

Literatura

- Fojtová, M., Van Houdt, H., Depicker, A., Kovařík, A. 2003. - Plant Physiol. In press.
 Kovařík, A., Van Houdt, H., Holý, A., Depicker, A. 2000. - FEBS Lett. **467**: 47.
 van Blokland, R., van der Geest, N., Mol, J.N., Kooter, J.M. 1994. - Plant J. **6**: 861.
 Van Houdt, H., Kovařík, A., Van Montagu, M., Depicker, A. 2000. - FEBS Lett. **467**: 41.

Poděkování. Práce byla financována granty GAČR 521/01/P042, GAČR 521/01/0037 a Z5004920.

Identifikace a izolace cDNA funkčně homologních genů

JAROSLAV FULNEČEK

Biofyzikální ústav AV ČR, Královopolská 135, 612 65 Brno, (tel.: 541 517 230, e-mail: fulnecek@ibp.cz)

Úvod

V molekulární biologii může nastat situace, že potřebujeme izolovat gen, jehož funkční homolog (ortholog) byl již dříve popsán u jiného organismu. Často k této situaci dochází u projektů zaměřených na rostliny, kdy nejjednodušším způsobem řešení je využití poznatků z živočišné říše, jejíž výzkum bývá zpravidla o krok dále. Nejsnazší situace je u *Arabidopsis thaliana* díky známé sekvenci jejího genomu. Zde lze poměrně jednoduše vyhledat orthologní geny a výzkum proteomu je u této rostliny nejdále. *A. thaliana* však není reprezentativní pro celou rostlinnou říši. Existuje řada procesů, které je vhodné studovat u jiných rostlinných druhů, například ty, jež souvisejí s produkcí zemědělských plodin. Studium biologických procesů u více rostlinných modelových systémů nám pomůže odhalit složitost dějů a upozorní na rozdíly, které doplní všeobecný model založený na výsledcích studia *A. thaliana*.

Je vhodné zvolit rostlinný druh, jehož genom je ve fázi sekvenování (<http://www.info-biogen.fr/services/deambulum/english/genomes1.html>). V budoucnu nám pak kompletní sekvenční data mohou ušetřit spoustu práce. Můžeme hledat příslušné orthology, identifikovat všechny členy genové rodiny a pseudogeny, studovat regulační oblasti atd. Rostlinný genom se však čte velmi dlouho, a proto zmíněné výhody mohou vyvážit dlouhodobé zkušenosti s jinou modelovou rostlinou. Navíc, jedná-li se o zemědělsky významnou plodinu, můžeme doufat ve snazší získání finančních prostředků.

Ve funkční genomice rostlin se již uplatňují přístupy reversní genetiky. Nejprve je popsán gen a poté se ověřuje jeho funkce. Lze použít několik přístupů k izolaci genu. O konkrétním postupu se zpravidla rozhoduje na základě dostupných informací a materiálu. Pokud chceme studovat jen několik genů, které jsou známy u jiných organismů, lze využít informací z biologických databází pro identifikaci a izolaci příslušných cDNA.

Podstata řešení

Na základě konzervativních oblastí aminokyselinových sekvencí proteinů kódovaných orthologními geny různých organismů můžeme nalézt gen pro podobný protein u jiného organismu. Pokud je naším modelovým organismem rostlina zařazená do sekvenačního projektu nebo pro ni existuje EST databáze (*expressed sequence tags*), lze k vyhledávání nukleotidové sekvence genu využít databáze bez nutnosti vlastního experimentu. K vyhledávání nukleotidových sekvencí na základě známých aminokyselinových sekvencí se používají například programy: TFastA, TFastX (GCG WISCONSIN PACKAGE, http://www.accelrys.com/products/gcg_wisconsin_package/index.html, <http://www.iccbnet.cz/Services.html>, <http://fasta.bioch.virginia.edu/> nebo fastx3, fasty3, (<http://www.ebiac.uk/fasta33/index.html>) nebo tblastn, <http://www.ncbi.nlm.nih.gov/BLAST/>).

