondrie, alebo a) bunkové jadro. V jadre bunky sa nachádza DNA organizovaná do c) chromozómov. Ribozómy f) sa podieľajú na aplikácii informácie z jadra do fyzických štruktúr bunky.

3. Štruktúra DNA a organizácia zápisu informácie

Kyselina deoxyribonukleová (v skratke DNA) patrí do skupiny nukleových kyselín. „Nukleové kyseliny sú biologické makromolekuly, ktorých úlohou je uchovávať a prenášať genetickú informáciu“[2]. DNA je tvorená atómami uhlíka, dusíka, fosforu, kyslíka a vodíka. Tento pomerne jednoduchý a stručný zoznam prvkov z počiatku nebudil dojem schopnosti spoľahlivo niesť rozsiahlu genetickú informáciu. „Mechanizmus, ktorým je dedičná informácia kopírovaná z bunky do bunky a ako

môžu byť proteíny určené inštrukciami v DNA, zostal tajomstvom až do roku 1953, kedy James Watson a Francis Crick určili štruktúru DNA“[3].

Primárna štruktúra DNA pozostáva z poradia nukleotidov zreťazených do polynukleotidového reťazca. Nukleotid je základná stavebná časť DNA. Skladá sa z kyslej fosfátovej skupiny, neutrálnej 2-deoxyribózy a zásaditej purínovej (adenín, guanín), alebo pyrimidínovej (tymín, cytozín) bázy. Ich molekulárna štruktúra je znázornená na obrázku 2. Fosfátová skupina sa prostredníctvom kyslíka viaže na voľný elektrón 5' uhlíka 2-deoxyribózy. Ďalším kyslíkom sa viaže na 3' uhlík inej 2-deoxyribózy. Väzba medzi fosfátovými zvyškami a 2-deoxyribózami sa takto opakuje a vytvára zreťazenie, kostru DNA.



Obr. 2. Atomárna štruktúra deoxyribonukleotidu.

Zásaditá časť nukleotidu sa pripája prostredníctvom atómu dusíka na 1' uhlík deoxyribózy. Nazýva sa báza a plní funkciu elementárnej jednotky kódovania informácie. Nukleotidové bázy sú postavené na základe pyrimidínu, alebo purínu. Ich naviazanie na zreťazenú kostru vytvára súvislý zápis informácie v abecede zostavenej zo štyroch písmen odvodených od názvov báz ACGT (1). V kyseline ribonukleovej RNA, jednoduchšej nukleovej kyseline, ktorá asistuje pri aplikácii informácie, sa namiesto tymínu viaže na ribózu pyrimidínová báza uracil (2). Funkcia F1 (3) popisuje jednoduchý preklad písmen z abecedy DNA do abecedy RNA.

$$\Sigma_{DNA} = A, C, G, T \quad (1)$$

$$\Sigma_{RNA} = A, C, G, U \quad (2)$$

$$F1(b) : \Sigma_{DNA} \rightarrow \Sigma_{RNA} = \begin{array}{l} A, b = A \\ C, b = C \\ G, b = G \\ U, b = T \end{array} \quad (3)$$

Všetky slová dĺžky 3 nad abecedou DNA, s výnimkou troch ukončovacích slov, reprezentujú aminokyseliny, stavebné jednotky bielkovín. „Proteíny tvoria väčšinu suchej hmotnosti bunky“[3], ich pomocou je potom definovaný každý organizmus. Slová dĺžky 3 v DNA kódujú 20 rôznych aminokyselín. Funkciu z množiny abecedy do množiny slov sa podarilo popísať pomocou pozorovania mRNA, nukleovej kyseliny, ktorá aplikuje informáciu z jadrovej DNA do bielkovinovej syntetizácie. Funkcia je určená tabuľkou číslo 1. Odborne sa slová dĺžky 3 v RNA nazývajú triplety (kodóny).

