

Molekulární struktura a organizace eukaryotického genomu

Prof. RNDr. STANISLAV ROSYPAL, DrSc., Přírodovědecká fakulta UJEP, Brno

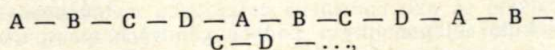
Klíčová slova: genové rodiny, opakující se skupiny genů, introny, exony, mezerníky, pseudogeny, upravené geny, transponovatelné elementy, imunoglobulinové geny

Struktura a organizace eukaryotického genomu je velmi široké téma zahrnující složitou problematiku, kterou v tak krátkém sdělení nelze uspokojivě v dostatečné míře vyloučit. Domnívám se proto, že pro účely tohoto článku je vhodnější zmínit se stručně jen o některých zvláštěnostech struktury a organizace eukaryotického genomu.

Z těchto zvláštností je to především způsob uspořádání genů na eukaryotické chromozomální DNA, existence transponovatelných elementů a zvláštnosti struktury genů pro imunoglobuliny.

Opakující se skupiny genů. Některé skupiny genů (geny kódující rRNA a histony) se na eukaryotické chromo-

zomální DNA mnohonásobně za sebou (v tandemu) opakují, takže se vyskytují v mnoha kopiích. V každé opakující se skupině mají geny zpravidla stejné pořadí, např.:



kde A, B, C a D je označení jednotlivých genů uvnitř opakující se skupiny; mezi jednotlivými skupinami genů, popř. i mezi jednotlivými geny se nacházejí tzv. mezeríky (viz dále).

Genové rodiny. Genovou rodinou se rozumí skupina příbuzných nikoli však identických genů. V příslušné genové rodině se geny částečně liší v sekvenci nukleotidů a kódují tedy bílkoviny s poněkud odlišnými sekvencemi aminokyselin. Jednotlivé geny téže rodiny se ve stejném vývojovém stadiu organismu nevyjadřují současně. Patří sem geny pro syntézu globulinů, aktinu, imunoglobulinů, vitelogeninu, ovalbuminu, chorionových bílkovin aj.

Opakující se geny rozptýlené po genomu. V eukaryotických buňkách se vyskytují též geny, které se neopakují v tandemu, ale jsou v genomu rozptýleny jednotlivě nebo po skupinách mezi jinými geny. Takové geny se označují jako rozptýleně opakované geny. Jsou to geny pro syntézu tRNA, snRNA a tzv. orfony. Orfony představují zvláštní případ rozptýleně opakovaných genů. Označují se tak kopie tandemově seskupených genů, které jsou jednotlivě rozptýleny mezi jinými geny. Orfony byly zjištěny u genů pro syntézu histonů a 5S-rRNA.

Introny, exony a mezeríky. U eukaryotických genů se často setkáváme s tím, že jsou uvnitř přerušeny krátkými sekvencemi, které nemají kódovací smysl; přepisují se do RNA, z níž se však během scelovacích procesů vyštěpí. Tyto sekvence se označují jako introny. Vlivem intronů může být gen na několika místech přerušen a rozdělen na několik kódujících oblastí, tzv. exonů. Exony se stejně jako introny přepisují do RNA, ale na rozdíl od intronů se jejich přepisy během scelovacích procesů spojí za vzniku celistvého přepisu genu.

Od intronů a exonů je třeba odlišovat mezeríky. Jsou to nekódující sekvence, které se nacházejí mezi geny nebo skupinami genů pro rRNA a histony. Jsou dvojího typu:

- mezeríky přepisované do pre-rRNA nebo hnRNA, které jsou pak v průběhu scelovacích procesů z pre-rRNA nebo hnRNA vyštěpeny, takže se pak již ve zralých molekulách rRNA nebo mRNA nevyskytují;
- mezeríky, které nejsou vůbec do pre-rRNA nebo hnRNA přepisovány.

Sekvence Alu. Sekvence Alu je citlivá k endonukleáze Alu I. Má délku 135 pb (u myši) a 300 pb (u člověka). V genomu člověka se vyskytuje v mnoha kopiích; na haploidní genom člověka připadá asi 400 000 kopií této sekvence. Většina těchto kopií je roztroušena po celém genomu, ale některé jsou též seskupeny do určitých oblastí genomu.

Sekvence Alu se přepisují do RNA a odtud pravděpodobně opět do DNA a ve formě sekvencí DNA se začleňují do chromozómu a tvoří pravděpodobně souhlasné sekvence v intronech.

Pseudogeny a upravené geny. Pseudogeny jsou neúplné duplikáty strukturálních genů, které prodělaly změny, v důsledku kterých se staly inaktivní. V eukaryotickém genomu se pseudogeny vyskytují často. Například některé globulinové geny se vyznačují inzercemi a delecemi, které způsobily, že jsou biologicky inaktivní a kódují funkčně inaktivní polypeptidové řetězce. Pseudogeny se obvykle vyskytují v genových rodinách.