Co je potřeba v laboratoři

Tento úkol lze řešit v laboratoři pro molekulární biologii, vybavené PCR cyklérem a aparaturami pro horizontální agarosovou a vertikální polyakrylamidovou elektroforézu. Z gelu izolované produkty PCR budou sekvenovány přímo nebo až po naklonování do bakteriálního vektoru. Oligonukleotidy budou zakoupeny. Bude nutné dodržet základní pravidla pro práci s RNA. Pro izolaci RNA a její reversní transkripci lze využít komerční soupravy. 3' RACE postup (*rapid amplification of cDNA ends*) pro získání 3' konce cDNA je možné provést i bez soupravy, některé 5' RACE postupy lze také provést bez soupravy, avšak pro izolaci 5' konce cDNA založené na využití čepičky mRNA bude vhodné soupravu zakoupit (*Invitrogen GeneRacer Kit* (5) 24 tis. Kč). Velice důležitá je možnost a dovednost prohledávat biologické databáze a pracovat se získanými sekvencemi.

Pracovní postup

Vyhledání známých sekvencí a jejich seřazení za účelem zjištění konzervativních domén vhodných pro navržení degenerovaných oligonukleotidů

Pravděpodobnost zachycení homologního genu roste s počtem již známých sekvencí, proto je velice důležité prohledat biologické databáze a získat co nejvíce informací o daném genu. Snažíme se získat nukleotidové a aminokyselinové sekvence byť i jen částečné nebo předpokládané a to bakteriální, živočišné i rostlinné. Nukleotidové sekvence převedeme na aminokyselinové a ty seřadíme pod sebe na základě homologií s použitím programu např. programy Clustal, (<http://www.ebi.ac.uk/clustalw/>), PileUp (GCG WISCONSIN PACKAGE, http://www.accelrys.com/products/gcg_wisconsin_package/index.html, <http://www.iccbnet.cz/Services.html>), apod.

Orthologní proteiny obvykle mají ve své aminokyselinové sekvenci vysoce konzervativní oblasti obsahující aminokyseliny, které se aktivně podílejí na plnění funkce proteinu. Mohou to být katalytická centra, domény zabezpečující vazbu s kofaktory nebo substrátem popřípadě s DNA nebo dalšími funkčními proteiny v multiproteinovém komplexu. Porovnáním bakteriálních, živočišných a rostlinných aminokyselinových sekvencí proteinu tyto oblasti zjistíme a sekvence z nejbližší příbuzných druhů využijeme pro navržení degenerovaných oligonukleotidů, které jsou poté použity při amplifikaci prvního vlákna cDNA k zachycení části hledané sekvence.

Navržení degenerovaných oligonukleotidů

Navržení degenerovaných oligonukleotidů je kritická část postupu. Konzervativní oblast proteinů (blok) se přepíše z pořadí aminokyselin zpět do pořadí nukleotidů podle genetického kódu, který je degenerovaný, a proto získáme sadu nukleotidových sekvencí jako výsledek kombinace různých kodonů jednotlivých aminokyselin. Navržený degenerovaný oligonukleotid je tedy ve skutečnosti směsí oligonukleotidů, která obsahuje velice málo molekul se správnou sekvencí. Ty jsou vyčerpány v prvních cyklech amplifikace. Nadbytkem oligonukleotidů s nesprávnou sekvencí dochází k nespecifické amplifikaci, která je navíc podpořena jak nastavením méně přísných podmínek amplifikace, což je na druhou stranu často nutné pro získání neznámého cílového fragmentu, tak použitím krátkých oligonukleotidů navržených podle krátkých bloků. Proto je často výsledkem vysoké pozadí nespecifických produktů, ve kterém nelze rozeznat správný fragment.

Jiným přístupem je použití oligonukleotidu s konsensus sekvencí, tedy oligonukleotidu, jehož jednotlivé nukleotidy odpovídají nejčastěji se vyskytujícím bazím v daném místě nukleotidových sekvencí konzervativního bloku. Ačkoli byla tato strategie s úspěchem použita pro izolaci velice konzervativních orthologních genů, je nevhodná pro izolaci vzdáleně příbuzných genů.

Řešení problémů obou metod spočívá v jejich kombinaci ve strategii CODEHOP (*consensus-degenerate hybrid oligonucleotide primer*) (Rose *et al.* 1998), která byla zpracována i jako program (<http://blocks.fhcr.org/codehop.html>). Hybridní oligonukleotid se skládá z poměrně krátké 3' degenerované aktivní oblasti, která je navržena podle 3 - 4 vysoce konzervativních aminokyselin a z delší nedegenerované 5' konsensus oblasti. Zkrácení degenerované oblasti na minimum sníží počet sekvencí vyskytujících se ve směsi. Její hybridizace s templátem je stabilizována vazbou s nedegenerovanou oblastí, což umožňuje zvýšení teploty nasednutí bez dalšího zvyšování degenerace ve směsi. V prvních cyklech amplifikace nasedá 3' oblast oligonukleotidu přesně, v 5' oblasti se případně mohou vyskytnout nepárované nukleotidy, které DNA polymerase nebrání v prodloužení řetězce. Díky jednotnosti sekvence v 5' oblasti nasedají oligonukleotidy na produkt amplifikace i v dalších cyklech, nepárované nukleotidy v 3' oblasti dále od konce nemusí bránit DNA polymerase v prodloužení řetězce a to v podstatě umožní využití téměř všech oligonukleotidů.