	U	C	A	G
U	UUU Fen	UCU Ser	UAU Tyr	UGU Cys
	UUC Fen	UCC Ser	UAC Tyr	UGC Cys
	UUA Fen	UCA Ser	UAG —	UGA —
	UUG Fen	UCG Ser	UAA —	UGG Try
C	CUU Leu	CUU Leu	CAU His	CGU Arg
	CUC Leu	CUC Leu	CAC His	CGC Arg
	CUA Leu	CUA Leu	CAA Gln	CGA Arg
	CUG Leu	CUG Leu	CAG Gln	CGG Arg
A	AUU Izo	ACU Tre	AAU Asn	AGU Ser
	AUC Izo	ACC Tre	AAC Asn	AGC Ser
	AUA Izo	ACG Tre	AAA Lyz	AGA Arg
	AUG Met	ACA Tre	AAG Lyz	AGG Arg
G	GUU Val	GCU Ala	GAU Asp	GGU Gly
	GUC Val	GCC Ala	GAC Asp	GGC Gly
	GUA Val	GCA Ala	GAA Glu	GGA Gly
	GUG Val	GCG Ala	GAG Glu	GGG Gly

V tabuľke 1 je zoznam všetkých slov dĺžky 3 a k nim sú priradené skratky aminokyselín, rozpísané v tabuľke 2.

Tab. 2. Zoznam skratiek.

Skratka	Význam
Fen	Fenylalanín
Leu	Leucín
Izo	Izoleucín
Val	Valín
Ser	Serín
Pro	Prolín
Tre	Treonín
Ala	Alanín
Tyr	Tyrozín
His	Histidín
Gln	Glutamín
Asn	Asparagín
Lyz	Lyzín
Asp	Kyselina asparágová
Glu	Kyselina glutámová
Cys	Cystein
Try	Tryptofán
Arg	Arginín
Gly	Glycín
Met	Metonín

Niektoré aminokyseliny sú zapísane viacerými tripletmi, iné kóduje len jeden kodón. Metonín reprezentovaný postupnosťou AUG plní navyše funkciu iniciačného kodómu a nachádza sa na začiatku súvislého zápisu. Terminačné triplety sú tie, ktoré sú v tabuľke bez popisu a boli pozorované na konci súvislého zápisu (UAG, UGA, UAA). Kódovanie samotných bielkovín slovami, ktoré reprezentujú aminokyseliny, predstavuje jazyk nad abecedou DNA. Celkový zápis genetickej informácie organizmu jedinca môžeme prirovnať k beletrii.

Sekundárna štruktúra DNA vystihuje podstatu pravotočivej dvojzávitnice znázornenej na obrázku 3. Dve dlhé polynukleotidové vlákna, reťazce nukleotidov naviazaných na kostru, „sú vzájomne spojené vodíkovými mostíkmi medzi bázami nukleotidov“[3]. Z priestorových a chemických dôvodov sa spolu viažu adenín a tymín dvoma vodíkovými mostíkmi a guanín s cytozínom tromi vodíkovými mostíkmi. Tento jav je pomenovaný ako komplementarita báz a popisuje ho funkcia komplementarity F2 (4).

$$F2(b) : \Sigma_{DNA} \rightarrow \Sigma_{DNA} = \begin{array}{l} A, b = T \\ C, b = G \\ G, b = C \\ T, b = A \end{array} \quad (4)$$

Predpokladá sa, že takáto organizácia zaručuje vysokú spoľahlivosť zápisu genómu, pretože sila vodíkových väzieb je vysoká a odolná voči vonkajším vplyvom. Informácia musí byť zaznamenaná bez nepriaznivých zmien veľmi dlhé obdobia. Navyše zápis do dvoch reťazcov naraz znižuje riziko zámieny nukleotidov počas replikácie pri bunkovom delení. Jednotlivé vlákna sú vzájomne komplementárne a antiparalelné. To znamená, že ak jeden reťazec začína na 5' uhlíku tak končí na 3' uhlíku, jeho antiparalelný reťazec začína na 3' uhlíku a končí na 5' uhlíku.