K pseudogenům patří též tzv. upravené geny (processed genes). Jsou to pseudogeny vznikající reverzní transkripcí posttranskripčně upravené hnRNA. Od svých funkčních a biologicky aktivních genových protějšků se liší v těchto vlastnostech:

- Neobsahují introny.
- Na 3'-konci obsahují signální sekvenci pro polyadenylaci.
- V některých případech za signální sekvencí pro polyadenylaci obsahují skupinu adenosinových zbytků připomínajících poly(A).

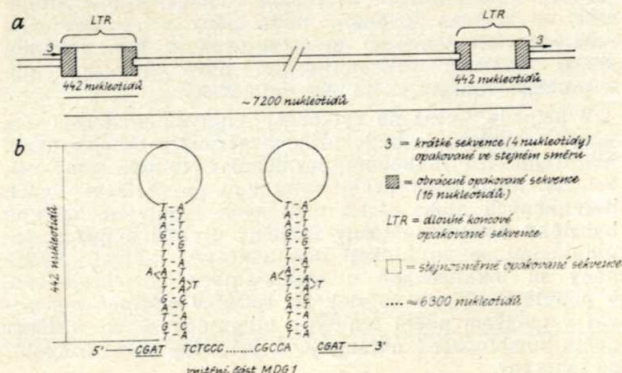
4. Na obou koncích jsou opatřeny sekvencemi dlouhými 10 až 20 pb, které se opakují v jednom směru (jednosměrně opakované sekvence). Touto vlastností se podobají sekvencím Alu a transponovatelným elementům.

Tvorba upravených genů zahrnuje tyto děje:

- Přepis strukturálního genu rozděleného introny do hnRNA.
- Posttranskripční úpravu hnRNA, při níž se vyloučí přepisy intronů.
- Přepis posttranskripčně upravené hnRNA, tj. mRNA, do DNA mechanismem podobajícím se nebo shodným s reverzní transkripcí. DNA, která touto transkripcí vznikne, představuje upravený gen.
- Začlenění upraveného genu do chromozómu. Začlenění je katalyzováno asi transpozázou jako transpozice transponovatelných elementů.

Transponovatelné elementy. Podobně jako u prokaryot jsou též u eukaryot některé úseky chromozomální DNA obsazeny transponovatelnými (přemístitelnými) elementy (transpozony). Z nich nejlépe jsou prostudovány transponovatelné elementy vyskytující se u *Saccharomyces cerevisiae*, *Drosophila melanogaster* a *Zea mays*.

Stejně jako u prokaryot též eukaryotické transponovatelné elementy bývají příčinou vysoké rychlosti mutace a přemísťují se na různá místa chromozómu. Na základě současných znalostí lze eukaryotické transponovatelné elementy charakterizovat takto (obr. 1):



Obr. 1. Schéma struktury transponovatelného elementu MDG1.

a) celkové schéma,

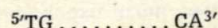
b) podrobnější schéma, v němž podtržené sekvence představují zdvojené sekvence hostitelské DNA v místě začlenění.

1. Délka eukaryotických transponovatelných elementů je 5000 až 7000 pb.

2. Stejný transponovatelný element se může opakovat v eukaryotickém genomu 10 až 200 krát, přičemž téměř všechny kopie téhož transponovatelného elementu jsou identické.

3. Konce eukaryotického transponovatelného elementu jsou tvořeny sekvencemi o délce 250 až 500 pb, které jsou orientovány ve stejném směru, tzv. stejnosměrně opakované sekvence. Uvnitř těchto sekvencí se nachází Hognessova sekvence, což má zřejmě rozhodující význam pro transpozici transponovatelného elementu. Části stejnosměrně opakovaných sekvencí jsou též obráceně opakované sekvence (u transpozónu MDG1, MDG3, copia a MDG 412 jsou stejnosměrně opakované sekvence ohraničené obráceně opakovanými sekvencemi).

4. Téměř všechny transponovatelné elementy jsou zakončeny dinukleotidy:



Tato skutečnost může být vyjádřením určité specifičnosti vzhledem k enzymům, které katalyzují transpozici.

5. K oběma koncům všech eukaryotických transponovatelných elementů bez výjimky přiléhají stejné sekvence 4 až 6 nukleotidů, které se vytvoří v hostitelském chromozómu duplikací 4 až 6 nukleotidů v místě začlenění transponovatelného elementu. U různých kopií téhož transponovatelného elementu jsou přilehlé koncové sekvence různé, tj. transponovatelný element se může začlenit do kteréhokoli místa na chromozómu. Počet nukleotidů ve zdvojené sekvenci je vždy konstantní.