Tímto přístupem navrhne oligonukleotidy podle všech konzervativních oblastí.

Příprava prvního vlákna cDNA

První vlákno cDNA se připraví reversní transkripcí celkové RNA nebo mRNA. Je užitečné si připravit několik cDNA použitím různých primerů např. oligo (dT), náhodných dodekamerů nebo hexamerů, oligo (dT) prodloužené o sekvenci dvou primerů podle 3' RACE protokolu viz. dále. Tyto cDNA se totiž liší vhodností pro amplifikaci různých oblastí. (Upozorňuji, že všechny oligonukleotidy, potenciálně použitelné pro reversní transkripci je nutné rozpouštět v roztocích prostých RNas).

Amplifikace neznámého cílového fragmentu z cDNA pomocí degenerovaných oligonukleotidů

V řadě amplifikací a následných uhnížděných amplifikací vyzkoušíme všechny vhodné kombinace navržených degenerovaných oligonukleotidů. (Zde nemá význam použít

oligo (dT) v kombinaci s degenerovanými oligonukleotidy, jejich T_m se výrazně liší, a proto nezískáme žádné fragmenty). Pokud neobdržíme ani jeden cílový fragment, navrhneme nové sekvence oligonukleotidů. Lze amplifikovat i genomovou DNA. Nedoporučuji ověřovat správnou funkci degenerovaných oligonukleotidů amplifikací genomové DNA nebo cDNA kontrolních druhů dříve než obdržíme cílový fragment. Hrozí totiž nebezpečí, že dojde ke kontaminaci kontrolním fragmentem, který posléze bude příčinou falešně pozitivních výsledků. Délka získaných fragmentů cDNA je dostačujícím vodítkem. Měla by souhlasit s předpokládanou délkou, určenou z homologních aminokyselinových sekvencí proteinů blízkých druhů. Pokud nečekáme délku amplifikovaného fragmentu větší než kilobázi, je jeho detekce citlivější v polyakrylamidovém gelu, protože agarosové gely mají vysoké pozadí vázaného ethidiumbromidu a malá množství fragmentu by v něm zanikla. Změnou podmínek amplifikace, nebo po extrakci fragmentu z gelu jeho reamplifikací, popřípadě uhnížděnou amplifikací s další sadou oligonukleotidů můžeme dosáhnout značného zvýšení množství získaného fragmentu.

Podaří-li se nám získat fragment odpovídající délky, zjistíme jeho sekvenci nukleotidů a převedeme ji do pořadí aminokyselin. Absence stop kodonu znamená, že se pravděpodobně jedná o kódující sekvenci. Porovnáním s odpovídajícími oblastmi orthologních proteinů zjistíme, zda se nám skutečně podařilo zachytit nového zástupce. Pořadí nukleotidů na tomto krátkém úseku využijeme pro navržení specifických oligonukleotidů, které budou dále použity pro zjištění sekvence na 3' konci, tak získáme delší sekvenci a jistotu, že se jedná o hledaný gen. Poté již zbývá jen investovat do odhalení sekvence na 5' konci a naklonování celé cDNA.

3'RACE

Zjištění sekvence na 3' konci provedeme pomocí 3' RACE (Frohman *et al.* 1988). K tomu není třeba kupovat soupravu, stačí použít cDNA připravenou reversní transkripcí celkové RNA nebo mRNA s využitím speciálního delšího oligo(dT). Ten kromě osmnácti T na 3' konci obsahuje na svém 5' konci ještě další sekvenci, v níž jsou sloučeny sekvence dvou speciálních oligonukleotidů, které spolu se dvěma oligonukleotidy specifickými pro hledaný gen slouží při amplifikaci a uhnížděné amplifikaci cDNA sekvence na 3' konci. Sekvenci tohoto oligo(dT) jako i dvou speciálních oligonukleotidů zjistíme v návodech pro 3' RACE soupravy například od firmy Invitrogen a necháme si je připravit na zakázku společně se dvěma oligonukleotidy specifickými pro gen. Po amplifikaci, případně až po uhnížděné amplifikaci, příslušný fragment sekvenujeme. Tímto způsobem často získáme několik různě dlouhých fragmentů, odpovídajících 3' konci mRNA sekvence. Jedná se o běžný jev způsobený výskytem několika motivů podobných klasickému polyadenylačnímu signálu na konci primárního transkriptu, a proto dochází k alternativnímu připojování poly(A) řetězce.