Obr. 3. Dvojzávitnica DNA.

Na obrázku 3. je možné sledovať zreťazenie častí popísaných v primárnej štruktúre DNA, nakreslených na obrázku 2. Kostra makromolekuly je tvorená zreťazenými pásmi fosfátových skupín (krátke biele pruhy) s 2-deoxyribózami (dlhé biele pruhy). Adenín a guanín sú oproti tymínu a cytozínu širšie tak, ako to vyplýva z ich molekulárnej štruktúry. Na jednu otočku dvojzávitnice vychádza približne 10,5 bázových párov s

výškou okolo 2,3nm. Šírka makromolekuly DNA je zhruba 1nm.

Terciárna štruktúra DNA popisuje stočenie dvojjávitnice v priestore do superhelixu, kde takto vzniká superšpiralizovaná DNA, chromozóm. Podstatu vyššej organizácie genetickej informácie sa snažia rozlúštiť tímy vedcov aj v súčasnosti. Spôsoby výskumu jednotlivých častí zápisu a ich súvisu medzi existujúcim organizmom sú materiálne a časovo náročné. V oblasti ľudského genómu prebieha zatiaľ len mapovanie postupnosti nukleotidov a ich vyššieho zaradenia do skupín. Vedecké manipulačné aktivity na ľudských jedincoch, nech majú akokoľvek ušľachtilý cieľ, sú v spoločnosti považované za absolútne nemorálne.

To poskytuje priestor aplikovanej informatike na uplatnenie svojich simulačných metód do praxe genetického inžinierstva. Z doposiaľ zistených faktov zlúčených s už zaužívanou terminológiou genetiky zavedenou ešte pred objavom DNA sa podarilo vystihnúť priestorová organizácia informácie v rámci veľmi dlhého reťazového zápisu. Reťazec makromolekuly DNA je u zložitejších organizmov dlhý desiatky až stovky miliónov bázových párov. Zistilo sa, že záznam obsahuje množstvo miest, kde sa, zatiaľ pre naše poznanie nezmyselne, niekoľko tisíc krát opakujú jednoduché postupnosti písmen DNA abecedy.

Postupnosti, ktoré sa javia ako zmysluplné sú separované práve týmito hluchými miestami a nazvali sa génmi. Gény tvoria súbor celkovej informácie, ktorá sa volá genóm. Z pohľadu fyzickej realizácie genómu, sa používa členenie na chromozómy. Zložitejšie organizmy majú makromolekulu príliš dlhú na to aby stála osamote, preto je rozčlenená na viacero častí, na chromozómy. Zatiaľ nebola objavená funkčná závislosť separácie informácie do chromozómov, preto sa zatiaľ táto separácia považuje za čisto fyzickú.

Ľudský genóm obsahuje približne $3,2 \times 10^9$ bázových párov, ktoré okrem hluchých miest tvoria okolo $2,5 \times 10^4$ génov, fyzicky organizovaných do 46 chromozómov. Ak toto množstvo vynásobíme počtom buniek 10^{14} dostávame približnú informačnú kapacitu priemerného ľudského tela v počte bázových párov.

4. Kopírovanie a aplikácia informácie

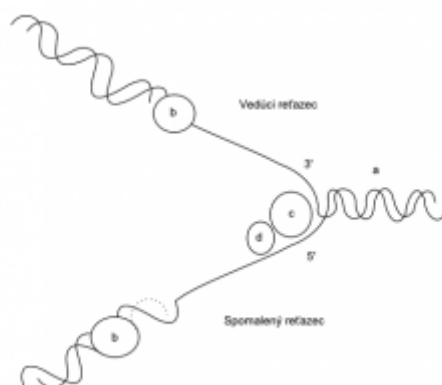
Replikácia DNA je proces kopírovania informácie prebiehajúci pri bunkovom delení. Sekundárna štruktúra DNA popisuje dvojjávitnicu makromolekuly ktorá je pevne spojená vodíkovými väzbami. Dva komplementárne reťazce vytvárajú dve šablóny pre genetickej informáciu. „Celý proces replikácie začínajú iniciačné proteíny, ktoré sa viažu na DNA a rozvíjajú jej dvojjávitnicu“[3]. Jedno vlákno sa otáča rýchlejšie a volá sa vedúce vlákno, to pomalšie sa nazýva spomalené vlákno. Na rad nastupuje DNA polymeráza, ktorá syntetizuje doplnok – komplement ku každému vláknu. Rýchlosť prepisu informácie v ľudskej bunke je približne 100 nukleotidov za sekundu. Celý genóm je potrebné prepísať čo najskôr, preto prepis informácie prebieha paralelne na viacerých miestach molekuly naraz.