6. Každý eukaryotický transponovatelný element podléhá intenzivní transkripci. Přepisují se oba řetězce, převládá však vždy transkripce jednoho řetězce. Předpokládá se, že transkripce eukaryotického transponovatelného elementu souvisí s jeho transpozicí, tj. přemístěním na jiné místo chromozómu. Transkripce vzniklá RNA se podobně jako u retrovirů přepíše reverzní transkriptázou do DNA, která pak přechází do kruhové formy a začlení se do chromozómu. Tato domněnka se opírá především o nápadnou shodu, co se týče struktury, mezi retroviry a transponovatelnými elementy. U retrovirů je kruhová forma DNA předpokladem jejich začlenění do genomu hostitelské buňky. Transponovatelné elementy copí se také nacházejí v kruhové formě. Lze se proto domnívat, že kruhová forma transponovatelných elementů je přechodným produktem jejich transpozice.

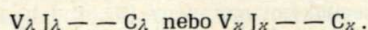
Transponovatelné elementy mají zřejmě význam při diferenciaci eukaryotických buněk. Lze jich též využít v genovém inženýrství. Příkladem v tomto směru je využití transponovatelného P-elementu u *Drosophila melanogaster*. P-element lze pomocí plazmidu pB322 používaného jako vektoru injikovat do embryí *D. melanogaster*. V embryích se P-element přemísť s vysokou účinností do chromozómu zárodečných buněk. Polovina z přežívajících embryí vede ke vzniku potomstva, jehož část se vyznačuje hypermutabilitou lokusu sn určujícího morfolologii chloupků a štětín kutikuly dospělé mouchy. DNA izolovaná z těchto mutantů obsahuje 1 až 5 P-elementů umístěných na různých místech chromozómu. Sekvence DNA umístěná mimo P-element na DNA, použité k injikaci embryí, se do chromozómu embryí nepřenáší. Z toho plyne, že jen P-elementy se mohou účinně a selektivně přemístit z plazmidové DNA do DNA chromozómu zárodečných linií embryonálních buněk *D. melanogaster*.

Jelikož klonované P-elementy lze přenášet plazmidy a začlenit do chromozomální DNA, může být do zárodečných linií *D. melanogaster* zaveden segment DNA obsahující jakýkoli gen, použije-li se P-element jako vektor. Vzhledem k tomu však, že segment DNA, vložený do P-elementu, může porušit jeho funkce, kterých je zapotřebí k transpozici, musí se do embryí s rekombinantním P-elementem (defektním vlivem vložené DNA) injikovat nedefektní standardní P-element, který poskytuje funkce potřebné k transpozici. Takto byl např. do zárodečných linií *D. melanogaster* ry^{-} (nekódujících xantindehydrogenázu) přenesen lokus ry^{+} (obr. 2).

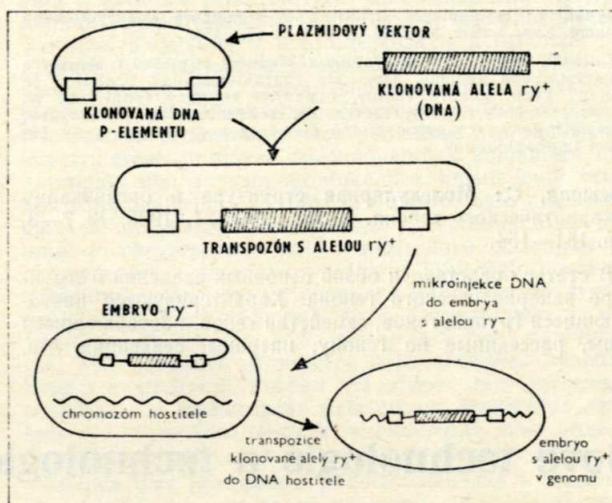
Strukturní geny kódující imunoglobulinové řetězce. Tyto geny jsou zajímavé v tom smyslu, že se sestavují až na úrovni svého přepisu. Genetická informace pro imunoglobuliny je totiž rozptýlena v exonech pro variabilní oblasti imunoglobulinů (V-exony), v J- a D-exonech a v exonech pro konstantní oblasti (C-exony). Základním zdrojem rozmanitosti protilátek je translokace V-, D-, J- a C-exonů, k níž dochází během diferenciaci B-lymfocytů. Touto translokací vzniká určitá kombinace V-, J- a C-exonů, nebo V-, D-, J- a C-exonů, která se přepíše do hnRNA, z níž se pak vyloučí přepisy intronů za vzniku celistvého (aktivního) přepisu strukturálního genu pro lehký nebo těžký imunoglobulinový řetězec.

