

P6. Evoluce proteinů

Genomová dynamika a evoluce proteinů jsou ve vztahu jako příčina a následek. *Genomová dynamika byla zdrojem modulární stavby genomů.* V tom spočívá její evoluční význam. I genomy prokaryot, které se globálně (například analýzou kinetiky reasociace denaturované DNA) jeví jako strukturně homogenní, při detailním rozboru vykazují modulární mikrostrukturu. Modularita se naplno projevuje v eukaryotických genomech.

Členění genomů do rozličných strukturních segmentů - chromosomů, doménového uspořádáním sekvencí v závislosti na obsahu párů C:G,(A:T) do tzv. isochorů (Salinas, Matassi et al. 1988; Bernardi 1989), hypometylovaných 5'-CpG-3' ostrůvků v genomech obratlovců (Gardiner-Garden and Frommer 1987), bloků různých typů sekvencí se strukturním kódem pro heterochromatinové a euchromatinové domény (chromatinový strukturní kód (Vogt 1990), rodin amplifikovaných genů (paralogních genů) a různých typů vmezeřených repetitivních a mobilních sekvencí; doménová struktura genů - introny a exony, subdomény kódujících sekvencí, regulační komponenty genomů, nakonec i tripletový zápis genetické informace.

☞ Modularita má počátky v prvotních sekvencích, vznikajících nekatalyzovanou replikací RNA v prebiotických podmínkách. Délka takových oligonukleotidů byla omezena málo účinnou kondensací prekursorů i vysokou frekvencí replikačních chyb. Řešním byly dělené genomy s vzájemně kooperujícími replikátory (viz hypotéza o hypercyklech).

Z tohoto předpokladu vyšel S. Ohno (Ohno 1987a) navrhl mechanismus tvorby delších ribonukleotidových řetězců tendemovým opakováním krátkých repetit (---▶▶▶▶---), optimálně o délce 5 nebo 7 bází.

Zde polymerace za katalýzy Zn^{++} , který umožňuje tvorbu korektních 3'-5' vazeb, s aktivací nukleosidtrifosfátů (NTP) imidazolem (Bridson and Orgel 1980):

$NTP + Zn^{++} + \text{imidazol} \Rightarrow \text{polymerace} \Rightarrow (\text{pentamer})_2$

GCCAAGCCAA \Rightarrow 1. replikace za katalýzy Zn^{++} \Rightarrow

GCCAAGCCAA \Rightarrow separace vláken a reasociace s posunem \Rightarrow
CGGUUCGGUU

GCCAAGCCAA \Rightarrow 2. replikace za katalýzy Zn^{++} \Rightarrow
CGGUUCGGUU

GCCAAGCCAA**GCCAA** \Rightarrow opakováním tohoto procesu postupné
CGGUUCGGUUCGGUU prodlužování řetězce o další pentamery

☞ *Oligopeptidy s periodickým opakováním kratších modulů umožňují vznik konformací s α -helixy a β -smyčkami, což jsou základní komponenty vazebných domén regulačních proteinů i katalytických center enzymů. Obecně vzato, pokud polypeptid dosáhne délky ~ 40 aminokyselinových zbytků, obsahuje dostatek vnitřních kohezivních sil, aby zaujal určitou preferovanou prostorovou konformaci.*

Například rodina rodopsinových proteinů, jejíž původ předchází divergenci prokaryot a eukaryot, má společný základ v opakování heptamerů bází a konservativní stavbu zahrnující

7 α -helixových transmembránových domén,
4 extracelulární smyčky,
4 intracytoplasmatických smyček.

Odtud například evolvoval prasečí muskarin-acetylcholinový receptor, v němž jsou původní heptamerové sekvence nejlépe zachovány v částech genu kódujících hydrofobní α -helixové transmembránové domény (Ohno 1987b).

Ohno uvádí i další přesvědčivý argument pro oligomerový původ pragenů na základě analýzy 630pb sekvence, kódující histon H1 duhového pstruha; tato sekvence obsahuje vysoce nenáhodná opakování krátkých modulů: 25 přesných kopií pentameru CCAAG, 5 kopií příbuzného pentameru TCAAG, 2 kopie pentadekamerů GCCCCAAGAAGCCC a jejich derivátů o různé délce. Nejstarší předchůdce genu pro histon H1 pstruha patrně pocházel ze základního modulu CCAAG.

