

Molekuly na povel V.

Čtení (genomu) na dobrou noc

ZUZANA STORCHOVÁ

Každý organismus potřebuje pro to, aby vypadal a fungoval tak, jak má, nějaké pokyny, informace. Tato informace je z podstatné části zapsána v pořadí bází nukleové kyseliny. Podle genů, tedy konkrétních úseků nukleové kyseliny, se vytvářejí bílkoviny – pořadí bází je pomocí určitého kódu, společného všem organismům, převedeno na posloupnost aminokyselin

v bílkovině. Pořadí neboli sekvence bází také do značné míry charakterizuje vlastnosti jednotlivých oblastí DNA – jinak vypadá místo, kde je zahajována replikace, a jiné sekvence najdeme na koncích chromozomu. Oblast DNA s určitou typickou sekvencí bází také například určuje, jak silně bude nějaký gen exprimován čili kolik molekul příslušné bílkoviny podle předpisu toho genu vznikne. Mnozí proto podleli dojmů, že právě sekvence genomu organismů je to jediné důležité, co nám o daném organismu řekne všechno. Ať už je toto tvrzení pravdivé nebo ne, je jisté, že nedávno zahájené velkoakce sloužící k zjišťování pořadí bází v chromozomech různých organismů poskytnou vědecké komunitě množství neocenitelných informací a komunitě laické řadu výhod (možná i nevýhod) plynoucích z možností využití získaných znalostí. K nejznámějším projektům patří akce HUGO, jejímž cílem je zjistit nukleotidovou sekvenci všech lidských chromozomů. Občas se v novinách objeví zpráva typu *Čtení genetického kódu* či *Genetický kód člověka rozlušťen* atp. Lžou vám. Genetický kód – čili kód, podle nějž se informace zapsaná pořadím nukleotidů v nukleových kyselinách převádí na informaci zapsanou pořadím aminokyselin v bílkovině – člověka i jakéhokoliv jiného organismu byl rozluštěn před dobrými 30 lety. Monstrózní projekt zmiňovaný v novinách se týká „pouze“ zjištění pořadí bází ve všech chromozomech člověka a řady dalších organismů. S genetickým kódem nemá nic společného.

Čtení chromozomů, čili zjišťování pořadí nukleotidů v daném úseku DNA, není nic jednoduchého. V projektu HUGO (a řadě dalších) je zapojeno množství odborných pracovníků, a přesto jednotlivé projekty trvají několik měsíců až let. Metod, které umožňují zjistit sekvenci nukleotidů v DNA, je několik, ale v poslední době převládla jediná. Jejím autorem je Frederick Sanger a byl za ni po právu odměněn Nobelovou cenou (spolu s Paulem Bergem a Walterem Gilbertem, kteří vypracovali méně používané techniky sekvencování). Je neobyčejně rafinovaná, a přesto elegantně jednoduchá (ostatně takové je v molekulární biologii všechno).

Co všechno k sekvencování potřebujeme

Sangerova enzymatická metoda zjišťování pořadí nukleotidů v molekule DNA vychází opět pouze a jediné z nám dobře známých fyzikálních a chemických vlastností DNA a z funkcí enzymů, které s DNA zacházejí. Pokud si vzpomínáte, syntézu DNA zahajuje enzym DNA-polymeráza pouze z primeru, tedy z krátké molekuly nukleové kyseliny. Na tento primer navěšuje nové nukleotidy v po-

Mgr. Zuzana Storchová (*1970) viz Vesmír 77, 15, 1998/1

1 CGTAGGTAAC TGGACCT
GCATCCATTGACCTGGACTAGCCAATATAGCATTAGGCATACGCCG **primer (17 nukleotidů) matrice**

1 rozdělíme do 4 zkumavek, do každé dáme jeden nukleotid ve stopovací i pokračovací formě

2 ve zkumavce G jsou pokračovací formy prekurzorů A, C a T a prekurzor G je ve formě stopovací i pokračovací v optimálním poměru, pokračovací forma prekurzoru A je radioaktivně značená

3 při enzymatické syntéze nového vlákna podle matrice se zařazují příslušné komplementární nukleotidy