Na obrázku 4 je pod písmenom a) označená pôvodná dvojjávitnica DNA makromolekuly, ktorá je prichytená na replikáciu po rozdelení na dve samostatné vlákna bunkovým enzýmom DNA helikázou. Ten svojimi chemickými pochodmi

spôľahlivo delí dvojzávitnicu bez toho aby menil štruktúru jednotlivých vlákien. Na obrázku je označený písmenom c). Na rad nastupuje DNA polymeráza b), ktorá syntetizuje doplnok - komplement ku každému vláknu podľa funkcie F2 (4) v smere vlákna od 5' uhlíku ku 3' uhlíku. Zrýchlené vlákno dokáže DNA polymeráza prepisovať plynule, tak ako ho DNA helikáza rozpletá.

Problém nastáva pri spomalenom vlákne, ktoré je voči zrýchlenému vláknu antiparalelné. Keďže DNA polymeráza je schopná prepisovať vlákno len v smere od 5' ku 3' uhlíku, je potrebný ďalší asistenčný aparát zabezpečujúci spoľahlivú a plynulú replikáciu spomaleného vlákna. Vedúci reťazec je syntetizovaný spojito v zmysle samostatných nukleotidov, no asistenčný aparát pri spomalenom vlákne spojito syntetizovať nemôže. Preto je reťazec DNA v smere od 3' ku 5' uhlíku syntetizovaný po častičkách.

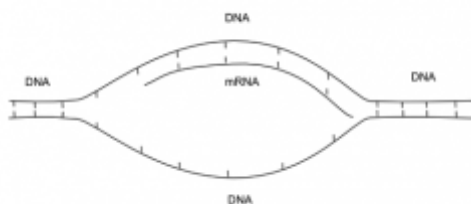
Pri spomalenom vlákne po DNA helikáze nastupuje na rad DNA primáza d), ktorá pripravuje krátke úseky RNA nazývané RNA primery. To sú práve spomínané častky po ktorých sa spomalené vlákno syntetizuje. Keď úsek RNA primeru dosiahne určitú dĺžku kompletárneho reťazca k spomalenému reťazcu, rádovo sto báz, dosyntetizuje sa k nemu kópia krátkej DNA nazývanej Okazakiho fragment. Na obrázku je tento fragment znázornený ako krátky čiarkovaný úsek vlákna.



Obr. 4. Replikácia DNA.

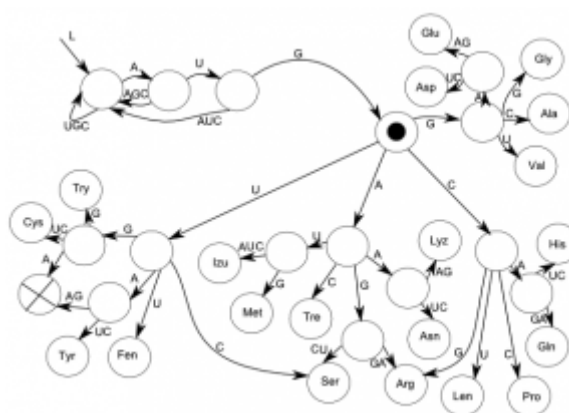
DNA polymeráza plní aj funkciu kontroly syntetizácie komplementárnych bázových párov. Zisťovanie chybných naviazaní je pomerne jednoznačné, pretože nukleotidy sa na seba viažu presne podľa priestorových pravidiel a podľa počtu vodíkových väzieb. Priemerná chyba replikácie DNA u človeka predstavuje 1 chybný replikovaný nukleotid k 10^9 správnym. Priemerne sa každá ľudská bunka rozdelí s 3 chybami.