Pokud délka modulů (pentamerů, heptamerů), není delitelná číslem 3, pak opakováním vzniká maximální kódovací potenciál pro tripletový kód. Vznikne „protogen“, který, v našem případě, může kódovat peptid s pentamerovou periodicitou ve všech třech fázích čtení:

--- (Ala-Lys-Pro-Ser-Gln)_n --- .

Modulární struktura protogenu tudíž poskytuje výhodu pro evoluci proteinů, protože při náhodném vzniku stop-kodonu v jedné fázi čtení umožňuje náhradní čtení ve druhé nebo třetí fázi. Vhodnými mutacemi definujícími počátek (iniciaci) a ukončení čtení může vzniknout konvenční gen. V dnešním H1 histonovém genu pstruha představuje kódující sekvence jeden čtecí rámec pro 210 aminokyselin. Kdyby ve dvou dalších čtecích rámcích náhodou vznikly iniciační kodony, oba rámce by produkovaly už jen zkrácené polypeptidy, neboť naakumulovaly stop kodony (po dvou stop-kodonech v každém z obou alternativních rámců Ohno 1987a).

„Rekapitulace“ nevyužitých alternativních možností translace jsou však i dnes možné. Pozoruhodný příklad rychlé adaptace na neobvyklé podmínky tímto mechanismem rovněž uvádí Ohno, 1984. Adaptace zahrnuje vznik enzymu, schopného štěpit nepřírodní zdroj uhlíku - oligomery nylonu. Nový enzym s 392 aminokyselinovými zbytky vznikl v plasmidu *Flavobacterium* jednonukleotidovým posunem čtecího rámce (insercí T) a zrušením původního, delšího čtecího rámce s 427 kodony. Strukturu původního proteinu lze odvodit z opakování dekameru --CAGCTGGCAGG--. Čtení téže genové sekvence v různých fázích zůstalo zachováno u některých prokaryot a virů: omezuje sice informační obsah - zejména jeho mutační flexibilitu, ale u méně komplexních forem umožňuje snížit komplexitu genomu.

Porovnání sekvencí archaických proteinů, například enzymů účastných v metabolismu sacharidů ukazuje, že moduly podmiňující specifické funkce jeví mimořádnou konservativitu, zatímco jiné části proteinu mohly při divergenci hromadit neutrální mutace. Tak glycerinaldehyd 3-fosfát dehydrogenasy obratlovce (prasete) a bakterie (*Bacillus stearothermophilis*) mají, přes evoluční vzdálenost ~3,5 miliardy let, 177 společných aminokyselin z 332.

☞ *Úroveň prebiotických oligonukleotidových mikro-modulů tvoří evolučně nejstarší vrstvu protein-kódujících sekvencí.*

☞ *Velmi efektivní mechanismus inovativní evoluce proteinů se zakládá na duplikaci (ampifikaci) genů a výsledné modulární stavbě genomů.*

Je účelné odlišit *tři různé mechanismy*, které se na této úrovni mohou uplatnit:

(a) *Vznik nových funkcí na základě postupných modifikací paralogních genů.*

Nereciproké genetické výměny mohou generovat rodiny sekvencí, odvozených od jedné základní předlohy. Zmnožení sekvencí pak dává prostor dalším genetickým změnám proto, že výchozí sekvence zůstává selekcí zachována pro původní funkci. Variantní sekvence se proto mohou šířit skrytě před selekcí mechanismem molekulárního drajvu.

Asimilace amplifikovaných sekvencí je omezena možným konfliktem s funkční integritou organismu. Nadbytečné sekvence jsou pak inaktivovány (např. umlčeny metylací cytosinu), degradovány delecemi a/nebo postupně eliminovány rekombinačními mechanismy. Pokud jsou zmnožené sekvence asimilovány, může následně dojít k jejich *adopci*: k *neo-funkcionalizaci* (sekvence nyní kódují nové

funkce), nebo k *sub-funkcionalizaci* díky diferenciálnímu působení různých regulačních sekvencí.