4 pokud se náhodou zařadí stopovací prekurzor, syntéza je ukončena

5 po proběhlé reakci získáme následující směs molekul:

CGTAGGTAAC TGGACCT **GATCGGTTATATCGTAATCCGTATGCGG**
CGTAGGTAAC TGGACCT **GATCGGTTATATCGTAATCCGTATGCG**
CGTAGGTAAC TGGACCT **GATCGGTTATATCGTAATCCGTATG**
CGTAGGTAAC TGGACCT **GATCGGTTATATCGTAATCCG**
CGTAGGTAAC TGGACCT **GATCGGTTATATCG**
CGTAGGTAAC TGGACCT **GATCGG**
CGTAGGTAAC TGGACCT **GATCGG**
CGTAGGTAAC TGGACCT **GATCG**
CGTAGGTAAC TGGACCT **GATCG**
CGTAGGTAAC TGGACCT **GATCG**
GCATCCATTGACCTGGACTAGCCAATATAGCATTAGGCATACGCCG

délka vlákna

45 nukleotidů
44 nukleotidy
42 nukleotidy
38 nukleotidů
31 nukleotid
23 nukleotidy
22 nukleotidy
18 nukleotidů
matrice

6 po elektroforetické analýze na polyakrylamidovém gelu se molekuly jednotlivých vláken od sebe oddělí

7 gel s molekulami rozdělenými podle velikosti vysušíme a přiložíme na citlivý film

8 po expozici a vyvolání uvidíme následující obrázek



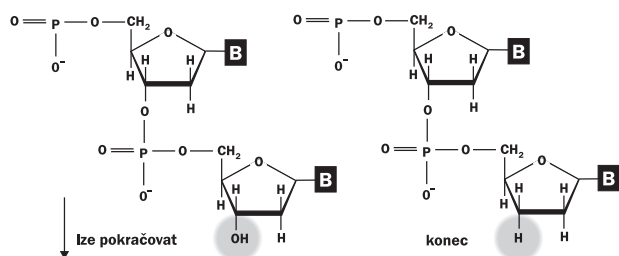
9 tedy: báze G (guanin) je v příslušném vlákně na 18., 22., 23., 31., 38., 42., 44. a 45. pozici, v komplementárním (matricovém) vlákně tomu odpovídá báze C (cytozin)

10 analogický výsledek získáme s každou další reakční směsí, v každé zkumavce pro jeden nukleotid

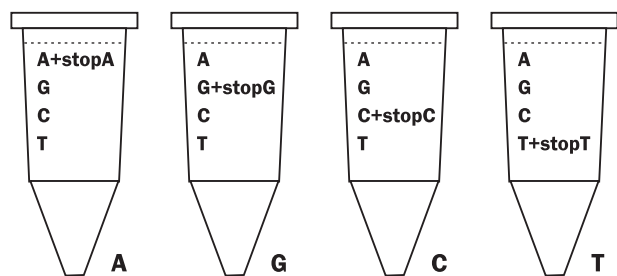
11 kombinací všech výsledků zjistíme úplné pořadí bází sekvencované molekuly

© VESMÍR

řadí od 5' konce k 3' konci, čili na skupinu OH předchozího nukleotidu napojí fosfátovou skupinu nukleotidu dalšího podle kopírované matrice. K tomu, aby DNA-polymeráza mohla pracovat, potřebuje tedy vhodné prostředí, primer, matricovou DNA a prekurzory syntézy, v tomto případě deoxyribonukleosidtrifosfáty (nenechte se vyděsit složitým názvem – důležité je, že z tohoto prekurzoru DNA-polymeráza vše nepotřebné odštěpne a na vznikající řetězec připojí tu správnou část). Představme si ale, že DNA-polymeráze nabídneme místo prekurzorů, které po připojení přes fosfátovou skupinu zůstanou do prostoru třet volným koncem OH připraveným k dalšímu spojování, nějaké prekurzory jiné, které už nebudou ochotny se s něčím dalším spojovat. Takovými prekurzory s odštěpenou skupinou OH jsou dideoxyribonukleosidtrifosfáty, a ty jsou také tím posledním, co na nově vznikající řetězec může DNA polymeráza připojit. Pro další syntézu totiž chybí ocásek OH, a tak není možné pokračovat.