Transkripcia DNA je prvým krokom aplikácie informácie a prebieha za asistencie mRNA (media RNA) v bunkovom jadre. „Prvým krokom pre uplatnenie genetickej informácie v bunke je prepísanie časti kódu DNA do nukleotidovej sekvencie RNA“[3]. V prípade potreby čítania informácie sa na určitom mieste za pomoci iniciačných proteínov rozdelí dvojzávitnica DNA. Na rozdelenom mieste sa syntetizuje komplementárny reťazec RNA tak, že na miesto tymínu sa doplní uracil, ktorý sa viaže na adenín. Po ukončení syntetizácie mRNA opúšťa novo vytvorený reťazec bunkové jadro a makromolekula DNA sa opäť spojí.



Obr. 5. Transkripcia DNA.

Translácia DNA je druhým krokom aplikácie informácie a odohráva sa v organelách bunky ribozómoch, mimo bunkového jadra. Na molekulu mRNA so zápisom konkrétnych aminokyselín v tripletoch sa naväzujú voľné aminokyseliny spojené s komplementárnym tripletom. Vytvára sa zhluk bielkovín, budúca bunková konštrukcia, ktorá nakoniec molekulu mRNA rozpustí. Funkcia ribozómu je jednoducho modelovateľná moorovým konečným automatom na obrázku 6.



Obr. 6. Funkčný model ribozómu.

Vstupom do automatu je postupnosť písmen abecedy DNA. Automat je skonštruovaný na základe tabuľky 1. Konečný stav je označený križikom a stavy označené skratkou aminokyseliny sú spojené so stavom označeným kruhom, no pre lepšiu prehľadnosť obrázku nie sú nakreslené.

5. Praktická aplikácia poznatkov

Poznatky genómu živých organizmov majú svoje uplatnenie už v súčasnosti. Jednoduché organizmy ako obilie sú pomerne dobre zmapované a ich relatívna jednoduchosť umožňuje pokusy s manipuláciou a pokročilé šľachtenie priamym zásahom do DNA. Darí sa pestovať obilie odolné na plesne a choroby, ktorého energetická výživová hodnota je vyššia ako u bežne používaného.

Omnoho zložitejšou problematikou je rozlúštenie kódu človeka, ktorá vyžaduje hromadnú medzinárodnú a medziodborovú spoluprácu. Okrem súčasnej jednoznačnej identifikácii jedinca pomocou vzorky DNA a predpovedanie niektorých dedičných chorôb, ponúka úplné predpovedanie dedičných chorôb s možnosťou ich liečenia ešte v štádiu jednobunkového embrya. Mnohe využitie týchto poznatkov je zatiaľ z oblasti fantázie, preto sa treba sústrediť na tie reálnejšie, dosiahnuteľnejšie ciele.

V minulosti boli založené medzinárodné organizácie zaoberajúce sa mapovaním ľudskej genetickej informácie ako HGP (Human genome project), HUGO (Human

genome organization), NIH, DOE a ďalšie. Omnoho reálnejší prínos technológie spracovania genetickej informácie je porovnávanie dvoch vzoriek, ktoré prebieha aj na Slovensku za účelom jednoznačnej identifikácie individua.

6. Použitá literatúra

1. Ferenčík M., Škárka B., Novák M., Turecký L., „Biochémia“, Bratislava 2000
 2. Jindra A., „Biochémia, Molekulárnobiologické a farmaceutické aspekty“, Osveta 1983
 3. Alberts B., Bray D., „Essential cell biology“, 1998
 4. Ušáková K., „Biológia pre gymnáziá 1“
-