(b) *Vznik nových funkcí přeskupením intragenových modulů.*

Modulární stavba genomů se odráží i v úrovni genů mozaikovým uspořádáním dvou typů sekvencí: kódujících *exonů* a „nekódujících“ *intronů*. Jednotlivé exony lze ve většině případů (ne však nutně) ztotožnit s dílčími funkčními moduly (doménami) proteinů. Modulární délky kódujících sekvencí ($\sim 10^2$ pb) a délky intronů se v dnešních genech eukaryot nacházejí v poměru 1:10 až 1:100.

Použití pojmů „kódující/nekódující“ se v běžně vztahuje k translaci: mozaice exonů odpovídá mozaika peptidů, které poskládány ve stejném pořadí jako exony, tvoří kontinuum bílkovinné molekuly; exony tedy kódují pořadí aminokyselin v bílkovině.

Ale introny rovněž obsahují kód: určují strukturu ribozymu nutného k jejich vystěpení z pre-mRNA (introny I typu) nebo i proteinové enzymy s reverzně-transkripázovou aktivitou pro jejich vlastní retrotransposici (introny II typu).

Dosud nerozhodnutou je otázka o původu intronů. Koexistují dvě hypotézy: (a) Introny jsou původním stavebním materiálem genomů; u prokaryot došlo postupem času ke ztrátě intronů (archebakterie si dosud introny zachovaly, alespoň jako součást primárních transkriptů tRNA a rRNA). Existuje určitá pozitivní korelace mezi obsahem intronů v genech a komplexitou organismů: Jednobuněčná kvasinka má více než 90% genů bez intronů, drozofila ~50%, v genomech obratlovců naopak více než 90% genů obsahuje introny. (b) Introny byly do kódujících sekvencí integrovány (transponovány) během evoluce eukaryotických genomů.

Varianta (a) se jeví jako pravděpodobnější; nepřímou informaci v tomto směru poskytují srovnání intronů-prostých genů současných prokaryot a homologních genů symbiogeneticky vzniklých organel (např. chloroplastů), které dosud obsahují introny.

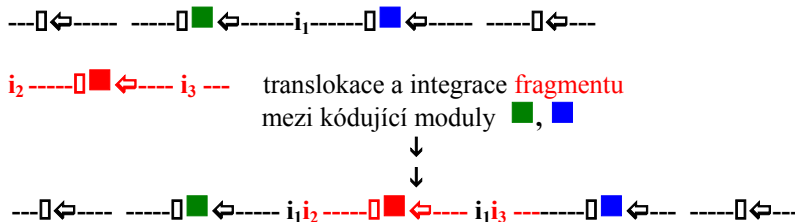
Varianta (b) naráží na problém porušení čtecí fáze insercí intronů do souvislých rámců. Pokud by se exony a introny pohybovaly navzájem zcela nezávisle, mohlo by náhodnou insercí intronů, a následnými translokacemi, docházet k nesmyslným situacím, např. u kódující sekvence **CCCGGGAACCC**:

- (1) **CCCGGGXYAAACCC**, inserce intronu XY mezi kodony neruší fázi čtení
- (2) **CCCGXYGGAAACCC**, inserce XY do kodonu **G↓GG** ruší fázi čtení
- (3) **CCCGGXYGAAACCC**, inserce XY do kodonu **GG↓G** ruší fázi čtení

Rekombinacemi mezi XY u sekvencí (2) a (3) nebo transposicemi podtržených motivů mezi (2) a (3) dojde k posunu fáze čtení a zrušení smyslu sekvence.

Přítomnost intronů ale rozšiřuje evoluční perspektivy proteinů.

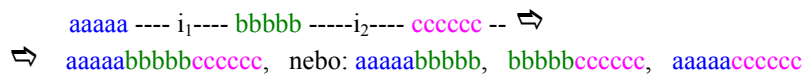
Pokud byly původní primitivní genomy mozaikou exonů a intronů od počátku, vznik sestřihu primárních transkriptů přenesl kombinační evolutivní „experimenty“ z úrovně genetické na úroveň RNA: varianty sestřihu, slučitelné s uspořádáním kódujících segmentů ve správné čtecí fázi vedly k emergenci nových proteinů.