Tvorba aktivního genu pro imunoglobulin a jeho přepis pro lehký nebo těžký řetězec zahrnuje v hrubých rysech tyto procesy (obr. 3):

1. Spojení na úrovni DNA příslušného exonu pro variabilní oblast (V-exon) imunoglobulinu s některým J-exonem a s exonem pro konstantní oblast (C-exon). Přitom se vyloučí úsek mezi V a J. Výsledný gen pro lehký řetězec má potom strukturu (pomlčkami jsou vyjádřeny introny):

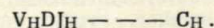


Jelikož při sestavování genů pro těžké řetězce přistupu-



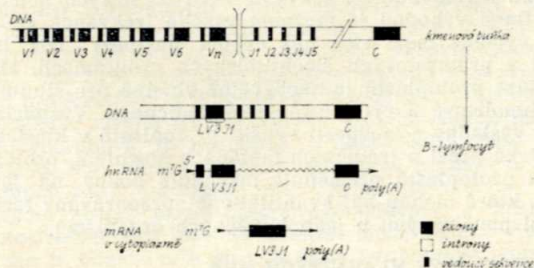
Obr. 2. Klonovaná kopie standardní alely ry^{+} je vložena do klonovaného P-elementu. Tento transpozón s alelou ry^{+} je pak injikován do embrya *Drosophila melanogaster* nesoucí alelu ry^{-} . Transpozázou syntézovaná injikovaným P-elementem katalyzuje transpozici ry^{+} z injikované DNA do chromozómu hostitele, v němž nahrazuje alelu ry^{-} .

je spojení mezi V_H a D, D a J_H , je třeba si strukturu tohoto genu představit takto:



2. Předpis takto spojených exonů do hnRNA.

3. Scelovací úpravy na úrovni hnRNA, při nichž dochází k vyloučení intronu mezi VJ a C. Vedoucí segment přepisu V se modifikuje čepičkou m^7G a na 3'-konec se napojuje poly(A). Vzniklá mRNA představuje přepis celistvého genu, v němž spojení mezi přepisem pro konstantní a variabilní oblast imunoglobulinového řetězce (lehkého nebo těžkého) není přerušeno. Přepis J-exonu je zde též obsažen a překládá se jako součást variabilní oblasti lehkého řetězce. Odpovídá asi 13 aminokyselinám variabilní oblasti.



Obr. 3. Znárodnění mechanismu vzniku aktivního genu pro imunoglobulin a jeho přepisu pro lehký a těžký řetězec. L-exony, tj. vedoucí sekvence, a jejich přepisy nejsou v textu uváděny.

Uvědomíme-li si, že k danému C-exonu lze přiřadit jakýkoli V-exon a jakýkoli J-exon, stojíme před velkým počtem kombinací, které vysvětlují protilátkovou rozmanitost. Tato rozmanitost je zvláště patrná v sestavě exonů pro těžké řetězce, kde dochází ke kombinaci exonů V, D, J a C.

Literatura

Podrobnou literaturu k této problematice lze nalézt v monografii Rosypal S.: Molekulární struktura a exprese eukaryotického genu. Universita J. E. Purkyně, Brno 1984.

Rosypal, S.: Molekulární struktura a organizace eukaryotického genomu. Kvas. prům. 31, 1985, č. 7—8, s. 151—154.

V článku je prezentován přehled základních poznatků o struktuře eukaryotického genomu. Jsou zde charakterizovány opakující se skupiny genů, genové rodiny, opakující se geny rozptýlené po genomu, introny, exony, mezerníky, sekvence Alu, pseudogeny, upravené geny, transponovatelné elementy a strukturní geny kódující imunoglobuliny.

Росыпал, С.: Молекулярная структура и организация эвкариотического генома. Квас. прум. 31, 1985, № 7—8, стр. 151—154.

В статье представлен обзор основных сведений о структуре эвкариотического генома. Характеризуются повторяющиеся группы генов, семейства генов, повторяющиеся гены, рассеянные по геному, интроны, секвенции Alu,

псевдогены, транспозоны и структурные гены кодирующие иммуноглобулины.

Rosypal, S.: Molecular Structure and Organisation of the Eukaryotic Genome. Kvas. prům. 31, 1985, No. 7—8, pp. 151—154.

In this article the survey of our knowledge on the structure of the eukaryotic genome is presented. The following components of the eukaryotic genome are characterized: repeated groups of genes, gene families, dispersed genes, introns, exons, spacers, Alu sequences, pseudogenes, processed genes, transposable elements, structural genes coding for immunoglobulins.

Rosypal, S.: Molekulare Struktur und Organisation des eukaryotischen Genoms. Kvas. prům. 31, 1985, Nr. 7—8, S. 151—154.

In diesem Artikel wurde eine Übersicht über unsere Kenntnisse des eukaryotischen Genoms präsentiert. Die folgenden Komponenten des eukaryotischen Genoms sind beschrieben: Repetitive Gengruppen, zerstreute Gene, Genfamilien, Intronen, Exonen, „Spacers“ Pseudogene, Immunoglobuline kodierende strukturelle Gene.