5'RACE

Pro zjištění sekvence na 5' konci cDNA existují v zásadě dvě možnosti. Klasickým vyhledáváním v cDNA knihovně se nám může podařit najít kompletní cDNA. Většinou však zachytíme několik různě dlouhých cDNA bez 5' sekvence a nezbývá než použít 5' RACE. Ta je založena na několika přístupech. Společnost TaKaRa zvolila reversní

transkripci z genově specifického oligonukleotidu, který je předem fosforylován na 5' konci (Maruyama *et al.* 1995). První vlákno cDNA se cirkularizuje T4 RNA ligasou a následuje reversní amplifikace s dalšími gen specifickými oligonukleotidy. 5' RACE systém firmy Invitrogen používá terminální deoxynukleotidyltransferasu, která prodlouží cDNA připravenou z genově specifického primeru o řetězec deoxycytidinmonofosfátů (Frohman 1990, 1993, Loh 1991). Ten slouží k nasednutí speciálního oligonukleotidu obsahujícího deoxyinosiny pro účinnější amplifikaci 5' konce. Avšak pokud dojde k předčasnému ukončení reversní transkripce, vytvoří se zkrácené cDNA a sekvenci na 5' konci stejně nezjistíme. Tento problém řeší systém soupravy GeneRacer firmy Invitrogen (Maruyama a Sugano 1994, Volloch *et al.* 1994, Schaefer 1995). Zde dochází k amplifikaci jen kompletních cDNA, což je docíleno využitím čepičky na 5' konci mRNA, která ochrání fosfátové skupiny na 5' koncích intaktních mRNA, zatímco fosfáty na 5' koncích degradovaných mRNA (bez čepičky) jsou odstraněny fosfatasou. Po odstranění čepičky na intaktních mRNA pomocí kyselé pyrofosfatasy z tabáku zůstává na jejich 5' konci zůstává fosfátová skupina. Ta je nutná při ligaci se speciálním RNA oligonukleotidem pomocí T4 RNA ligasy. Tak získáme známou RNA sekvenci na 5' konci, kterou po reversní transkripci s oligonukleotidem specifickým ke genu využijeme pro amplifikaci 5' konce cDNA.

Klonování a sekvenování kompletní cDNA

V okamžiku, kdy zjistíme kompletní sekvenci cDNA, naklonujeme celý úsek a ověříme jeho sekvenci přečtením několika klonů z alespoň tří nezávislých amplifikací. Poté můžeme s klidným svědomím sekvenci poslat do databází a začít s charakterizací genu a s řadou dalších jistě velice zajímavých studií.

Závěr

Pro úspěch je klíčové nalezení první krátké specifické sekvence v RT-PCR experimentech. S trochou štěstí by to snad nemělo trvat déle než půl roku. Přeji hodně zdaru.

V přednášce bych se chtěl věnovat získaným výsledkům při identifikaci cDNA tří základních typů rostlinných DNA-(cytosin-5)-methyltransferas u *Nicotiana sylvestris*.

Literatura

- | | |
|---|--|
| Frohman, M.A. 1990. - In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (eds): PCR Protocols: A Guide to Methods and Applications. S. 28. Academic Press, San Diego. | Maruyama, I.N., Rakow, T.L., Maruyama, H.I. 1995. - Nucleic Acids Res. 23 : 3796. |
| Frohman, M.A. 1993. - Meth. Enzymol. 218 : 340. | Maruyama, K., Sugano, S. 1994. - Gene 138 : 171. |
| Frohman, M.A., Dush, M.K., Martin, G.R. 1988. - Proc. Natl. Acad. Sci. USA 85 : 8998. | Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., Henikoff, S. 1998. - Nucleic Acids Res. 26 : 1628. |
| Loh, E. 1991. - Methods 2 : 11. | Schaefer, B.C. 1995. - Anal. Biochem. 227 : 255. |
| | Volloch, V., Schweitzer, B., Rits, S. 1994. - Nucleic Acids Res. 22 : 2507. |

Poděkování. Experimenty, zmíněné v tomto článku byly a jsou podporovány prostředky z grantů GAČR 204/99/D001 a 204/03/P104.