Tento experiment je vám nepochybně jasný a stejně nepochybně asi nechápete, k čemu jsou takové triky dobré. Inu, pro sekvencování! Vezměme si úsek DNA, jehož pořadí bází chceme znát, a primer, z něž můžeme syntézu nového řetězce, komplementárního k původnímu, začít. Matricovou DNA (tu, kterou chceme číst) denaturujeme a v jednořetězcovém stavu smísíme s primerem. Primer nasedne na komplementární místo a přidáme-li do směsi i DNA-polymerázu a všechny prekurzory, syntéza může začít. To my ovšem neuděláme, neboť by nám to bylo k ničemu. Místo toho takto připravenou směs rozdělíme do čtyř zkumavek, do každé z nich přidáme polymerázu a prekurzory, ale pozor! Do každé zkumavky dáme vždy jenom tři správné prekurzory a 4. prekurzor (pro každou zkumavku jiný) připravíme tak, že bude např. 90 % toho správného a 10 % toho, co se na něj už nedá nic navěsit.



Proč bude čtvrtý prekurzor v reakční směsi přítomen v obou formách? Kdybychom už od začátku přidali do zkumavky pouze onen stopovací prekurzor, zastavila by se nám výroba nového vlákna DNA na tomto prvním zařazeném nukleotidu. Pokud však připravíme prekurzor „stopovací“ a prekurzor „pokračovací“ ve správném poměru, zařazuje se stop-prekurzor jenom občas. A tak vznikají během syntézy vlákna nejrůznějších délek, která však vždy končí tím příslušným „stopovacím“ prekurzorem. Nastavíme-li poměry obou typů prekurzorů správně, získá-

me vlákna všech délek vytvořená podle naší oblíbené DNA, jejíž sekvenci si chceme po večerech číst. A protože máme 4 zkumavky a v každé je jiný stopovací prekurzor, získáme ve velmi krátké době:

- A) všechna možná různě dlouhá vlákna DNA vytvořená z primeru podle části matricové molekuly, v nichž poslední báze je adenin,
- G) všechna možná různě dlouhá vlákna DNA vytvořená z primeru podle části matricové molekuly, v nichž poslední báze je guanin,
- C) všechna možná různě dlouhá vlákna DNA vytvořená z primeru podle části matricové molekuly, v nichž poslední báze je cytozin a
- T) všechna možná různě dlouhá vlákna DNA vytvořená z primeru podle části matricové molekuly, v nichž poslední báze je thymin.

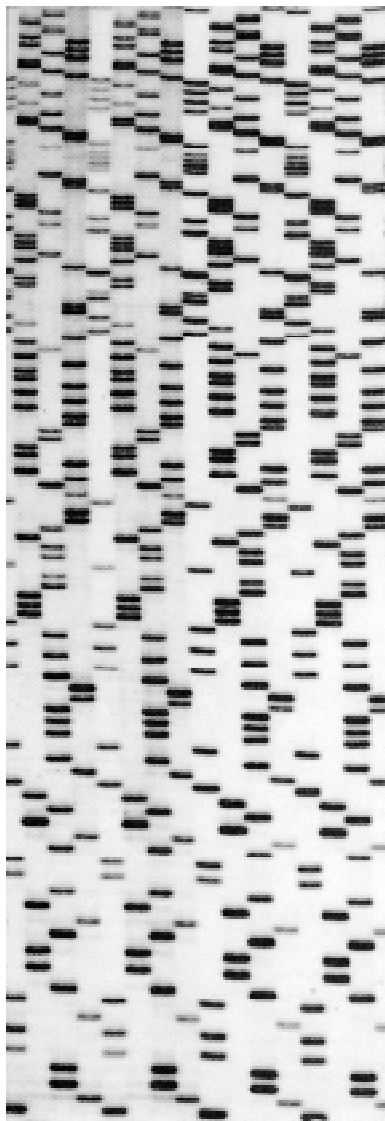
Pokud vám to přece jenom není úplně jasné, prohlédněte si důkladně obrázek 1, kde je detailně zpravená situace z případu G.