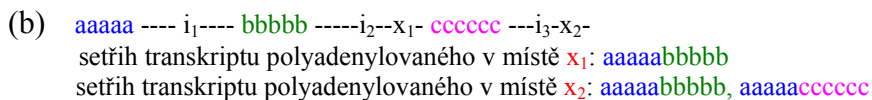
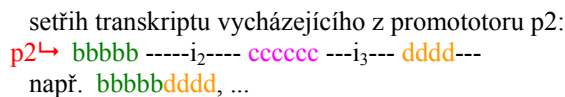
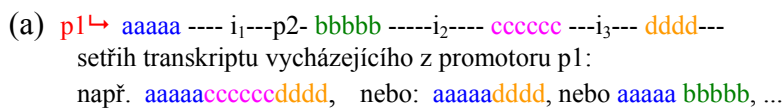
sekvence specifické pro sestřih: \leftarrow , 5'; \square , 3';
Kódující moduly si po translokaci zachovávají korektní čtecí rámec

Další možnosti vzniku variantních molekul proteinů jsou dány alternativním sestřihem primárních transkriptů:

alternativní (diferenciální) sestřih umožňoval produkci různých forem proteinů na základě téže, neměnné předlohy DNA, *při zachování pořadí exonů*:



Možnosti sestřihu koevolvovaly spolu s geny, které časem získaly alternativní promotory a/nebo alternativní signály pro polyadenylaci 3'-konců transkriptů:



Konstrukce různých forem proteinů alternativním sestřihem transkriptů z téhož genu jsou dosud zachovány a využívány během embryonálního vývoje a diferenciaci buněk.

Při porovnání struktur nehomologních proteinů s velmi odlišnými funkcemi objevíme přítomnost rozličných domén stejného původu v různých kombinacích. Je to důsledek jejich přeskupování („shuffling“) v ranné evoluci geonomů předcházející divergenci prokaryot a eukaryot; Na genových strukturách s informačně „prázdnými“ úseky se mohly uplatnit nehomologní výměny vedoucí k translokaci exonů. I v tomto případě náhodný „tvořivý“ rekombinační proces nesl významný podíl kombinací, které neměly funkční smysl. Rozhodným arbitrem nových funkcí byla selekce.

✌ **Závěrem:** modulární struktura genomů a genů umožnila efektivní evoluci proteinů s novými funkcemi díky rozličným způsobům přeskupování sekvencí („shuffling“). Na evoluci katalytických center a vazebných domén se podílela i graduální akumulace mutací v paralogních genech.

Doporučená četba:

- R. F. Doolittle and P. Bork, „*Evolutionarily mobile modules in proteins*“, Scientific American, October 1993: 34-40.
L. Patthy, „*Protein Evolution*“, Blackwell, 1999.
P. Turner, A. McLenna, A. Bates, M. White, „*BIOS Instatnt Notes - Molecular Biology*“, Taylor and Francis Ed., 2005.
B. Lewin, „*Genes IX*“, Jones and Bartlett, 2008.

Literatura

- Bernardi, G.(1989). "The isochore organization of the human genome." Annu. Rev. Genet. 23(637-61).
- Bridson, P. K. and L. E. Orgel (1980). "Catalysis of accurate poly (C) directed synthesis of 3'-5' linked oligoguanynes by Zn⁺²." J. Mol. Biol. 144: 567-577.
- Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J. Mol. Biol. 196: 261-282.

Ohno, S. (1984). "Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence." Proc. Natl. Acad. Sci. USA 81: 2421-2425.

Ohno, S. (1987a). "Evolution from primordial oligomeric repeats to modern sequences." J. Mol. Evol. 25: 325-329.

Ohno, S. (1987b). "Early genes that were oligomeric repeats generated a number of divergent domains on their own." Proc. Natl. Acad. Sci. USA 84: 6486-6490.

Salinas, J., G. Matassi, et al. (1988). "Compositional compartmentalization and compositional patterns in the nuclear genome of plants." Nucleic Acids Res. 16: 4269-4285.

Vogt, P. (1990). "Potential genetic functions of tandem repeated DNA sequence blocs in the human genome are based on a highly conserved "chromatin folding code"." Hum. Genet. 84: 301-336.