A jdeme číst!

Máme čtyři různé reakční směsi a v každé z nich je kromě matricového vlákna spousta nových, různě dlouhých vláken, a ta vždy v každé zkumavce končí jedním konkrétním nukleotidem – buď A, nebo G, nebo C, nebo T. Délky jednotlivých vláken ze všech směsí se při sekvencování liší o jeden jediný nukleotid! Musíme tedy připravit takový systém, abychom od sebe oddělili i vlákna, která se délkou liší o tento nejkratší možný úsek. To lze provést při speciální polyakrylamidové elektroforéze (viz Vesmír 77, 372, 1998/7), kde se jednořetězcové molekuly DNA pohybují v mimořádně tenkém gelu od záporné elektrody ke kladné a dělí se podle své velikosti – nejrychleji putuje nejmenší molekula, zatímco vlákno dlouhé se líně batoumá gelovými zákrutami. Na elektroforéze necháme proběhnout naše reakční směsi – každou zvlášť – při napětí obvykle kolem 1500–2000 V. V těchto podmínkách se od sebe oddělí i molekuly, jejichž velikost se liší o jediný nukleotid. Zviditelnění jednotlivých proužků však není jednoduché. Nemůžeme použít klasické barvení nějakou málo citlivou interkalační barvičkou, jak se to dělá při běžné elektroforéze, neboť sekvenační elektroforézu podstupují jednořetězcová vlákna DNA a interkalační barvička se nemá kam vmezerit (interkalovat). Proto se už při syntéze používají prekurzory, které jsou radioaktivně označeny (k dostatečnému signálu stačí použít jenom jeden typ prekurzoru radioaktivně značený) a vyzářují ionizující záření, takže každé vzniklé vlákno a následně i každý proužek DNA určité velikosti na gelu „svítí“. Gel se pak po elektroforéze „jednoduše“ plácne na film a kde je na gelu molekula vlákna příslušné délky, tam vznikne na filmu černý proužek (obr. 2). Po vyvolání filmu si pak badatel vezme silnou lampu, pravítko a zapisovatele, načež proužek po proužku odečítá jednotlivé báze v molekule. Co na tom, že odečtená sekvence neodpovídá původní molekule! Zato čteme báze k matici komplementární – ve dvouřetězcové DNA nám stačí znát jeden řetězec, druhý si podle zákona párování vymyslíme sami. Důležité je, že jsme opět pouze s využitím enzymů metabolismu nukleových kyselin a jednoduchých fyzikálně-chemických vlastností DNA získali informaci o pořadí bází v dané molekule. (Pro štouraly: postup používaný v praxi se maličko liší od výše uvedeného, princip však zůstává stejný.)

Až budeme mít všechno přečtené ...

Ohromný rozvoj molekulární biologie a genetiky v posledních letech vede k získávání mimořádného



2. Sekvence přečtená pomocí ručního sekvenování. V každé dráze byla rozdělena jedna reakční směs, čtyři dráhy společně poskytují informaci pro celou sekvenovanou molekulu.

množství nových poznatků a tím na principu pozitivní zpětné vazby umožňuje další zrychlení sebe sama. Téměř každý nový objev totiž může být využit k dalším studiím, které přinesou nový objev, který může být využit k novým studiím, které... Bouřlivý vývoj molekulární biologie přivedl badatele do stavu, kdy se i velmi smělé plány zdají být realizovatelné. Jedním z nejambicióznějších programů je HUGO, mezinárodní program zjištění úplné sekvence lidského genomu, k němuž se připojila řada dílčích podprogramů sekvenování genomů mnoha dalších organismů (viz Vesmír 76, 163, 1997/3). Při řešení tohoto úkolu se objevují problémy nejen vědecké a technické, ale i společenské, s nimiž si již dnes, několik let před dokončením programu, lámou hlavy právníci, sociologové i pojišťovací agenti (a tēm je taky, alespoň zatím, přenecháme). Náročné technické problémy, které se při řešení tohoto úkolu vynořily, donutily badatele zdokonalit metodu sekvenování až k hranicím představitelnosti.

Výše popsané klasické ruční sekvenování je totiž značně časově náročné a může je provádět pouze dovedný molekulární bio-

log. Na jednom gelu je možné zpravidla analyzovat 6–8 sekvencí dlouhých od 200 do 500 nukleotidů – podle konkrétních podmínek. Lidský genom obsahuje zhruba 5×10^9 párů bází, a aby byla zjištěna sekvence dostatečně hodnověrná, je třeba ji správně přečíst nejméně dvakrát. V optimálním případě by tedy bylo zapotřebí provést 3,5 milionu gelových analýz sekvence. Evidentní nemožnost tohoto úkolu pomohla na svět automatickému sekvenátoru – přístroji, který dělá (skoro) všechno sám. Princip sekvenování zůstává stejný, nepoužíváme však radioaktivní značení, ale každý stopovací prekurzor je označen

3. Výsledek sekvenování pomocí automatického sekvenátoru. Počítač data shromažďuje a ihned vyhodnocuje. Případné nejednoznačnosti (např. překrývají-li se dva vrcholy atp.) musí ovšem rozhodnout člověk nebo další sekvenování.



jinou fluorescenční barvičkou. Reakce probíhá se všemi prekurzory („stopovacími“ i „pokračovacími“) v jediné zkumavce, a poté je směs nanášena do jediné kapiláry s předpřipraveným gelem optimálních dělicích vlastností. Vzniklé jednotlivé různě dlouhé a různě „barevně“ označené molekuly pak putují v elektrickém poli od – k + a na konci z kapiláry, správně podle velikosti rozdělené, vyputují. Přímou v bodě, kde kapiláru opouštějí, však musí označené molekuly projít svazkem laserových paprsků, který umožní detektoru „odečíst“ jejich barvu a poslat ji počítači. Ten pak každé „barvě“ sám přiřadí příslušný nukleotid a badateli nahlásí rovnou celou sekvenci (obr. 3). Akce je tak zjednodušena a zrychlena, že se ze sekvenování stává příjemná odpolední zábava místo vysilujícího celodenního programu s nejistým výsledkem.

Ani automatický sekvenátor však neumožňuje sekvenovat najednou libovolně dlouhou molekulu. Jed-

PE Biosystems

A DIVISION OF PERKIN-ELMER

Přední světový výrobce přístrojů a chemikálií pro oblast molekulární biologie nabízí ze své produkce:

- **automatické genetické analyzátoři na bázi kapilární elektroforézy**
- **vysoce výkonné sekvenátory DNA**
- **systém TaqMan™ pro kvantitativní stanovení produktů polymerázové řetězové reakce**
- **termocyklyery a reagentie pro PCR**

ABI PRISM

Perkin Elmer s.r.o., Nad Ostrovem 7, 147 00 Praha 4
Tel: 02/61222164-7 Fax: 02/61222168
E-mail: PE-Czechia@perkin-elmer.com

ním z nejnáročnějších úkolů se tak při čtení pořadí nukleotidů v DNA stává příprava fragmentů pro sekvenování včetně udržení pořádku při sestavování získaných sekvencí dohromady. V praxi to vypadá tak, že každá laboratoř zapojená do projektu „obdrží“ k sekvenování určitý úsek DNA, který si rozdělí na řadu dalších vzájemně se překrývajících molekul DNA. Ty pak postupně sekvenuje a překrývající části využívá k sestavování celé molekuly dohromady. Do projektu HUGO a řady dalších sekvenačních projektů je celosvětově zapojeno množství pracovišť. Informace, které díky nim získáme, bude možno využít nejen v diagnostice dědičných chorob, k objevu nových genů či nových regulačních oblastí v genomu. Zjištěná data jsou neméně důležitá a přínosná pro systematické a evoluční biology, kteří mohou využít sekvence DNA jako další ze znaků pro určování příbuznosti druhů. Neočekávané jsou údaje, které ze sekvence DNA zjišťují matematici – objevují další informační roviny v molekule DNA, které mohou její fungování značně ovlivnit. Asi málokdo si dovede do všech detailů představit důsledky tohoto monstrózního projektu. Jedno je však skoro jisté – životu díky tomu lépe rozumět nebudeme. □